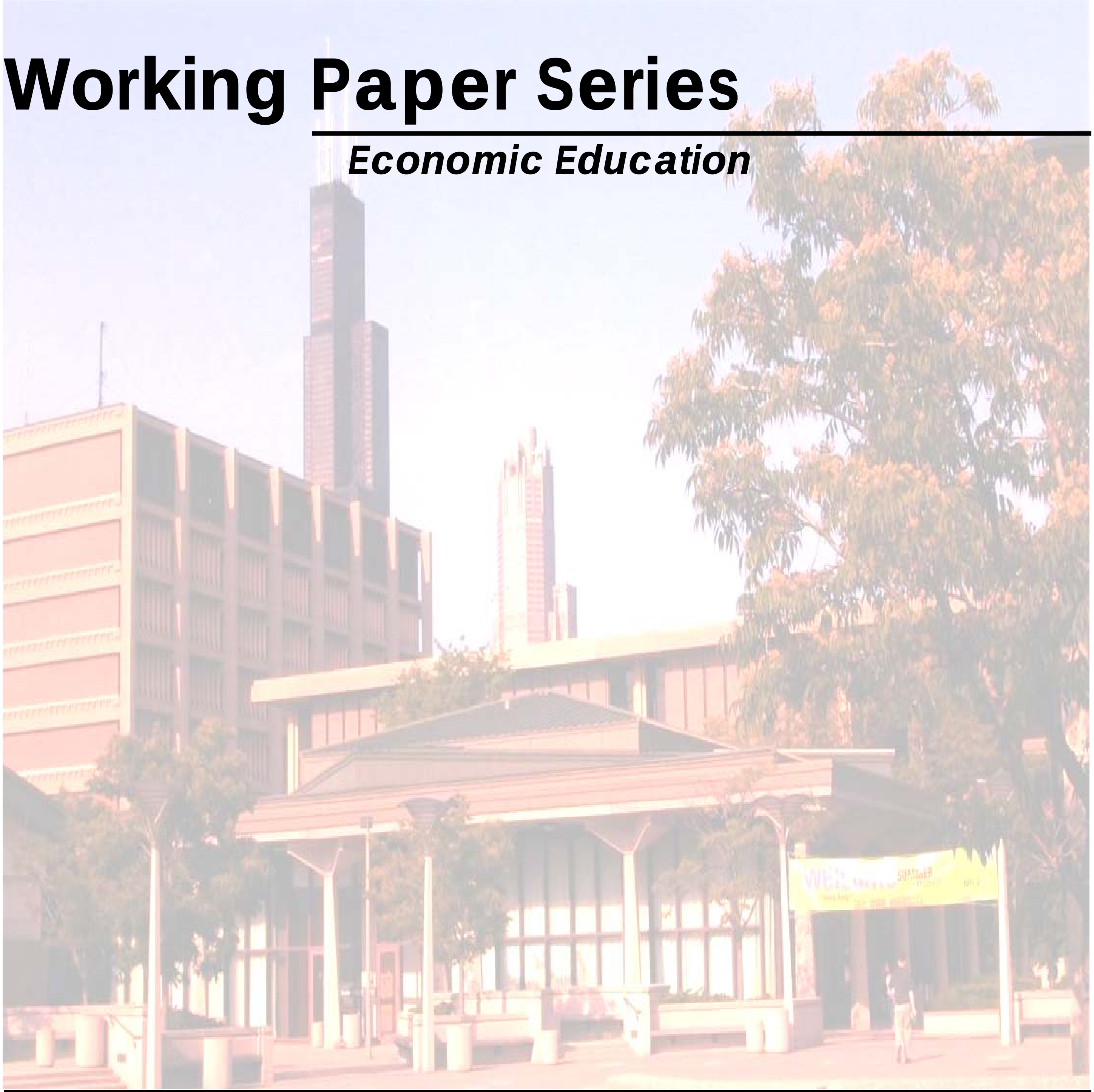
Working Paper Series

Economic Education



No. 004

New Selection Indices for University Admissions: A Quantile Approach

Mo-Yin S. Tam, Professor
University of Illinois at Chicago

Gilbert W. Bassett Jr.
University of Illinois at Chicago

Uday Sukhatme
University at Buffalo, SUNY

Center for Economic Education

Department of Economics (m/c144)
University of Illinois at Chicago
601 South Morgan Street
Chicago, Illinois 60607
Tel 312-996-2684
Fax 312-996-3344
http://cee.econ.uic.edu

New Selection Indices for University Admissions: A Quantile Approach

Mo-Yin S. Tam, Gilbert W. Bassett Jr., Uday Sukhatme

Abstract. All universities seek to admit a freshmen cohort of a specified size who will be successful in graduating from college. Previous work has shown a high correlation between students' eventual success and their ability to do well in their first term. This objective translates into admitting students whose first term grade point average (GPA) exceeds a minimum acceptable level. Current admissions decisions are based however on a selection index that is constructed from a GPA regression that estimates expected GPA. We consider a new approach in which the selection index is based directly on the quantiles of the GPA distribution. The new approach realistically assumes that student characteristics have differential impacts at different parts of the GPA distribution. Since impacts usually vary for the low, middle and upper parts of the GPA distribution, the quantile approach provides additional information. The quantile method also provides admissions officers the flexibility to target different GPA properties for the freshman class. For example, they can directly implement a criterion that selects students whose characteristics imply a maximal probability of a first term GPA of better than any specified value. We illustrate the quantile method by application to actual admission practices at the University of Illinois at Chicago.

1. Introduction

