Does Affirmative Action Lead to Mismatch? A New Test and Evidence[*]

Peter Arcidiacono[†] Esteban M. Aucejo[‡] Hanming Fang[§] Kenneth I. Spenner[¶]

July 27, 2011

Abstract

We argue that, once we take into account the students' rational enrollment decisions, mismatch in the sense that the intended beneficiary of affirmative action admission policies are made worse off *ex ante* can only occur if selective universities possess private information. Ex ante mismatch occurs when revelation of this information would have changed the student's choice of school. This necessary condition for mismatch provides the basis for a new test. The test is implemented using data from the Campus Life and Learning Project (CLL) at Duke University. Evidence shows that Duke does possess private information that is a statistically significant predictor of the students' post-enrollment academic performance. Further, this private information is shown to affect the subjective measures of students' satisfaction as well as their persistence in more difficult majors. We also propose strategies to evaluate more conclusively whether the presence of Duke private information has generated mismatch.

Keywords: Mismatch; Private information; Affirmative Action

JEL Classification Codes: D8, I28, J15.

---

[*]We would like to thank Editor Orazio Attanasio, two anonymous referees, Judy Chevalier, Joe Hotz, Caroline Hoxby, Christine Jolls, Jon Levin, Jim Levinsohn, Tong Li, Linda Loury, Jeff Smith, Justin Wolfers and seminar participants at Brown, Emory, South Carolina, Vanderbilt, Penn, Stanford GSB, NBER Public Economics and Higher Education Meetings and ERID Law and Economics Conference at Duke for helpful comments. The authors gratefully acknowledge support for this research provided by grants from the Andrew W. Mellon Foundation and Duke University. All remaining errors are ours.

[†]Department of Economics, Duke University and NBER; Email: psarcidi@econ.duke.edu.

[‡]Department of Economics, Duke University. Email: esteban.aucejo@duke.edu.

[§]Department of Economics, University of Pennsylvania; and the NBER. Email: hanming.fang@econ.upenn.edu.

[¶]Department of Sociology, Duke University. Email: kspen@soc.duke.edu.

1 Introduction

The use of racial preferences in college and university admissions has generated much debate. Proponents of racial preferences argue that race-conscious admissions are important both for helping minorities overcome the legacy of the institutionalized discrimination and for majority students to receive the benefits from diverse classrooms.[1] Opponents of racial preferences assert that race-conscious admissions are unfair and may actually be damaging to the intended beneficiaries by placing them at institutions where they are unlikely to succeed.[2]

Recently the controversy over race-conscious admission policies has increasingly moved from a normative to a positive perspective. On one front, several papers attempted to empirically examine the educational benefits of attending racially diverse colleges. For example, Black, Daniels and Smith (2001) found a positive relationship between the proportion of blacks in the college attended and the post-graduate earnings in the National Longitudinal Survey of Youth; Arcidiacono and Vigdor (2010), using information on graduates of 30 selective universities in the College and Beyond data, found only weak evidence of any relationship between collegiate racial composition and the post-graduation outcomes of white or Asian students.[3] Duncan *et. al.* (2006), exploiting conditionally random roommate assignment at one large public university, found that cross-racial exposure influences individual attitudes and friendship patterns.

A second front, spurred by the provocative article of Sander (2004) and followed up by Ayres and Brooks (2005), Ho (2005), Chambers *et. al.* (2005), Barnes (2007) and Rothstein and Yoon (2008), attempts to empirically examine whether the effects of affirmative action policies on the intended beneficiaries is positive or negative. These papers essentially test for the so-called "*mismatch hypothesis*," i.e. whether the outcomes of minority students might have been worsened as a result of attending a selective university relative to attending a less selective school.

But even if some of the *ex post* outcomes for minority students are worse under affirmative action, it still may be the case that minority students are better off *ex ante* under affirmative action. To illustrate this point, suppose that one can convincingly establish that blacks are less likely to pass bar exams after attending an elite law school. Does this necessarily mean that blacks are worse off in an *ex ante* expected utility sense? If attending an elite university also makes it possible for blacks to be high-profile judges, and if the outcome of being a high-profile judge is valued by blacks much higher than just passing the bar exam, blacks could still be better off *ex ante* under affirmative action. Alternatively, it is possible that elite universities may provide amenities to minority students that more than compensate for the worse outcome measures that are examined by the researcher, thus making the minority students better off *ex ante* in an expected utility sense.

In this paper we take a new and complementary viewpoint to the above-mentioned literature on mismatch by bringing to the center the *rational enrollment decision* of the minority students who are offered

[1] In both *Regents of University of California v. Bakke* 438 U.S. 265 (1978) and more recently in *Grutter v. Bollinger*, 539 U.S. 306 (2003), the Supreme Court ruled that the educational benefits of a diverse student body is a compelling state interest that can justify using race in university admissions.

[2] See Kellough (2006) for a concise introduction to various arguments for and against affirmative action.

[3] Arcidiacono, Khan and Vigdor (2011) also suggest that affirmative action actually leads to *less* inter-racial interaction due to the exacerbation of the within-school gap of academic backgrounds between minority and majority students.

admission to a selective school, possibly due to affirmative action policies. The question we ask is, why would students be willing to enroll themselves at schools where they cannot succeed, as the mismatch hypothesis stipulates? Posing the question in this way immediately leads us to focus our attention to the role of asymmetric information. We show in a simple model that students' rational enrollment decisions have important implications regarding the possibility of *ex ante* mismatch due to affirmative actions. We argue that a *necessary condition* for *ex ante* mismatch to occur once we take into account the minority students' rational enrollment decisions is that the selective university has private information about the treatment effect (or value added) for the students.[4, 5] In the absence of the selective university's private information about the student's treatment effect (or value added) relative to attending a non-selective university, a minority student will choose to enroll in the selective university only if her treatment effect is positive, thus there is no room for *ex ante* mismatch to occur. However, when the selective university has private information about a minority student's treatment effect, it is possible that a minority student with a negative treatment effect may end up enrolling in the selective university if offered admission. The reason is simple: when the minority student decides whether to enroll in the selective university, she can only condition her decision on the event that her treatment is above its admission threshold. When the selective university's admission threshold for the minority student is negative, due to its desire to satisfy a diversity constraint for example, it may still be optimal for a minority student with a negative treatment effect to enroll as long as the expected treatment effect conditional on admission is higher than that from the non-selective university.

The central message from the simple model is that the presence of private information by the selective university regarding the students' treatment effect is a necessary condition for mismatch effect as a result of affirmative action. This simple observation leads to a novel test for a necessary condition for mismatch, which is a test for whether selective universities possess private information regarding the students they admit. We will emphasize that our test is only a test for *necessary condition*: if we find strong evidence for asymmetric information, it does not necessarily imply that mismatch has occurred. However, if we find no evidence for asymmetric information, then we can rule out mismatch without having to rely on strong unverifiable assumptions needed for the assessment of counterfactual outcomes.

We propose a non-parametric method to test for asymmetric information. We assume that the researcher has access to the elite university's assessment of the applicants, the applicants' subjective expectation about their post-enrollment performance in the selective university and their actual performance. We show that the celebrated Kotlarski (1967) theorem can be used to decompose the private information possessed by the applicant, the private information possessed by the selective university, and the information common to the selective university and the applicant but unobserved to the researcher.[6] We propose

[4]Our focus is on *ex ante* mismatch as this is what the university has control over. Clearly *ex post* mismatch can result from the uncertainty in students' educational outcomes alone.

[5]There is some evidence in the literature that students' expectations about their performance are inaccurate and are updated over time. Stinebrickner and Stinebrickner (2008) have information at multiple points during the student's college career from Berea college. They find strong evidence that students update their expectations over time and make decisions (such as the decision to drop out) based upon the new information they receive through their grades.

[6]Kotlarski theorem has been applied in the economics literature in Cunha, Heckman and Navarro (2005) and Krasnokutskaya (2009).

an estimation method after the Kotlarski decomposition to test whether the selective university possess private information important for the prediction of the students' actual post-enrollment outcomes.

We use data from the Campus Life and Learning Project (CLL), which surveys two recent consecutive cohorts of Duke University students before and during college. The survey was administered to all under-represented minorities in each of the cohorts as well as a random sample of whites and Asians. The CLL provides information about the participants' college expectations, social and family background, and satisfaction measures as well as providing confidential access to students' academic records. The key features of the data for our purposes are that we have Duke Admission Office's ranking of the applicants as well as the student's pre-enrollment expectations about their grade point averages. We also have a rich set of control variables about the students' family and high school background.

We test whether Duke's private information is important to outcomes such as grade point average after conditioning on what is in the student's information set, including the private information in the student's expected grade point average. Not only is Duke's private information important for both grades and graduation rates even after conditioning on the student's information set, but we also find that students have virtually no private information on their probabilities of succeeding. That is, once we condition on Duke's information set, the student's expected grade point average is virtually uncorrelated with their grades. Duke's private information is also shown to affect both the subjective measures of students' satisfaction and their persistence in more difficult majors.

The remainder of the paper is structured as follows. In Section 2 we discuss the mismatch literature. In Section 3 we present a simple model of a selective university's admission problem with rational students to clarify the key concepts of mismatch in our framework, and illustrate that the selective university's private information is a necessary condition for mismatch to occur. In Section 4 we describe the Campus Life and Learning (CLL) Project data used in our application to test for private information. In Section 5 we provide some baseline regression results to provide some preliminary bounds on the importance of Duke and student private information in predicting the students' performance at Duke. In Section 6 we describe a non-parametric empirical method to identify private information and present our main empirical results. In Section 7 we show that the private information Duke has on students' future performance also affects outcomes besides just grade point average. In Section 8 we discuss three potential avenues to provide more conclusive evidence for mismatch and conclude.

