A Test of Fair Treatment with Application to University Admissions (Incomplete draft)

Debopam Bhattacharya, Shin Kanaya, Margaret Stevens

University of Oxford

October, 2011.

Abstract: Admissions to selective colleges – especially publicly funded ones – are often criticized for "unfair" admission policies that favour specific socioeconomic groups. In this paper, we develop a statistical framework for testing whether current admission practices maximize expected academic performance of the entering cohort. Such maximization would imply that the expected performance of the marginal admitted candidate –the admission threshold– is equalized across socioeconomic groups. For biased practices, the marginal candidate in the discriminated group will have higher expected outcomes. If we observe all pre-admission applicant characteristics available to the university and the post admission performance for admitted candidates, then the admission threshold can be identified as long as the density of the conditional expected performance is positive around that threshold. Our key identifying assumption is that the median (but not necessarily the distribution) of unobserved heterogeneity affecting admissions is independent of observed covariates, as in Manski's maximum score setting. The fair treatment hypothesis can then be tested by checking if the identified thresholds are equal across demographic groups. We further show that under a stronger assumption of independence between unobserved heterogeneity and observed covariates, one can calculate the welfare loss – in terms of current academic achievement foregone – resulting from the use of different thresholds as would happen under

affirmative actions. Applying these methods to two recent cohorts of UK-based applications to Oxford's flagship PPE degree, we find that male applicants face a higher threshold than female applicants and that there is no difference between thresholds faced by state-school and private school pupils.

Keywords: University admissions, affirmative action, selection on observables, conditional median restriction, welfare loss estimation.

1 Introduction

Every year, thousands of students around the world apply for admission to colleges and universities. The colleges' decisions on whom to admit are typically based on high-school academic performance as well as scores on standardized aptitude tests, such as the ACT and SAT in the US and the Cambridge assessments in the UK. In addition, some colleges take into account the applicants' demographic backgrounds and some conduct expensive admission interviews. In this context, it becomes important from a policy perspective to explore two related questions regarding the admissions process – (a) how the observed applicant characteristics and test-scores should be weighed in the admission process and (b) how to test whether prevalent admission procedures are "biased" against applicants from disadvantaged backgrounds.

There is a substantial literature on the first question in educational statistics based on "validity" studies which analyze the relative importance of aptitude test scores vis-a-vis measures of high school academic performance and/or socioeconomic status (SES, henceforth) in predicting post-enrolment academic outcomes. See, for instance, Kobrin et al (2004), Noble (2003), Parks et al (2011), Sawyer (1996), Sackett et al (2009) and the references therein as well as a large number of staff research papers made available online by the College Board and the act.org in the United States and the Institute of Education in the UK. However, less attention seems to have been given to the second issue – that of evaluating current admission policies – in this literature.

The motivation for studying the second question is that widening access to good universities is generally considered to be a key mechanism for promoting social mobility. Therefore, under-representation of low SES communities among the entering students, especially at relatively elite colleges, is flagged as suggestive of unfair admission policies.[1] Colleges sometimes respond to such allegations by saying that they intend to admit students with the best academic potential, irrespective of their SES and such objectives can lead to uneven demographic compositions of entering classes. Testing whether this claim is correct is impor-

[1] For the UK, see for instance, the recent newspaper stories: http://www.guardian.co.uk/education/2011/mar/20/oxford-universit,y-entrance-requirements-row, or http://www.guardian.co.uk/education/2009/aug/19/oxford-university-men-places-women

tant from a policy perspective because the conclusion would inform what steps are necessary to promote fair access. For example, if admissions are indeed academic outcome-oriented, then good students from lower SES should be encouraged to apply in order to increase their presence in elite colleges. Conversely, if admissions are explicitly or implicitly "taste-based" – i.e., use SES or demographics in admission to an extent exceeding their effects on expected academic outcome – then additional interventions may be needed in the admission process itself to promote fairness. This issue is substantively distinct from (i) the differential performance of different SES or demographic groups in aptitude tests (lower SES applicants perform worse on average), (ii) lower aptitude test scores proxying for lower SES in explaining post-enrolment outcomes (mixed evidence, c.f., Rothstein (2004) and Sackett et al (2009)) and (iii) the differential average college-outcomes of different demographic groups at the same value of aptitude test scores (black applicants have lower expected scores than Caucasian ones (Noble (2003)). These latter issues pertain to the average outcomes and are not directly relevant to testing academic fairness because academically fair – i.e., outcome oriented admissions – have implications for the outcomes of the marginal, but not the average, admitted candidates across demographic or SES categories. To see why, note that if the marginal male admit has a lower expected outcome than the marginal female admit then one can admit the "next best" female candidate instead of the marginal male admit and increase the expected overall outcome. So outcome-oriented admissions must equate the expected outcome across marginal admitted candidates in every demographic/SES group. Violations of such inequality suggest that the college must place emphasis on non-academic attributes such as gender, race or SES of the applicant *beyond these attributes' impacts on his/her potential academic outcomes* and, in that sense, is "taste-based". On the other hand, the average outcome across all admitted males bears no relation to that across all admitted females, no matter whether admissions is outcome-oriented or not. This key but subtle distinction between the average and the marginal, to the best of our knowledge, is not recognized or used in the educational statistics literature and in public debates about unfair access to elite universities. But such distinctions are well-recognized in studies of discrimination in economics (c.f., Becker (1971) for an early discussion and Persico (2009) for a recent survey). In this paper, we utilize, adapt and extend these insights from the economics

literature to analyze the question of testing academic fairness of admissions. In the empirical implementation of our methods using admissions data, we pay particular attention to the modelling of unobserved heterogeneity whose presence makes it challenging (c.f., Heckman (1999)) to test fairness of treatment regimes in general, and college admissions in particular.

Specifically, in this paper we construct an econometric model of admissions involving observed covariates, unobserved heterogeneity across admission officers and outcomes of past and present admitted students. We allow for the fact that not all admission offers translate into enrolment because applicants may get alternative offers or fail to satisfy conditions specified in the offer, such as securing a certain mark in the school-leaving public examination. Our first contribution in this setting is to show that under reasonable behavioral assumptions, the admission thresholds faced by different categories of applicants can be identified from data on admission outcomes of current applicants and post-enrolment performance of just the admitted students; it is not necessary to identify the (unobserved) distribution of potential outcomes for non-admitted applicants. These thresholds, in turn, will equal the expected outcome of the marginal admit in every category and a test of outcome-oriented admissions can then be based on testing the equality of these thresholds across the categories. Our second contribution is to develop a method of calculating the efficiency loss – in terms of individual academic achievement foregone by the college – which would result from the use of unequal thresholds across demographic subgroups due to, say, adoption of affirmative action policies.

