University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Economics Department Faculty Publications | Economics Department

4-1-2004

Sample Selection in Models of Academic Performance

Matthew Cushing
University of Nebraksa - Lincoln, MCUSHING1@UNL.EDU

Mary McGarvey
University of Nebraska - Lincoln, MMCGARVEY1@UNL.EDU

Follow this and additional works at: http://digitalcommons.unl.edu/econfacpub

Part of the Economics Commons

Cushing, Matthew and McGarvey, Mary, "Sample Selection in Models of Academic Performance" (2004). *Economics Department Faculty Publications.* Paper 42.
http://digitalcommons.unl.edu/econfacpub/42

This Article is brought to you for free and open access by the Economics Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Economics Department Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Published in *Economic Inquiry* 42:2 (April 2004), pp. 319-322; doi 10.1093/ei/cbh063
Copyright © 2004 Western Economic Association International. Used by permission.

The authors thank an anonymous referee for *Economic Inquiry* for useful suggestions.

Sample Selection in Models of Academic Performance

Matthew J. Cushing and Mary G. McGarvey

University of Nebraska–Lincoln

M. J. Cushing is Associate Professor, University of Nebraska–Lincoln, Lincoln, NE 68588;
tel 402 472-2323, fax 402 472-9700, email mcushing1@unl.edu

M. G. McGarvey is Associate Professor, University of Nebraska–Lincoln, Lincoln, NE 68588;
tel 402 472-9415, fax 402 472-9700, email mmcgarveyl@unl.edu

Abstract
This article shows how admission and enrollment processes affect the interpretation of simple validation studies of academic performance. In a competitive market for students, optimal behavior of admissions committees and applicants drives the simple correlation between test scores and performance toward zero, regardless of the relationship in the population of prospective students. Data from our university's MBA program support the prediction that applicants exhibit a higher correlation between test scores and undergraduate GPAs than do current students. This suggests that standard validation studies will understate the importance of GMAT scores in predicting performance of potential MBA students.

Abbreviations: GMAT, Graduate Management Admissions Test; GPA, Grade Point Average; GRE, Graduate Record Examination; MBA, Master's of Business Administration

I. Introduction

Virtually all graduate programs in the United States require applicants to submit scores on standardized tests, such as the Graduate Record Examination (GRE) or the Graduate Management Admissions Test (GMAT). The usefulness of these tests as predictors of graduate student performance, however, has never been firmly established. In fact, formal statistical investigations (validation studies) typically find that standardized tests have surprisingly little predictive content. Hanson and Harrell (1985) report a negative correlation between GMAT scores and subsequent compensation, Hansen (1971) reports very little correlation between GRE scores and graduate grade point averages (GPAs) in economics, and Sternberg and Williams (1997) find GRE scores of limited use in predicting performance in psychology. Even a recent study by the GRE Board of the Educational Testing Service (1998) shows relatively small correlations between first-year graduate GPA and general GRE scores. Why do admissions committees pay so much attention to these scores when formal evidence suggests that they have little or no predictive content?

In this article, we suggest that a form of sample selection bias makes interpretation of standard validation studies problematic. As Darlington (1998) and Cornell (1998) point out, validation studies use data on students who have matriculated at a particular institution. The admission procedure and acceptance decisions make these data a censored sample from the population of prospective students. Institutions occasionally admit students with atypically low test scores, but only if these students present other, countervailing evidence of high ability. Students with atypically high test scores occasionally will choose to enroll, but only if countervailing evidence of low ability precluded their admission to higher-ranking institutions. Given that both high and low scorers exhibit some other, countervailing evidence on ability, it is not surprising that standardized test scores by themselves are not strong predictors of performance.

This article provides a formal model of how the selection process influences the results of standard validation studies. We find that in a competitive market for students, optimal behavior on the part of admissions committees and on the part of applicants drives the simple correlation between test scores and perfor-

mance toward zero, regardless of the relationship in the population of prospective students. The key features of the model are that institutions select students (and students self-select) based on more information than test scores alone and that a competitive market for students results in institutions enrolling students of relatively homogenous predicted ability. The sample selection process generates a negative correlation between test scores and other observable information, which, as the market becomes more competitive, drives the simple correlation between test scores and performance toward zero.