2 Mismatch Literature

The mismatch literature to date has focused on comparing the "outcome" (e.g., GPA, bar passage, post-graduate earnings etc.) of the minority students enrolled in elite universities relative to the corresponding *counterfactual* outcome when these minority students attend less selective universities. As well summarized in Rothstein and Yoon (2008), the papers differ in how the counterfactual outcomes are assessed. For example, Sander (2004) first used a comparison of black and white students with the same *observable* credentials, who typically attend different law schools because of affirmative action, to estimate a negative effect of selectivity on law school grades; he then included both selectivity and grades in a regression for graduation and bar passage where he found that both selectivity and grades have positive coefficient, with the latter much larger than the former.[7] Combining these two findings, he concluded that, on net, preferences in law school admission in favor of black students depressed black outcomes because such preferences led black students into more selective schools, lowering their law school grades, which swamps the positive effective of attending a selective school on their graduation and passing the bar.

Ayres and Brooks (2005), Ho (2005), Chambers *et. al.* (2005) and Barnes (2007), however, used versions of *selective-nonselective comparison,* i.e., comparing students of the same race and same observable admission credentials who attend more- and less-selective schools to assess whether attending more selective schools has negative effects.[8] All strategies used above to assess the counterfactual outcome are likely to yield biased estimates when there are applicant characteristics that are considered in admission but *unobserved* by researchers. For example, the selective-unselective comparison used by Ayres and Brooks (2005), Ho (2005), Chambers *et. al.* (2005) and Barnes (2007) are likely to underestimate mismatch effect because those who are admitted to more selective schools are likely to have better unobserved credentials.[9] In contrast, Sanders (2004), by attributing the blacks' lower grades in selective schools to school selectivity instead of potential unobserved credentials, is likely to overstate the mismatch effect.

Finally, Rothstein and Yoon (2008) used both the selective-unselective and the black-white comparisons to provide bounds for the mismatch effect in law school. They find no evidence of mismatch effects on any students' employment outcomes and on the graduation or bar passage rates of black students with moderate or strong entering credentials, a group that makes up 25% of the sample. However, they could not conclusively find effects for the bottom 75% of the distribution due to not having enough whites with similar credentials. In the spirit of our approach, we argue that the success of the top 25% is *necessary* for mismatch to occur if blacks at least know the overall relationship between credentials and success while only have expectations on their own credentials. Namely, if all blacks were mismatched then there would be no scope for students making rational decisions to attend schools where they were mismatched: there has to be some non-mismatched black students in order for rational mismatch to occur.

To summarize, the existing literature on the mismatch effect differs in the empirical strategy used to assess the counterfactual outcome of minority students attending less selective universities; and the evidence is mixed. We want to recast the mismatch problem in the context of rational decision making which, as show in the next section, points us towards examining whether selective universities have private information on the future success of their students.

---

[7] Loury and Garman (1995) appears to be the predecessor of the "mismatch" literature. They found that college selectivity and performance at college both have significant effects on earnings. The earnings gain by black students from attending selective colleges are offset by worse college performance for those Black students whose own SAT scores are significantly below the median of the college they attended, i.e. those "mismatched" blacks.

[8] Barnes (2007) also explains that the performance for black students may suffer in a selective school both because of mismatch, i.e., they are over-placed in such selective schools, or because there are race-based barriers to effective learning in selective schools.

[9] Dale and Krueger (2002) proposed and applied a strategy to control for the unobservable credentials in estimating the treatment effect of attending highly selective colleges by comparing students attending highly selective colleges with others admitted to these schools but enrolled elsewhere. Ayres and Brooks (2005) and Sanders (2005b) also attempted to approximately apply the Dale and Krueger strategy by comparing law students who reported attending their first choice schools with those who reported attending their second choices because their first choices were too expensive or too far from home.

3 A Simple Model

In this section, we present a simple model to clarify the notion of *ex ante* mismatch, and show that a necessary condition for *ex ante* mismatch to occur when students make rational enrollment decisions is that the elite university possesses private information regarding the students' treatment effect.

Consider two universities that differ in selectiveness. For convenience, suppose that only one university is selective, which we refer to as the elite university. The elite university has an enrollment capacity C; but the non-selective university, which essentially encompasses all the other options for the students in our model, does not have a capacity constraint.

Students belong to one of two racial groups, referred to as "White (w)" and "Black (b)." The total number of race-r applicants is given by N_r for $r \in \{w, b\}$. Let $T_r \in R$ denote the "treatment effect" of a student with race-r from attending the elite university. The "treatment effect" measures the difference in a student's outcome from attending the elite university instead of her second option (which in this model is the non-elite university). Importantly, this treatment effect is determined by the quality of matching between the student's own characteristics and the university's characteristics. To the extent that the non-elite university is better suited to some students, T_r could be negative. In the population of race r students, T_r is distributed according to a continuous CDF F_r.

Let the payoff that the elite university receives from admitting a race-r student with treatment effect T_r be denoted by $\pi_U(T_r, r; \lambda)$, where $\lambda \geq 0$ indexes the elite university's concerns for diversity with $\lambda = 0$ denoting no diversity concern and higher λ indicating stronger diversity concerns.[10] The elite university's admission decision is denoted by $a_\lambda(T_r, r) \in \{0, 1\}$ where $a_\lambda(T_r, r) = 1$ indicates admission and $a_\lambda(T_r, r) = 0$ otherwise. Using the above notation, the objective of the elite university is:

$$\max_{\{a_\lambda(\cdot)\}} \sum_{r \in \{w,b\}} N_r \int \pi_U(T_r, r; \lambda)\, a_\lambda(T_r, r)\, dF_r(T_r)$$

subject to two constraints. The first is the enrollment *capacity* constraint, which is given by:

$$\sum_{r \in \{w,b\}} N_r \int a_\lambda(T_r, r)\, dF_r(T_r) \leq C.$$

The second constraint is the *rational enrollment* constraint by the students. However, the expression of this constraint depends on the information of the student:

- **Symmetric Information Case:** If the student's treatment effect T_r is symmetrically known between the student and the elite university, then the rational enrollment constraint is simply:

$$T_r \geq 0 \text{ for } r \in \{w, b\}. \tag{1}$$

  That is, no student will enroll in the elite university if she knows that her treatment effect is negative.

[10] The exact form of the elite university's payoff function is not important for our argument. For example, it can be a weighted average of the total treatment effects and a measure of racial diversity of the enrolled students, which is the case studied in detail in our working paper, Arcidiacono *et. al.* (2009).

- **Asymmetric Information Case:** If the elite university knows about T_r but the student does not, then the student's rational enrollment constraint is

$$E\left[T_r | T_r \in \Omega_{\lambda,r}\right] \geq 0 \tag{2}$$

where

$$\Omega_{\lambda,r} \equiv \{T_r : a_\lambda (T_r, r) = 1\} \text{ for } r \in \{w, b\}$$

is the *admission set* for race-r students when the elite university has diversity concern λ. That is, the student will enroll in the elite university only if her expected treatment effect, conditional on being admitted, is positive.

The Effect of Affirmative Action on the Admission Sets. We capture the effect of affirmative action policies on the admission sets with the following assumption: the optimal admission sets of the elite university are such that, if $\lambda \geq \lambda'$, then

$$\Omega_{\lambda,w} \subseteq \Omega_{\lambda',w} \text{ and } \Omega_{\lambda,b} \supseteq \Omega_{\lambda',b}.$$

That is, the set of admitted white students (weakly) shrinks and the set of admitted black students (weakly) expands as the elite university's diversity concern increases (i.e. when λ increases).[11] Of course, the optimal admission sets also depend on the enrollment capacity C of the elite university.

Note that under symmetric information, the rational enrollment constraint (1) ensures that no black student with negative treatment effect will enroll in the elite universities, even if they were admitted due to affirmative action. However, if T_r is the private information of the elite university, then the rational enrollment constraint (2) can be satisfied even for some students whose T_r is actually negative, resulting in what we will refer to as "local mismatch." Formally, we say that affirmative action admission policy by the elite university leads to a *local mismatch effect* for blacks if *some* black students with negative treatment effects are admitted and enroll; that is, if there exists $T_b \in \Omega_{\lambda,b}$ such that $T_b < 0$ but (2) is satisfied.

A stronger notion of mismatch is "global mismatch," which happens when blacks as *a group* are made worse off as a result of affirmative action policies by the elite university. To state this more formally, define

$$\Phi_r(\lambda) = \int_{\Omega_{\lambda,r}} T_r dF_r(T_r).$$

as the total treatment effect for race-r students when the elite university's diversity concern is λ. The total treatment effect for blacks without affirmative action is simply $\Phi_b(0)$. We say that affirmative action admission policy by the elite university leads to a *global mismatch effect* for blacks if black students as a whole are made worse off in expectation, i.e., $\Phi_b(\lambda) < \Phi_b(0)$ for some λ.[12] Clearly, global mismatch

[11] In Arcidiacono *et. al.* (2009), we provide a specific specification for $\pi_U(\cdot)$ that delivers such predictions. But these predictions are consistent with many alternative models.

[12] In Arcidiacono *et. al.* (2009), we assumed that the objective function for the elite university is to maximize the total treatment effect of enrolled students. Under this specification, we fully characterize the admission sets and show that global mismatch can indeed exist while the rational enrollment constraints are respected. Specifically, we show that, in the presence of the capacity constraints, $\Phi_b(\lambda)$ increases in λ when λ is small, but decreases with λ when λ is large. When λ is sufficiently large, $\Phi_b(\lambda)$ becomes lower than $\Phi_b(0)$, implying global mismatch.

implies local mismatch. It is also useful to note that the existing mismatch literature reviewed in Section 2 examines only global mismatch effect.