2 A Benchmark Model

We start our analysis by laying out a benchmark economic model of optimization to help fix ideas. In the next section, we develop an analogous econometric model, incorporating unobserved heterogeneity, based on this benchmark economic model.

Let W denote an applicant's covariates which are observed by the university. Let $\mathcal{W}$ denote the support of W, $P_W(\cdot)$ denote the C.D.F. of W and let G denote the discrete component of W capturing the group identity of the applicant (such as sex or type of high school attended). For an applicant i, let Y_i denote his/her future academic performance if

admitted to the university. Let $\mu(w)$ denote a w-type student's expected performance if he/she enrols, as perceived by the university and let $\alpha(w)$ denote the probability that a w-type student offered admission eventually enrols. Let $c \in (0,1)$ denotes the fraction of applicants who can be admitted, given the number of available spaces.

Admission protocols: We can define an admission protocol as a probability $p(\cdot) : \mathcal{W} \mapsto [0,1]$ such that an applicant with characteristics w is offered admission with probability $p(w)$. A generic objective of the university may be described as

$$\max_{p(\cdot)} \int p(w)\,\alpha(w)\,[h(w)\,\mu(w)]\,dP_W(w), \text{ s.t. } \int p(w)\,\alpha(w)\,dP_W(w) \leq c.$$
$$\max_{p(\cdot)} \int p(w)\,\alpha(w)\,\mu(w)\,dP_W(w), \text{ s.t. } \int p(w)\,\alpha(w)\,dP_W(w) \leq c.$$

Here, $h(w)$ denotes a non-negative welfare weight, capturing how much the outcome of a w-type applicant is worth to the university. For affirmative action policies, $h(\cdot)$ will be larger for applicants from disadvantaged socioeconomic backgrounds or under-represented demographic groups. The overall objective is thus to maximize mean weighted outcome among the admitted applicants, subject to a budget constraint. The solution to the above problem takes the form (c.f., Bhattacharya (2011) and proof in appendix):

$$p^{opt}(w) = \begin{cases} 1 \text{ if } \beta(w) > \gamma, \\ q \text{ if } \beta(w) = \gamma, \\ 0 \text{ if } \beta(w) < \gamma, \end{cases} \tag{1}$$

where

$$\begin{aligned} \beta(w) &: = h(w)\,\mu(w), \\ \gamma &: = \inf\left\{ r : \int \alpha(w)\,1\{\beta(w) > r\}\,dF_W(w) \leq c \right\}, \end{aligned}$$

and $q \in [0,1]$ satisfies

$$\int \alpha(w)\,(\{1(\beta(w) > \gamma) + q1(\beta(w) = \gamma)\})\,dF_W(w) = c.$$

The result basically says that the planner should order individuals by their values of $\beta(W)$ and first admit applicants with those values of W for which $\beta(W)$ is the largest, then to those for whom it is the next largest and so on till the budget is exhausted. If the

distribution of $\beta(W)$ has point masses, then there could be a tie at the margin, which is then broken by randomization (hence the probability q). In the absence of any point masses, the optimal protocol is of a simple threshold-crossing form $p^{opt}(w) = 1(\beta(W) \geq \gamma)$. For the rest of the paper, we will assume that this is the case.

Academically fair admissions: We define an academically fair admission protocol as one which maximizes expected performance of the incoming cohort subject to the restriction on the number of vacant places. Such an objective is fair in the sense that the expected performance criterion gives equal weight to the *outcomes* of all applicants, regardless of their value of W, i.e., $h(w)$ is a constant. In this case, the previous solution takes the form $p^{opt}(w) = 1\{\mu(w) \geq \gamma\}$, where γ solves

$$c = \int_{w \in \boldsymbol{w}} \alpha(w) 1(\mu(w) \geq \gamma) dP_W(w).$$

The key feature of the above rule is that the value of W affects the decision only via its effect on $\mu(W)$ and γ does not depend on W. To get some intuition on this, consider the case where one of the covariates in W is gender and assume that the admission threshold for women, γ_f is strictly smaller than the threshold for men, γ_m. Then the marginal female admit with $w = (x_f, f)$, contributes $\alpha(x_f, f)\gamma_f$ to the expected aggregate outcome and takes up $\alpha(x_f, f)$ places, implying a contribution of γ_f $(=\alpha(x_f, f)\gamma_f/\alpha(x_f, f))$ to the objective of average realized outcome. Similarly, the marginal male reject's contribution to the average outcome would be $\gamma_m > \gamma_f$. Therefore, we can increase the average outcome if we replaced the marginal female admit with the marginal male reject. Thus different thresholds cannot be consistent with the objective of maximizing the mean outcome.

Lastly, to make our requirement of common thresholds logically possible, we need to assume that there exists a common feasible threshold. Toward that end, let us rewrite $W = (X, G)$ and let Ω_g, Υ_g denote respectively the support for the distribution of X and of $\mu(X, G)$, given $G = g$. Let γ_g be the admission threshold for group g, i.e., an applicant of type g is admitted if and only if $\mu(X, g) \geq \gamma_g$. We now introduce a formal assumption which would imply that equality of admission thresholds is equivalent to outcome-oriented admissions – which forms the basis of our testing approach.

Assumption 1 $\Upsilon_g \equiv \Upsilon$ *for all* g; Υ *contains an interval* I *such the density of* $\mu(X,g)$ *conditional on* $G = g$ *is strictly positive on* I *and* γ_g *lies in* I *for each* g. *Further, the common outcome maximizing threshold* γ *lies in* Υ.

To interpret the above assumption, consider the case where G denotes gender and X contains one or more continuous variables like previous test scores. Then the above assumption basically says that the (conditional) expected outcome for men and that for women take values in the same set. So given any value $x \in \Omega_{male}$, there exists an $x' \in \Omega_{female}$ such that $\mu(x, male) = \mu(x', female)$. It seems very unlikely that there will be a range of possible values of the (conditional) expected outcome for men where no women's conditional expected outcome lies (or vice versa). This can only possibly happen at the very extremes of the former's support but, since most universities admit at least 5% of applicants, it is unlikely that any admission threshold can lie in such extreme regions.[2] Note also that we do not require Ω_g to be identical across g. Generically, one would expect the admission threshold to be an "interior" point in the support of $\mu(X,g)$, such as a quantile in the marginal distribution of $\mu(X,G)$, so that the second part of the above assumption seems natural if $\mu(X,G)$ is continuously distributed.