Consider a university's problem of choosing its freshmen cohort of students from a pool of N applicants. It is desired to select n=rN students from the pool, where (for simplicity) it is assumed accepted students actually enroll. In addition the university wants to select students who will be successful and graduate.[1] Previous work has shown that success is associated with achieving a first semester grade point average (GPA) that exceeds a critical threshold, call it g_0.[2] Students with a GPA<g_0 are unlikely to ever graduate and those with a GPA>g_0 are likely to continue to do well in their studies. The selection objective is to admit students with a high probability of a GPA greater than g_0.

Let X be a vector of applicant characteristics. We denote the proportion of applicants with characteristics X by h(X).

Let GPA(X) denote the (random variable), first term grade point average. Let the associated cumulative distribution function be denoted by F(z|X). F(z|X) is the probability that an applicant with characteristics X would, if admitted, have a GPA≤z. The above mentioned success criterion translates into selecting students so that $F(g_0|X)=Pr(GPA(X)<g_0)$ is as small as possible, or $Pr(GPA(X)>g_0)$ is as large as possible.

In the admissions literature a selection index has come to be defined as a scoring rule mapping applicant characteristics X to some statistic of the GPA(X) distribution. Current admission procedure is usually based on an index equal to *expected* GPA. Let us denote this, $SI^E(X)$ (so $SI^E(X)=E(GPA(X))$).[3] Students are then admitted if, $SI^E(X)>g$ where g is chosen to satisfy the enrollment (n=rN) constraint. We will call this mapping from applicants to admitted students, SM^E. While SM^E maximizes the minimum expected GPA of freshmen students (given the enrollment constraint) it need *not* be consistent with the objective of maximizing the chance that admitted students achieve a certain threshold GPA.[4]

This paper presents a new approach to the admission decision problem. Instead of being based on E(GPA(X)) (or on the implicit assumption that X affects GPA in an identical way at all parts of the GPA distribution), it is based on explicitly maximizing the chance that admitted students achieve g_0. It is derived from a conditional quantile model,[5] which is used to construct a selection index achieving the target enrollment while maximizing the

---

[1] In many public universities, there might be significant number of students who would leave the university when they entered as freshmen in good standing and transfer to another university. These students would also be counted as being successful in college.

