CENTER FOR LABOR ECONOMICS
UNIVERSITY OF CALIFORNIA, BERKELEY
WORKING PAPER NO. 45

College Performance Predictions and the SAT*

Jesse M. Rothstein
Department of Economics
University of California, Berkeley
jrothst@econ.berkeley.edu

March 2002

Abstract

Previous studies of the SAT's contribution to predictions of academic performance ignore many of the other variables available for prediction and are consistent only under unrealistic sample selection assumptions. I propose a new estimator of the SAT's contribution that is consistent under plausible assumptions and I include high school demographic variables as predictors. In University of California (UC) data, the SAT's contribution is about 75% lower than the usual estimates, a result equally attributable to the two innovations. One important implication is that the SAT functions as a proxy for omitted background characteristics in sparse prediction models, and that this serves to inflate the SAT's apparent contribution. An application evaluates the UC's new "Four Percent Plan," which enhances academic quality but about which the usual models are overly optimistic.

JEL Codes: C24, C53, I20, J78
Keywords: *SAT validity, college admissions, sample selection, omitted variables*

---

* I thank Jared Bernstein, David Lee, Darren Lubotsky, Justin McCrary, Jack Porter, Eugene Smolensky, the participants of the Berkeley and Princeton Labor Lunches, and especially David Card for helpful comments. Errors are my own. This material is based upon work supported under a National Science Foundation Graduate Research Fellowship and a grant from the University of California's Center for Studies of Higher Education (CSHE). Opinions, findings, conclusions and recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the NSF or the CSHE.

1 Introduction

The SAT college entrance exam has been widely criticized: as an impediment to minority students' college admission, as too expensive, and as the cause of distortions to high school curricula (Lemann, 1999; Nairn & Associates, 1980; Leonard & Jiang, 1999; Slack & Porter, 1980; Jencks & Crouse, 1982). In recent years, several prominent colleges have de-emphasized the SAT score in admissions. The University of California, for example, recently enacted an admissions rule whereby students ranked near the top of their high school class are admitted without regard to their SAT scores; a proposal to abandon the SAT entirely—replacing it with subject exams tied to California's official high school curriculum—is under consideration (Atkinson, 2001). [1]

The defense of the SAT has been simple: The SAT helps predict academic performance (Camara & Echternacht, 2000; Stricker, 1991; Willingham et al., 1990; Camara, 2001; Izumi, 2001). In this view, the acknowledged correlation between SAT scores and student socioeconomic status is an unfortunate side effect of educational inequality in the United States: Students from disadvantaged backgrounds are simply not as well prepared to succeed in college, and the SAT should be credited, not blamed, for measuring this shortfall (Caperton, 2001). The College Board, publisher of the SAT, has the following description on its web site:

> The SAT I measures developed verbal and math reasoning abilities related to successful performance in college. It provides a standard by which the skills of students applying to colleges and universities can be compared. Studies show that using both SAT I scores together with high school records provides a more accurate prediction of future academic success than using either alone.
> (`http://www.collegeboard.org/sat/html/admissions/about001.html`)

This defense relies heavily on the demonstration that the SAT contributes to predictions. A typical study uses freshman grade point averages of students at a single college as a metric of academic success, demonstrating first that the correlation between SAT scores and freshman grades is large and positive and second that the correlation of both with high school grades does not fully

[1] Throughout, I use "SAT" to denote the widely-known test that was once officially the "Scholastic Aptitude Test" and is, since 1994, the "SAT I." The College Board's SAT testing program also encompasses the lesser-known SAT II subject exams, formerly "Achievement" tests. The UC proposal would in the short term replace the SAT I with the SAT II exams.

account for the bivariate relationship (Breland, 1979; Bridgeman et al., 2000; Ford & Campos, 1977; Willingham et al., 1990).

I argue here that the econometric models in the SAT validity literature are misspecified and misestimated. I focus on two important omissions. Each has the effect of biasing upward the SAT's measured predictive contribution.

First, few analyses control for student characteristics beyond high school grades and racial indicator variables in estimating the SAT's contribution (partial exceptions include Bowen & Bok, 1998; Bridgeman, 1991; Bridgeman et al., 1992; Crouse & Trusheim, 1988; Hunter & Hunter, 1984). College admissions offices have much more information than this about applicants. They observe several nonquantitative indicators of student quality—essays, recommendations, etc.—and, especially at public universities that primarily attract students from a single state, they can easily gather information about students' high schools. If the question of interest concerns the SAT's contribution to colleges' ability to predict academic success, as the above College Board quotation suggests, studies that omit available predictors are likely to overestimate the SAT's importance. I present evidence that the SAT's correlation with future performance arises primarily from its across-high-school variation, and I include school-level predictor variables in the baseline specification to which SAT-based predictions are compared.

A second omission in the SAT validity literature concerns the highly selected samples on which performance prediction models are estimated. Because grading standards are thought to differ across colleges, studies typically use samples drawn from a single college (Breland, 1979). The estimator used in the literature is consistent for population parameters only under unrealistic assumptions about the process by which students are selected into these samples. I propose an alternative estimator that is consistent under more reasonable—although still quite restrictive—assumptions, and I apply this estimator to data in which the required assumptions are plausibly satisfied.

My preferred estimate of the SAT's contribution to prediction is about 75% smaller than is indicated by the models and methods used in the literature, a result approximately equally attributable to my two innovations. I find that the correlation between SAT scores and student demographic variables is far from spurious, but is rather a primary source of the predictive power usually attributed to the SAT.

The relatively saturated prediction models estimated here are not intended to translate directly into admissions policies. Unsurprisingly, students from socioeconomically advantaged high schools are found to outperform students from disadvantaged schools, even controlling for individual high school grades and SAT scores. If estimated academic ability is the sole admissions variable, then, these models would suggest giving explicit advantages in admissions to students from schools with "good" demographics. Typically, of course, admissions offices have other objectives that militate against what might be called "affirmative action for high-SES children."

Nevertheless, even if philosophical or legal constraints rule out admissions based on demographic variables, it is worth examining the SAT's predictive validity in light of the full set of available predictors. The discussion in the literature fails to make clear just how much of the estimated SAT contribution derives from the exclusion of school-level demographic variables. My results suggest a substantially different interpretation of the SAT's role in admissions than is implied in the literature and by the College Board.

The paper proceeds as follows: Section 2 sets out the econometric model underlying grade predictions, introduces my "omitted variables" estimator, and demonstrates its superiority to the usual estimator in selected samples. Section 3 introduces the University of California (UC) data used in my empirical work and describes the construction of a sample in which my estimator is plausibly consistent but the usual estimator is not. Section 4 presents models estimated on this sample, treating the eight UC campuses as a single college. Models in Section 5 disaggregate the data into individual campuses, using Heckman's (1979) two-step estimator with geographic instruments to correct for endogenous campus assignment. Results indicate that parameters vary substantially across campuses. However, the pooled-sample estimates are near the center of the range spanned by the individual campus estimates, justifying my interpretive focus on the former results.

Section 6 discusses, paying special attention to the interpretation of models that include demographic predictor variables considered inappropriate for use in admissions decisions. An Appendix applies the richer prediction models from Sections 4 and 5 to a real-world problem, that of forecasting the academic performance of students admitted under the University of California's new class-rank-based "Four Percent Plan." This plan seems unambiguously to improve the quality of admitted students, but the models estimated here suggest that the effect is smaller than would be implied by the models in the literature.

2 Econometric Model

Assume that college j has access to applicant i's SAT score and to a vector X_i of other characteristics; the college hopes to use this information to predict y_{ij}, a measure of the student's future academic achievement.[2] Let the relationship between predictor variables and latent outcomes satisfy

$$y_{ij} = \alpha_j + SAT_i\beta_j + X_i\gamma_j + \varepsilon_{ij}. \tag{1}$$

This is best understood as a linear projection rather than a causal model: The SAT is more reasonably a proxy for otherwise unobserved ability than a direct cause of y_{ij}.[3] However, I follow the literature and assume that the population relationship is linear, additive and homoskedastic:

[2] I assume throughout that the college attempts to minimize mean squared prediction errors, but this is a drastic simplification. The college probably cares only about Type I and Type II errors in the admissions decision, implying a highly nonlinear objective function: prediction errors that do not cross the admissions threshold are irrelevant to the University's objectives. Semiparametric models and simulations of the admissions decision, not reported here, offer little evidence that the SAT's contribution at the admissions margin differs from its average contribution.

[3] A reasonable justification for the importance of y_{ij} is that the college uses it as a realization of the ability, y_i^*, that is of direct concern. This implies certain restrictions on the parameters of (1)—namely, that β_j/γ_j should not depend on j—and is taken up briefly in Section 5.

$\mathsf{E}\left[\varepsilon_{ij} \mid SAT_i,\, X_i\right] = 0$ and $\mathsf{E}\left[\varepsilon_{ij}^2 \mid SAT_i,\, X_i\right] = \sigma_\varepsilon^2$.

We are interested in how much predictive accuracy would be lost if observations on SAT_i were not available. A reasonable measure of this is the difference between the fit of model (1) and that of a restricted projection that excludes the SAT:

$$y_{ij} = \delta_j + X_i\theta_j + \nu_{ij}. \tag{2}$$

For this projection to be meaningful, it must be that $\mathsf{E}\left[\nu_{ij}\right] = 0$ and $\mathsf{E}\left[X_i'\nu_{ij}\right] = 0$. These imply a relationship between the parameters of (2) and those of (1). Let Σ be the variance-covariance matrix of $(SAT_i,\, X_i)$, partitioned so that $\Sigma_{22} \equiv \mathsf{V}\left[X_i\right]$ and $\Sigma_{21} \equiv \mathsf{cov}\left(X_i',\, SAT_i\right)$. Then

$$\begin{aligned} \delta_j &= \alpha_j + \left(\mathsf{E}\left[SAT_i\right] - \mathsf{E}\left[X_i\right]\Sigma_{22}^{-1}\Sigma_{21}\right)\beta_j, \\ \theta_j &= \gamma_j + \Sigma_{22}^{-1}\Sigma_{21}\beta_j, \text{ and} \\ \nu_{ij} &= \varepsilon_{ij} + \omega_i\beta_j, \end{aligned} \tag{3}$$

where

$$\omega_i \equiv \left(SAT_i - \mathsf{E}\left[SAT_i\right]\right) - \left(X_i - \mathsf{E}\left[X_i\right]\right)\Sigma_{22}^{-1}\Sigma_{21} \tag{4}$$

is the residual from a projection of SAT onto a constant and X.