Under the above assumption, for every fixed g, there will exist $x^*(g) \in int(\Omega_g)$ such that $\mu(x^*(g), g) = \gamma$; so we can define individuals in subgroup g with $X = x^*(g)$ to be the "ideally marginal" admits in subgroup g. If the protocol is outcome-oriented, then for every g, $\mu(x^*(g), g) = \gamma$. However, if the protocol is not outcome oriented and the marginal admits are denoted by $x(g)$ in group g, then $\mu(x(g), g) \equiv \gamma_g$ will differ across g. The question then is how to identify $x(g)$ and then how to identify $\mu(x(g), g)$, i.e., the expected outcome of the marginal admit for various values of g. If the common support assumption did not hold, then it would be possible that admission is outcome oriented with a common optimal γ which lies within the support of $\mu(X, g_1)$ conditional on $G = g_1$ but not of $\mu(X, g_2)$ given $G = g_2$. In that case, for group 1, we will have equality at the margin but for group 2, the marginal admits will have expected outcome exceeding the threshold if γ lies in a "hole" with

[2]The above assumption can be weakened a bit by requiring that the intersection of Υ_g's contains a nonempty interval and γ belong to it. But this mathematical generalization seems to us to be substantively a bit artificial and hence we do not pursue it here.

respect to the support of $\mu(X, g_2)$ given $G = g_2$. Figure 1 illustrates the point. To be clear, this common support assumption has nothing to do with the identification of group-specific thresholds, analyzed in the following section. Instead, the purpose of this assumption is that it enables us to interpret the inequality between group-specific thresholds as symptomatic of non outcome-oriented admissions.

3 Econometric Model and Identification

The economic model above takes the entire university as a single decision-making entity whereas in reality, admission decisions are made by individual officials who apply at least some personal discretion and/or display heterogeneity in taste or knowledge in making the decision. This heterogeneity is likely to introduce idiosyncratic variation in the individual decisions around a baseline university-wide policy. Further, if the model of the previous section literally generated the data, then admission would be completely deterministic, given every value of X and G, which is almost never the case in practice. With these motivations, we extend the previous economic model into an econometric one, which incorporates heterogeneity across admission officers and forms the basis of our empirical analysis.

To set up the empirical framework, we assume that we have data on pre-admission characteristics (including prior test-scores) on a pool of applicants, drawn in an I.I.D. fashion from a distribution of potential applicants. For each applicant, we also observe whether they were made an offer of admission, for each offer we observe whether it was accepted and finally, for each accepted offer, we observe the outcome of interest (e.g., examination score after 1st year of college). We have such data for several years. When referring to variables from past years or expectations calculated on the basis of past variables, we will use the superscript P.

For a current applicant i in the data, let X_i, G_i denote his/her values of pre-admission characteristics and let D_i denote the dummy for whether (s)he received an offer and the dummy A_i equals 1 if and only if i was made an offer and accepted it. We observe (X_i, G_i, D_i, A_i) for every individual and the outcome Y_i if $A_i = 1$. Let $\mu^P(x, g) := E\left(Y^P | X = x, G = g, A^P\right.$ denote the conditional expected outcome in previous years for admits with characteris-

tics x, g. We assume that when admission officers decide on whether to admit an (x, g) type student, they base it on $\mu^P(x, g)$ which they calculate using data on (x, g) type students who were admitted in previous years. Note that $\mu^P(x, g)$ is in general different from $E\left(Y^P|X = x, G = g\right)$ which is typically unknown to admission officers in universities (or loan officers in banks).[3] Indeed, a large literature in educational statistics on so-called "validation studies" use predicted performance of *admitted* candidates to infer the relative predictive ability of aptitude test scores vis-a-vis high school grades and socioeconomic indicators and prescribe policies based on this analysis. See for example, the large number of staff research papers available online at the College Board, ACT.org and ETS in the US and the central institute of education in the UK. While several such studies are aware of sample selection biases, they do not appear to (be able to) "correct" for it. Since our analysis evaluates what educational institutions actually do – rather than what they should have done – using $\mu^P(x, g)$ rather than $E\left(Y^P|X = x, G = g\right)$ is the correct approach here.

Toward that end, let $\mathcal{X}_g$ denote the support of X conditional on $G = g$ and $A^P = 1$, i.e., $\mathcal{X}_g = \left\{x : \Pr\left(A^P = 1|X = x, G = g\right) > 0\right\}$. These are the values of X which occur among the admits of type g in past years and so one can calculate the values of $\mu^P(x, g)$ when $x \in \mathcal{X}_g$. We assume that a g-type applicant with a value of X outside $\mathcal{X}_g$ will be not be offered admission with probability 1 (such as applicants with very low admission test scores) in the current year. On the other hand, an applicant i with $G_i = g$ and $X_i = x \in \mathcal{X}_g$ is offered admission if and only if $\mu^P(x, g) \geq \gamma_g + \varepsilon_i$, where ε_i denotes heterogeneity across admission officials and γ_g denotes the university-prescribed admission threshold for applicants of demographic type g. Thus we may summarize the admission process as

[3] If there existed trial data where admissions (respectively, loan approvals) were randomized, then the latter can be calculated and used instead of $\mu^P(x, g)$. Alternatively, if Y^P were independent of A^P, given X, G (the so-called selection-on-observables case), then the two would be identical. However, from a policy perspective, detecting whether universities are outcome oriented requires the use whichever expectation the officers can base their decision on, and not $E\left(Y^P|X = x, G = g\right)$. In some special cases, the two can be identical but this is somewhat irrelevant to the task at hand.

Assumption 2

$$D_i = \begin{cases} \mathbf{1}\left\{\mu^P\left(X_i, G_i\right) \geq \gamma_{G_i} + \varepsilon_i\right\}, \text{ if } X_i \in \mathcal{X}_{G_i} \\ 0, \text{ if } X_i \notin \mathcal{X}_{G_i} \end{cases}.$$

The above assumption implies that the set $\{x : \Pr\left(D = 1 | X = x, G = g\right) > 0\}$ is a subset of $\mathcal{X}_g$. Therefore, for any x such that $\Pr\left(D = 1 | X = x, G = g\right) > 0$, we can compute $\mu^P\left(x, g\right)$ as $E\left(Y^P | X = x, G = g, A^P = 1\right)$, since the conditioning set is non-empty.