[2] See Tam and Sukhatme (2002).

[3] This is done in many public universities including the University of Illinois, see [].

[4] Note that while the threshold GPA described earlier is selected to ensure eventual success of the students, the cutoff value of the SI is to picked to satisfy the constraint of the size of the admitted students.

[5] For a discussion of the quantile model, see Koenker and Bassett (1978) and Koenker and Hallock (2001).

probability of success. It does not assume that the impacts of student characteristics are constant along the entire GPA distribution. Hence, it is more general and realistic since impacts can vary for the low, middle and upper parts of the GPA distribution. Since the quantile model maximizes the chance of being successful, it exactly captures the intention of an admission policy that seeks to admit students whose GPA will be greater than g_0.

Section 2 presents the linear quantile and expectation models. In that section, attention is restricted to the two characteristics, namely ACT score and high school percentile rank (HSPR), which have traditionally been used to construct a selection index. The models are then used to express the corresponding selection indices and selection mechanisms. Section 3 presents estimates for the model parameters using the 1994 UIC (University of Illinois at Chicago) data, first for the two characteristics mentioned above. The estimation of the models is then extended to include a third variable, the average high school ACT score (HSACT) as an indicator of the quality of high school.[6] Section 4 uses the UIC data to illustrate differences in the characteristics of the admitted classes using the alternative approaches. Discussion of the results and topics for additional research is in the concluding section.

2. The GPA(X) Model, Selection Indices, Selection Methods

2.1 GPA(X) Model

Given F(z|X), the associated quantile (inverse) function is denoted by Q(θ|X), 0≤θ≤1. It will be assumed that GPA(X) is determined by a linear quantile model,

$$Q(\theta|X)=\alpha(\theta)+ \beta_{ACT}(\theta)\ ACT + \beta_{HSPR}(\theta)\ HSPR \qquad 0\le\theta\le1 \qquad (1)$$

The set of possible (ACT, HSPR) comes from a discrete set with ACT values 12, 13,…,36, and HSPR takes values, 1,2,…,100.

This quantile specification is more flexible than the more common conditional expectation model. With the standard approach the scale of GPA does not vary with the quantiles and,

$$E[GPA(X)] = \alpha^E + \beta^E_{ACT}\ ACT + \beta^E_{HSPR}\ HSPR \qquad (2)$$

This model takes the impacts of student characteristics on GPA to be uniform over the GPA distribution, identical to the impact at the E(GPA(X). This is a special case of the quantile model in which there are no quantile effects, $\alpha^E = \alpha(\theta)$, $\beta_{ACT}(\theta) = \beta^E_{ACT}$ and $\beta_{HSPR}(\theta) = \beta^E_{HSPR}$ for all θ. Under the general quantile model the coefficients on ACT and HSPR depend on θ. This means the impacts of student characteristics on GPA can vary for the low, middle and upper parts of the GPA distribution.

[6] Tam and Sukhatme (2002) find that HSACT has significant impact on GPA.

2.2. Selection Indices and Selection Mechanisms

The admission literature refers to a *selection index* as a mapping from X to some statistic of the GPA(X) distribution. Students are then admitted if their selection index value is high enough. The standard index is, $SI^{E}(X) = E[GPA(X)]$.

The associated *selection mechanism* will be denoted by SM^E. This maps applicants to the set of admitted students where: for all admitted students, $SI^E(X)>s^E$, and s^E is determined so that

$$\sum_{X|SI^E(X)>s^E} h(X) = \frac{n}{N} = r\,. \qquad (3)$$

That is, there are n students in the admitted class.

The selection indices proposed in this paper are based on a different feature of GPA than its expected value. We consider a selection index defined by the θ quantile, $SI^{\theta}(X) = Q(\theta|X)$, $0 \le \theta \le 1$. Note that $SI^{\theta}(X)$ represents a set of selection indices, one for each value of θ. Each of the selection indices $SI^{\theta}(X)$ maps X to the quantile value of GPA given by (2).

