

The Science of Political Science Graduate Admissions*

Gary King, *Harvard University*

John M. Bruce, *Georgetown University*

Michael Gilligan, *New York University*

As political scientists, we spend much time teaching and doing scholarly research, and more time than we may wish to remember on university committees. However, just as many of us believe that teaching and research are not fundamentally different activities, we also need not use fundamentally different standards of inference when studying government, policy, and politics than when participating in the governance of departments and universities. In this article, we describe our attempts to bring somewhat more systematic methods to the process and policies of graduate admissions.

We had a role in the graduate admissions process at the Department of Government at Harvard University at different times over the past half-decade.¹ We conducted a study of the admissions committee's policies and attempted to bring some of the modern methods of statistical inference, common in political science research, to the task of choosing among applicants to our graduate program. We report here our experience, our statistical studies, and our improvements to the process, as well as a variety of information that may be of use to scholars and administrators at other universities in similar circumstances.

Admissions committee decisions represent an interesting combination of judgments based on quantitative and qualitative information. Until our changes, virtually all decisions were made using only qualitative (or "clinical") methods, even though some of the data on applicants were quantitative, such as grades and standardized test scores. We speculated that this pointed to an inefficiency in our admissions process since "a search of the literature fails to reveal any studies in which clinical judgment has been shown to be superior to statistical prediction when both are based on the same codable input variables" (Dawes 1982, 394). Because some of the information available to admissions committees is quantifiable, it seemed only reasonable that using quantitative methods would help improve our decision making, if appropriately combined with relevant qualitative information (see King, Keohane, and Verba n.d.).

We begin with a brief summary of the Harvard admissions process, prior to our involvement, and provide some generally useful information about standardized tests. We then outline our first statistical study, which we conducted before making any changes to the system. In this study, we demonstrate that the most

common folk wisdom about admissions processes is wrong: admissions committees *are* able accurately to distinguish which students will do best if admitted. This is followed with a section that reports on our changes to the admissions process, based on further statistical studies, designed to reduce the workload of the committee while still improving the quality of its decisions.

Throughout, we are more vague than usual about the specific numerical results of our statistical analyses in order to protect the confidentiality of our applicants, graduate students, and certain parts of our admissions process. For example, we avoided presenting results that would enable prospective applicants to calculate the probability of admission or expected grades in graduate school.

An Outline of the Harvard Admissions Process

In recent years, the Department of Government has received between 600 and 700 applications to its Ph.D. program. From this, we accept 40-50 students (about 7%). As these statistics indicate, the admissions process is extremely competitive, and those admitted have every reason to feel

proud. (However, there is no truth to the rumor that doorways at Harvard are shaped like large keyholes so faculty and students can fit their big heads into their offices.) The process is also far from perfect. Many students who have gone on to become eminent political scientists have been rejected from our program, including a healthy number of our own faculty. (We have no comment on Type I errors!)

From informal checks with chairs of admissions committees at other universities, it appears that the number of applicants to our program is at least as large as any other political science Ph.D. program in America.² The process by which potential applicants self-select into our pool may therefore have a relatively minor effect on many of our subsequent statistical analyses, perhaps, making generalizations to programs at other universities somewhat safer. Of course, like other programs, we do seem to get a number of applicants who do not apply elsewhere, some who prefer to be here in any program than in some particular field, and others who are geographically constrained.

Each application file contains a copy of the application, a written statement of purpose, at least three letters of recommendation, Graduate Record Examination (GRE) scores, undergraduate transcripts, and the Test Of English as a Foreign Language (TOEFL), where appropriate. The letter of recommendation form also asks reviewers to score applicants from 1 (above average) to 9 (below average) on various attributes, including intellectual ability, oral and written expression, emotional maturity, imagination and probable creativity, potential as a teacher, motivation and background for the proposed program of study, and an overall score for promise as a graduate student. Application files from protected minorities are identified with a bright green sticker on the outside and a pink status sheet inside.³

Although fist-fights often break out among the crush of faculty desperately vying to join the Admissions Committee, only about six faculty members serve each year. In fact, committee members take their

jobs very seriously, putting in many hundreds of hours carefully considering the applications. In the years before we implemented our changes (which we describe below), two faculty members read each admissions file and scored applicants on an overall scale from 1 to 14, where higher scores indicated applicants closer to the (unspecified) ideal graduate student. After each file was read by two faculty, we proceeded through several stages where we would weed out some applications, and additional faculty would read and score the remaining files. Eventually, we were left with about 100 applications, each of which was read by all or nearly all members of the admissions committee. Minorities were not weeded.