II. A Simple Analytic Model

Assume that a simple scalar variable, y, measures scholastic ability in the population of potential students to a particular institution. This ability variable will be revealed perfectly (as performance) if the student matriculates but cannot be observed before matriculation. The admissions committee at this and other institutions must make acceptance and rejection decisions based on observable variables that may be correlated with this performance variable. The committee observes characteristics, x_1 and x_2, which, for concreteness, we interpret as the score on a standardized test and an index of undergraduate performance, respectively. Assuming joint normality of three variables, y, x_1, and x_2, in the population of prospective students, expected ability (and performance) will be linear in the two observable variables,

$$\mathrm{E}(y \mid 1, x_1, x_2) = \mu + x_1\beta_1 + x_2\beta_2 \tag{1}$$

The observed characteristics are measured so that β_1 and β_2 are positive.

Consider first, the (infeasible) validation study using the population of potential students at a particular institution. The validation study would reveal predicted performance y, conditional on x_1 (standardized test scores), with expectations taken over the population of potential students,

$$\mathrm{E}(y \mid 1, x_1) = \mu + x_1\beta_1 + \mathrm{E}(x_2 \mid 1, x_1)\beta_2 \tag{2}$$

$$= \mu' + x_1(\beta_1 + b^p_{21}\beta_2) \tag{3}$$

where $b^p_{21} = \mathrm{cov}(x_1, x_2)/\mathrm{var}(x_1)$. The coefficient on test scores, $B^p_1 = \beta_1 + b^p_{21}\beta_2$, reflects not only the true marginal effect of test scores on performance but also the marginal effect of x_2 and the relationship between test scores and x_2 in the population. Because we presume that test scores measure ability, we would expect a positive correlation between test scores and undergraduate performance in the population, ($b^p_{21} > 0$). Therefore, B^p_1 would be an *upwardly* biased measure of the marginal effect of test scores, β_1. The low correlations between performance and test scores in typical validation studies would appear to indicate that the marginal effect of test scores on performance is small and that institutions should give little weight to test scores in the admissions decision.

Now consider actual validation studies, conducted only on matriculating students. In the population of matriculating students, the relationship between mean performance and x_1 is

$$E_T(y \mid 1, x_1) = \mu + x_1\beta_1 + E_T(x_2 \mid 1, x_1)\beta_2 \tag{4}$$

with expectations taken over the truncated subpopulation of matriculating students at institution i. Because the truncated distribution of x_1 and x_2 is not normal, the conditional expectations in (4) are not linear in x_1. Letting $b^s_{21} = \partial E_T[x_2 \mid 1, x_1]/\partial x_1 \mid_{x_1=\bar{x}_1}$, we can approximate (4) with

$$E_T(y \mid 1, x_1) \approx \mu'' + x_1(\beta_1 + b^s_{21}\beta_2) \tag{5}$$

Comparing (5) and (3) tells us that the results of the actual validation study will differ from the hypothetical study to the extent that the relationship between x_1 and x_2 in the truncated subpopulation differs from that in the population of potential students.

To properly interpret equation (5), we provide a simple model of the selection process. Matriculating students are a subpopulation of the group of potential students that have passed through an admissions filter and have then chosen to attend the institution. Let institution i's acceptance criterion be expressed as the following rule:

$$\text{If } \gamma_1 x_1 + \gamma_2 x_2 > M_i, \text{ accept student.}$$
$$\text{Otherwise, reject student.} \tag{6}$$

The variables in the admissions criterion are measured so that the γ's are positive. The next higher-ranking institution has the same decision rule, but with a higher admissions standard, $M_{i+1} > M_i$. Assuming students will always choose to attend the highest-ranking institution for which they qualify, the student's acceptance criterion for the i-th institution is

$$\text{If } \gamma_1 x_1 + \gamma_2 x_2 > M_{i+1}, \text{ reject offer.} \quad \text{Otherwise, accept offer.} \tag{7}$$