We are interested in estimating the threshold γ_g for various values of g and testing if they are identical across g. For the identification/estimation of γ_g, we impose the following conditions:

Assumption 3 *(i)* $\{(X_i, G_i, D_i, A_i, Y_i A_i)\}_{i=1}^n$ *is an i.i.d. sequence; (ii)* $\text{median}(\varepsilon_i \mid X_i, G_i) = 0$ *and (iii) the distribution of* ε_i *has a strictly positive density (with respect to the Lebesgue measure) around* 0*, given* X_i *and* G_i*.*

Discussion: The error term ε_i can be interpreted as the error in subjective expectation of the specific admission officer handling applicant i's case and γ_g the university-specified cut-off for applicants of type g. The presence of ε allows admission to be non-deterministic, given X and G. In other words, the admission cut-off (for expected performance) faced by applicants of type g varies randomly within g-type applicants (depending on which officer handled their file) with a median value of γ_g stipulated by university-wide admission policy. For academically fair admissions, as explained above, one would expect γ_g to be identical across g. Part (ii) of Assumption 3 – which allows the scale of ε to depend on X, G – will hold when the systematic determinants of admission, such as past test scores, interview grades and demographic characteristics are commonly observed but idiosyncratic preferences and/or errors in the subjective expectations of individual admission officers are incorporated in ε. Part (iii) of Assumption 3 is a regularity condition that aids the proof of identification.

3.1 Identification of γ_g

The basic idea is to use for each fixed g, the median restriction and the observed $\Pr\left(D = 1 | X = x, G = g\right)$ to identify the values of X defining the marginal admits, viz., those x for which $\Pr\left(D = 1 | X = x, G = g\right) =$

1/2 and then average $\mu^P(x,g)$ – separately identified from admitted students in previous years – across these marginal admits to yield γ_g.

Our identification is facilitated by the following assumptions:

Assumption 4 *For each value g in the support of G, the distribution of the random variable* $\mu^P(X,G)$ *conditional on* $G=g$ *has a strictly positive density (with respect to the Lebesgue measure) on an open interval around* γ_g.

The above assumption guarantees that there exists some $x \in \mathcal{X}_g$, such that $\Pr(D=1|X=x,G=g) = 1/2$. It will hold as long as $\mu^P(X,G)$ varies with X and X has a rich enough support.[4] To complete the identification argument, note that if there exists an x such that $\Pr(D=1|X=x,G=g)=1/2$, then we must have that

$$\Pr\left[\mu^P(x,g) - \gamma_g \geq \varepsilon | X=x, G=g\right] = 1/2,$$

implying that

$$\mu^P(x,g) - \gamma_g = 0,$$

by (ii) and (iii) of Assumption 3. Therefore, it would hold that

$$E\left[\mu^P(x,g) - \gamma_g \mid \Pr(D=1|X=x,G=g) = 1/2\right] = 0. \tag{2}$$

Observe that Assumption 3 lets us identify the marginal admits as those x for which the probability of admission is equal to a half. Then 4 guarantees that for every fixed g, the set $\Pi_g = \left\{x \in \mathcal{X}_g : \mu^P(x,g) = \gamma_g\right\}$ – identical to the observable set of $x \in \mathcal{X}_g$ satisfying $\Pr(D=1|X=x,G=g) = 1/2$ – is nonempty. Finally, $x \in \mathcal{X}_g$ guarantees that we can compute $\mu^P(\cdot,g)$ for each $x \in \Pi_g$ from past cohorts simply by computing $E\left(Y^P|D=1,X=x,G=g\right)$, and these can then be averaged over $x \in \Pi_g$ to yield γ_g. We summarize these discussions into the following proposition.

Proposition 1 *Assume that assumptions 3 and 4 hold. Then for each g, the threshold* γ_g *is point-identified.*

[4]Since γ_g is likely to be close to a quantile of $\mu^P(X,g)$ (see above), it is quite likely to be located in the interior of the support of $\mu^P(X,g)$ and so a "large" support for X is not necessary to guarantee the above condition.

Once we have identified the group-specific thresholds γ_g, we can test if admission is outcome-oriented by testing the equality of γ_g across g. This implication is facilitated by our common support assumption (1) in the previous section for $\mu^* (\cdot, \cdot)$ which is equal to $\mu^P (\cdot, \cdot)$, by definition.

Assumption 5 *The past and current cohorts are drawn from the same population distribution of applicants and the admission thresholds have not changed over the entire period of analysis.*

In view of the stationarity assumption 5, we drop the P superscript from now on and combine all years of data for estimating conditional means and probabilities in order to maximize precision.

4 Estimation and Inference

Motivated by the restriction of (2), we now present an estimator of γ_g. Basically, if X contains continuous components, then the estimated probability of getting an offer will never be exactly 1/2; so we will consider those (x, g)'s for which the predicted probability is close to a half. The extent of closeness is guided by a bandwidth h. More formally, in a first step, we estimate $\mu (x, g)$ and $p (x, g) = \Pr [D_i = 1 | (X_i, G_i) = (x, g)]$ and then in a second step, we estimate γ_g by a weighted average of $\hat{\mu} (X_i, G_i)$, where the weights are a decreasing function of the distance between $\hat{p} (X_i, G_i)$ and 1/2,

$$\hat{\gamma}_g = \frac{\sum_{i=1}^{n} K_h (\hat{p} (X_i, G_i) - 1/2) \hat{\mu} (X_i, G_i) \mathbf{1} \{G_i = g\}}{\sum_{l=1}^{n} K_h (\hat{p} (X_l, G_l) - 1/2) \mathbf{1} \{G_l = g\}}.$$

where $K_h (z) = K (z/h) / h$; $K (\cdot)$ is a kernel function; h is a smoothing parameter (bandwidth); $\hat{p} (x, g)$ and $\hat{\mu} (x, g)$ are first-step estimators of $p (x, g)$ $(:= \Pr [D_i = 1 | (X_i, G_i) = (x, g)])$ and $\mu (x, g)$, respectively. This is exactly the average predicted outcome for those (x, g) types for whom the predicted probability of getting an offer, i.e., $p (x, g)$ is close to a half.

In order to develop a distribution theory for conducting inference, we consider the operationally convenient case where the first step estimators $\hat{p} (x, g)$ and $\hat{\mu} (x, g)$ are parametric.