Because both the local and global notions of mismatch require that the admission set for blacks, $\Omega_{\lambda,b}$, includes black students with negative treatment effects, and we know that students with negative treatment effects will choose to attend the elite university only when they are not fully knowledgeable about their treatment effects, we conclude that a *necessary* condition for mismatch to occur is that the elite university has private information regarding the students' treatment effects. Thus, we have shown that a necessary condition for either local or global mismatch to result from affirmative action admission policy is that the elite university has private information about the students' treatment effects.

The argument above can be substantially generalized to situations in which the students may also have private information. Denote $\mathbf{X}_S$ and $\mathbf{X}_U$ respectively as the list of variables observed by the student and the university that will affect their payoffs if the student attends the elite university. Let $u_S(\mathbf{X}_S, \mathbf{X}_U)$ be the student's payoff from attending the elite university (relative to the second option). Again consider two cases. The first is the case that the elite university does not have private information, i.e., when $\mathbf{X}_U$ is a sub-list of $\mathbf{X}_S$. Then the student's rational enrollment constraint is

$$E\left[u_S(\mathbf{X}_S, \mathbf{X}_U)\,|\mathbf{X}_S, a_\lambda(\mathbf{X}_U) = 1\right] \geq 0. \tag{3}$$

But since $\mathbf{X}_U$ is a sub-list of $\mathbf{X}_S$, we have $E\left[u_S(\mathbf{X}_S, \mathbf{X}_U)\,|\mathbf{x}_S, a_\lambda(\mathbf{X}_U) = 1\right] = E\left[u_S(\mathbf{X}_S)\,|\mathbf{X}_S\right]$. Thus (3) ensures that the blacks will never be made worse off *ex ante* when the elite university's admission set for them expands as a result of affirmative action. The second case is when the elite university has private information, i.e., when $\mathbf{X}_S$ contains variables that are not in $\mathbf{X}_U$. In this case, the rational enrollment constraint for the student, i.e., $E\left[u_S(\mathbf{X}_S, \mathbf{X}_U)\,|\mathbf{X}_S, a_\lambda(\mathbf{X}_U) = 1\right] \geq 0$, can be consistent with events where $u_S(\mathbf{x}_S, \mathbf{x}_U) < 0$ for some realizations of $\mathbf{X}_S = \mathbf{x}_S$ and $\mathbf{X}_U = \mathbf{x}_U$, i.e., events with *ex ante* local mismatch.

Finally, our simple theoretical framework also points toward an alternative test for the presence of mismatch. Specifically, our model clearly shows that local mismatch occurs whether the student's enrollment decisions would have been different if she has access to the elite university's private information about her treatment effect. We discuss several potential approaches to implement this novel test of mismatch in Section 8.

4 The Campus Life and Learning (CLL) Project Data

In Section 3, we argued that once we take into account the students' rational matriculation decisions, a *necessary* condition for either local or global mismatch to arise is that the elite university has private information about the students' treatment effects. In our empirical section, we propose tests for private information by the elite university. If our tests reject the presence of private information by the elite university, then we can conclude that mismatch does not arise as a result of affirmative action admission policies; however, if we detect private information, it is *not sufficient* to establish that mismatch occurred.

In this section, we describe data from the Campus Life and Learning Project (CLL) at Duke University that will allow us to test whether or not Duke has private information regarding the future success of their students.[13] CLL is a multi-year prospective panel study of consecutive cohorts of students enrolled at Duke University in 2001 and 2002 (graduating classes of 2005 and 2006).[14] The target population of the CLL project included all undergraduate students in Duke's Trinity College of Arts & Sciences and Pratt School of Engineering. Using the students' self-reported racial ethnic group from their Duke Admissions application form, the sampling design randomly selected about 356 and 246 white students from the 2001 and 2002 cohorts respectively, all black and Latino students, about two thirds of Asian students and about one third of Bi-/Multi-racial students in each cohort. The final design across both cohorts contains a total of 1536 students, including 602 white, 290 Asian, 340 black, 237 Latino and 67 Bi-/Multi-racial students.

Each cohort was surveyed via mail in the summer before their initial enrollment at Duke, in which they were also asked to sign an informed consent document, as well as given option of providing confidential access to their student information records at Duke. About 78 percent of sample members ($n = 1185$) completed the pre-college mail questionnaire; with 91 percent of these respondents providing signed release of their institutional records for the study. In the Spring semester of the first, second and fourth college year, each cohort was again surveyed by mail.[15] However, response rates declined in the years following enrollment with 71, 65 and 59 percent responding in the first, second and fourth years of college, respectively.[16]

The pre-college survey provides detailed measurement of the students' social and family background, prior school experiences, social networks, and expectations of their college performance. In particular, students were asked:[17]

"What do you realistically expect will be your cumulative GPA at Duke after your first year?"

We can then relate this measure to the student's actual first year grade point average (GPA). The in-college surveys contain data on social networks, performance attributions, choice of major, residential and social life, perception of campus climate and plans for the future.

For those who released access to their institutional records, we also have information about their grades, graduation outcomes, test scores (SAT and ACT) and financial aid and support. Further, we have the Duke Admission Officers' rankings of their applications on six measures: achievement, curriculum, essay, personal qualities, recommendations and test scores. Each of these rankings are reported on a five point scale. It is these rankings coupled with student expected performance that will be used to disentangle what the student knows from what the institution knows about how well the student will perform in college.

Table 1 contains summary statistics for the key variables in the CLL data set by race. The first rows reveal that there is a substantial amount of variation in entering credentials among students of difference

[13] A description of the CLL Project and its survey instruments can be found at `http://www.soc.duke.edu/undergraduate/cll/`, where one can also find the reports by Bryant *et. al.* (2006, 2007).

[14] Duke is among the most selective national universities with about 6,000 undergraduate students. Duke's acceptance rate for its regular applications is typically less than 20 percent.

[15] The survey was not conducted in the third year as many Duke students study abroad during that year.

[16] In the Appendix we examine who attrits and test for non-response bias.

[17] Asking students about the probabilities associated with different grade outcomes (as opposed to just the mean) would have been particularly advantageous for other purposes. Namely, those individuals who are less certain have more potential to be mismatched as these individuals would benefit the most from more information.

Table 1: Summary Statistics of Key Variables by Race.

Variable	White ($N = 419$)		Black ($N = 174$)		Asian ($N = 178$)		Latino ($N = 169$)	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Duke Admission Office Evaluations:								
Achievement	4.34	0.87	3.75	0.80	4.67	0.58	4.13	0.81
Curriculum	4.71	0.56	4.46	0.62	4.91	0.37	4.72	0.50
Essay	3.52	0.61	3.26	0.48	3.58	0.66	3.31	0.52
Personal Qualities	3.57	0.64	3.34	0.51	3.52	0.63	3.30	0.51
Recommendations	3.97	0.68	3.55	0.60	4.06	0.57	3.55	0.57
Test Scores	3.69	1.18	2.09	1.04	4.10	1.13	2.79	1.23
Male	0.52	0.50	0.28	0.45	0.51	0.50	0.51	0.50
Student's Expected First Year GPA	3.51	0.31	3.44	0.34	3.67	0.28	3.53	0.31
Actual First Year Cum. GPA	3.33	0.46	2.90	0.48	3.40	0.42	3.13	0.46
Graduate Early or On Time	0.90	0.30	0.88	0.33	0.92	0.27	0.87	0.34
Private High School	0.31	0.46	0.25	0.44	0.25	0.44	0.39	0.49
SAT (Math + Verbal)	1417	100	1281	93	1464	90	1349	102
Family Income Less than $49,999	0.10	0.30	0.32	0.47	0.19	0.39	0.22	0.42
Family Income $50K to $99,999	0.19	0.40	0.30	0.46	0.24	0.43	0.23	0.42
Family Income Above $100K	0.71	0.46	0.37	0.49	0.57	0.50	0.54	0.50
Father Education BA or Higher	0.91	0.29	0.63	0.48	0.90	0.30	0.78	0.42
Mother Education BA or Higher	0.82	0.39	0.61	0.49	0.74	0.44	0.70	0.46

races. Asians and Whites tend to have higher evaluations by Duke Admission Officers in all six categories than black and Latino students, with test score showing by far the largest gap. Despite these differences in credentials, black and white students have quite similar expectations about their GPA during their first year in college (3.51 for whites and 3.44 for blacks); a t-test *cannot* reject the null hypothesis of equal means.

However, Table 1 shows that there is a significant racial difference in the actual first year cumulative GPAs. The actual GPA for blacks is on average 2.90, in contrast to that for whites (3.33) and for Asians (3.40). In fact, a t-test rejects the null hypothesis of equal means. Notice that, for all races, the students' actual first year GPAs are on average lower than their expected GPAs. This suggests that all students have over-optimistic expectations. However, this optimism bias is much stronger for black (0.54) and Latino (0.4) students than for white (0.18) and Asian (0.27) students. Again, a t-test rejects the null hypothesis of equality of means. Note that these overly optimistic beliefs may be consistent with rational beliefs in the population as a whole, including those who chose not to attend Duke. There may be a 'winner's curse' whereby individuals who receive the best signals are the ones who choose to enroll. Hence, even if on average individuals are receiving signals centered around the correct mean, those who chose to attend will have beliefs that are on average above the actual average of first year GPAs.

Of course, part of the actual GPA differences across races are predicted by observable differences across races in their entering credentials. For example, Table 1 shows Asians and whites have substantially higher (more than one standard deviation) SAT scores than Latino and black students. Average family income for Black students tend to be lower than Asians and Latinos, which in turn are lower than the whites. The parents of white students tend to have higher educational attainment than blacks.

The key question is then, why do the black and Latino students suffer a worse bias in their expectation about their academic performance at Duke? Does Duke Admission Office's evaluation of their application contain valuable information that would have been useful in help these students form more realistic expectations? If the black and Latino students were able to form more realistic expectations about their academic performance at Duke, would they have reconsidered their decisions to enroll at Duke? These are the key empirical questions related to the mismatch hypothesis.