The fit statistic used in the psychometrics literature is R, the square root of the economist's usual R^2. The SAT's contribution to prediction, then, is alternately $\Delta R \equiv R_1 - R_2$ or $\Delta R^2 \equiv R_1^2 - R_2^2$, where subscripts indicate the model described. My discussion focuses on ΔR^2, but empirical results also report ΔR to enable comparison with the literature.

The $\Delta R \,/\, \Delta R^2$ summary statistics are preferable to the alternatives, the t-statistic on the SAT variable, which increases with the sample size, and the SAT coefficient β_j, which is of little substantive significance without knowledge of $\Sigma_{22}^{-1}\Sigma_{21}$. However, ΔR and ΔR^2, like the alternatives, measure the SAT's contribution in light of the other X variables in the model; if X_i does not include all the information available for prediction of y_{ij}, (1) may be misspecified and the incremental improvement to goodness-of-fit conferred by the SAT score may not be an interesting parameter.

2.1 Sample Selection

Unfortunately for the researcher, y_{ij} is only observed for students who enroll at college j, likely not a representative group. The interest in fit statistics rather than regression coefficients complicates inference from a selected sample. Even when coefficients can be estimated without bias, within-sample fit statistics may be biased for the population goodness-of-fit. Moreover, estimates of both R_1^2 and R_2^2 are required; estimators that are consistent for the former may not be for the latter.

The problem to be solved can be broken into two distinct steps.[4] I omit j subscripts hereafter where there is no ambiguity:

- The coefficients from (1) and (2), and the residual variance σ_ε^2, must be estimated.
- These estimates must be used to construct an estimate of the population ΔR^2.

The second of these steps is the easiest solved. Note that

$$\Delta R_p^2 = R_{1,p}^2 - R_{2,p}^2 = \frac{\mathsf{V}_p\left(SAT_i\beta + X_i\gamma\right) - \mathsf{V}_p\left(X_i\theta\right)}{\mathsf{V}_p\left(y_i\right)} \tag{5}$$

where the p subscript indicates population statistics, which will in general differ from their within-sample analogues. Assume consistent estimates $\hat{\beta}$, $\hat{\gamma}$, $\hat{\theta}$ and $\hat{\sigma}_\varepsilon^2$. Then, using the population variance matrix Σ, which may be presumed known,

$$\widehat{\Delta R_p^2} \equiv \frac{\begin{pmatrix}\hat{\beta} & \hat{\gamma}\end{pmatrix} \Sigma \begin{pmatrix}\hat{\beta} & \hat{\gamma}\end{pmatrix}' - \begin{pmatrix}0 & \hat{\theta}\end{pmatrix} \Sigma \begin{pmatrix}0 & \hat{\theta}\end{pmatrix}'}{\begin{pmatrix}\hat{\beta} & \hat{\gamma}\end{pmatrix} \Sigma \begin{pmatrix}\hat{\beta} & \hat{\gamma}\end{pmatrix}' + \hat{\sigma}_\varepsilon^2} \xrightarrow{p} \Delta R_p^2. \tag{6}$$

The within-sample ΔR_s^2 can be written similarly, with the sample variance $\mathsf{V}_s\left(SAT_i,\, X_i\right)$ replacing Σ everywhere that it appears. The calculation (6) is thus known in the SAT validity literature as a "correction for restriction of range" (Camara & Echternacht, 2000; Bridgeman et al., 1992; Willingham et al., 1990; Hezlett et al., 2001).

[4] The first step may not strictly be required; ΔR_p^2 can be derived strictly as an expression of the population moments of y, X, and SAT. However, any algorithm to compute it directly will likely imply coefficient estimates; for ease of exposition, I treat these as a distinct first step.

2.2 Alternative estimators for θ

The range correction (6) is consistent for the population statistic ΔR_p^2 only when $\hat{\gamma}$, $\hat{\beta}$, $\hat{\theta}$ and $\hat{\sigma}_\varepsilon^2$ are all consistent, a requirement that has been largely ignored in the literature. α and δ are nuisance parameters, as they do not appear in (6).

For the moment, I assume that β and γ are estimated by within-sample OLS, and denote the estimates $\hat{\beta}_{OLS}$ and $\hat{\gamma}_{OLS}$. The estimation of θ is where my approach differs from that used by previous authors.

The estimator universally used in the literature is the OLS coefficient $\hat{\theta}_{OLS}$. The omitted variables formula (3) gives:

$$\hat{\theta}_{OLS} = \hat{\gamma}_{OLS} + \mathsf{V}_s^{-1}(X_i)\,\mathsf{cov}_s(X_i', SAT_i)\,\hat{\beta}_{OLS}. \tag{7}$$

My proposed estimator replaces the within-sample $\mathsf{V}_s^{-1}(X_i)\,\mathsf{cov}_s(X_i', SAT_i)$ in (7) with its population analogue. Thus,

$$\hat{\theta}_{O.V.} \equiv \hat{\gamma}_{OLS} + \Sigma_{22}^{-1}\Sigma_{21}\hat{\beta}_{OLS}. \tag{8}$$

I refer to this, and to the resulting $\widehat{\Delta R_p^2}$, as "omitted variables" estimators.

2.3 Selection bias results

My basic results are simple: $\hat{\beta}_{OLS}$ and $\hat{\gamma}_{OLS}$—and therefore $\hat{\theta}_{O.V.}$—are unbiased whenever selection is purely on observables; $\hat{\theta}_{OLS}$ is unbiased only under the stronger condition that selection is on X alone. However, a good deal of notation is required to establish these facts.

Unbiasedness of within-sample OLS coefficients in (1) and (2) requires, respectively, that $\mathsf{E}[\varepsilon_i \mid SAT_i, X_i; y_i \text{ is observed}] = 0$ and $\mathsf{E}[\nu_i \mid X_i; y_i \text{ is observed}] = 0$. Formalize the selection process as follows: Let y_i be observed iff $Z_i^* \geq 0$, where Z_i^* is a latent variable satisfying

$$Z_i^* = \psi_0 + SAT_i\psi_1 + X_i\psi_2 + \mu_i. \tag{9}$$

μ_i is assumed independent of SAT_i and X_i but may carry information about ε_i. Let $f(c) \equiv \mathsf{E}[\varepsilon_i \mid \mu_i \geq c]$; independence of ε and μ would thus imply $f(c) \equiv 0$.[5]

When selection is of this form,

$$\mathsf{E}[\varepsilon_i \mid SAT_i,\, X_i;\, Z_i^* \geq 0] = f(-\psi_0 - SAT_i\psi_1 - X_i\psi_2)\,. \tag{10}$$

The conditional expectation of ν_i is not as simple. Recall that $\nu_i = \omega_i\beta + \varepsilon_i$ and note that Z_i^* can be expressed as

$$\begin{aligned} Z_i^* &= \left(\psi_0 + \left(\mathsf{E}[SAT_i] - \mathsf{E}[X_i]\,\Sigma_{22}^{-1}\Sigma_{21}\right)\psi_1\right) + X_i\left(\psi_2 + \Sigma_{22}^{-1}\Sigma_{21}\psi_1\right) + \omega_i\psi_1 + \mu_i \\ &\equiv \tilde{\psi}_0 + X_i\tilde{\psi}_2 + \omega_i\psi_1 + \mu_i. \end{aligned} \tag{11}$$

Thus,

$$\begin{aligned} \mathsf{E}[\nu_i \mid X_i;\, Z_i^* \geq 0] &= \mathsf{E}_\omega\left[f\left(-\tilde{\psi}_0 - X_i\tilde{\psi}_2 - \omega_i\psi_1\right)\right] \\ &\quad + \mathsf{E}_\mu\left[\omega_i \mid \omega_i\psi_1 \geq -\tilde{\psi}_0 - X_i\tilde{\psi}_2 - \mu_i\right]\beta, \end{aligned} \tag{12}$$

where subscripts indicate the distribution over which expectations are to be taken. The first term arises from non-independence of ε_i and μ_i, and is therefore analogous to (10). The second term captures differences between the sample and population conditional SAT distribution. It is nonzero whenever $\psi_1 \neq 0$; that is, whenever the selection variable Z_i^* covaries with SAT_i conditional on X_i.

Equations (10) and (12) establish sufficient conditions for unbiasedness of the various estimators.[6] From (10), $\hat{\beta}_{OLS}$ and $\hat{\gamma}_{OLS}$ are unbiased when the $f(\cdot)$ function is identically zero (or in the less interesting case that $\psi_1 = \psi_2 = 0$). Because $\hat{\theta}_{O.V.}$ relies on the sample only for these parameters, in which it is linear, it is unbiased under the same conditions. However, $\hat{\theta}_{OLS}$ may be biased, as the second term of (12) may be nonzero. An additional assumption, that $\psi_1 = 0$, is needed.

[5] I impose linearity and additivity in (9) for ease of exposition. The basic results would hold in a more general case, for an arbitrary function $h(SAT_i,\, X_i) \equiv \mathsf{E}[Z_i^* \mid SAT_i,\, X_i]$ and $\mu_i \equiv Z_i^* - h(SAT_i,\, X_i)$.

[6] These conditions are very nearly necessary; only pathological distributions for ε_i, μ_i, and ω_i could make the estimators consistent when the conditions are violated.

It is helpful to consider how these conditions translate into assumptions about the admissions process. I consider three situations, beginning with the most restrictive:

I. Selection on X_i: $f(\cdot) \equiv 0$ & $\psi_1 = 0$.

   Here, the college considers only the X variables in admissions, and individual application and enrollment decisions are uninformative about ability or the SAT score conditional on X. Neither θ_{OLS} nor $\theta_{O.V.}$ is biased.

   This form of selection is immediately suspect, however, in the context of SAT validity studies: Any college that collects SAT scores from its applicants likely does so in order to consider them in admissions decisions. Moreover, even if college j does not admit on the SAT score, intercollegiate competition for students would likely mean that the students attending j are those whose SAT scores are not good enough to gain admission at a more selective college.