The selection mechanism corresponding to the quantiles will be designated by a GPA value, g_0. SM^{g0} is defined as the mapping from applicants to an admitted class so that the probability of exceeding g_0 is as large as possible subject to an entering class of n students. This means, $\Pr(GPA(X)>g_0)=1-F(g_0|X)=1-\theta_0$ is as large as possible, subject to $\sum_{X|SI^{\theta_0}(X)>g_0} h(X) = \frac{n}{N} = r$. It is clear that the value of θ_0 in the selection index depends on the level of the threshold GPA level, g_0. Specifically, the larger the threshold GPA (g_0),the larger is the value of θ_0 [and the smaller is $(1-\theta_0)$], or the smaller the probability of exceeding g_0.

3. Conditional Expectation and Quantile Estimates

We use student data of the University of Illinois at Chicago to show how the impacts of student characteristics on GPA vary at different points of the GPA distribution. We also compare the expectation selection index approach to the quantile selection index .As indicated above, we will first consider the two-characteristic (ACT and HSPR) models and then extend the analysis to include a third characteristic, HSACT.

Estimates are based on data for the 1994 freshmen cohort of University of Illinois at Chicago (UIC).[7] UIC is a comprehensive 4-year public university with an undergraduate

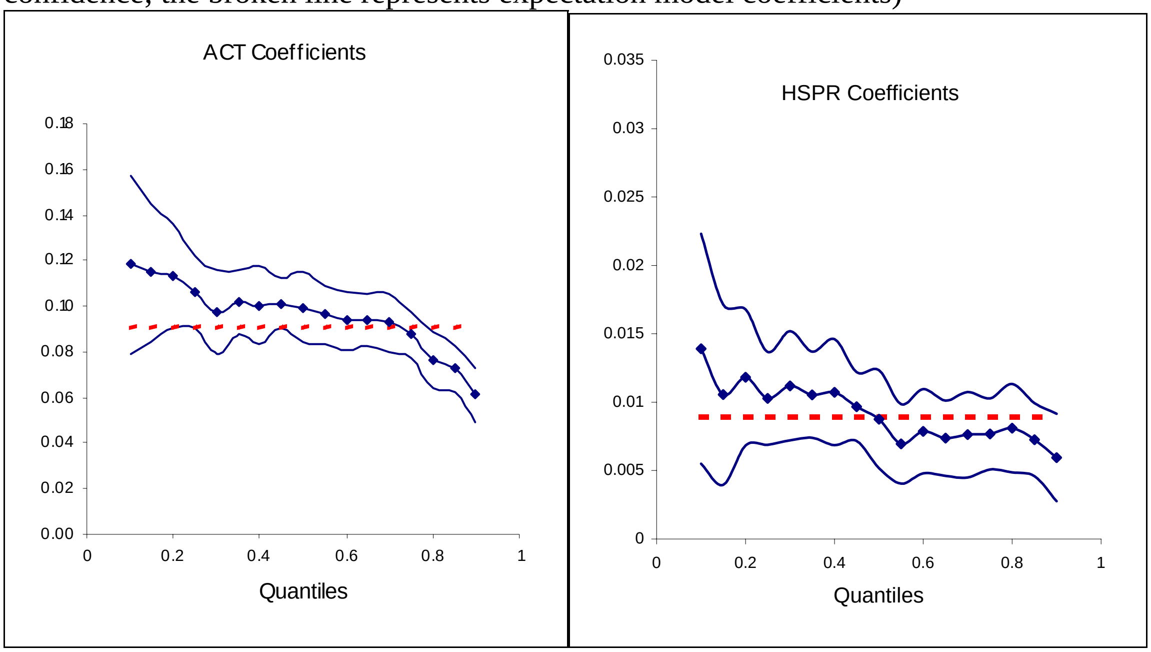
[7] Fall 1994 data is picked because we would like to trace the eventual success of the admitted students for six years since admission.

enrollment of about 16,000 students. Its 1994 freshmen cohort is little over 1,600 students. The average ACT score and HSPR of the admitted students in fall of 1994 are respectively 21 and 74.