5 Baseline Regressions

While the CLL data set has the advantage of reporting both information from the students regarding their expected grades and information from Duke regarding the ranking of the applicant, disentangling Duke's private information from what the student knows is challenging. We begin by running some baseline regressions which may *bound* the amount of private information by the student and Duke regarding the student's future performance.

We begin by examining the difference between the student's expected GPA for their freshman year, ExpGPA, and their actual cumulative GPA for their freshman year, GPA. Specifically we see how forecastable this difference is with variables the student should know the effects of, such as their race and SAT scores. Let $\mathbf{Z}$ denote this set of variables. We then add variables the student might only have partial information about such as Duke's ranking of the student (DukeEv). The forecast error for student i is

Table 2: The Components of Students' Forecasting Error.

Variable	(1)	(2)
Constant	-0.256**	-0.883***
	(0.046)	(0.370)
Male	-0.120***	-0.093***
	(0.037)	(0.036)
Black	-0.131**	-0.110*
	(0.060)	(0.063)
White	0.144***	0.118**
	(0.048)	(0.051)
Asian	-0.010	-0.051
	(0.057)	(0.059)
Adjusted SAT	0.106***	0.061***
	(0.022)	(0.023)
Controls for Duke Eval?	No	Yes
R^2	0.088	0.148

Notes: Dependent variables is (GPA−ExpGPA); N = 938. Adjusted SAT is the SAT score normalized to have zero mean and a standard deviation of one. The coefficients on the Duke evaluation rankings are reported in Column 1 of Table A.1 in the Appendix. *, ** and *** indicate that the coefficient is significant at 10%, 5% and 1% respectively.

then:

$$\mathrm{GPA}_i - \mathrm{ExpGPA}_i = \mathbf{Z}_i\boldsymbol{\alpha}_1 + \epsilon_{i1} \tag{4}$$

$$\mathrm{GPA}_i - \mathrm{ExpGPA}_i = \mathbf{Z}_i\boldsymbol{\alpha}_2 + \mathrm{DukeEv}_i\boldsymbol{\beta}_2 + \epsilon_{i2} \tag{5}$$

where the ϵ's are the projection errors.

Results from regressions (4) and (5) are reported in Table 2.[18] For ease of interpretation, we adjusted the SAT score such that it has zero mean and a standard deviation of one. Column 1 of Table 2 shows that students underestimate the relationship between their SAT score and performance. Virtually all groups on average over-predict their performance, with the one exception being white females with SAT scores more than one standard deviation above the mean. As expected given the descriptive statistics in Table 1, blacks significantly overestimate their performance relative to the other racial groups. Further, the variance of (GPA-ExpGPA) is 0.27 and is actually *higher* than the variance of first year GPA, which is 0.22. Clearly if we assume that the student's only information about their future performance is captured in their expected GPA, then there is a lot of information that the university possesses and a significant amount of noise in expected GPA. Moreover, the statistically significant coefficient estimates on the Adjusted GAP and race variables indicates that the student does not accurately know how these characteristics translate into their

[18] We have also experimented with specifications that include high school characteristics (private, public, religious etc.) in the regressions. Their coefficients are not significant and they neither affect the other coefficient estimates, nor significantly increase the R^2 of the regressions.

Table 3: Baseline Tests of Private Information.

Variable	(1)	(2)	(3)
Constant	3.309***	2.792***	1.828***
	(0.039)	(0.187)	(0.325)
Male	-0.080**	-0.086***	-0.043
	(0.037)	(0.036)	(0.029)
Black	-0.191***	-0.182***	-0.158***
	(0.051)	(0.052)	(0.053)
White	0.047	0.061	0.030
	(0.041)	(0.041)	(0.042)
Asian	0.037	0.030	-0.010
	(0.049)	(0.049)	(0.049)
Adjusted SAT	0.178***	0.167***	0.103***
	(0.018)	(0.018)	(0.017)
Expected GPA		0.145***	0.050
		(0.052)	(0.047)
Controls for Duke Eval?	No	No	Yes
R^2	0.188	0.196	0.321

Notes: Dependent variables is GPA; N = 938. Adjusted SAT is the SAT score normalized to have zero mean and a standard deviation of one. The coefficients on the Duke evaluation rankings are reported in Column 2 of Table A.1 in the Appendix. *, ** and *** indicate that the coefficient is significant at 10%, 5% and 1% respectively.

future performance. Duke, however, is likely to know more accurately about the relationship between characteristics and performance. Column 2 in Table 2 adds controls for Duke's evaluation rankings of the students. The R^2 increases from 0.088 to 0.148 when we include Duke's rankings, again suggesting that Duke has either private information about the student's future performance or in how information known to both the student and Duke translates into future performance.

The expected GPA of the student, however, may not reflect the student's true information set. We now test whether the university has private information under a more restrictive setting. Namely, we assume students know how their SAT scores and other demographics translate into future performance. The information set for both the student and the university then contains this common observed information plus common information that is unobserved to the researcher. Under these assumptions, running the regression

$$\text{GPA}_i = \mathbf{Z}_i \boldsymbol{\alpha}_3 + \epsilon_{i3}, \tag{6}$$

and calculating the R^2 then leads to a lower bound on the amount of common information that the student and the university have regarding the student's future performance as it does not include common unobserved information. Results from this regression are reported in Column 1 of Table 3.[19] Close to 19

[19]Adding up to the fourth-order polynomial of the student's expected GPA, ExpGPA, barely increases the R^2s. Also,

percent of the variation in grades can be explained by these observables. Comparing this result with that in Column 1 of Table 2 suggests that students underestimate the relationship between SAT scores and performance by more than 50 percent.

To this baseline regression, we add the student's expected GPA:

$$\text{GPA}_i = \mathbf{Z}_i \boldsymbol{\alpha}_4 + \text{ExpGPA}_i \delta_4 + \epsilon_{i4}. \tag{7}$$

The difference in R^2 between (6) and (7) should provide an *upper* bound on the student's private information as it includes not only the student's private information, but also common unobserved information that is correlated with student's private information. These results are reported in Column 2 of Table 3. The differences in R^2 between Column 2 and Column 1 in Table 3 indicates that including the expected GPA of the student increases the R^2 by less than 0.01, which provides an upper bound of the importance of student's private information.

Finally, we add Duke's evaluation rankings of the students:

$$\text{GPA}_i = \mathbf{Z}_i \boldsymbol{\alpha}_5 + \text{ExpGPA}_i \delta_5 + \text{DukeEv}_i \boldsymbol{\beta}_5 + \epsilon_{i5}. \tag{8}$$

The difference in R^2 between (7) and (8) should provide a *lower* bound on the importance of Duke's private information. Notice from Column 3 that controlling for Duke's rankings increase the R^2 by more than 0.12, again suggesting substantial Duke private information.[20] Note that this still leaves two-thirds of the variation in GPA unexplained, perhaps due to course selection and shocks to how students respond to college life.

A drawback of this empirical strategy is that we do not fully observe common information; as a consequence, there is no guarantee that the reported bounds are in fact the real ones. Further, measurement error in the expected GPA variable may be contaminating the results. In the following section, we implement a different strategy that overcomes these limitations; and it allows us to identify private and common information in order to perform a more accurate variance decomposition analysis. However, it is worth mentioning that both strategies provide surprisingly similar results.

6 Non-Parametric Identification of Private Information

There is a large existing economics literature that tests for asymmetric information particularly for adverse selection in the empirical analysis of a variety of insurance markets.[21] Most of these papers test whether the data supports a positive association between insurance coverage and *ex post* risk occurrence,

---

adding additional variables such as family income and mother's education had little effect on the R^2 but did lead to some attrition. These results are available from the authors upon request.

[20]One can also reverse the order of the regressions such that we first control for Duke's evaluation rankings and then add student's expected GPA. The addition of the student's expected GPA in this order increases the R^2 by only 0.001, with an insignificant coefficient estimate on expected GPA.

[21]The rapidly growing literature includes Cawley and Philipson (1999) for life insurance market, Chiappori and Salanie (2000) for auto insurance market, Cardon and Hendel (2001) for health insurance market, Finkelstein and Poterba (2004) for annuity market, Finkelstein and McGarry (2006) for long-term care insurance market and Fang, Keane and Silverman (2008) for Medigap insurance market.

a robust prediction of the classical models of insurance market developed by Arrow (1963), Pauly (1974), Rothschild and Stiglitz (1976) and Wilson (1977).[22]

Our setting substantially differs from the insurance market setting studied in the existing literature. The empirical insurance literature assumes that private information is possessed by one-side of the market – the potentially insured – and it is manifested through their insurance purchase and their *ex post* risk occurrence. In our setting, there is presumably private information about the treatment effect by both the student and the university. Moreover, the empirical insurance literature typically assumes either to have access to observations for individuals with and without insurance and their risk realizations, or to have access to observations for individuals with different amount of coverage and their risk realizations. In our setting, if a student does not attend the elite university, we will not observe the student's outcome had he attended it; or if the student attends the elite university, we will not observe the student's outcome had he not attended. For these reasons, we describe below a new empirical strategy to identify private information in our setting.

6.1 Available Data and Assumptions

As we mention in section 3, we have data about an **observed student outcome** Y (i.e. first year cumulative GPA, denoted by GPA). Conceptually, we assume that Y is a linear function of X_U, X_S and X_C where X_U denotes the unobserved university's private information about student performance, X_S denotes the unobserved student's private information and X_C denotes the information that is common to both students and the university but unobserved by the researcher. Of course, we can also include a set of variables $\mathbf{Z}$ that are common information to the university and the students and are observed by researchers, such as observed family and high school characteristics. For simplicity, we will ignore $\mathbf{Z}$ in the discussion below.