II. Selection on observables: $f(\cdot) \equiv 0$ but $\psi_1 \neq 0$.

   This situation might arise at a college where admissions decisions are a function solely of the X variables and SAT. It requires still, however, that enrollment decisions are independent of ε_i. $\theta_{O.V.}$ remains unbiased, but θ_{OLS} does not. In the most plausible situation (ψ_1 and β positive, and the elements of ψ_2 and Σ_{21} having the same sign as the elements of θ), $\mathsf{E}\left[\hat{\theta}_{OLS}\right] < \theta$, so $\operatorname{plim} \Delta R^2_{p,OLS} > \Delta R^2_p$.

   The selection-on-observables assumption is somewhat more realistic than selection-on-X, but it still requires a somewhat peculiar college that does not consider non-X variables—X will typically not include information gleaned from essays, recommendation letters, and extracurricular activities—in admissions. I argue below that the UC system has just this sort of admissions rule.

III. Selection on unobservables: No restrictions on (9).

Most colleges admit on the basis of variables that are not realistically included in the X vector. At these colleges, $f(\cdot) \neq 0$, and both θ estimators are biased. However, even in this situation there is reason to prefer $\hat{\theta}_{O.V.}$ to $\hat{\theta}_{OLS}$: If a selection correction enables consistent estimation of β and γ, the former will be unbiased but the latter may not. In Section 5, I use Heckman's (1979) algorithm to estimate β and γ under unrestricted campus selection; $\hat{\theta}_{O.V.}$ differs substantially from the Heckman model's direct estimate.

Consideration of these cases highlights the unrealistic assumptions implicit in the usual methodology: College j must collect the SAT score from applicants, else the data would not be available, but must not consider it in admissions, else $\hat{\theta}_{OLS}$ would be inconsistent. Moreover, when these assumptions are violated, the usual estimator for θ is likely to be biased downward, reducing the estimated R_2^2 and producing an upward-biased estimated SAT contribution.

In the next section I discuss the construction of a University of California sample in which the restrictive assumptions needed for consistency of $\hat{\theta}_{O.V.}$ are plausibly satisfied. Results presented in Section 4 indicate that the bias introduced by the usual methodology is quite large in this sample.

3 Data

The current analysis is made possible by access to an unusually large and rich data set extracted from University of California administrative records.[7] It contains longitudinal academic information on all 22,526 California residents from the 1993 high school class who applied to, were admitted by, and enrolled as freshmen at any of the UC campuses.

I construct three outcome measures to summarize students' academic progress at the UC. The first is the freshman grade point average, *FGPA*: the average grade achieved in all courses attempted during the student's first year at the UC. The second, *5YRGPA*, is the overall grade average in all courses attempted during the five years covered by the data, without adjustment for dropouts

[7] I thank Saul Geiser and Roger Studley of the UC Office of the President for providing me these data, which are not publicly available and to which I have access under a restricted use agreement.

or periods of nonenrollment.[8] Finally, a dichotomous variable *GRADIN5* indicates whether the student is observed to graduate within five years.[9] The three outcomes are taken as alternative realizations of the academic capacity that the SAT is meant to predict. For comparability with the literature, *FGPA* is the primary focus here although the other outcomes are arguably more important to the university's objectives.

My empirical models use several other variables from the UC database: students' self-reported high school GPAs (*HSGPA*); their SAT scores (though not sub-scores from the math and verbal sections nor SAT II scores, neither of which are available in the data); and their official majors at the end of each academic year. The UC data are also supplemented in several ways: with demographic characteristics of students' high schools, with geographic information, and with an auxiliary data set representing the California SAT-taker population.

761 public high schools in California are represented in the UC data, along with 275 California private schools, 886 out-of-state high schools, and 157 foreign schools. Information about the California public high schools is taken from the California Department of Education's 1999 Academic Performance Index (API) database. I extract from these data five school-level demographic variables (the fraction of students in each of three racial groups, the average education of students' parents, and the fraction of students receiving subsidized meals) and one outcome variable (the mean score on the API test battery).[10] I construct a crosswalk between the API data and the UC database using the school identifiers in the National Center for Education Statistics' Common Core of Data and a link file from the College Board. Students are assigned to the high school that they

[8]Course grades are optional and infrequently assigned at the Santa Cruz campus. Models for *FGPA* and *5YRGPA* therefore exclude Santa Cruz observations.

[9]This five year window, dictated by data availability, is somewhat shorter than that usually used (the *U.S. News and World Report* college rankings use the six year graduation rate). Historically, about 90% of students who graduate from the UC within six years have done so by the end of the fifth year.

[10]There are two potentially serious sources of measurement error in the API data. First, I use the API database from 1999, the first year in which it was compiled; this may not perfectly measure 1993 high school characteristics. Second, the parental education variable comes from a survey with very low response rates. However, the correlation between school mean parental education and the school API score is 0.92, suggesting that concerns about within-year data unreliability are probably unwarranted (Technical Design Group, 2000).

report having attended most; when the API database does not have complete information about this school, I assign students instead to the school from which they report having graduated. 83% of the California public high schools successfully match to complete API information, providing school-level information for 88% of the in-state public high school graduates and 72% of all students in the UC data.

I approximate each high school's location by the centroid of the county in which the school is located. Latitude and longitude information from the US Geological Survey's Geographic Names Information System (`http://geonames.usgs.gov/gnishome.html`) is used to calculate the distance between this point and each of the eight UC campuses; these distance variables are used in Section 5 as instruments for students' choice of campuses.

Finally, both range corrections and my omitted variables estimator require an estimate of Σ, the population variance matrix of SAT scores and other predictor variables. The population of interest is the set of potential applicants to the University of California, as these are the students for whom performance might need to be predicted. This population is difficult to identify, and SAT validity studies typically approximate it by the universe of SAT takers. SAT-takers are not representative of all high school graduates, but it seems reasonable that all California students who contemplate attending a four-year college will have taken the SAT. I use a College Board data set consisting of observations on all California students from the 1994-1998 high school classes who took the SAT exam.[11] A crosswalk table is used to transform the College Board database's post-1994 "recentered" SAT scores to the pre-1994 scale used in the UC data. Students in the College Board database are matched to API data, using a procedure similar to that described above, and the observations with non-missing data are taken to be the population of interest.

[11] I thank David Card and Alan Krueger, the Mellon Foundation, and the College Board for permitting me access to these data.

3.1 Sample Construction

In Section 2, I discussed the nonrandom sample selection that can be introduced by the admissions process and by students' application and enrollment decisions. I exploit an institutional feature of the University of California admissions process to construct an analysis sample in which the ordinarily intractable selection-on-unobservables is arguably reduced to selection-on-observables: Students may apply to as many of the eight UC campuses as they like, but a systemwide eligibility determination is the first step in each campus' admissions process. The minimum eligibility criteria are public information and, in 1993, were primarily a function of the SAT score and high school GPA. Eligible students denied admission by all of the campuses to which they applied were permitted to enroll at one of the less selective campuses, while students who did not meet the criteria were not ordinarily eligible for admission at any campus.[12]

My pooled analysis sample consists of the 13,363 students from the eight UC campuses who have non-missing individual and school data and who appear to have met the 1994 eligibility criteria. For students in this sample, admission was guaranteed; selection came only from their decisions to apply and to accept an admission offer that may not have been at the student's preferred campus. It seems plausible that these decisions are uninformative about unobserved ability. I impose this assumption throughout my empirical analysis: When (SAT_i, $HSGPA_i$) meet the eligibility requirements, ε_{ij} is assumed to satisfy

$$\mathsf{E}\left[\varepsilon_{ij} \mid SAT_i,\, X_i;\, i \text{ attends a UC campus}\right] = \mathsf{E}\left[\varepsilon_{ij} \mid SAT_i,\, X_i\right] = 0. \tag{13}$$

Table 1 presents summary statistics for the resulting UC sample and for the population of SAT-takers. Mean SAT scores in the UC sample are more than 200 points higher than among California SAT-takers, and mean *HSGPA*s are more than half a point higher. UC students' high schools, however, are only slightly demographically advantaged relative to those of typical SAT

[12] Leonard & Jiang (1999) note the utility of the same institutional feature for validity studies.

takers. It is clear that the sample of UC students is a considerably selected subset of all SAT-takers, but it seems plausible that this selection is primarily driven by the SAT score and *HSGPA*. Additional columns in Table 1 present Σ_{21}, the population correlation between the SAT score and X, and its within-sample counterpart. These two measures differ substantially, indicating that the within-sample $\hat{\theta}_{OLS}$ is likely biased for the population θ.

Figure 1 presents another view of the selection process. It displays kernel estimates of the empirical SAT distribution: A solid line describes the constructed analysis sample and a dashed line the population of SAT-takers.[13] A third line describes the subset of SAT-takers who had their scores sent to one of the UC campuses, a required part of the UC application. The UC applicant distribution again suggests that assumption (13) may not be far off, at least as regards the application decision: Applicants are only slightly positively selected (with respect to the SAT score; Figure 1 does not speak directly to selection on unobservables) from among SAT takers.

4 Estimation and Results: Pooled Model

The UC eligibility rules mean that selection into my analysis sample is arguably a function of observable variables. However, the sample includes students at eight UC campuses, while the presentation in Section 2 treats each campus separately. I deal with this discrepancy in two ways: In this section, I impose assumptions that permit a single specification for the pooled UC sample. To test the robustness of the findings, in Section 5 I abandon these assumptions and estimate individual campus parameters with corrections for endogenous sorting.