This case is different from standard semiparametric estimators where the first step is nonparametric and the second step involves some form of averaging of the first step estimators, leading eventually to$\sqrt{n}$-normal estimates. Here, kernel smoothing at the second step is unavoidable and so, even if the first step is parametric, one cannot estimate γ_g at the parametric rate. We may use any parametric model satisfying some mild (c.f., assumptions 11 and 12, below) e.g., a probit or logit model for $p(w)$; and a linear (regression) model for $\mu(w)$.

Define $\tilde{\gamma}_g$ as the infeasible estimator that would result if the true values $\mu(X_i, g)$ and $p(X_i, g)$ were used instead of their estimates:

$$\tilde{\gamma}_g := \frac{\sum_{i=1}^{n} K_h\left(p(X_i, g) - 1/2\right) \mu(X_i, g) \mathbf{1}\{G_i = g\}}{\sum_{l=1}^{n} K_h\left(p(X_l, g) - 1/2\right) \mathbf{1}\{G_i = g\}}, \tag{3}$$

for each $g \in \{f, m\}$. We will show in the appendix (see theorem 1) that our semiparametric estimator $\hat{\gamma}_{sp}(g)$ has the same asymptotic distribution as $\tilde{\gamma}_g$, under the assumption of correct specification. Since $\tilde{\gamma}_g$ is a nonparametric regression of $\mu(X_i, g)$ on $p(X_i, g)$ evaluated at $p(X_i, g) = 1/2$, its distribution follows from standard nonparametric regression theory. The distributional equivalence of $\hat{\gamma}_{sp}(g)$ and $\tilde{\gamma}_g$ and an explicit form for the asymptotic distribution are provided in the appendix.

Choosing Bandwidths: Note that our parameter of interest, γ_g is essentially the conditional mean $E(\mu(X, G) | p(X, G) = 1/2, G = g)$. So we recommend the standard method based on cross-validation, which uses a global goodness of fit criterion for the conditional mean $E(\mu(X, G) | p(X, G), G = g)$. In the present context, cross-validation is achieved by minimizing the leave-one-out criterion

$$\begin{aligned} R(h) &= \sum_{j=1}^{n} 1(G_j = g) \times \left[\bar{\mu}\left(X_j, g; \hat{\theta}_\mu\right) - m_{-j}\left(\bar{p}\left(X_j, G_j; \hat{\theta}_p\right), g\right); h\right]^2, \text{ where} \\ m_{-j}(a, g; h) &= \frac{\sum_{\substack{i=1 \\ i \neq j}}^{n} K_h\left(\bar{p}\left(X_i, G_i; \hat{\theta}_p\right) - a\right) \bar{\mu}\left(X_i, G_i; \hat{\theta}_\mu\right) \mathbf{1}\{G_i = g\}}{\sum_{\substack{i=1 \\ i \neq j}}^{n} K_h\left(\bar{p}\left(X_i, G_i; \hat{\theta}_p\right) - a\right) \mathbf{1}\{G_i = g\}} \end{aligned}$$

is an estimator of $E(\mu(X, G) | p(X, G) = a, G = g)$, calculated using the bandwidth h. We then use as optimal bandwidth the minimizer of the CV criterion divided by $(\ln n)$. This

undersmoothing, as is well-known from the econometric literature, serves to reduce the asymptotic bias and makes it possible to construct confidence intervals for γ_g.

5 Efficiency Cost of Affirmative Action (Incomplete...)

Set-up: Many universities intentionally apply a lower threshold for applicants from historically disadvantaged backgrounds – such as women or racial minorities in the US, applicants from state-run schools in the UK and so on. From a policy perspective, it is important to cast the social benefits of such a policy against the short-term loss in productive efficiency – measured as the academic achievement foregone by the university – which results from it. Toward that end, suppose that progressive motives lead a university to choose a lower threshold for female candidates, i.e., $\gamma_f < \gamma_m$. Then, the efficiency cost of this policy may be computed as the achievement foregone by replacing high ability males with low ability females relative to the efficient situation where a common threshold $\tilde{\gamma} \in (\gamma_f, \gamma_m)$ is used. This intermediate $\tilde{\gamma}$ can be chosen so that the aggregate number of admits, if the common $\tilde{\gamma}$ is used, will remain identical to current number of admits in expectation.

The model of admission is now given by: admit a candidate of type $X = x$ and $G = g$ according to the rule

$$D = \begin{cases} \mathbf{1}\{\mu(x,g) \geq \tilde{\gamma} + \varepsilon\}, \text{ if } x \in \mathcal{X}_g \\ 0, \text{ if } x \notin \mathcal{X}_g. \end{cases}.$$

Note that this allows changes in admission policies such that the set $\mathcal{X}_g$ remains the same; so those characteristics for which the probability of admission was zero before the change will still lead to outright rejection. It's only individuals with characteristics in $\mathcal{X}_g$ who will now be treated differently, corresponding to the use of a new threshold $\tilde{\gamma}$.

Let $\mu_g(x) = E(Y|A=1, X=x, G=g)$, $\alpha_g(x) = \Pr(A=1|X=x, G=g)$ and $p_g(x) = \Pr(D=1|X=x, G=g)$. We maintain the stationarity assumption throughout this section, so that we will drop the P superscript everywhere. Let π_m denote the fraction of applicants

who are male. Define

$$\begin{aligned} M &= \left\{ x \in \mathcal{X}_m : \mu(x,m) + \gamma_m - \tilde{\gamma} < \max_{x \in \mathcal{X}_m} \mu(x,m) \right\} \\ F &= \left\{ x \in \mathcal{X}_f : \mu(x,f) + \gamma_f - \tilde{\gamma} < \min_{z \in \mathcal{X}_f} \mu(z,f) \right\}. \end{aligned}$$

Assumption 6 *ε is independent of X, given G; i.e., for all g, the conditional c.d.f. of ε satisfies $F_{\varepsilon|X=x,G=g}(\cdot) = F_{\varepsilon|G=g}(\cdot)$.*

Assumption 7 *Conditional on each g in the support of G, the random variable $\mu(X,G)$ assumes every value within $\max_{z \in \mathcal{X}_g} \mu(z,g)$ and $\min_{z \in \mathcal{X}_g} \mu(z,g)$ with positive probability. That is, for any $a \in \left(\min_{z \in \mathcal{X}_g} \mu(z,g), \max_{z \in \mathcal{X}_g} \mu(z,g)\right)$, there exists $x \in \Omega_g$ such that $\mu(x,g) = a$.*