Specifically, suppose that

$$Y = X_C\gamma_C + X_U\gamma_U + X_S\gamma_S + \varepsilon, \tag{9}$$

where ε is noise. By construction, and thus without loss of generality, we assume that X_C, X_U, X_S and ε are independent.

Suppose that we also have access to *two additional variables*: a variable, denoted by W_U, that measures the selective university's assessment about the student's treatment effect given its private knowledge about the match between the student and the university X_U, as well as the common information X_C; and another variable denoted by W_C that measures the student's own performance expectation in the selective university given the common information X_C and her own private information X_S.[23] We assume that (W_U, W_S) are related to X_C, X_U and X_S as follows:

$$\begin{aligned} W_U &= X_C + X_U, & (10)\\ W_S &= X_C + X_S. & (11)\end{aligned}$$

To summarize, suppose that we observe a data set consisting $\{W_U, W_S, Y\}$ and assume that there exists independent variables X_C, X_U, X_S and ε such that $\{W_U, W_S, Y\}$ are generated by (9)-(11). The

[22]See Chiappori *et. al.* (2006) for a general derivation of the positive association property.

[23]We will describe in Subsection 6.2 below the empirical counterparts of W_U and W_S in our setting.

question we are interested in is, how do we estimate the coefficients γ_C, γ_U and γ_S, and/or decompose the importance of common information X_C, student private information X_S, university private information X_U and noise ε in explaining the variation of Y in the data?

The key to answering this question is that, even though X_C, X_S, and X_U are unobserved, it is possible to recover their marginal distributions from the observed joint distribution of (W_U, W_S). This result is due to Kotlarski (1967):[24]

Theorem 1 ***(Kotlarski's Theorem)*** *Let X_C, X_U and X_S be three independent real-valued random variables. Suppose W_U and W_S are generated as in (10) and (11). Then the joint distribution of (W_U, W_S) determines the marginal distribution of X_C, X_U, X_S up to a change of the location as long as the characteristic function of (W_U, W_S) does not vanish (i.e., it does not turn into zero on any non-empty interval of the real line).*

The intuition for Kotlarski's Theorem can be heuristically conveyed from thinking about how all the moments of the three variables X_C, X_U and X_S can be recovered from the joint distribution of (W_U, W_S) under the assumption that X_C, X_U and X_S are independent. For example, the covariance of W_U and W_S tells us about the second moment of X_C. The covariance of W_U^2 and W_S tells us about the third moment of X_C, and so on.

The proof of the theorem also shows how the marginal distributions for X_C, X_S and X_U can be constructed. Once we have the marginal distributions of X_C, X_S, and X_U, we can generate data where the i-th observation is a triple $\{X_{Ci}, X_{Si}, X_{Ui}\}$. Equations (10) and (11) can then be used to obtain $\{W_{Ui}, W_{Si}\}$.

If we then could generate data on Y, we could use (9) to estimate γ_C, γ_U, and γ_S, as well as assess their importance in determining Y by performing a variance decomposition. Given the observed joint distribution of $\{Y, W_U, W_S\}$, we can use multiple imputation to generate Y_i. Details on how to recover the marginal distributions of of X_C, X_S, and X_U as well as how the multiple imputation was conducted can be found in the appendix.

Remarks. We have assumed in equation (9) that the student outcome Y is a linear function of X_C, X_U, X_S. This is for simplicity only. With the pseudo data sets we simulated in Step 3, we can also estimate Y as a nonlinear function of these variables, or even non-parametrically estimate their relations.

It is also worth noting in specification (10) and (11), we interpret X_U and X_S are respectively the true private information for the university and the student, and assume away noise in the measurement of the variables W_U and W_S. If instead the variables we extract in step 1 contain the true private information of the university and students contaminated by noise, then we will have, in step 4, a mismeasured independent variables in the regressions. This may bias our coefficient estimates for γ_U and γ_S downward, but when we do variance decomposition for Y, we should still be able to recover the importance of the true private information of the university and the student in explaining the variance of the outcome variable Y.

[24] The proof can also be found in Rao (1992, pp 7-8). Kotlarski theorem has been widely used in measurement error models in econometrics (e.g., Li and Vuong 1998). It has been applied elsewhere in economics, e.g., Cunha, Heckman and Navarro (2005) used it to distinguish uncertainty from heterogeneity in their analysis of life-cycle earnings, and Krasnokutskaya (2008) used it in the context of identifying and estimating auction models with unobserved auction heterogeneity.

Table 4: Regressing GPA on Duke Private Information, Student Private Information and Common Information.

	(1)			(2)		
	$W_U = \text{GPA}_U, W_S = \text{GPA}_S$			$W_U = \text{GPA}_U, W_S = \text{ExpGPA}$		
	Coef.	Std. Err.	R^2	Coef.	Std. Err.	R^2
Duke Priv. Inf. (X_U)	1.070***	0.037	0.091	0.957***	0.023	0.235
Student Priv. Inf. (X_S)	0.066*	0.040	0.0004	0.037*	0.022	0.0005
Common Inf. (X_C)	0.993***	0.018	0.265	0.994***	0.044	0.081
Total			0.356			0.317

Notes: *, ** and *** indicate that the coefficient is significant at 10%, 5% and 1% respectively.

6.2 Data and Results

As we have already mentioned, it is necessary to have access to (at least) two variables $\{W_U, W_S\}$ in order to apply Kotlarski's decomposition. Here we provide the details of these variables in our empirical application.

W_U is specified as Duke's predicted first year GPA for the student, which we denote by GPA_U. Specifically, GPA_U is predicted student GPA from the estimated regression

$$\text{GPA}_i = \mathbf{Z}_i \boldsymbol{\alpha}_6 + \text{DukeEv}_i \boldsymbol{\beta}_6 + \epsilon_{i6},$$

where GPA_i denotes the actual first year GPA. Recall that $\mathbf{Z}_i$ are the observed SAT scores and demographics and DukeEv refers to the Duke ranking variables.

For W_S, we consider two alternative specifications. The first specification for W_S is the student's predicted GPA, which we denote by GPA_S, predicted from the estimated regression equation (7) from the previous section:

$$\text{GPA}_i = \mathbf{Z}_i \boldsymbol{\alpha}_4 + \text{ExpGPA}_i \beta_4 + \epsilon_{i4}.$$

This specification implies that students have an accurate idea about how to weigh each informational variable (e.g. SAT) when they predict their performance. The second specification for W_S is the expected GPA (ExpGPA) reported by the student before coming to Duke in the CLL survey. To the extent that the students may not properly weigh the effect of the observable variables on their actual GPA, as documented in Table 2, we will be attributing some of the students' wrong weighting on the importance of common information X_C to Duke private information.

Following the discussion in the previous section, we generated data on Y_i and the triple $\{X_{Ci}, X_{Ui}, X_{Si}\}$. Table 4 reports the variance decomposition of GPA following two different specifications for W_S as described above. Specification (1) assumes that students know how to weight the available information when they predict their performance; results show that Duke's private information explains 9.1 percent of the variance in the students' actual first year cumulative GPA; the student's private information explains no more than 0.1 percent and the common information 26.5 percent. Specification (2) allows that students may not know how to weigh the information, as a consequence, the fraction of the variance in GPA explained by Duke private information increases to 23.5 percent, that by common information declines to 8.1 percent, but the fraction explained by student private information remains about 0.1 percent.

It is worth noting how the R^2 changes depending on the specification. First the total R^2 in specification (2) is smaller than in specification (1), possibly due to the loss of valuable information when students do not correctly weigh the available information or to students reporting expected GPA with error. Second, there is an important change, similar in magnitudes but in opposite directions, of the proportion of the variance that could be explained by common information and Duke private information. This seems to suggest that the size of Duke private information not only depends on what information is not available to the students, but also how they weigh the information available to them in forecasting their performance at Duke.

6.3 Measurement Error in Student Expectations

Assuming that students are rational implies that coefficient on students' private information should be equal one; however as we can see from Table 4, this is not the case. One possible explanation to this discrepancy is that students may report with error their expected GPA. The attenuation bias from the measurement error might drive the small R^2 we found for the students private information reported in Table 4. However, if we assume that the discrepancy between estimated $\hat{\gamma}_S$ and the postulated value $\gamma_S = 1$ under rational expectations is due completely to measurement error, we can easily provide an upper bound of the variance of the student private information without measurement error.

In particular, note that the Kortlaski result allows us to separate student expected GPA and university expected GPA into three orthogonal components. Imbedded in X_u is then both the student's private information as well as their measurement error. Denote X_S^* as the student's true private information and ϵ_u as measurement error. Equation (9) then uses X_S^* rather than X_S:

$$Y = X_C\gamma_C + X_U\gamma_U + X_S^*\gamma_S + \varepsilon, \tag{12}$$

Assuming that ϵ_u is orthogonal to the common, university, and student information and is classical measurement error, we have:

$$\widehat{\gamma}_S = \gamma_S \frac{\text{Var}(X_S^*)}{\text{Var}(X_S)}$$

where $\text{Var}(X_S^*)$ is the variance of the student private information when it is purged of measurement error, and $\text{Var}(X_S)$ is the variance of X_S^* measured with error. Given that we know $\hat{\gamma}_S, \gamma_S$ (which is equal 1 under the rational expectation assumption) and $\text{Var}(X_S)$, then we can recover $\text{Var}(X_S^*)$, which will allow us to obtain an upper bound of the R^2 for student private information.

Once we correct for measurement error, the upper bounds of the R^2 for student private information under specifications (1) and (2) are respectively equal to 0.0068 and 0.0013 respectively; again, they are substantially smaller in magnitude than the private information possessed by Duke. The results obtained in this section are then quite similar to those obtained from the baseline regressions. Thus, the conclusion that Duke does possess private information that can predict the students' post-enrollment performance is robust to different empirical strategies including those that account for measurement error.