Thus far, my notation has allowed grading standards (described by α_j, β_j, γ_j, δ_j, and θ_j) to vary freely across colleges. The potential for such variation has led many authors (see, e.g., Breland, 1979, p. 4) to recommend against pooling data from colleges with potentially divergent

[13]The density estimates use an Epanechnikov kernel with a bandwidth of 40 SAT points. This oversmooths the SAT-taker data, but is chosen to eliminate much of the artificial clumpiness introduced by the inexact, many-to-one conversion used to transform post-1994 SAT scores to the pre-recentering scale.

grading standards. The UC campuses, being part of the same system, might plausibly have similar grading standards. In this section, I impose the restriction that

$$(\beta_j, \gamma_j, \theta_j) = (\beta_k, \gamma_k, \theta_k) \equiv (\beta, \gamma, \theta) \tag{14}$$

for j, k campuses of the UC system. This permits grading standards to differ only by a constant location parameter across campuses, with the parameter perhaps chosen to allow each campus to use the usual 4-point scale.[14] Models (1) and (2) become

$$y_{ij} = \alpha_j + SAT_i\beta + X_i\gamma + \varepsilon_{ij} \text{ and} \tag{1'}$$

$$y_{ij} = \delta_j + X_i\theta + \nu_{ij}. \tag{2'}$$

With the additional imposed assumption that the distribution of students across campuses is uninformative about ε_{ij}, OLS with campus fixed effects consistently estimates (1'). I require only estimates of β and γ—α_j is a nuisance parameter that is interpreted as a characteristic of the college, rather than of the individual.

Columns A, D, and F of Table 2 report estimates of these parameters from the pooled analysis sample, using each of the three outcome measures as the dependent variable and the high school GPA as the only X variable. Columns A and D include fixed effects for end-of-year major, in addition to the previously discussed campus fixed effects, to control for differences in grading standards across departments. Column F excludes both sets of fixed effects, under the assumption that differential grading standards should not affect graduation rates, and uses a linear probability model to facilitate computation of goodness-of-fit statistics.

The lower portion of Table 2 reports the sample and population goodness-of-fit measures R_s^2 and R_p^2. These are calculated as described in equation (5), with estimated campus/major fixed

[14]When modeling the *GRADIN5* outcome variable, for which grading standards are plausibly irrelevant, I impose the additional assumption that α and δ are constant across campuses. My discussion here focuses on the less restrictive case.

effects excluded both from coefficient vectors and from $\mathsf{V}(y_i)$; calculations of R_s^2 replace population variances with their sample analogues.

Column B presents estimates of the restricted model and SAT contribution using the usual methodology (OLS for equation (2') and corrections for restriction of range). The results are in line with those found in the literature: Incremental R is 0.082 in the sample and 0.073 in the population. Looking at the ΔR^2 statistics, the SAT incrementally explains about 7.5% of the within-sample *FGPA* variance, and the range correction suggests that it explains 9.4% of the population variance. However, because SAT scores enter the sample selection rule for my pooled sample but do not enter equation (2'), $\hat{\theta}_{OLS}$ is not consistent for θ and these estimates of the SAT's contribution are biased.

Columns C, E, and G use estimates of Σ from the College Board database ($\Sigma_{22}^{-1}\Sigma_{21} = 0.187$, with a standard error of 0.0005, when X is just *HSGPA*) to consistently estimate θ for each of the three outcome variables. The fixed effect parameters in my pooled model present a problem for the omitted variables correction, as students who never enrolled in the UC are not assigned to a campus or major. However, because the fixed effects are interpreted as differences in grading standards rather than as realizations of individual ability, variation in α_j and δ_j should not be counted as explained variance and the population projection of SAT onto X is performed without fixed effects.[15] When θ is estimated using the omitted variables correction, residuals need not be orthogonal ot X in the sample. R_s^2 is therefore not analogous to the usual R^2 in columns C, E, and G; it is presented nevertheless for completeness.

Comparison of the OLS model in column B with the consistent estimates in column C is instructive. As expected, $\hat{\theta}_{OLS}$ is lower than $\hat{\theta}_{O.V.}$, by about 10%. This produces population goodness-of-fit statistics that are lower than the consistent statistics, inflating estimates of the SAT's contribution. The biased ΔR_p and ΔR_p^2 reported in column B are roughly in line with those

[15]This might not strictly identify θ if campus and/or major choice varies with ω_i. However, it provides a reasonable estimate of the SAT's contribution to within-campus/major grade prediction, abstracting out the SAT's contribution to predictions of campus and major. Range corrections require a similar approach (e.g. Young, 1990), regardless of the θ estimator used.

found in the literature; the consistent estimates of the same statistics in column C are over 40% lower. Similar patterns hold for the *5YRGPA* and *GRADIN5* outcome variables, although Table 2 does not report the OLS estimates of the restricted model for these outcomes.[16]

However, the models in Table 2 answer a question of limited substantive interest: What would be the SAT's contribution to prediction if *HSGPA* were the only other available predictor variable? They do not answer the more interesting question: What is the SAT's contribution to predictions based on the full set of other variables available to admissions offices?

Before turning to this question, I present in Table 3 an instructive variant of the models in Table 2. Here, *HSGPA* and the SAT score are broken into two parts each: The within-sample high school mean, and the deviation from that mean. Restricted models exclude both the mean SAT score and the deviation from mean, and are estimated with the omitted variables correction. The estimated coefficients indicate that the SAT's predictive power comes primarily from its variation across schools; within-school variation in scores is much less indicative of future performance.[17] The *HSGPA* variable is the opposite: the coefficient on the mean *HSGPA* at the high school is small and, in models including SAT, insignificantly different from zero; the deviation from school mean has a large coefficient. These results are consistent with the SAT's claimed role as a corrective for differences in grading standards across high schools. They suggest, however, that the inclusion of variables characterizing students' high schools in X might have a substantial effect on SAT's estimated coefficient and predictive contribution. Admissions offices typically have access to such information; at private colleges school-level measures are often explicit admissions variables. A fair assessment of the SAT's contribution to pure prediction should not permit the SAT score to stand in for these measures.

[16] Some authors (discussed, e.g., in Breland, 1979; Camara & Echternacht, 2000) have found that predictions are different for long-term grade averages like *5YRGPA* than for *FGPA*. Table 2 offers some support for this conclusion, but indicates that the consistently-estimated SAT contribution is the about same for the two measures.

[17] One explanation for this result might be that any measurement error in SAT scores is concentrated into the deviation from school mean SAT score. However, simple simulations suggest that the measurement error would have to be substantial—corresponding to a SAT reliability of about 0.5, well below the College Board's claim of 0.9—to account for the sizeable difference in coefficients seen here.

4.1 The Effect of School-Level Demographic Variables

Table 4 adds several school-level demographic measures to the X vector. Consistent with Table 3, the school-level variables have a relatively small effect on the *HSGPA* coefficients but cause the *SAT* coefficient to fall substantially from Table 2. The change reduces the SAT's measured contribution to grade prediction to about half of that seen in Table 2: Consistently estimated ΔR_p^2 was 0.055 for *FGPA* and 0.053 for *5YRGPA* in Table 2; it falls to 0.026 and 0.021, respectively, when the school demographic variables are included in X. The SAT's contribution to prediction of *GRADIN5* falls by a factor of eight from Table 2, to near zero.

To test whether the school-level variables are merely standing in for individual versions of the same variables, additional unreported regressions included individual race indicators along with the school averages. In these models, the school averages retain substantial predictive power and the SAT's contribution is, if anything, slightly higher than in Table 4.

However, demographic measures are only a subset of the school-level variables that might predict performance. Demographics might be considered inputs to the high school's production function; to the extent that high schools differ in quality, understood as the efficiency with which inputs are turned into outputs, measures of educational output should also be predictive of college performance. Table 5 restricts the focus to the *FGPA* dependent variable and considers three exam-based measures of high school output: The mean score among all students at the high school on the API exam battery, the mean SAT score among all SAT-takers at the high school, and the mean SAT score among UC matriculants in my sample at the high school. These measures are progressively more precisely targeted at the subpopulation within the school for whom outcomes are to be predicted. Each is significantly predictive of *FGPA*, and each adds somewhat to predictive accuracy. The mean SAT variables, in particular, reduce the predictive contribution of the individual SAT score. These specifications may understate the SAT's contribution to prediction—under what circumstances can we imagine colleges having access to the high school's mean SAT score if individual SAT scores aren't required for admission?—but are suggestive of what might be seen if scores on other standardized tests were averaged over the appropriate subpopulation within the school.[18]

Taken together, Tables 4 and 5 indicate that Table 2 overstates the SAT's role in prediction: Somewhere around half of the consistently-estimated SAT contribution to grade prediction—and a much larger share of its contribution to graduation prediction—is more reasonably attributed instead to the school-level variables considered here. College admissions offices have access to measures of school demographics and quality, and many colleges even consider them in admissions decisions (though not typically for the purpose of improving predictions). If the SAT score were not available, colleges could use school demographic and quality measures to achieve about the same predictive accuracy as is provided by the SAT.

5 Campus-by-Campus Modeling

The above results rely on the assumptions that the distribution of students across the UC campuses is ignorable and that coefficients are constant across campuses. The former assumption, in particular, seems implausible: One would have to believe that the campus admissions committees fail to uncover any information about future performance in students' essays, extracurricular activities, or letters of recommendation that is not already available from the high school GPA and SAT score.

A crude attempt to test the pooling assumptions was made by estimating a version of the model from Table 2, Column A, in which the SAT and $HSGPA$ effects are allowed to vary by campus. Tests for equality of either coefficient across campuses decisively reject these hypotheses (F[6; 12,957]=10.7 and 3.8, respectively). A slightly weaker hypothesis, that the ratio of β_j to

[18] Of course, Table 5 does not prove the importance of school quality, as the variation in mean test scores may simply reflect differences in unobserved inputs. In particular, the sub-school-level mean SAT scores may reflect demographic variation among different subpopulations within the school.

γ_j is constant across campuses—this would arise if, for example, $y_i^* = SAT_i\beta + X_i\gamma + \varepsilon_i$ is some unobserved individual ability and $y_{ij} = \alpha_j + y_i^*\kappa_j$—is also rejected, with an F statistic of 6.4. These tests indicate either that coefficients vary across campuses, violating assumption (14), or that students sort nonignorably into campuses.

Consistent estimation thus requires modeling prediction at the individual campus level. Because the individual campus admissions rules, unlike the systemwide eligibility determination, consider variables unobserved in my data, I can no longer assume that $\mathsf{E}\left[\varepsilon_{ij} \mid SAT_i,\, X_i;\, \text{student } i \text{ attends campus } j\right] = 0$. OLS estimates of β_j and γ_j are therefore potentially biased. The context is of course different, but the problem here is conceptually similar to that of estimating industry-specific wage equations with Roy-model differences in unobserved skill across industries (Roy, 1951; Gibbons & Katz, 1992).