Assumption 8 *Acceptance of an offer is independent of ε, given X and G, i.e., corresponding to any pair of offer thresholds $\gamma, \tilde{\gamma}$, the corresponding acceptance variables A and $\tilde{A}$ satisfy*

$$\begin{aligned} & E(Y|A=1, \varepsilon \leq \mu(X,G) - \gamma, X=x, G=g) \\ = \; & E\left(Y|\tilde{A}=1, \varepsilon \leq \mu(X,G) - \tilde{\gamma}, X=x, G=g\right). \end{aligned}$$

Discussion: Assumption 6 strengthens the conditional median independence of assumption 3 (ii) to full distributional independence. Assumption 7 strengthens the support condition above to require that there are no "holes" in the (conditional on $G = g$) support of $\mu(X,g)$. These two assumptions enable us to compute the (counterfactual) probabilities $\Pr_{\varepsilon|X=x,G=g}\left[\tilde{\gamma} \leq \mu(x,m) - \varepsilon < \gamma_g\right]$ for $G = m$ and $G = f$, which are ingredients of the welfare change calculations, as illustrated below. Assumption 8 says that two candidates with identical values of X and G who are offered admission have the same expected outcome, irrespective of the offer generating mechanism. This assumption is implied by the stronger condition that ε is a purely officer-specific error and applicant files are randomly allocated among admission officers. Note that assumption 8 still allows acceptance to be correlated with (X,G) and with Y and thus includes cases where more able students (who also have higher test-scores on average) get outside offers and do not accept the university's offer or less able students cannot satisfy the conditions of the offer and tend to accept the current offer.

Welfare bounds: For $x \in M$, define $\tilde{x}(x)$ as the solution to: $\mu(\tilde{x}(x), m) = \mu(x, m) + \gamma_m - \tilde{\gamma}$. For $x \in F$ define $\bar{x}(x)$ as the solution to $\mu(\bar{x}(x), f) = \mu(x, f) + \gamma_f - \tilde{\gamma}$. Since $\gamma_f \leq \tilde{\gamma} \leq \gamma_m$, by assumption 7 , such an $\tilde{x}(x)$ must necessarily exist in $\mathcal{X}_m$ and $\bar{x}(x)$ in $\mathcal{X}_f$.

- $\omega_m(\tilde{\gamma}) \in [\omega_{ml}(\tilde{\gamma}), \omega_{mu}(\tilde{\gamma})]$ with

$$\begin{aligned}\omega_{mu}(\tilde{\gamma}) &= \pi_m \int_{x \in M} \mu_m(x)\alpha_m(x)\{p_m(\tilde{x}(x)) - p_m(x)\} dF(x|G=m) \\ &\quad + \pi_m \int_{x \in \mathcal{X}_m \cap M^c} \mu_m(x)\alpha_m(x)\{1 - p_m(x)\} dF(x|G=m),\end{aligned}$$

$$\begin{aligned}\omega_{ml}(\tilde{\gamma}) &= \pi_m \int_{x \in M} \mu_m(x)\alpha_m(x)\{p_m(\tilde{x}(x)) - p_m(x)\} dF(x|G=m) \\ &\quad + \pi_m \int_{x \in \mathcal{X}_m \cap M^c} \mu_m(x)\alpha_m(x)\left\{\max_{z \in \mathcal{X}_m} p_m(z) - p_m(x)\right\} dF(x|G=m).\end{aligned}$$

- $\omega_f(\tilde{\gamma}) \in [\omega_{fl}(\tilde{\gamma}), \omega_{fu}(\tilde{\gamma})]$

$$\begin{aligned}&\omega_{fl}(\tilde{\gamma}) \\ &= (1-\pi_m) \int_{x \in F} \mu_f(x)\alpha_f(x)\{p_f(x) - p_f(\bar{x}(x))\} dF(x|G=f) \\ &\quad + \pi_m \int_{x \in \mathcal{X}_f \cap F^c} \mu_f(x)\alpha_f(x)\left\{p_f(x) - \min_{z \in \mathcal{X}_f} p_f(z)\right\} dF(x|G=f).\end{aligned}$$

$$\begin{aligned}&\omega_{fu}(\tilde{\gamma}) \\ &= (1-\pi_m) \int_{x \in F} \mu_f(x)\alpha_f(x)\{p_f(x) - p_f(\bar{x}(x))\} dF(x|G=f) \\ &\quad + \pi_m \int_{x \in \mathcal{X}_f \cap F^c} \mu_f(x)\alpha_f(x)p_f(x) dF(x|G=f).\end{aligned}$$

- If $\min_{z \in \mathcal{X}_f} p_f(z) = 0$ and $\max_{z \in \mathcal{X}_m} p_m(z) = 1$, then the bounds shrink to singletons

$$\begin{aligned}\omega_m(\tilde{\gamma}) &= \omega_{mu}(\tilde{\gamma}) = \omega_{mu}(\tilde{\gamma}) \\ \omega_f(\tilde{\gamma}) &= \omega_{fl}(\tilde{\gamma}) = \omega_{fu}(\tilde{\gamma}).\end{aligned}$$

6 Empirical Application

Our application is based on admission data from Oxford University's flagship undergraduate degree in Philosophy, Politics and Economics (PPE)...Admission process...

Our current data cover 3 cohorts of Oxford PPE students who started their study in 2008, 2009 and 2010 respectively. We focus on those applicants who had taken the TSA (an aptitude test comparable with the SAT in the US), and the GCSE – the standardized school-leaving examination in the UK and those who were deemed strong enough to be called to interview. Table 0 summarizes for each gender the key variables. Of these, col1inter is the interview score at admissions, got_in is a dummy for whether the candidate was offered admission, independent_school is a dummy for whether the applicant was educated at a fee-paying (i.e., private) school. The variable prelim_tot is the overall score after the first year examinations, available for admitted candidates only, which will be our outcome of interest.

Table 0 shows that many more males apply to read PPE, males applicants have better TSA distributions, worse interview scores and male admits score an average of 4 points (out of 300) higher in the first year exams. These differences are statistically significant at 5%. There is no significant difference in acceptance rates between male and female candidates.