7 Information Sets and Student Outcomes

We have shown that Duke private information can help students to obtain more accurate expectations about their future performance. However, academic performance is an intermediate outcome and might not be the treatment that guides individual decision making. Therefore, this section will present evidence that links measures of Duke private information with more "treatment related" outcomes.

We focuses on two sets of outcomes. First, we consider changes in beliefs or decisions as a result of the new information, providing corroborating evidence of Duke's private information. Second, we consider the student's overall satisfaction with their Duke experience, including using information on whether students as seniors, knowing what they know now, would have made the decision to attend Duke. For both sets of outcomes, we show clear evidence that the new information conveyed in the individual's first year performance is a significant predictor of future outcomes.

7.1 Changes in Beliefs and Choices

In order to examine how new information changes beliefs and choices, we again partition information into what the student knows and what the university knows. Unfortunately, as our measures are discrete, the Kotlarski decomposition is not as useful here.[25] Recall that in section 6 freshman GPA was partitioned into three parts: student information, Duke information, and common information. Here the partitioning occurs only on student information and Duke information with Duke information constructed to be orthogonal to student information.

We work with two different assumptions on what is the student's information set. We begin by assuming that the students knows much more than what they report in their expected grade point average. Namely, as in equation (7), we regress actual grade point average on information that is available to the student plus their expected grade point average:

$$\mathrm{GPA}_i = \mathbf{Z}_i \boldsymbol{\alpha}_4 + \textsc{Exp}\mathrm{GPA}_i \delta_4 + \epsilon_{i4}. \tag{13}$$

The predicted values from this regression are used as the information set of the student. We then regress the residuals from (13) on Duke information, with the remaining error attributed to noise. Note that our measure of Duke information is then, by construction, orthogonal to our measure of student information and to the shock.

Our second measure of the student's information set is to regress actual grade point average solely on expected grade point average. We then create Duke's private information by regressing the residual on Duke's information, with the error from this second stage regression again attributed to noise. The second measure recognizes the fact that students may not know how to weigh certain information (e.g. SAT, gender and race) which should have already been taken into account when they reported their expected GPA. The remaining variation is considered to be the shock: information available neither to the student or Duke.

[25] Experimentation using this method for these discrete outcomes showed that the results were very sensitive to the tuning parameters. The disadvantage of not using the Korlarski decomposition is that we no longer can deal explicitly with measurement error. However, given how little measurement error affected the results in the previous section, we feel comfortable ignoring it here.

Table 5: Ordered Probit Estimates of Performance Relative to Expectations

	Information Set Specification I		Information Set Specification II	
	Coef.	Std. Err.	Coef.	Std. Err.
Shock	1.929***	0.128	1.928***	0.133
Student Information	0.725***	0.153	-0.481	0.453
Duke Information	1.479***	0.364	1.192***	0.187
R^2	0.1394		0.1436	

Notes: Dependent variable is the students' subjective satisfaction measured in their senior year; it takes one of five values with higher values associated with performance exceeding expectations (see text for details). N = 748. *, ** and *** indicate that the coefficient is significant at 10%, 5% and 1% respectively. Information set I assumes individuals weight their observed characteristics appropriately when forming their expected GPA's. Information set II allows students to use wrong weights when forming their expectations. See text for details.

We first relate these information sets to self-reported measures of whether the student exceeded or under-performed relative to their expectations. In particular, the CLL senior year data contain responses to the following question:

> "Which statement best describes how your academic achievement so far compares to your expectations for yourself when you arrived at Duke? 1. I am doing much better than expected; 2. I am doing a little better than expected; 3. I am doing as well as I expected; 4. I am doing a little worse than I expected; 5. I am doing much worse than I expected."

Table 5 presents ordered Probit results from regressing student response to the question above on the two sets of information variables. The left panel shows results when students are assumed to know the effects of observed variables like their SAT scores. The estimated coefficients of all three measures of information, that of the student, Duke, and the shock, are all positive and significant, implying that higher values for these variables makes it more likely that students report that they exceeded their expectations. If the student had the information in these variables ahead of time, then they should have had no effect on the outcome of whether the student performed better or worse than expected. That the coefficient on student information is positive and significant suggests that students were surprised about how their observed characteristics translated into their academic achievement.

Column 2 reports the results when we take the student's expected grade point average as their information. The estimated coefficient on student information changes from positive to negative but is not statistically significant. A negative sign is indicative of students having bad information and then realizing the information is bad when responding to whether they performed better or worse than they expected. The coefficient on Duke information is once again positive and significant, providing corroborating evidence that Duke has information about the student that may be useful for their decision-making.

We next see if Duke has information that can predict changes in subjective measures of student happiness. The CLL survey asked students during the year before coming to Duke and at the end of the Spring semester of their freshmen year to indicate the level of agreement with the following statement: *"On the*

Table 6: Change in Student Satisfaction and Information

	Information Set Specification I		Information Set Specification II	
	Coef.	Std. Err.	Coef.	Std. Err.
Shock	0.376***	0.114	0.369***	0.118
Student Information	0.114	0.171	-0.874*	0.510
Duke Information	0.400	0.307	0.350**	0.146
R^2	0.0081		0.0104	

Notes: Ordered Probit estimates. See text for details of the construction of the variables. N = 751 *, ** and *** indicate that the coefficient is significant at 10%, 5% and 1% respectively. Standard Errors were bootstrapped. Information set I assumes individuals weight their observed characteristics appropriately when forming their expected GPA's. Information set II allows students to use wrong weights when forming their expectations. See text for details.

whole, I am satisfied with myself". The possible answers are 5 categories that go from strongly disagree to strongly agree. We difference the two measures to get a measure of change in student satisfaction.

Ordered Probit estimates are reported in Table 6. These results mirror those in Table 5. Both positive shocks to grades and Duke information have positive effects on changes in student happiness. Further, the coefficient on student information moves from positive to negative as the information on the true relationship between grades and student characteristics is removed from the student's information set.

Finally, we examine how information affects major choice. Accurate predictions about persistence in certain majors constitute highly valuable information given that majoring in the humanities instead of the natural sciences may lead to significantly lower earnings. For instance, James *et. al.* (1989) argue that "...while sending your child to Harvard appears to be a good investment, sending him to your local state university to major engineering, to take a lots of math, and preferably to attain a high GPA, is an even better private investment." Both Loury and Garman (1995) and Arcidiacono (2004) document large differences in earnings across majors. We split majors into two categories: natural sciences/engineering and humanities/social sciences.[26] Table 7 shows major switching patterns across races, taking the initial major from the pre-college survey. Conditional on the initial major being in the natural science category, more black students actually switch their major than stay in the natural sciences. This is in contrast to whites where those who initially chose one of the natural sciences were two and a half time more likely to stay in the natural sciences than to switch.

Very few students decided to start majoring in humanities/social sciences but later change into the natural sciences, which may indicate that information is more important to success in these fields. Majors in the natural sciences or engineering are, in general, characterized by a more rigorous curricula (Babcock and Marks 2010) and grading policy (Babcock 2010, Johnson 2003). A direct consequence of these differences is considerable variation in the distribution of grades across departments. For instance, the CLL data show that the median grade in Chemistry courses at Duke University is a B while in English it is an A-. Grading disparities together with over-optimistic expectations about schooling performance constitute a possible

[26]See Section A.3 in the Appendix for the full list of majors in each category.

Table 7: Major Choice Patterns by Race

Initial Major	Final Major	White (%)	Black (%)	Asian (%)	Latino (%)
Natural Sci	Natural Sci	25.4	15.7	41.2	18.4
Natural Sci	Humanities/Soc Sci	10	22.2	9.1	15.1
Humanities/Soc Sci	Humanities/Soc Sci	30.6	41.6	20.9	36.3
Humanities/Soc Sci	Natural Sci	0.9	0.5	2.7	0.6
Don't know	Natural Sci	6.8	4.3	11.8	2.8
Don't know	Humanities/Soc Sci	26.3	15.7	14.4	28.8
Observations		468	185	187	179

Note: See the appendix for how majors were partitioned.

Table 8: Estimates of the Probability of Persisting in the Natural Sciences

	Information Set Specification I		Information Set Specification II	
	Coef.	Std. Err.	Coef.	Std. Err.
Shock	0.471***	0.147	0.481***	0.149
Student Information	1.467***	0.325	-0.551	0.827
Duke Information	1.931***	0.510	1.889***	0.306
Constant	-4.310***	1.075	2.444	2.764
R^2	0.0841		0.0886	

Notes: N = 279. Estimated on only those whose initial major was in the natural sciences or engineering *, ** and *** indicate that the coefficient is significant at 10%, 5% and 1% respectively. Standard Errors were bootstrapped. Information set I assumes individuals weight their observed characteristics appropriately when forming their expected GPA's. Information set II allows students to use wrong weights when forming their expectations. See text for details.

explanation for the intensive emigration patterns that are observed from hard science majors each year.

We then examine how information affects whether the individual persists in the natural sciences. Table 8 presents Probit regressions in which the dependent variable is a dummy that takes value 0 if a student reports as expected major natural sciences or engineering but later change into humanities or social sciences, and takes value 1 if the student remains in the natural sciences. The patterns in Tables 6 and 7 are again repeated here. The estimated coefficient student information again changes sign when the true relationship between grades and student characteristics is removed from the student's information set. The results show that higher values of Duke information make persistence more likely. Using the results from information set specification II, a one standard deviation increase in Duke information increases the probability of persisting in a natural science majors by around 8%.

In sum, Duke's private information on student performance is important beyond the student's grade point average, affecting both the student's subjective well-being as well as the likelihood of persisting in a difficult major. It is important to note, however, we have still only established that some of the necessary conditions for mismatch are met. In order for mismatch to exist, information must not only be important but the revelation of the information must change the institution the individual would have chosen to attend. Hence, even if individuals finds out they are not as academically strong as they thought, the information still may not have changed their decisions.