3.1 Two-Characteristic Models

The results of the conditional expectation model and quantile regression estimates for the two-characteristic models are illustrated in Figure 1.

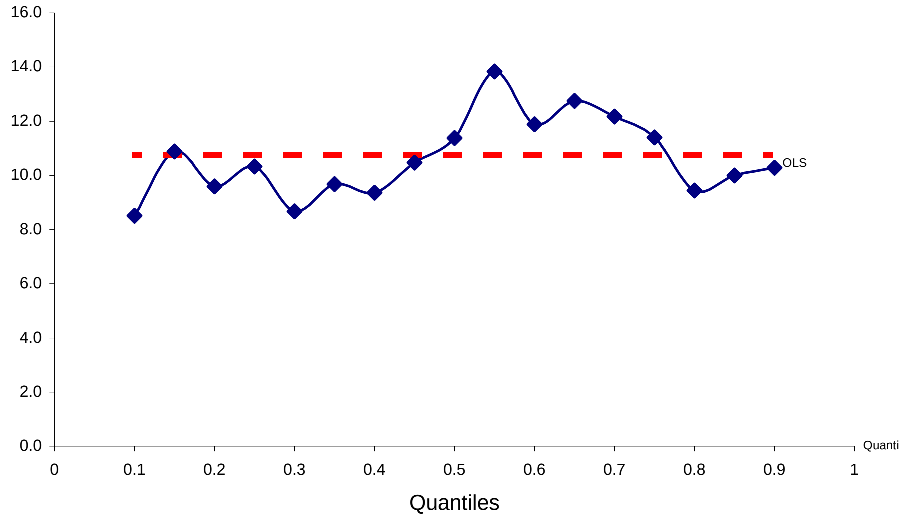
Figure 1: **Expectation and Quantile Model Estimates for the two-Characteristic Models** (solid curves with dotes indicate the quantile coefficients with corresponding 95% confidence; the broken line represents expectation model coefficients)



The figure shows that ACT and HSPR have a positive effect on GPA with quantile coefficients that vary with θ. For both the ACT and HSPR variables, they are larger at the lower quantiles. The impacts of student characteristics on GPA are therefore greater at the lower GPA levels. However, the variations are not large and the expectation model coefficients are within the 95% confidence intervals for most of the quantile coefficients.

A different view of the coefficients is presented in Figure 2, which shows the ACT/HSPR tradeoffs for $SI^{\theta}(X)$ as a function of θ. At the expected value the ratio of 10.8 indicates one ACT point is worth about 11 points of class rank for SI^{E}. The tradeoffs at the quantiles vary slightly from the expectation, and are largest in the middle of the distribution.

Figure 2: **ACT/HSPR Tradeoffs** (solid curves with dotes indicate the quantile ACT/HSPR tradeoffs; the broken line represents expectation model ACT/HSPR tradeoff)