We now test if the marginal admitted male and the marginal admitted female student have identical expected first year scores. To do this, for each gender, we compute the expected score as a linear function of age, GCSE score, overall TSA, interview score and whether the applicant came from an independent school. Using the 0 median restriction on expectation errors, as explained above, we calculate the threshold faced by each gender as the average of expected first year scores for admitted applicants whose probability of being admitted is predicted (through a probit) to be close to 0.5. To choose the bandwidth for defining "closeness", we use the leave-one-out cross-validation. The CV criterion is plotted in figure 2 for the four cases of male/female and state-school/independent-school. The numerical minimum of this criterion over a grid of 0.01 to 0.1 is taken to be the optimal bandwidth. Table 1 lists the optimal bandwidth for various categories.

In table 2, we show the estimated difference for a range of bandwidths (which define "closeness to 0.5") and an Epanechnikov kernel. The middle bandwidth is the optimal one, according to the cross validation criterion calculated above. The other entries are 0.5 times the optimal one and 10 times the optimal one, respectively. This table strongly suggests that male applicants face a threshold that is higher than that faced by female applicants by about 6-8 points out of 300. The p-value corresponds to a one-sided test of no gender

difference in thresholds against a one-sided alternative of a positive male-female difference. From table 2, it becomes clear that although male and female acceptance rates are nearly equal, male candidates have to cross a higher barrier of expected performance than females in order to get admitted.

We then repeat the analysis reversing the roles of gender and school background, i.e., we use gender as an explanatory variable and test if applicants from independent school face a higher threshold than their counterparts who apply from state-funded schools. The results are reported in table 3. Now, we see that there is hardly any difference in the thresholds faced by the two groups. Although the estimated thresholds for private school applicants are higher, the difference is not statistically significant.

7 Conclusion:

To follow...

Table 0

	Female			Male			
Variable	# Obs	Mean	Std. Dev.	# Obs	Mean	Std. Dev.	pvalue
age	609	17.99	1.24	923	18.12	1.62	0.99
gcsescore	609	3.82	0.26	923	3.75	0.35	0.0001
tsaoverall	609	61.96	7.72	923	65.46	7.94	0.00001
col1inter	609	62.46	11.03	923	63.46	11.44	0.0214
indep_school	609	0.34	0.47	923	0.39	0.49	0.4777
got_in	609	0.28	0.45	923	0.31	0.46	0.199
prelim_tot	150	183.71	15.76	247	187.62	19.05	0.0392

Table 1:Optimal bandwidth

State	0.451
Indep	0.11
Male	0.46
Female	0.146

Table 2: Threshold Difference by Gender

scale	male thld	female thld	male-fem	p-value
0.50	185.76	176.4	9.36	0.01
1.00	182.44	173.58	8.86	0.03
10.00	186.88	178.79	8.09	0.00

Table 3: Threshold Difference by high_school type

scale	Indep thld	state thld	indep-state	pvalue
0.50	183.00	179.75	3.25	0.22
1.00	179.79	177.16	2.63	0.37
10.00	181.63	177.16	4.47	0.54

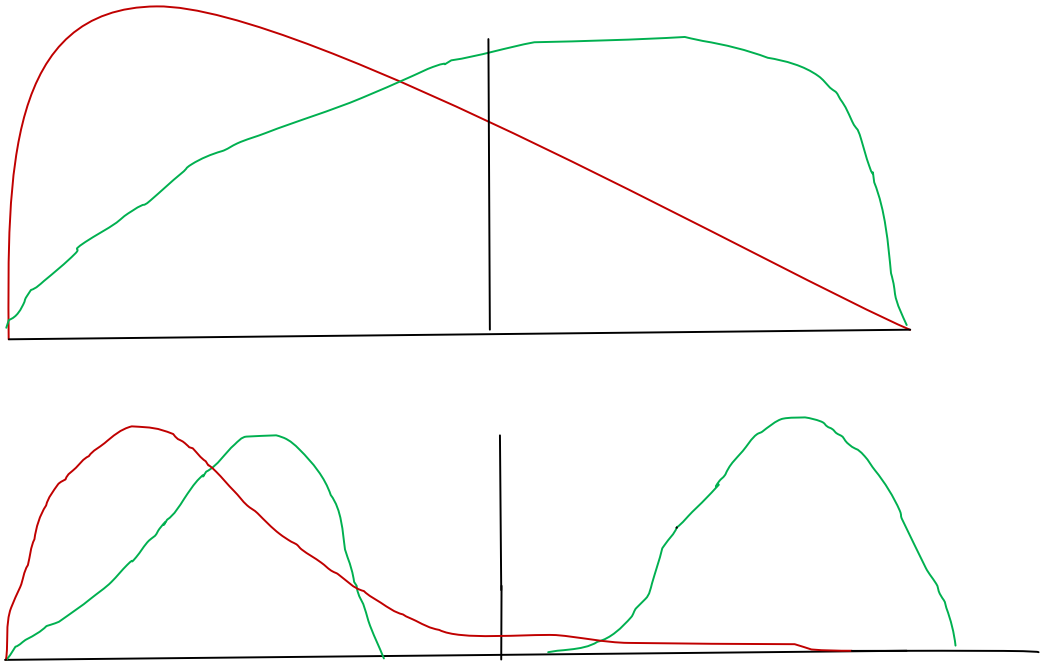


Figure 1: The red curve represents a fictitious conditional density of μ(X,male) and the green curve the density of μ(X,female). In the top panel, they have the same support and the common treatment threshold gamma is shown by the vertical line. In the bottom panel, the common threshold lies in the "hole" of the support of μ(X,female). So there is no x in the support of X for females where μ(X,female) can equal the common threshold.