7.2 Information and Satisfaction

The previous section showed that new information affected the student's beliefs and choice of major. Here we see if the new information provided in freshmen grades has an impact on how the student values the Duke experience as a senior. In particular, the senior survey asked how satisfied the student was with their Duke experience and also with their choice of major. Of course the new information, if it was bad, could lead to less satisfaction regardless of the individual attended Duke. But one final question was asked that gets at this issue. Namely, students were asked whether, given the information they have now, they would have made the same decision to attend Duke. To the extent that the new information affects answers to this question, we have a direct measure of mismatch. The three survey items we use are then:

1. *On a scale of 0 to 10, your overall satisfaction with your Duke experience.*

2. *On a scale of 0 to 10, your overall satisfaction with your major.*

3. *Knowing what you know now, all things considered, wold you come to Duke if you were doing it all over again?*

Since these measures do not correspond to changes in behavior or beliefs, we use additional controls such as sex, race, and initial major. In addition, power is an issue here; so rather than separating out freshmen grades into three sources, we focus on the difference between the student's expected freshman GPA from their actual freshmen GPA. This is equivalent to constraining the coefficient on Duke information to be equal to the coefficient on the shock in the previous subsections. Finally, we also control for expected GPA directly to take into account that those who have higher expected GPA's may be more optimistic in general.

Results are presented in Table 9. Each cell corresponds to a separate regression with the rows giving the three dependent variables. The results indicate that the shock to a student's GPA does influence their satisfaction with both their school and their major: positive information is associated with higher levels of satisfaction. Even more importantly in terms of testing for mismatch, negative information is associated with regret about attending Duke.

8 Discussion and Conclusion

We have argued that once we take into account the students' rational enrollment decisions, mismatch in the sense that the intended beneficiary of affirmative action admission policies are made worse off could occur only if selective universities possess private information about students' post-enrollment treatment effects. This necessary condition for mismatch provides the basis for a new test. We propose an empirical methodology to test for private information in such a setting. The test is implemented using data

Table 9: The Relationship Between Freshmen GPA Shock and Satisfaction

Dependent Variable	Specification (1)	(2)	(3)
Overall Satisfaction	0.3288**	0.3041**	0.5376***
	(0.1409)	(0.1491)	(0.1580)
Overall Satisfaction in Major	0.2970*	0.2536	0.4657**
	(0.1610)	(0.1708)	(0.1950)
Would attend Duke again	0.0490*	0.0477	0.0678**
	(0.0273)	(0.0292)	(0.0331)
Controls for:			
Family Background and Major	No	Yes	Yes
Expected GPA	No	No	Yes

Notes: Controls for major refers to an indicator variable for majoring in engineering or natural sciences. Family background characteristics denotes gender and race.

from Campus Life and Learning Project (CLL) at Duke. The evidence shows that Duke does possess private information that is a statistically significant predictor of the students' post-enrollment academic performance.

However, even though we have shown substantial evidence that Duke does indeed possess private information about the student's future performance, we can not conclude that there is mismatch. We now propose three potential avenues that may lead to a more conclusive test of mismatch. All three proposed avenues are suggested by our simple theoretical framework in Section 3; namely, mismatch occurs only if the release of the private information about the student's treatment effect leads to a different enrollment decision.

The first potential avenue exploits a particular group of admitted students who are actually aware of their rankings by the admission officers, namely those who are admitted off the *waiting list*. Suppose that researchers have access to the complete admission list, including those admitted off the waitlist, as well as their subsequent matriculation decisions. Then to the extent that the students admitted off the waitlist and those admitted in the initial round but close to the initial-round admission margin in the admission officers rankings are similar, we can compare their enrollment rates to infer the presence of mismatch. If the enrollment rate for those admitted off the waitlist is smaller than that of similar students admitted in the initial round, then it is likely that the knowledge of the admission officers' ranking for those admitted off the waitlist leads to a change in the student's enrollment decision, which is evidence for mismatch.

The second potential avenue requires the cooperation of the Office of Admissions at a selective university. After the admission decisions are made, the Office of Admissions could randomly assign admitted minority students into two groups: the first group will receive the standard admission letter; and the second group will receive the standard admission letter together with information about average freshmen performance as well as the student's expected performance given the experience of past cohorts. Then observing that the enrollment rate for the second group is smaller than the first group will prove that the university's private information may have generated mismatch.

The third potential avenue to test for mismatch is to ask the admitted students two questions:

Q1. "What do you realistically expect will be your cumulative GPA at Duke after your first year?"

Q2. "Suppose your expected GPA at Duke was X. Would you still have chosen Duke?"

where X is filled in by the researcher as the student's predicted cumulative GPA based on the private information of the Office of Admissions. If a significant fraction of students responds "no" to Q2 , then there is mismatch.[27]

However, it is important to note that even if one cannot conclusively prove the existence of mismatch, evidence that a selective university possesses valuable *ex ante* information could be used in *preventing* mismatch. To the extent that a university with active affirmative action programs is concerned about potential mismatch, it suggests that releasing more information to their applicants about how the admission officers feel about their fit with the university will minimize possibilities for actual mismatch.[28] More transparency and more effective communication with the students, and possibly pre-enrollment sit-ins in college classrooms, etc. can help minority students enrolling in an elite university find out that they might have potentially been better off elsewhere.

A Appendix

In this appendix, we examine the CLL data for drop-out bias and non-response bias. We also report the coefficients for the Duke ranking measures from Tables 5 and 3 as well as how majors were partitioned.

A.1 Drop-out Bias and Non-Response Bias

The Registrar's Office data provided information on students who were not enrolled at the end semester in each survey year. Non-enrollment might occur for multiple reasons including academic or disciplinary probation, medical or personal leave of absence, dismissal or voluntary (including a small number of transfers) or involuntary withdrawal. Fewer than one percent of students ($n = 12$) were not enrolled at the end of the first year; about three percent by the end of the second year ($n = 48$) and just over five percent ($n = 81$) by the end of the senior year. We combined all of these reasons and tested for differences in selected admissions file information of those enrolled versus not enrolled at the end of each survey year. The test variables included racial ethnic group, SAT verbal and mathematics score, high school rank (where available), overall admission rating (a composite of five different measures), parental education, financial aid applicant, public-private non-religious-private religious high school and US citizenship. Of over 40 statistical tests, only two produced significant differences (with p-value less than 0.05): (1). At the end of the first year, dropouts had SAT-verbal scores of 734 versus 680 for non-dropouts; (2). by the end of the fourth year, those who had left college had an overall admissions rating of 46.0 (on a 0-60 scale) while

[27] Note that in both cases we would be testing for local mismatch rather than global mismatch.

[28] Releasing this information reveals not only information about the student's unobserved ability but also how the student's observed characteristics translate into success in a college environment.

those in college had an average rating of 49.7. No other differences were significant. We conclude that our data contain very little drop-out bias.

We conducted similar tests for respondents versus non-respondents for each wave for the same variable set plus college major (in 4 categories: engineering, natural science/mathematics, social science, humanities), whether or not the student was a legacy admission, and GPA in the semester previous to the survey semester. Seven variables show no significant differences or only a few small sporadic differences (one wave but not others), including racial ethnic category, high school rank, admissions rating, legacy, citizenship, financial aid applicant, and major group. However, several other variables show more systematic differences:

- Non-respondents at every wave have lower SAT scores (math: 9-15 points lower, roughly one-tenth to one-fifth of a standard deviation; verbal: 18-22 points lower, roughly one-third of a standard deviation).
- Non-respondents have slightly better educated parents at waves one and three, but not waves two and four.
- Non-respondents at every wave are less likely to be from a public high school and somewhat more likely to be from a private (non-religious) high school.
- Non-respondents have somewhat lower GPA in the previous semester compared with respondents (by about one-quarter of a letter grade).

These differences are somewhat inconsistent in that they include lower SAT and GPA for non-respondents, but higher parental education and private (more expensive) high schools. In general, the non-response bias is largest in the pre-college wave and smaller in the in-college waves even though the largest response rates are in the pre-college wave. In general, we judge the non-response bias as relatively minor on most variables and perhaps modest on SAT measures.

A.2 Omitted Coefficients For Duke Evaluation Rankings in Tables 5 and 3

Here we report coefficients for the Duke ranking variables that were omitted from Tables 5 and 3. Column 1 shows the coefficients when the dependent variable is GPA-ExpGPA, the omitted coefficients from column 2, Table 5. Column 2 shows the coefficients when the dependent variable is GPA, the omitted coefficients from column 3, Table 3. The Admission Officer's ranking of the student's achievement and personal qualities are very significant in both regressions suggesting that they may be the key variable for Duke's private information. Recommendations, however, are only significant in the second column, suggesting that student's may have some idea of the informational content of their recommendation letters.

A.3 Partitioning of Majors

Majors were partitioned according to Table A.2.

Table A.1: Coefficients on Duke Evaluation Rankings.

	(1) (GPA-ExpGPA)	(2) GPA
Achievement_3	0.217 (0.137)	0.227** (0.103)
Achievement_4	0.256* (0.138)	0.305*** (0.105)
Achievement_5	0.448*** (0.135)	0.520*** (0.102)
Curriculum_3	0.307 (0.275)	0.301 (0.224)
Curriculum_4	0.246 (0.258)	0.400* (0.212)
Curriculum_5	0.273 (0.259)	0.452** (0.213)
Essay_3	-0.086 (0.105)	-0.104 (0.103)
Essay_4	-0.026 (0.107)	-0.038 (0.104)
Essay_5	-0.124 (0.137)	-0.196 (0.129)
Personal Qualities_3	0.053 (0.198)	0.116 (0.168)
Personal Qualities_4	0.047 (0.198)	0.118 (0.168)
Personal Qualities_5	0.213 (0.209	0.305 (0.175)
Recommendations_3	0.010 (0.210)	0.393** (0.168)
Recommendations_4	0.026 (0.217)	0.423** (0.173)
Recommendations_5	0.014 (0.221)	0.427** (0.176)

Notes: Base category for each evaluation measure is 2, none of the sample had 1's for any of these measures. Column 1 refers to the omitted coefficients in Table 5 (Column 2), Column 2 refers to the omitted coefficients in Table 3 (Column 3). *, ** and *** indicate that the coefficient is significant at 10%, 5% and 1% respectively.