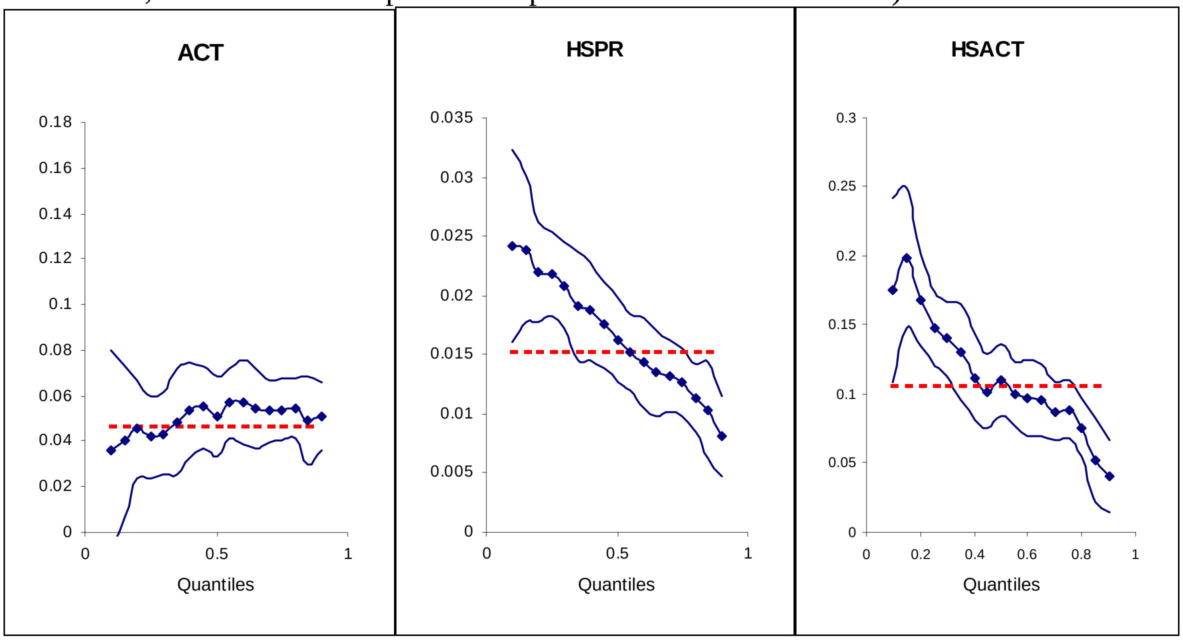
3.2 OTHER CHARACTERISTICS

Other student characteristics and their impact on GPA have been examined in the literature.[8] Tam and Sukhatme (2002) have found that the average high school ACT score (HSACT) has significant impact on GPA.[9] In this section we include this variable in the comparison of the expectation and quantile estimates, again using the UIC data. The results are illustrated in Figure 3.

[8] These include studies of the academic rigor of a student's high school program [Young & Barrett, (1992)], academic records [Touron, (1983)], mathematics testing [Haeck, Yeld, Conradie, Robertson & Shall (1997)] and district performance indices [Bennett, Wesley & Dana-Wesley (1999)].

[9] HSACT is just one of several plausible indicators of high school quality. Other possible quantitative attributes of quality which we do not investigate in this paper include the percentage of students going to college and the number of AP (Advanced Placement) courses offered.

Figure 3: **Expectation and Quantile Model Estimates for the three-Characteristic Models** (solid curves with dotes indicate the quantile coefficients with corresponding 95% confidence; the broken line represents expectation model coefficients)



Figures 2 and 3 for the coefficient estimates of the same characteristics are represented on the same scale for comparison. Controlling for the impact of high school quality (HSACT), ACT has a much smaller impact on GPA. The coefficient estimates for ACT in the three-characteristic model are roughly half of the value of those in the two-characteristic variable. On the other hand, impact of HSPR on GPA increases after high school quality in taken into consideration. However, it remains to be much smaller relatively to the ACT impact.

Among the three characteristics, HSACT has the strongest impact on GPA. The value of the HSACT coefficient estimates is about twice as large as the ACT coefficient estimates and about six times as large as the HSPR coefficient estimates.

When quantile coefficient estimates are compared to the expectation coefficient estimates for the same variable, the variations of quantile coefficient estimates remain small and the expectation model estimate is still within the 95% confidence interval of the quantile estimates for ACT when HSACT is included in the models. For HSPR, however, bigger variations of quantile estimates occur in the three-characteristic quantile model compared to the two-characteristic one. Also, the expectation model coefficient estimate for HSPR lies outside the 95% confidence interval of the quantile estimates at upper and lower values of θ, when HSACT is included in the model. Even larger variations exist among the quantile estimates of HSACT and its expectation estimate is also outside the 95% confidence intervals of the quantile estimates at the upper and lower values of θ.