Unlike in the industry wage differential literature, however, I have plausible instruments for students' choice of campus: Students are more likely to attend campuses close to their homes. For each student i and campus j, I construct three geographic variables: an indicator for whether student i's high school is in the same county as campus j, another indicator for whether the high school is in the same county as campus $k \neq j$, and a continuous measure of the distance between i's high school and campus j.

I use Heckman's (1979) two-step estimator to obtain estimates of campus-specific β_j and γ_j's that are consistent in the presence of endogenous sorting. I retain the assumption that selection into the sample of UC matriculants is ignorable—$\mathsf{E}\left[\varepsilon_{ij} \mid SAT_i,\, X_i;\, i \text{ attends the UC}\right] = 0$—and thus need only model the choice among the eight campuses. Let W_{ij} be the 3-vector of geographic instruments, and let

$$Z_{ij}^* \equiv \pi_{0j} + SAT_i\pi_{1j} + X_i\pi_{2j} + W_i\pi_{3j} + \eta_{ij} \tag{15}$$

be a latent variable that determines selection into campus j: i attends j iff $Z_{ij}^* \geq 0$.[19] Assume that

[19]This Z_{ij}^* plays a role similar to, but distinct from, that of the similar variable used in Section 2, which can be

within the UC sample, $\mathsf{V}[\eta_{ij}] = 1$, η_{ij} and ε_{ij} are independent of SAT_i, X_i, and W_{ij}, and η_{ij} and ε_{ij} are jointly normally distributed with correlation ρ_j.

Under these assumptions, $\pi_j \equiv (\pi_{0j}, \pi_{1j}, \pi_{2j}, \pi_{3j})'$ can be estimated by a probit model for the dichotomous dependent variable $Z_{ij} \equiv \mathbf{1}(i \text{ attends } j)$ in the sample of UC students. Heckman (1979) proves that (1) then implies

$$\begin{aligned} \mathsf{E}\left[y_{ij} \mid SAT_i, X_i, W_{ij}; Z_{ij}^* \geq 0\right] &= \alpha_j + SAT_i\beta_j + X_i\gamma_j \\ &\quad + \rho_j\sigma_\varepsilon\lambda\left((1, SAT_i, X_i, W_{ij})\pi_j\right), \end{aligned} \tag{16}$$

where $\lambda(c) \equiv \phi(c)/\Phi(c)$. Heckman further shows that when the inverse-Mill's-ratio term is constructed using estimated π_j from the probit model for selection, OLS is consistent for (16). Heckman (1979) and Greene (1981) provide asymptotically correct standard errors. σ_{ε_j} is always positive, so the sign of the fourth coefficient is the same as the sign of ρ_j: A positive coefficient thus means that campus j attracts students with high ε_{ij}'s; a negative coefficient that the students are negatively selected.

Table 6 reports the results when the Heckman model is estimated for the *FGPA* outcome variable at the San Diego campus (UCSD). Panel A reports OLS coefficients, Panel B1 the probit coefficients π_j, and Panel B2 the selection-adjusted *FGPA* coefficients from equation (16).[20] Columns A and D estimate equation (1); in Column A the *HSGPA* is the only X variable, while in Column D the school-level variables from Table 4 are also included in all three panels. The exclusion of W_{ij} from the outcome model is especially plausible in Column D, as school demographic variables might absorb any geographic variations in unobserved ability.

The coefficients in Panel B1 show that the geographic variables are strongly predictive of

thought to govern selection into the UC system as a whole. Equation (15) applies only to students already in the UC system. It is admittedly somewhat misspecified: Students in the system must attend one of the eight campuses, but (15) cannot hold for each of the eight. I ignore this complication and treat each of the eight selection decisions as separate and independent.

[20] Because some majors are found at only one of the UC campuses, a full set of major dummies would perfectly predict campus choice for some students. To avoid this, the freshman majors are collapsed into five broad groups with fixed effects for four of these groups included in every model in Table 6.

selection into the UCSD campus with the expected signs: Students from San Diego County are more likely than others to attend UCSD; students from counties containing other UC campuses are less likely; and students from far-away counties are less likely to attend than are students from nearby. In both A and D, the Mill's ratio coefficient in Panel B2 is significantly positive, indicating that at this campus, the third most selective in the UC system, students with above-average residual ability (high ε_{ij}) are more likely to be admitted and enroll (high η_{ij}).

Columns B and C present two approaches to estimation of the restricted model (2). In Column B, θ is estimated using a restricted version of (16) in which I constrain $\pi_{1j} = \beta_j = 0$; this is analogous to $\hat{\theta}_{OLS}$ in the pooled sample. In column C, the omitted variables approach is used, with estimates $\hat{\gamma}_j$ and $\hat{\beta}_j$ taken from the Heckman-corrected model in Column A. These two approaches are repeated for the expanded X vector in columns E and F. In both cases, the omitted variables estimator yields a higher *HSGPA* coefficient and $R^2_{2,p}$ than the direct Heckman estimation of the restricted model.

Selection into the UCSD subsample occurs in two stages: Students are selected into the UC sample, then students in this sample are selected into UCSD. The Heckman estimator recovers UC-sample projection coefficients from the subsample of students attending UCSD; the conditions in Section 2 must still be met for these UC sample coefficients to equal their population counterparts. As in Section 4, I assume that $f(\cdot) \equiv 0$ but $\psi_1 \neq 0$ in the UC selection model (9); as a result, Heckman-model estimates of β_j and γ_j are consistent but the restricted Heckman model is not consistent for θ_j. As before, the omitted variables estimates in columns C and F are not biased by selection on SAT.

The Heckman model permits estimation of the implied residual variance of y in the population from which the sample was selected. This, along with the Heckman-model estimates $\hat{\beta}_j$ and $\hat{\gamma}_j$ and the omitted variables estimate $\hat{\theta}_{j,O.V.}$, is sufficient to calculate the $\widehat{\Delta R^2_p}$ estimator from (6). UCSD $\widehat{\Delta R^2_p}$'s are slightly lower than in the pooled sample; as in that sample, however, the omitted variables estimator's consistent estimates in Columns C and F are substantially smaller than the conventional, selection-biased estimates in B and E.

Compare the usual methodology's estimate of the SAT's contribution to prediction of *FGPA* at UCSD—$\Delta R_p^2 = 0.150$ in Panel A, Column B—with the corrected estimates elsewhere in Table 6. The estimated ΔR_p^2 falls to 0.068 (Panel A, Column C) when the omitted variables estimator adjusts for selection-on-SAT into the UC system; again to 0.054 (Panel B2, Column C) when the Heckman model permits endogenous sorting into the UCSD subsample; and yet again to 0.024 (Panel B2, Column F) when school-level variables are included in the X vector. The final estimate is roughly one-sixth as large as would be implied by the conventional methodology.

The coefficients of the selection-adjusted model are quite unstable across campuses, however. Space constraints prevent reporting a complete analogue of Table 6 for each of the UC campuses, but Table 7 presents selected statistics from these models. Panel A reports naive estimates—ignoring the problem of endogenous selection into campuses—for both the sparse model (X includes just *HSGPA*) and the school demographics model (X includes the variables used in Table 4). Panel B reports Heckman-corrected estimates for the same two models. All models in both panels use the omitted variables estimator for θ and ΔR_p^2.

At most campuses, there is strong evidence for endogenous selection, but the SAT's estimated contribution in the Heckman model is not substantially different from that indicated by OLS. The pattern of coefficients on the Mill's ratio is interesting. Samples seem to be positively selected at San Diego, Riverside, Los Angeles, and Santa Barbara, and negatively selected at Davis. This perhaps indicates that the former campuses do a better job of admitting and recruiting students who will outperform their numerical qualifications.

The Berkeley campus warrants a special note: There, the SAT coefficient in the outcome equation is unusually low, and it actually increases when the school demographic variables are added to X. Berkeley is easily the most selective campus in the UC system, a fact which suggests several potential explanations for its anomalous estimates.[21] Two of the most plausible are that there are declining returns to the SAT score or that competition with elite private colleges produces negative selection among the most qualified students at Berkeley. Either of these hypotheses implies a higher estimated SAT coefficient in subsamples that exclude students with the highest SAT scores. The Berkeley campus models were repeated on a sample that artificially excluded students with SAT scores above 1,400. I found no evidence for different SAT coefficients in the truncated sample. Further study is clearly needed to understand the Berkeley result; the implied conclusion that Berkeley grades are much noisier than at other campuses seems implausible.

6 Implications for Admissions Policy

The current analysis extends the literature on SAT validity in two important ways: By addressing sample selection issues that are commonly ignored and by including a broader set of nonacademic predictor variables than are typically included in grade prediction models. Figure 2 displays the effect of these changes on campus-specific and pooled-sample estimates of the SAT's contribution (I use ΔR_p here for comparison to the statistics usually reported in the literature). Each extension substantially reduces the indicated contribution of the SAT to predictive accuracy.

The discussion of sample selection corrects an important methodological oversight in the existing literature. The samples ordinarily used in SAT validity studies are too highly selected to support the conclusions that have been drawn from raw within-sample correlations. Models that allow for a more reasonable selection model than than is implicitly imposed in the literature halve the estimated SAT contribution.

This study's other innovation, the inclusion of school-level variables in prediction equations, has an equally large impact on the stylized findings. The SAT is shown to retain predictive power

[21] Berkeley students in my analysis sample have a mean SAT score of 1,213, and 13% of them score above 1,400; the nearest other campus on both measures is UCLA, where the mean score is 1,150 and only 3.7% have scores above 1,400.

for college academic success when nonacademic school characteristics are included in prediction models, but considerably less than is indicated by the sparse models used in the literature. It is worth noting that the school-level variables used here are not difficult to obtain—the API database is available on the California Department of Education's web site—and are therefore readily available to any admissions office hoping to use them. If pure prediction is the goal, my results suggest that accuracy is maximized when nonacademic variables are used in combination with the HSGPA and SAT score.