Table A.2: Assignment of Majors

Humanities / Social Sciences	Natural Sciences / Engineering
African / African American Studies	Biological Anthropology & Anatomy
Art History	Biology
Asian & African Language / Literature	Biomedical Engineering
Classical Civilization	Chemistry
Cultural Anthropology	Civil Engineering
Economics	Computer Science
English	Earth & Ocean Sciences
German	Computer Engineering
History	Electrical Engineering
International Comparative Studies	Environmental Sciences
Interdept. Literature / History	Mathematics
Italian & Euro Studies	Mechanical Engineering
Linguistics	Physics
Literature	
Medieval & Renaiss. Studies	
Music	
Philosophy	
Political Science	
Psychology	
Public Policy Studies	
Religion	
Sociology	
Spanish	
Theater Studies	
Women's Studies	

References

[1] Arcidiacono, Peter and Jacob Vigdor (2010), "Does the River Spill Over? Estimating the Economic Returns to Attending a Racially Diverse College." *Economic Inquiry*, Vol. 47(3).

[2] Arcidiacono, Peter, Shakeeb Khan, and Jacob Vigdor (2011), "Representation versus Assimilation: How do Preference in College Admissions Affect Social Interactions? "forthcoming, *Journal of Public Economics*.

[3] Arcidiacono, Peter (2004), "Ability Sorting and the Returns to College Major." *Journal of Econometrics*, Vol. 121, Nos. 1-2, August, 343-375.

[4] Arcidiacono, Peter, Esteban Aucejo, Hanming Fang and Kenneth Spenner (2009), "Does Affirmative Action Lead to Mismatch? A New Test and Evidence." NBER Working Paper 14885.

[5] Arrow, Kenneth (1963), "Uncertainty and the Welfare Economics of Medicare Care." *American Economic Review,* Vol. 53, No. 6, 941-973.

[6] Ayres, Ian, and Richard Brooks (2005), "Does Affirmative Action Reduce the Number of Black Lawyers?" *Stanford Law Review*, Vol. 57, No. 6: 1807-1854.

[7] Babcock, Philip (2010), "Real Costs of Nominal Grade Inflation" *Economic Inquiry*, forthcoming.

[8] Babcock, Philip and Mindy Marks (2010), "The Falling Time Cost of College: Evidence from Half a Century of Time Use Data" *Review of Economics and Statistics*, forthcoming.

[9] Barnes, Katherine Y. (2007), "Is Affirmative Action Responsible for the Achievement Gap Between Black and White Law Students?" *Northwestern University Law Review*, Vol. 101, No. 4, Fall: 1759-1808.

[10] Black, D., K. Daniel and J. Smith (2001), "Racial Differences in the Effects of College Quality and Student Body Diversity on Wages." in *Diversity Challenged*, Harvard Educational Review.

[11] Bryant, Anita-Yvonne, Kenneth I. Spenner and Nathan Martin, with Alexandra Rollins and Rebecca Tippett (2006), "The Campus Life and Learning Project: A Report on the First Two College Years." Available at http://www.soc.duke.edu/undergraduate/cll/final_report.pdf

[12] Bryant, Anita-Yvonne, Kenneth I. Spenner and Nathan Martin, with Jessica M. Sautter (2007), "The Campus Life and Learning Project: A Report on the College Career." Available at http://www.soc.duke.edu/undergraduate/cll/2nd_report.pdf

[13] Cardon, James H. and Igal Hendel (2001), "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey." *Rand Journal of Economics,* Vol. 32 (Autumn): 408-427.

[14] Cawley, John and Thomas Philipson (1999), "An Empirical Examination of Information Barriers to Trade in Insurance." *American Economic Review,* Vol. 89, No. 5: 827-846.

[15] Chambers, David L., Timothy T. Clydesdale, William C. Kidder, and Richard O. Lempert (2005), "The Real Impact of Eliminating Affirmative Action in American Law Schools: An Empirical Critique of Richard Sander's Study." *Stanford Law Review*, Vol. 57, No. 6: 1855-1898.

[16] Chiappori, Pierre-André and Bernard Salanié (2000), "Testing for Asymmetric Information in Insurance Markets." *Journal of Political Economy,* Vol. 108, No. 2: 56-78.

[17] Chiappori, Pierre-André (2001), "Econometric Models of Insurance under Asymmetric Information." In *Handbook of Insurance*, edited by Georges Dionne. Springer.

[18] Chiappori, Pierre-André, Bruno Jullien, Bernard Salanié and Francois Salanié (2006), "Asymmetric Information in Insurance: General Testable Implications." *Rand Journal of Economics*, Vol. 37 (Winter): 783-798.

[19] Cunha, Flavio, James J. Heckman and Salvador Navarro (2005), "Separating Uncertainty from Heterogeneity in Life Cycle Earnings." *Oxford Economic Papers* (2004 Hicks Lecture), Vol. 57, No. 2: 191-261.

[20] Cureton, Edward E. and Ralph B. D'Agostino (1983), *Factor Analysis: An Applied Approach.* Lawrence Erlbaum Associates, Inc. Publishers: New Jersey.

[21] Dale, Stacy Berg, and Alan B. Krueger (2002), "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables." *Quarterly Journal of Economics,* Vol. 117, No. 4: 1491-1527.

[22] Duncan, G.J., J. Boisjoly, D.M. Levy, M. Kremer, and J. Eccles (2006), "Empathy or Antipathy? The Impact of Diversity." *American Economic Review*, Vol. 96, No. 6: 1890-1905.

[23] Fang, Hanming, Michael P. Keane and Dan Silverman (2008), "Sources of Advantageous Selection: Evidence from the Medigap Insurance Market." *Journal of Political Economy,* Vol. 116, No. 2, 303-350.

[24] Finkelstein, Amy and Kathleen McGarry (2006), "Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market." *American Economic Review*, Vol. 96, No. 5: 938-958.

[25] Finkelstein, Amy and James Poterba (2004), "Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Market." *Journal of Political Economy*, Vol. 112, No. 1: 183-208.

[26] Ho, Daniel E. (2005), "Why Affirmative Action Does Not Cause Black Students to Fail the Bar." *Yale Law Journal,* Vol. 114, No. 8, 1997-2004.

[27] Horowitz, Joel (1998), *Semiparametric Methods in Econometrics.* Springer.

[28] James, Estelle, Nabeel Alsalam, Joseph C. Conaty and Duc-Le To (1989), "College Quality and Future Earnings: Where Should You Send Your Child to College?" *American Economic Review,* Vol. 79, No 2, 247-252.

[29] Johnson, Valen (2003), *Grade Inflation: A Crisis in College Education.* Springer.

[30] Kellough, J. Edward (2006), *Understanding Affirmative Action: Politics, Discrimination and the Search for Justice.* Georgetown University Press: Washington D.C.

[31] Kim, Jae-On and Charles W. Mueller (1994), "Introduction to Factor Analysis: What It Is and How to Do It." in Michael S. Lewis-Beck ed., *Factor Analysis and Related Techniques* (International Handbooks of Quantitative Applications in the Social Sciences, Vol. 5): Sage Publications.

[32] Kotlarski, Ignacy (1967), "On Characterizing the Gamma and Normal Distribution." *Pacific Journal of Mathematics*, 20, 729-738.

[33] Krasnokutskaya, Elena (2009), "Identification and Estimation in Highway Procurement Auctions Under Unobserved Auction heterogeneity." forthcoming, *Review of Economic Studies.*

[34] Li, Tong and Quong Vuong (1998), "Nonparametric Estimation of Measurement Error Model Using Multiple Indicators." *Journal of Multivariate Analysis,* Vol 65, 135-169.

[35] Loury, Linda D. and David Garman (1995)., "College Selectivity and Earnings." *Journal of Labor Economics*, Vol. 13, No. 2: 289-308.

[36] Pauly, Mark V. (1974), "Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection." *Quarterly Journal of Economics,* Vol. 88, No. 1: 44-62.

[37] Rao, B.L.S. Prakasa (1992), *Identifiability in Stochastic Models: Characterization of Probability Distributions.* Academic Press: New York.

[38] Rothschild, Michael and Joseph E. Stiglitz (1976), "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *Quarterly Journal of Economics,* 90 (November): 629-649.

[39] Rothstein, Jesse and Albert Yoon (2008), "Mismatch in Law School." mimeo, Princeton University.

[40] Rubin, Donald B. (1987). *Multiple Imputation for Nonresponse in Surveys.* Wiley, New York.

[41] Sander, Richard H. (2004), "A Systemic Analysis of Affirmative Action in American Law Schools." *Stanford Law Review*, Vol. 57, No. 2, 367-483.

[42] Sander, Richard H. (2005a), "Mismeasuring the Mismatch: A Response to Ho." *Yale Law Journal*, Vol. 114, No. 8: 2005-2010.

[43] Sander, Richard H. (2005b), "Reply: A Reply to Critics." *Stanford Law Review*, Vol. 57, No. 6, 1963-2016.

[44] Stinebrickner, Todd and Ralph Stinebrickner (2008). "Learning about academic ability and the college drop-out decision." mimeo, University of Western Ontario.

[45] Wilson, Charles (1977), "A Model of Insurance Markets with Incomplete Information." *Journal of Economic Theory,* Vol. 16 (December): 167-207.