The acceptance and rejection decisions on the part of the i-th institution and the applicants make the subpopulation of matriculating students a truncated multivariate normal with truncation rule,

$$M_i < \gamma_1 x_1 + \gamma_2 x_2 < M_{i+1} \tag{8}$$

The relationship between x_1 and x_2 induced by the sample selection rules follows from the results of Kotz *et al.* (2000). The conditional expectation of x_2 doubly truncated by the rule, $M_1 < \gamma_1 x_1 + \gamma_2 x_2 < M_2$ can be expressed as

$$E_T(x_2 \mid 1, x_1) = \mu''' + b^p_{21} x_1 + \sigma(\varphi[\alpha_1] - \varphi[\alpha_2])/(\Phi[\alpha_2] - \Phi[\alpha_1]) \tag{9}$$

where $\varphi(\cdot)$ and $\Phi(\cdot)$ are the PDF and CDF, respectively, of a standard normal, with $\alpha_1 = \sigma^{-1}([M_1 - x_1\gamma_1]/\gamma_2 - x_1 b^p_{21})$, $\alpha_2 = \sigma^{-1}([M_2 - x_1\gamma_1]/\gamma_2 - x_1 b^p_{21})$ and σ is the conditional standard deviation of x_2, absent truncation.

Recall that we defined b^s_{21} as the derivative of the conditional expectation in (9), evaluated at the mean of x_1. This derivative can be shown to be

$$b^s_{21} = \partial E_T[x_2 \mid 1, x_1]/\partial x_1 \big|_{x_1=\bar{x}_1} = [1 - \delta(\bar{\alpha}_1, \bar{\alpha}_2)]\, b^p_{21} + \delta(\bar{\alpha}_1, \bar{\alpha}_2)(-\gamma_1/\gamma_2) \tag{10}$$

where $\bar{\alpha}_1 = \sigma^{-1}([M_1 - \bar{x}_1\gamma_1]/\gamma_2 - \bar{x}_1 b^p_{21})$, $\bar{\alpha}_2 = \sigma^{-1}([M_2 - \bar{x}_1\gamma_1]/\gamma_2 - \bar{x}_1 b^p_{21})$ and $\delta(\bar{\alpha}_1, \bar{\alpha}_2) = 1 - \mathrm{var}_T(x_2 \mid 1, x_1)\big|_{x_1=\bar{x}_1}/\mathrm{var}(x_2 \mid 1, x_1)$. In other words, b^s_{21} is a convex combination of the parameter b^p_{21} that describes the relationship between x_1 and x_2 in the population of potential students and the weights, $-\gamma_1/\gamma_2$, assigned to x_1 and x_2 in the admissions rule. The weights in this convex combination depend on the relative variances of x_2 in the population of potential students and the population of matriculating students.

With these results, we can compare standard validation studies with the ideal study conducted on the population of potential students. The traditional validation study gives a coefficient on test scores of

$$B^s_1 = \beta_1 + ([1 - \delta(\alpha_1, \alpha_2)]\, b^p_{21} + \delta(\alpha_1, \alpha_2)(-\gamma_1/\gamma_2))\, \beta_2 \tag{11}$$

$$= B^p_1 - (\delta[\alpha_1, \alpha_2][b^p_{21} + \gamma_1/\gamma_2])\, \beta_2 \tag{12}$$

The second term in (12) reflects the effects of truncation.

At one extreme, consider institutions with relatively open admissions policies that do not face competition from more highly ranked institutions. For these institutions, the population of students is a random sample from the population of potential students. The conditional variance of x_2 in the matriculating population is close to that of the applicant population, so $\delta(\alpha_1, \alpha_2) \to 0$, $b^s_{21} \to b^p_{21}$ and $B^s_1 \to B^p_1$. For the open-admissions institution faced with little competition, a validation study on the matriculating population will generally provide an upwardly biased estimate of β_1, the true marginal effect of test scores, but provide an unbiased estimate of B^p_1 the predictive content of test scores alone.

At the other extreme, consider a selective institution in a highly competitive market for students. Here, the gap between the admissions standard at competing institutions may be very small, $M_i \to M_{i+1}$. As this gap narrows, $\alpha_1 \to \alpha_2$, $\delta(\alpha_1, \alpha_2) \to 1$, and $b^2_{21} \to -\gamma_1/\gamma_2$. The coefficient on x_1 from the validation study will be a downwardly biased estimate of the true marginal effect of test scores, $B^s_1 \to \beta_1 - (\gamma_1/\gamma_2)\beta_2$, and a downwardly biased estimate of the predictive content of test scores in the population, $B^s_1 \to B^p_1 - (b^p_{21} + \gamma_1/\gamma_2)\beta_2$.