4. Comparing Selection Methods

In this section the usual selection index is compared to the quantile-based methods. We consider a real-world application and show how the applicants would have differed at UIC if the quantile-based methods had been used in place of the standard expectation based approach.

The outcome of a selection method depends on the distribution of characteristics in the applicant population. Since ACT and HSPR are traditionally used in the construction of SI, we only consider the two-characteristic model. For the simulation we took the distribution of applicant (ACT, HSPR) to be normal, but truncated so that ACT was in 15 to 36, and HSPR was in in 30 to 100. The parameters of the distribution were calculated so that they yield an admitted class (with the standard selection index) whose summary statistics for GPA, ACT, and HSPR match the values of the actual admitted class at UIC in 1994.[10]

We present results for r=.48; similar results were obtained using r=.74, and r=.89. The parameters of the conditional expected and quantile models are the ones presented in the previous section.

The base case is SM^E. The selection index with r=.48 then leads to a selection rule such that, SI^E=E(GPA(X))>3.0. All students with E(GPA(X))>3.0 are admitted (and those with E(GPA)<3.0 are rejected).

Quantile-based selection, SM^g is considered for g values of 2.5, 3.0 and 3.5. Each achieves the same proportion of admitted applicants, equal to r=.48. The admission target is achieved in such a way that an admitted student has, Pr(GPA>g)>(1-θ) where g is alternatively, 2.5, 3.0, 3.5, and (1-θ) is as large as possible.

Table 1 shows, for each g, the quantile value determined by each SM^g. For example, the .485.515 value in the first column for $SM^{3.0}$ means, Pr(GPA>3.0)>51.5%, for all admitted students. Similarly, the .300 value in the first column for $SM^{3.5}$ means that PR(GPA>3.5)>30%. The greater the threshold GPA, the lower is the chance of the admitted students having GPA above that threshold.

From the point of view of admitting students whose probability of exceeding a g-threshold is as large a possible the SM^g method is optimal, and the SM^E makes two types of mistakes. Comparing the SM^E with the $SM^{3.5}$, some students are rejected under SM^E even though their probability of a 3.0 is greater than 51.1%. The other type of mistake occurs when students are accepted by SM^E even though their probability of a 3.0 is below 51.5%.

One measure of the size of these mistakes is shown in the last two columns of Table 1. The value in the column, Max Pr(SM^E|Reject), shows the highest probability of exceeding g for students who are (incorrectly) rejected by SM^E. The value in, Min Pr(SM^E|Accept), show the lowest probability for students (incorrectly) accepted.

[10] The correlation of ACT and HSPR among admitted UIC students was .04 and for the example ACT and HSPR were taken to be uncorrelated.

At the 3.5-threshold, for example, $SM^{3.5}$ has every student with a better than 30% chance getting a 3.5 or better, whereas as SM^E will admit some students whose chance of success is only 29% and reject some students whose chance of success is 35%.

Table 1

	Optimal	Max Pr(SM^E|Reject)	Min Pr(SM^E|Admit)
$SM^{2.5}$	.700	.72	.69
$SM^{3.0}$	.515	.65	.50
$SMI^{3.5}$	.300	.35	.29

In spite of the differences in Table1, the overall impact on the admitted class using SM^E and SM^g is not too large for the UIC example. This is seen in the Table2, which shows the extent to which the quantile-based and expectation-based methods result in different cohorts of students. The results are similar at the different g values, and only $SM^{3.5}$ presented. It shows the proportions of applicants who are admitted with $SM^{3.5}$ but rejected by SI^E, and conversely. (The off-diagonal proportions are not quite identical due to the discrete values of ACT and HSPR used for the example). It shows a difference of 1.0% relative to all applicants, or about 2% relative t the size of the admitted class, since r=.48.