However, in the real world, pure prediction is not exactly the goal. Admissions offices do not use all available variables, although they surely know that by doing so they could generate more accurate predictions. An inquiry into the reasons for this decision is well beyond the scope of this paper, but it is perhaps fair to say that we have made a social decision to forgo the predictive accuracy that certain "disallowed" variables would permit. Thus, for example, public universities may not give explicit preferences to white students over blacks, Hispanics, and Asians on the basis of a conclusive demonstration that this would produce a higher-performing class of admittees. The legalities of admissions rules that condition on high school characteristics are—at least to this non-legally-trained economist—less clear, but it is difficult to imagine any university granting the preferences implied by Table 4 without great public criticism.

What do these restrictions imply about the interpretation of models that contain disallowed variables? The answer depends on the question that is being asked. For the moment, consider the question that, in my reading, motivates the validity literature: Does the SAT provide information about how well students will perform in college? To illustrate the relevance of the models in Table 4 to this question, I consider an admittedly contrived example.

Suppose that an educator offered colleges a new variable to use in admissions, suggesting that a count of students' dental cavities is a good measure of students' "developed math and verbal abilities related to successful performance in college." Suppose further that in a hypothetical data set he demonstrated a strong positive correlation between the number of cavities and the freshman grade average, and that the correlation remained in regressions controlling for $HSGPA$.

Would this be considered evidence that the number of cavities is an important predictor of performance? Most likely not. A plausible interpretation of the result is that cavities are serving as a proxy for omitted socioeconomic measures. One might imagine that lower-income families devote comparatively few resources to children's dental care, leading to an SES gradient in the number of cavities. If SES predicts college performance, this could produce the hypothesized results.

This objection rests on a testable hypothesis: that dental health has no relationship to academic performance once SES is controlled. A researcher might estimate a regression containing both the number of cavities and more direct measures of socioeconomic status. If the coefficient on the number of cavities were unchanged by the inclusion of SES variables, the hypothesis would be rejected, and the researcher would be forced to conclude that the number of cavities is potentially a measure of academic potential. If, on the other hand, the coefficient on the number of cavities fell to zero when SES variables were added, the researcher would reject the claims made on behalf of the cavity variable, concluding that it serves primarily as a proxy for SES in sparse models. One can also imagine an intermediate case, in which the cavity coefficient falls substantially, though not to zero, when SES is controlled directly. Interpretation is slightly trickier here, but a reasonable conclusion would be that the cavity measure containes some independent information about student quality, though not as much as had been indicated in the sparse model. Put slightly differently, this result would imply that the portion of the cavities measure that can be predicted from SES is more highly correlated with academic outcomes than is the residual portion.

Of course, the SAT score is different from the number of cavities. We have theoretical reasons to think that SAT scores measure academic ability: An examination of the SAT instrument reveals that it requires skills—literacy, logical thinking, arithmetic—that are directly related to the skills required for college success. The cavity variable lacks what psychometricians call "face validity" (Anastasi, 1988). The point, however, is that an empirical evaluation of this theoretical argument risks overestimating the information contained in SAT scores if other available variables are artificially excluded from the models estimated.

Comparison of Tables 2 and 4 reveals that the inclusion of school-level demographic variables reduces the SAT's coefficient by 25 to 65 percent (for *FGPA* and *GRADIN5* prediction, respectively) and its contribution to predictive accuracy by 50 to 85 percent. The smaller estimated contribution in Table 4 is the more appropriate measure of the amount of unique information contained in SAT scores, and it is this contribution that should be counted on the benefits side of a societal cost-benefit analysis of the SAT.

But what do these results say to a college that must decide whom to admit and is prohibited from using demographic variables directly? After all, while much of the information contained in the SAT score is not unique, the SAT score may be the only source that the college is permitted to use. The import of my results thus depends on the college's attitude about the prohibited variables.

If the college regrets the prohibition, and would be willing to use any permitted variables that help predict performance, it will ignore Table 4. It will generate for each student a predicted performance using weights derived from sparse models like those in the literature and in Table 2, and admit on that basis. This will weight the SAT more heavily than in the richer models presented in Tables 4 and 5, in effect allowing it to stand in for the school-level demographic and outcome variables that are excluded from the admissions decision. (In my hypothetical example, such a college might also grant preferences to students with few cavities, as a permissible route toward a demographically desirable class of admittees.) The resulting class will be less able, on average, than what would result from a complete information best-prediction rule, but the SAT-based rule will discriminate less effectively against students from low SES high schools. This is analogous, though in the reverse direction, to Chan and Eyster's (2000) result: admissions offices that wish to use affirmative action but are prohibited from doing so will randomize admissions to obtain more minority admissions at the cost of weakened ability standards.

On the other hand, a college that is committed to the spirit, and not just to the letter, of the prohibition concerning demographic variables may take a different view. Such a college may not be willing to exploit the predictive power of these demographic variables, either explicitly or implicitly through the SAT. One approach for such a university might be to estimate saturated models like those in Table 4, but then conduct admissions on the basis of a rule that "zeros out" coefficients on disallowed variables while retaining the estimated coefficients on academic variables. A college that pursued this policy might be willing to consider cavities as well, but only to the extent that they retain predictive power after controlling directly for SES. Another, more radical, approach might be to regress SAT scores and *HSGPA*s on the disallowed variables, then consider only the residuals from these regressions in admissions (Studley, 2001). This, too, would weight the academic variables as in Table 4, but would only make use of the portion of those variables that is orthogonal to the disallowed variables. Either approach would sacrifice mean predicted performance relative to best-predictor rules but would admit more students from schools with disadvantageous demographics.

A Appendix: Evaluating the "4% Plan"

The UC eligibility rules changed in two important ways in 2001. First, the eligibility thresholds were changed to incorporate a new variable, the SAT II score.[22] Second, the "Eligibility in the Local Context" rule, known popularly as the "Four Percent Plan," was introduced. Under this rule, students who ranked in the top four percent of their high school classes became eligible for the UC, even if their SAT scores fell below the usual thresholds. The evaluation of students admitted under this rule provides an interesting application for grade prediction models.

The Four Percent Plan (hereafter, FPP) was explicitly intended to admit students from low-performing high schools, where even the best students achieve low SAT scores. (At all but the worst schools, well over four percent of students meet the regular UC eligibility thresholds, which are chosen to admit 12.5% of high school graduates in California.) Grade prediction models that do not take account of differences between schools might be expected to err in predicting these students' performance. The models estimated in this study take explicit account of some of the school characteristics that correlate with both SAT scores and college performance, and are therefore more appropriate predictors of FPP students' performance.

The College Board database of all SAT-takers in California in the 1994 through 1998 cohorts is used to simulate admissions under both regular and FPP rules. I use the National Center for Education Statistics' Common Core of Data to obtain a measure of school size, and I use that measure to construct a class rank under the assumption that all SAT-takers at each high school outrank all non-takers. This assumption probably leads me to overstate the number of students made eligible by the FPP. On the other hand, any behavioral response to the change in eligibility rules would expand the pool of FPP students beyond those I can identify.

[22]The UC data used here do not include SAT II scores, so do not permit an evaluation of this change. However, although the validity of SAT II scores has not been widely studied, in the College Board database SAT II scores are less predictable from the demographic variables considered here than are SAT I scores. This suggests at least the possibility that the SAT II can contribute more to prediction than the SAT I.

From the universe of SAT-takers in the 1994-1998 cohorts, 344 students are identified who would have been eligible under the FPP, but not otherwise, had the 2001 criteria been in effect when they applied to colleges. A natural comparison group is the 299 students who would have been eligible under the regular rules had the thresholds been 10 SAT points lower. The students in the FPP group have considerably higher *HSGPA*s and lower SAT scores than do students in the comparison group. The FPP group also includes more blacks, Hispanics, Asians, and women, and on average its members come from schools with larger minority populations, more students getting free lunches, lower parental education, and lower API scores.

Table A.1 compares the predictions generated by sparse and richer models for $FGPA$ prediction. Coefficients are drawn from the pooled-sample models in Tables 2 and 4 and from models for the San Diego and Riverside campuses summarized in Tables 6 and 7.[23] Each coefficient vector predicts that the FPP group will dramatically outperform the comparison group, a result driven by the sizeable gap in *HSGPA*s. However, the models that account for school characteristics predict a smaller performance gap than do the sparse models, which do not adequately capture the effects of the low-SES schools from which the FPP students are drawn.

The clear gap in predicted performance between FPP and comparison groups suggests that the FPP could be made considerably larger without diluting the quality of the admitted students. Because students brought in by the FPP come disproportionately from schools with characteristics that are predictive of poor performance in college, models that ignore these characteristics are likely to suggest overexpansion of the program relative to its performance-maximizing size. The final row of Table A.1 describes the optimal size of the percent plan implied by each prediction model, defined as the size at which the average student brought in by a 0.5% expansion of the percent plan will perform as well as the average student in the comparison group. The sparse models indicate that

[23]Riverside is the least selective UC campus and, therefore, the campus most likely to accept students at the eligibility margin. San Diego will likely accept fewer marginally eligible students, but is chosen because its prediction coefficients seem reasonably close to the middle of the range seen in Table 7.

the percent plan should reach up to 1.5% deeper into the class rank distribution than is suggested by the models incorporating school-level measures.

Equally interesting, there is considerable variance among the different prediction coefficients used: The Riverside model suggests a percent plan reaching deeper than that implied by either the pooled sample or the UCSD coefficients. Because Riverside is likely to enroll most of the marginally eligible students, perhaps coefficients for that campus should be given more weight in this decision; in any case, the divergence among different models, like the variation in coefficients in Table 7, raises a warning flag about the generalizability of validity models.

References

Anastasi, A. (1988). *Psychological Testing* (6 ed.). New York: Macmillan Publishing Company.

Atkinson, R. C. (2001). Standardized tests and access to American universities. 2001 Robert H. Atwell Distinguished Lecture to the 83rd Annual Meeting of the American Council of Education.

Bowen, W. G. & Bok, D. (1998). *The Shape of the River: Long-Term Consequences of Considering Race in College and University Admissions*. Princeton, NJ: Princeton University Press.

Breland, H. M. (1979). *Population Validity and College Entrance Measures*. New York: College Board.