For selective institutions in a competitive market for students, the interpretation of B^s_1 depends on the relative weights attached to x_1 and x_2 in the admissions process, γ_1/γ_2. The optimal weights would presumably be proportional to the relative marginal contributions to ability. If institutions place relatively too much weight on test scores ($\gamma_1/\gamma_2 > \beta_1/\beta_2$), B^s_1 will be negative, whereas if they place relatively too little weight on test scores, B^s_1 will be positive. If institutions weight test scores optimally ($\gamma_1/\gamma_2 = \beta_1/\beta_2$), B^s_1

should be close to zero. In other words, a small correlation between test scores and performance need not indicate that test scores are useless as predictors of performance. Instead a small correlation may simply indicate that institutions are using test scores optimally.

III. Discussion and Conclusion

This article provided a formal model of how the admission and enrollment process affects the interpretation of simple validation studies. Under plausible assumptions, we found that the admission and acceptance process systematically reduces the correlation between test scores and performance relative to the relationship in the population of potential students. Traditional validation studies may mislead institutions into ignoring useful information about the potential productivity of prospective students.

The model assumed that institutions seek to predict performance using undergraduate performance in addition to test scores. In practice, admissions committees often examine a host of qualitative information (*e.g.,* quality of recommendation letters, undergraduate coursework), whereas validation studies typically use a small number of easily quantifiable variables. Of course, admissions decisions are sometimes based on considerations other than expected ability, and admissions based on these other considerations would tend to mitigate the bias emphasized in this study.

The model also assumed that each institution operates in a competitive environment where applicants base enrollment decisions solely on the ranking of institutions. All but the very top institutions face competition from more highly ranked programs. Of course, to the extent that applicants weigh other factors (*e.g.,* location, tuition) in their enrollment decisions, the hypothesized bias due to the selection process will be mitigated.

One implication of our model is potentially testable. The correlation between test scores and other productivity characteristics should be systematically different in the sample of matriculating students and the population of potential students. We tested this proposition on the sample of students and applicants to the master's of business administration (MBA) program at the University of Nebraska–Lincoln. We calculated the coefficient, b^s_{21} from a regression of undergraduate GPA on GMAT scores from currently enrolled students and compared it to the regression coefficient, b^p_{21} using the pool of applicants as the population of potential students.

For male applicants, we obtained $\hat{b}^s_{21} = 0.0038$ with standard error 0.00097; for males currently in the program, $\hat{b}^s_{21} = 0.0023$ with standard error 0.00425. For females, as is often the case, we found the correlations between test scores and performance to be lower for both the matriculating and applicant sample. The estimate of b^p_{21} for female applicants is 0.0014 with standard error 0.00095, and the estimate of b^s_{21} for female students in the program is 0.0007 with standard error 0.00099.

The selection process at the MBA program was not sufficient to render the correlation between test scores and undergraduate GPAs negative. The University of Nebraska-Lincoln MBA program draws largely from a regional population and does not face direct competition from higher-ranking institutions. Further, knowledge of the selection rule at UNL and competing institutions may make the pool of applicants a truncated sample of the pool of potential students. Nevertheless, the correlation between undergraduate GPAs and GMAT scores is almost twice as high in the applicant pool as in the pool of matriculating students. Standard validation studies will therefore understate the importance of GMAT scores in predicting performance of potential students to the MBA program.

References

Cornell, D. G. "Residual Validity Is Not Meaningful." *American Psychologist,* 53, 1998, 575-76.

Darlington, R. B. "Range Restriction and the Graduate Record Examination." *American Psychologist,* 53, 1998, 572-73.

Graduate Record Examinations Board of the Educational Testing Service. "Guide to the Use of Scores." Princeton, NJ, 1998.

Hansen, W. L. "Prediction of Graduate Performance in Economics." *Journal of Economic Education,* Fall 1971, 49-53.

Hanson, B. A,, and T. W. Harrell. "Predictors of Business Success over Two Decades: An MBA Longitudinal Study." Stanford Graduate School of Business Research Paper No. 788, 1985.

Kotz, S., N. Balakrishnan, and N. Johnson. *Continuous Multivariate Distributions, Volume 1, Models and Applications.* New York: Wiley, 2000.

Sternberg, R. J., and W. M. Williams. "Does the Graduate Record Examination Predict Meaningful Success in the Graduate Training of Psychologists?" *American Psychologist,* 52(6), 1997, 630-41.