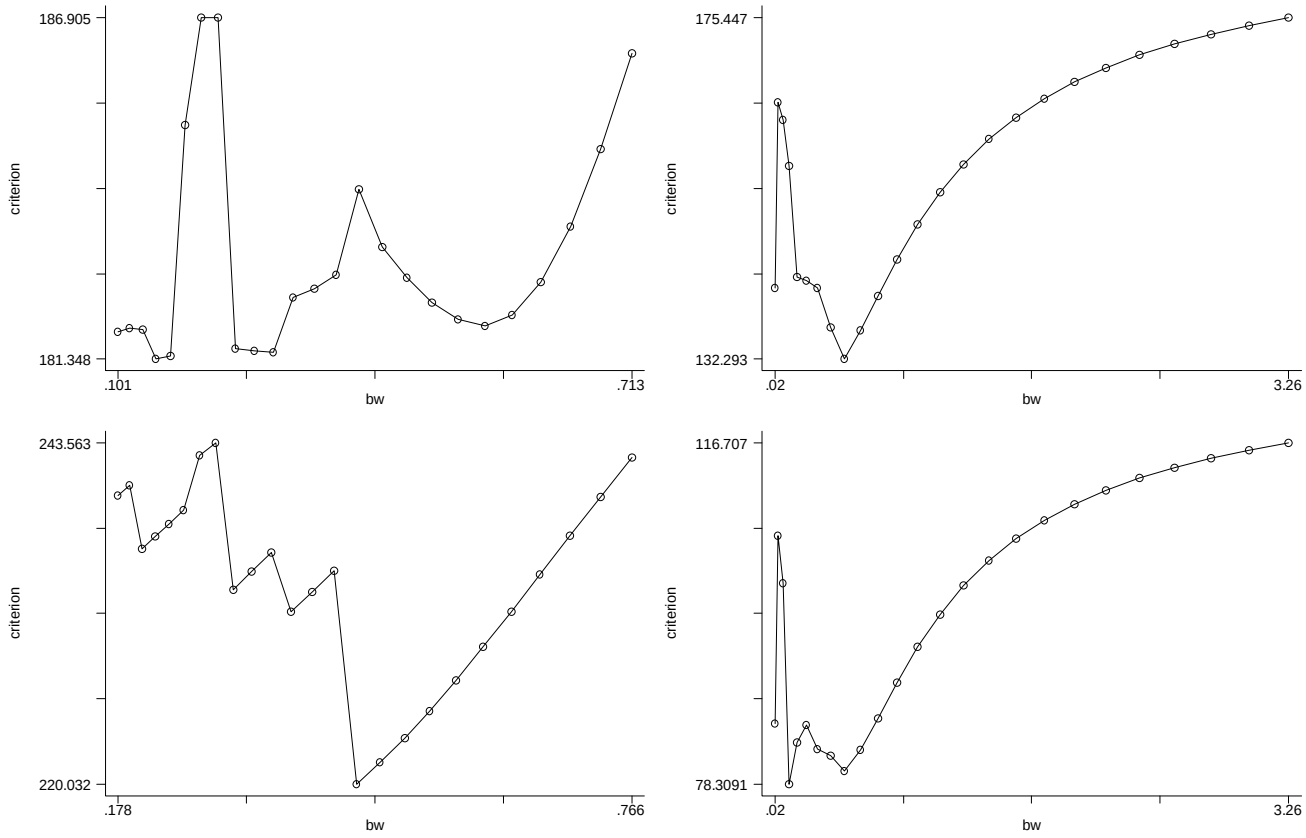


Figure 2: Graphs showing cross-validation criterion for (clockwise from top left) female, male, independent and state school applicants, respectively.

References

[1] Ahn, H. & J.L. Powell (1993) Semiparametric estimation of censored selection models with a nonparametric selection mechanism, *Journal of Econometrics* 53, 3-29.

[2] Becker, Gary (1957). The economics of discrimination, University of Chicago Press.

[3] Bhattacharya, D. (2009). Inferring Optimal Peer Assignment from Experimental Data. Journal of the American Statistical Association, Jun 2009, Vol. 104, No. 486: 486–500.

[4] Bhattacharya, D. & Dupas, P. (2010). Inferring Efficient Treatment Assignment under Budget Constraints", NBER working paper number 14447.

[5] Bouezmarni, T. & O. Scaillet (2005) Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data, *Econometric Theory* 21, 390-412.

[6] Gospodinov, N. & M. Hirukawa (2010) Nonparametric estimation of scalar diffusion processes of interest rates using asymmetric kernels, Working Paper 08-011, Department of Economics, Concordia University.

[7] Heckman, J. (1998). Detecting discrimination, Journal of Economic Perspectives-Volume 12, Number 2, 101-116.

[8] Ichimura, H. & P.E. Todd (2007) Implementing nonparametric and semiparametric estimators, *Handbook of Econometrics* Vol. 6-2, Ch. 74, 5369-5468, Ed. by J.J. Heckman and E.E. Leamer, North-Holland.

[9] Kanaya, S & D. Kristensen (2011) Optimal Sampling and Bandwidth Selection for Kernel Estimators of Diffusion Processes, Working paper, University of Oxford.

[10] Kim T.Y. and D.D. Cox (1996) Bandwidth Selection in Kernel Smoothing of Time Series, *Journal of Time Series Analysis* 17, 49-63.

[11] Kobrin et al (2008). Validity of the SAT for Predicting First-year College Grade Point Average, College Board, New York.

[12] Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. Psychological Bulletin, 127, 162-181.

[13] Li, Q. & J.S. Racine (2007) *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.

[14] Lindskog, F., McNeil, A., & Schmock, U. (2004). Kendalls tau for elliptical distributions.

[15] Mack, Y.P. (1981) Local Properties of k-NN Regression Estimates, *SIAM Journal on Algebraic and Discrete Methods* 2-3, 311-323.

[16] Marron, J. S. & W. Härdle (1986) Random Approximations to Some Measures of Accuracy in Nonparametric Curve Estimation, *Journal of Multivariate Analysis* 20, 91-113.

[17] Masry, E. (1996) Multivariate local polynomial regression for time series: uniform strong consistency and rates, *Journal of Time Series Analysis* 17, 571-599.

[18] Pagan, A. & A. Ullah (1999) *Nonparametric Econometrics*. Cambridge University Press.

[19] Parks, G. (2011). Academic Performance of International baccalaureate students at Cambridge, by school, http://www.cam.ac.uk/admissions/undergraduate/research/docs/ib_performance.pdf

[20] Persico, N (2009). Racial Profiling? Detecting Bias Using Statistical Evidence. Annual Review of Economics, volume 1.

[21] Rothstein (2004). College Performance Predictions and the SAT, Journal of Econometrics 121(1-2), July-August 2004, 297-317.

[22] Sawyer, R. (1996). Decision Theory Models for Validating Course Placement Tests. Journal of Educational Measurement, Vol. 33, Issue 3, 271-290.

[23] Sawyer, R. (2010). Usefulness of High School Average and ACT Scores in Making College Admission Decisions, available online at: http://www.act.org/research/researchers/reports/pdf/ACT_RR2010-2.pdf

[24] Sackett, P., Kuncel, N., Arneson, J., Cooper, G., Waters, S. (2009). Socioeconomic Status and the Relationship Between the SAT and Freshman GPA - An Analysis of Data from 41 Colleges and Universities, available online at: http://professionals.collegeboard.com/data-reports-research/cb/SES-SAT-FreshmanGPA