Bridgeman, B. (1991). Essays and multiple-choice tests as predictors of college freshman GPA. Research Report RR-91-3, Educational Testing Service, Princeton, New Jersey.

Bridgeman, B., Hale, G. A., Lewis, C., Pollack, J., & Wang, M.-m. (1992). Placement validity of a prototype SAT with an essay. Research Report RR-92-28, Educational Testing Service, Princeton, New Jersey.

Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). Predictions of freshman grade point average from the revised and recentered SAT I: Reasoning test. Research Report 2000-1, The College Board, New York.

Camara, W. J. (2001). There is no mystery when it comes to the SAT I. College Board Press Release.

Camara, W. J. & Echternacht, G. (2000). The SAT I and high school grades: Utility in predicting success in college. Research Notes RN-10, The College Board, Office of Research and Development.

Caperton, G. (2001). Toward a holistic definition of merit: A response to the UC proposal. College Board Press Release.

Chan, J. & Eyster, E. (2000). Does banning affirmative action harm college student quality? unpublished manuscript.

Crouse, J. & Trusheim, D. (1988). *The Case Against the SAT*. Chicago: The University of Chicago Press.

Ford, S. F. & Campos, S. (1977). Summary of validity data from the admissions testing program validity study service. Advisory Panel on the Scholastic Aptitude Test Score Decline, College Board and Educational Testing Service.

Gibbons, R. & Katz, L. (1992). Does unmeasured ability explain inter-industry wage differentials? *The Review of Economic Studies*, *59*(3), 515–535.

Greene, W. H. (1981). Sample selection bias as a specification error: A comment. *Econometrica*, *49*(3), 795–798.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153–162.

Hezlett, S. A., Kuncel, N. R., Vey, M. A., Ahart, A. M., Ones, D. S., Campbell, J. P., & Camara, W. (2001). The predictive validity of the SAT: A meta-analysis. In Ones, D. S. & Hezlett, S. A. (Eds.), *Predicting Performance: The Interface of I-O Psychology and Educational Research*, 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego.

Hunter, J. E. & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*(1), 72–98.

Izumi, L. (2001). Keep the SAT as a UC admissions requirement. Action Alert, Pacific Research Institute for Public Policy.

Jencks, C. & Crouse, J. (1982). Should we relabel the SAT . . . or replace it? *Phi Delta Kappan*, *63*, 659–663.

Lemann, N. (1999). *The Big Test: The Secret History of the American Meritocracy.* New York: Farrar, Straus & Giroux.

Leonard, D. K. & Jiang, J. (1999). Gender bias and the college predictions of the SATs: A cry of despair. *Research in Higher Education*, *40*(4), 375–407.

Nairn, A. & Associates (1980). The reign of ETS: The corporation that makes up minds. The Ralph Nader Report on the Educational Testing Service.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, *3*, 135–146.

Slack, W. V. & Porter, D. (1980). The Scholastic Aptitude Test: A critical appraisal. *Harvard Educational Review*, *50*(2), 154–175.

Stricker, L. J. (1991). Current validity of 1975 and 1985 SATs: Implications for validity trends since the mid-1970's. *Journal of Educational Measurement*, *28*(2), 93–98.

Studley, R. (2001). Inequality and college admissions. Unpublished paper presented at UC Berkeley Labor Lunch, Feb. 2, 2001.

Technical Design Group (2000). Construction of California's 1999 School Characteristics Index and Similar Schools Ranks. PSAA Technical Report 00-1, Advisory Committee for the Public Schools Accountability Act of 1999.

Willingham, W. W., Lewis, C., Morgan, R., & Ramist, L. (1990). *Predicting College Grades: An Analysis of Institutional Trends Over Two Decades.* Princeton, NJ: Educational Testing Service.

Young, J. W. (1990). Are validity coefficients understated due to correctable defects in the GPA. *Research in Higher Education*, *31*(4), 319–325.

Table 1: Summary Statistics for the UC Analysis Sample and the California SAT-taker Population

	UC Sample			CA SAT-takers		
Variable	Mean	S.D.	corr $(SAT_i, \cdot)$	Mean	S.D.	corr $(SAT_i, \cdot)$
Individual Academic Variables						
HSGPA	3.82	0.39	0.38	3.25	0.63	0.52
SAT	1,100	173	1.00	898	228	1.00
High School Demographic Characteristics						
Frac. Free Lunch	0.25	0.22	-0.33	0.29	0.23	-0.36
Frac. Black	0.07	0.10	-0.16	0.08	0.11	-0.18
Frac. Asian	0.21	0.18	0.12	0.17	0.17	0.10
Frac. Hispanic	0.27	0.22	-0.31	0.31	0.25	-0.34
Avg. Parental Ed.	14.4	1.3	0.37	14.1	1.3	0.41
Individual Outcome Variables						
FGPA	2.89	0.62	0.33			
5YRGPA	2.98	0.58	0.31			
GRADIN5	0.70	0.46	0.15			
N		13,363			435,890	

Notes: *FGPA* and *5YRGPA* statistics exclude 366 observations from the Santa Cruz campus. See Section 3 for description of variables.

Table 2: Basic Prediction Models for Pooled Sample

	FGPA			5YRGPA		GRADIN5	
	OLS	OLS	O.V. Est.	OLS	O.V. Est.	OLS	O.V. Est.
Variable	(A)	(B)	(C)	(D)	(E)	(F)	(G)
HSGPA	0.58	0.68	0.75	0.49	0.64	0.20	0.24
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
SAT/1000	0.90	–	–	0.79	–	0.22	–
	(0.03)			(0.03)		(0.02)	
Campus/Major FEs	6/19	6/19		6/21		–	
R^2_{OLS}	0.219	0.173		0.222		0.046	
Goodness-of-fit and SAT contribution							
R^2_s	0.249	0.174	0.205	0.224	0.183	0.045	0.041
R^2_p	0.455	0.362	0.400	0.420	0.367	0.110	0.103
ΔR^2_s		0.075	0.044		0.041		0.005
ΔR^2_p		0.094	0.055		0.053		0.008
ΔR_s		0.082	0.046		0.046		0.011
ΔR_p		0.073	0.042		0.043		0.012

Note: See text for descriptions of *O.V. Estimator* (eq. 8) and R^2_p (eq. 5).

Table 3: Within- and Between-School Variables in Prediction Models for Pooled Sample

	FGPA		5YRGPA		GRADIN5	
	(A)	(B)	(C)	(D)	(E)	(F)
High School Means						
HSGPA	0.03	0.20	-0.00	0.16	0.02	0.09
	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
SAT/1000	1.73	–	1.69	–	0.76	–
	(0.06)		(0.05)		(0.04)	
Deviations from High School Means						
HSGPA	0.74	0.85	0.66	0.74	0.28	0.28
	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)
SAT/1000	0.46	–	0.33	–	-0.06	–
	(0.04)		(0.03)		(0.03)	
Campus/Major FEs	6/19		6/21		–	
R^2_{OLS}	0.257		0.266		0.065	
Goodness-of-fit and SAT contribution						
R^2_p	0.525	0.472	0.504	0.445	0.157	0.132
ΔR^2_p		0.053		0.059		0.025
ΔR_p		0.038		0.043		0.033

Notes: Columns B, D, and F calculated by omitted variables correction; see text for details. School means calculated within sample.

Table 4: Prediction Models Incorporating High School Demographics, Pooled Sample

	FGPA		5YRGPA		GRADIN5	
Variable	(A)	(B)	(C)	(D)	(E)	(F)
HSGPA	0.62	0.74	0.54	0.63	0.23	0.24
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
SAT/1000	0.69	–	0.56	–	0.07	–
	(0.04)		(0.03)		(0.03)	
High School Demographic Characteristics						
Frac. Free Lunch	-0.01	-0.01	-0.02	-0.02	-0.10	-0.10
	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
Frac. Black	-0.32	-0.43	-0.31	-0.40	-0.15	-0.16
	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)
Frac. Asian	0.16	0.20	0.14	0.17	0.11	0.12
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
Frac. Hispanic	-0.18	-0.17	-0.18	-0.17	-0.07	-0.07
	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
Avg. Parental Ed./100	1.67	5.80	2.48	5.81	1.03	1.42
	(0.90)	(0.90)	(0.83)	(0.82)	(0.72)	(0.71)
Campus/Major FEs	6/19		6/21		–	
R^2_{OLS}	0.235		0.242		0.060	
Goodness-of-fit and SAT contribution						
R^2_p	0.472	0.446	0.445	0.424	0.133	0.132
ΔR^2_p		0.026		0.021		0.001
ΔR_p		0.019		0.016		0.001

Note: Columns B, D, and F calculated by omitted variables correction.

Table 5: High School Input and Output Variables in *FGPA* Prediction

Variable	(A)	(B)	(C)	(D)	(E)	(F)
HSGPA	0.62	0.74	0.63	0.74	0.64	0.73
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
SAT/1000	0.68	–	0.62	–	0.55	–
	(0.04)		(0.04)		(0.04)	
High School Demographic Characteristics						
Frac. Free Lunch	-0.01	-0.01	0.00	0.00	0.02	0.02
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Frac. Black	-0.18	-0.23	-0.05	-0.00	-0.08	-0.04
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Frac. Asian	0.11	0.13	0.09	0.10	0.08	0.09
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Frac. Hispanic	-0.13	-0.11	-0.15	-0.13	-0.16	-0.13
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Avg. Parental Ed./100	-2.15	0.06	-5.47	-5.55	-5.24	-5.30
	(1.20)	(1.20)	(1.26)	(1.26)	(1.26)	(1.26)
High School Output Measures						
API Score/1000	0.62	0.89	0.25	0.33	0.30	0.38
	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)
$\overline{SAT}$/1000, all SAT-takers	–	–	1.00	1.51	0.47	0.91
			(0.12)	(0.12)	(0.15)	(0.15)
$\overline{SAT}$/1000, UC matriculants	–	–	–	–	0.70	0.72
					(0.11)	(0.11)
R^2_{OLS}	0.236		0.240		0.242	
Goodness-of-fit and SAT contribution						
R^2_p	0.474	0.449	0.479	0.459	0.477	0.461
ΔR^2_p		0.025		0.019		0.016
ΔR_p		0.018		0.014		0.012

Notes: All columns include fixed effects for 6 campuses and 19 end-of-year majors. Columns B, D, and F calculated by omitted variables correction.