Table 2

		Proportions		
		$SM^{3.5}$		
		Admit	Reject	
SM^E	Admit	0.732	0.010	0.742
	Reject	0.013	0.245	0.258
		0.745	0.255	

The small difference reflects the fact that there are not dramatic differences in the conditional quantile coefficients as illustrated in Figure 2. It also occurs because the probability differences shown in Table 2 tend to occur for only a small proportion of the applicant population.

When HSACT is included in the models, bigger variations occur in the conditional quantile coefficients for HSPR and HSACT as illustrated in Figure 3. Hence bigger impact on the admitted class is expected using SM^E and SM^g for the three-characteristic model.

5. Discussion

The selection methods considered in this paper are defined by the requirement that *all* students meet a minimal GPA standard whether it be an expected or quantile value. Such methods may be compared to those designed to achieve objectives for the unconditional GPA distribution of the admitted class. This distinction does not usually arise because with the standard methods there is no difference.

Under SM^E all admitted students have an expected GPA greater than or equal to s^E (and $E(GPA(X) < s^E$ for all rejected students). It follows that SM^E generates an entering class of n students with maximal GPA. (Any alternative SM that admits n students requires replacing a student whose $E(GPA(X) \geq s^E$ with one whose $E(GPA(X) < s^E$, thus lowering overall GPA). The corresponding situation for the quantiles is more complicated.

Let L(z) denote the cumulative GPA distribution of the entering class. (We suppress the fact that this distribution depends on r, SM, h, and F(z|X)). This unconditional distribution is obtained by integrating GPA over the distribution of characteristics of admitted students . This can be written as,

$$L(z)=\sum_x I\{SI(x) > s\}h(x)F(z \mid x) \qquad (4)$$

where I{} is the indicator function.

Consider, SM^g, so, $F(g|X)>\theta_g$ for all admitted students. Plugging values into (4) gives the resulting GPA distribution of the entering class. Does this yield a class of n students so that 1-L(g) is as large as possible? Is there some g' such that $SM^{g'}$ makes 1-L(g) as large as possible?

The answer to both questions is, "no". The problem is that the unconditional quantiles associated with L cannot be simply expressed in terms of the conditional quantiles associated with F(z |X). This does not mean that we cannot maximize 1-L for a given g and r, but that in practical applications numerical methods are required. Still, a topic for further investigation involves the relation between the selection methods described in the paper and selection methods designed to optimize properties of the unconditional GPA distribution of admitted students.

References

Bennett, D. T., Wesley, H. & Dana-Wesley, M. (1999). Planning for Imminent Change in College Admissions: Research on Alternative Admissions Criteria. *Journal of College Student Retention,* 1, 83-92.

Haeck, W., Yeld, N., Conradie, J., Robertson, N. & Shall, A. (1997). A Developmental Approach to Mathematics Testing for University Admissions and Course Placement. *Educational Studies in Mathematics*, 33, 71-92.

Koenker, R. & Bassett, G. (1978). Regression Quantiles. *Econometrica*, 46, 33-50.

Koenker, R. & Hallock K. (2001) Quantile Regression: an Introduction, *Journal of Economic Perspective*, Fall, 15-4, 143-161.

Stern, D. & Briggs, D. (2001). Changing Admissions Policies: Mounting Pressures, New Developments, Key Questions. *Change*, 33, 34-41.

Tam, M.S. & Sukhatme U. (2002).The Importance of High School Quality in University Admission Decisions. Working Paper.

Touron, J. (1983). The Determination of Factors Related to Academic Achievement in the University: Implications for the Selection and Counseling of Students. *Higher Education*, 12, 399-410.

Young, J. & Barrett, C. (1992). Analyzing High School Transcripts to Improve Prediction of College Performance. *Journal of College Admissions*, 137 , 25-29.