Table 6: OLS and Selection-Adjusted Models for FGPA at San Diego Campus

	Sparse Model			School Demographics Model		
Estimator	OLS	OLS	O.V.	OLS	OLS	O.V.
	(A)	(B)	(C)	(D)	(E)	(F)
Panel A: Models without Campus Selection Correction						
HSGPA	0.64	0.68	0.83	0.68	0.72	0.81
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
SAT/1000	1.02			0.77		
	(0.09)			(0.09)		
R_p^2	0.553	0.403	0.484	0.569	0.482	0.537
ΔR_p^2		0.150	0.068		0.086	0.032
Panel B1: Probit Coefficients for Selection into UCSD Subsample						
HSGPA	0.24	0.34		0.29	0.36	
	(0.04)	(0.04)		(0.04)	(0.04)	
SAT/1000	0.62			0.44		
	(0.10)			(0.10)		
HS in S.D. County	0.73	0.74		0.63	0.62	
	(0.07)	(0.07)		(0.07)	(0.07)	
HS in other UC Cnty	-0.16	-0.15		-0.16	-0.15	
	(0.05)	(0.05)		(0.05)	(0.05)	
Miles to UCSD/100	-0.05	-0.05		-0.06	-0.06	
	(0.01)	(0.01)		(0.02)	(0.01)	
Avg. Log Likelihood	-0.343	-0.345		-0.341	-0.342	
Panel B2: Selection-Adjusted Models						
HSGPA	0.75	0.84	0.93	0.80	0.87	0.93
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
SAT/1000	1.01			0.76		
	(0.09)			(0.10)		
Inverse Mill's Ratio	0.33	0.43		0.37	0.44	
	(0.04)	(0.05)		(0.05)	(0.05)	
R_p^2	0.538	0.409	0.484	0.550	0.473	0.526
ΔR_p^2		0.129	0.054		0.077	0.024

Notes: Sample for Panel B1 is the UC sample from Table 1; Panels A and B2 use the subsample of 1,621 San Diego campus observations. All models in all panels include fixed effects for four major groups; Columns D through F also include the five school demographic measures used in Table 4.

Table 7: Summary of Campus-Specific OLS and Heckman Models for FGPA

	Campus						
	San Diego	Irvine	Riverside	Los Angeles	Santa Barbara	Berkeley	Davis
N at campus	1,621	1,644	804	2,468	1,890	2,402	2,168
Panel A: Models without Campus Selection Correction							
Sparse Model							
SAT Coeff.	1.02	1.12	0.80	1.11	1.08	0.41	0.96
	(0.09)	(0.09)	(0.14)	(0.07)	(0.09)	(0.08)	(0.07)
ΔR_p^2	0.068	0.080	0.034	0.087	0.082	0.013	0.053
School Demographics Model							
SAT Coeff.	0.77	0.85	0.55	0.83	0.89	0.27	0.75
	(0.09)	(0.10)	(0.15)	(0.08)	(0.09)	(0.09)	(0.08)
ΔR_p^2	0.032	0.037	0.012	0.041	0.044	0.004	0.025
Panel B: Heckman Selection Models with Geographic Instruments							
Sparse Model							
SAT Coeff.	1.01	1.18	0.54	1.19	0.85	0.23	1.02
	(0.09)	(0.10)	(0.15)	(0.08)	(0.13)	(0.13)	(0.08)
Inverse Mill's Ratio	0.33	-0.06	0.17	0.19	0.25	-0.12	-0.09
	(0.04)	(0.04)	(0.05)	(0.05)	(0.10)	(0.06)	(0.03)
ΔR_p^2	0.054	0.086	0.016	0.085	0.062	0.004	0.058
School Demographics Model							
SAT Coeff.	0.76	0.88	0.42	0.94	0.72	0.32	0.80
	(0.10)	(0.11)	(0.16)	(0.09)	(0.15)	(0.15)	(0.08)
Inverse Mill's Ratio	0.37	-0.03	0.09	0.14	0.14	0.04	-0.09
	(0.05)	(0.04)	(0.05)	(0.05)	(0.10)	(0.08)	(0.03)
ΔR_p^2	0.024	0.039	0.008	0.047	0.033	0.006	0.028

Notes: All models are analogous to those in Table 6. ΔR_p^2 is calculated using the omitted variables correction (8).

Table A.1: Prospective Evaluation of the University of California's Four Percent Plan

		Mean Predicted *FGPA*					
FGPA Prediction Model:		Sparse			School Demographics		
Sample:		Pooled	UCR	UCSD	Pooled	UCR	UCSD
	N	(A)	(B)	(C)	(D)	(E)	(F)
Regular Eligibility Rules							
Inframarginal	150,620	2.80	2.64	2.09	2.80	2.81	1.96
Marginal (Comparison)	299	2.30	2.18	1.48	2.30	2.32	1.37
Not Regularly Eligible, by Class Rank (noncumulative)							
0.5%	19	2.65	2.59	1.88	2.59	2.64	1.72
1.0%	29	2.53	2.47	1.74	2.48	2.52	1.55
1.5%	39	2.56	2.49	1.77	2.52	2.57	1.59
2.0%	29	2.51	2.45	1.72	2.46	2.51	1.53
2.5%	46	2.47	2.41	1.67	2.43	2.49	1.50
3.0%	54	2.46	2.40	1.66	2.43	2.48	1.49
3.5%	61	2.47	2.40	1.67	2.43	2.49	1.50
4.0%	67	2.45	2.38	1.64	2.41	2.47	1.47
Total, Top 4% (FPP)	*344*	*2.49*	*2.43*	*1.70*	*2.45*	*2.51*	*1.52*
4.5%	76	2.43	2.36	1.62	2.42	2.47	1.46
5.0%	70	2.38	2.31	1.57	2.35	2.42	1.38
5.5%	74	2.37	2.31	1.56	2.34	2.40	1.39
6.0%	83	2.33	2.26	1.51	2.30	2.37	1.34
6.5%	95	2.30	2.23	1.47	2.26	2.31	1.30
7.0%	95	2.33	2.25	1.50	2.29	2.33	1.34
7.5%	96	2.29	2.21	1.46	2.25	2.32	1.28
8.0%	83	2.24	2.16	1.40	2.20	2.25	1.22
8.5%	117	2.26	2.18	1.42	2.22	2.27	1.25
9.0%	123	2.23	2.14	1.39	2.17	2.23	1.18
9.5%	125	2.22	2.14	1.38	2.17	2.23	1.19
10.0%	120	2.19	2.10	1.34	2.13	2.19	1.14
10.5%	140	2.18	2.09	1.33	2.12	2.17	1.13
11.0%	154	2.18	2.09	1.33	2.13	2.19	1.14
11.5%	155	2.16	2.07	1.30	2.10	2.14	1.10
12.0%	164	2.17	2.07	1.32	2.10	2.15	1.11
12.5%	208	2.16	2.06	1.30	2.09	2.14	1.09
Implied optimal % plan size		7.0%	7.5%	7.0%	6.0%	7.5%	5.5%

Notes: Marginal group would be eligible with 10 extra SAT points. *Optimal % plan size* is size at which an additional 0.5% expansion would bring in students with average predicted *FGPA*s lower than in the marginal group.

Figure 1: Kernel Estimates of SAT Distribution in Three Samples

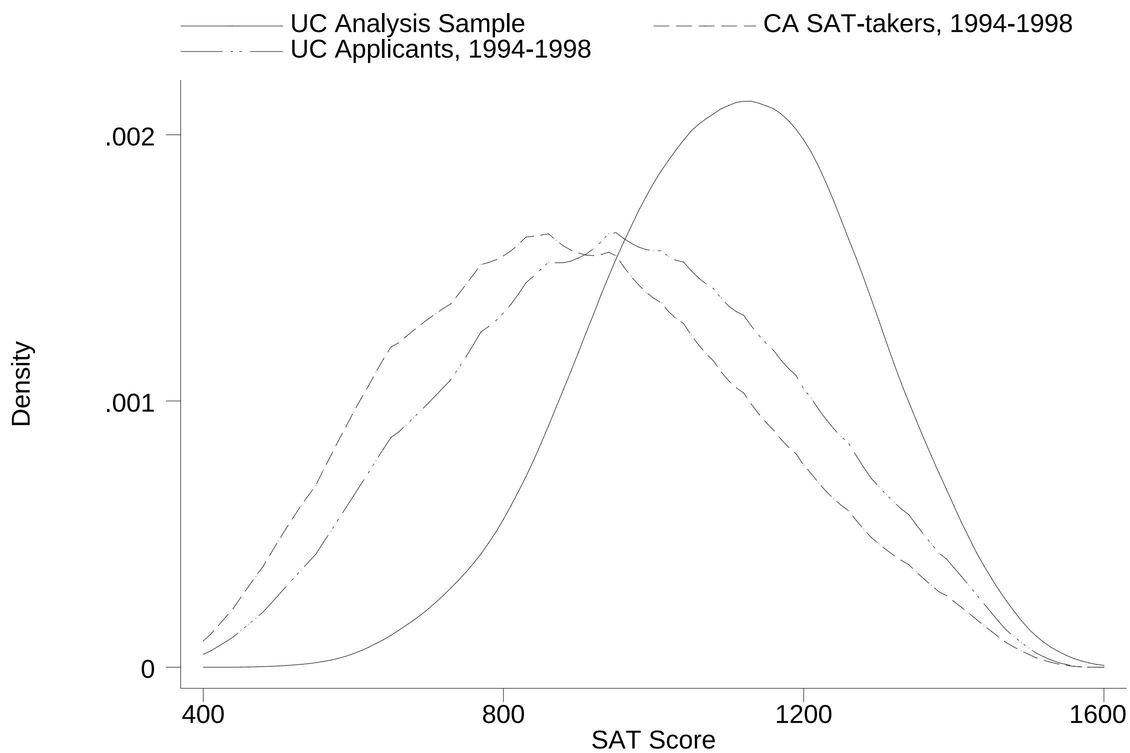

Figure 2: Effect of Estimator and Included Variables on Estimated SAT Validity