Finally, the committee deliberates in person for an entire day to narrow down this list of about 100 to approximately 45 who are admitted. We make a preliminary set of decisions based as much as possible on potential to succeed in our program, balanced by intended field of study. We have no strict quotas by field, and the proportion admitted in each changes over the years according to our beliefs about the relative quality of each year's applicants. Although for some smaller fields we do frequently admit zero students if none rank high enough, we do try to ensure that a minimum number are admitted in each of the department's four major fields of study (American, comparative, international relations, and theory, not necessarily in that order).

An important variable in our analyses is whether an applicant is admitted to the program. However, we modify this variable to compensate for features of Harvard University's Affirmative Action and financial aid policies, as it affects the work of the admissions committee.⁴ We give a brief overview of these policies and then our analytical decisions.

In most years, minority applicants appear on this primary list of admits, that is, even before the implementation of Harvard's Affirmative Action policy. After the primary list is complete, we go through the admissions files of all remaining minority applicants to ensure that we do not miss

anyone who meets these same criteria. Then, according to departmental custom, we admit, in a separate Affirmative Action category, any minority applicant who we believe would complete the program if admitted.

We dealt with the separate Affirmative Action list by first duplicating our analyses within each list. The results of these analyses were similar enough so that we were able to merge the two lists for further analyses, with the addition of an indicator for minority status. As a benchmark for why we needed this indicator, if we applied the same rule we are required to use for our Affirmative Action list (admitting those we think would graduate) to all applicants, we would admit 200-300 students a year.

Financial aid decisions are also governed by "Harvard policy." Harvard's policies for graduate admissions are sometimes thought to be "need-blind," but are in fact described in all of its literature as "merit and need-based." That is, the financial status of applicants is unknown to the admissions committee at the time it ranks students. However, these committee rankings are violated for two groups of applicants: First, for students on the margin of being admitted, Harvard accepts those who have resources to attend, even if the admissions committee ranks them below impecunious students we reject. Second, all minorities receive our maximum financial aid package regardless of need.⁵ These policies affect roughly 3-10 students from our primary list and all minorities every year.

In all of our analyses, below, we consider an applicant "admitted" if the admissions committee ranked the candidate on the primary or Affirmative Action lists. Because of our financial aid policy, some of those we consider "admitted" for the purposes of this paper actually do get rejected from the program. About 70% of students who are admitted eventually attend, although we usually receive all or almost all of our top choices and Affirmative Action admittees.

Standardized Tests

There are three GRE scores—analytical, verbal, and quantitative. Each score ranges from 200 to 800, but reported scores have several sources of error. According to the Educational Testing Service (ETS), the standard error of measurement of each of these scores is about 40, which means that the chances are about 95% that the (unobserved) “true score” ranges from 80 points below to 80 points above the reported score. In order to be reasonably confident that their true scores differ, then, two students must have observed scores that differ by at least 112 points (the 95% confidence interval for the *difference* between two scores). In addition, if the same student takes the test a second time, information which is always available, the score increases by an average of about 30 points. (GRE scores are not usually criticized for being biased against certain minority groups, as have SAT scores.)

TOEFL scores range from 200 to 677, with a standard error of measurement of 14.8. This yields a 95% confidence interval of the reported score plus or minus about 30 points. It is reasonable to assume that two students' true scores differ (with 95% accuracy) if one student's observed score is at least 41 points higher than the other. We find that the TOEFL score is a good check on the applicant's abilities in English, but a better test is their written statement of purpose.

Note that these calculations refer only to the standard error of *measurement*. These standard errors are useful in judging differences in (unobserved) true scores from the reported scores. Even if there were no other sources of error, rigidly applying a 95% confidence interval will cause one to make mistakes about 5% of the time, which is a substantial number of individual mistakes in a pool as large as 600-700. Moreover, the standard errors in *predicting* success in graduate school are much larger, and even ETS recommends not using their standardized scores as the sole criteria for admission.⁶

Based on our analyses, the quantitative GRE score is a better pre-

dictor of grades in graduate school than the other two scores, regardless of the field of study within the Government Department. This is not because mathematics is essential to a graduate education at Harvard, but, because mathematics questions are by their nature less ambiguous, it appears to be a better test. Moreover, quantitative, verbal, and analytical skills correlate very highly, and so a high score on a math test, even for an aspiring political philosopher, often predicts success in his or her field.

From the perspective of making a decision about an *individual* student applicant, these figures advise extreme caution when interpreting standardized test scores. A GRE difference between two students of 112 points on a single test is a very large interval, meaning that all students who receive scores above 688 are indistinguishable. About 15% of our applicants receive scores above this figure on all three tests. If we add to this the likely prediction error (which will vary from program to program and student to student), and realize that some students will fall outside even this much larger interval, it becomes critical not to judge any one student too strongly on the basis of this test.

However, admissions committees make decisions for the entire *group* of applicants. Although they probably prefer to reach a just decision in any individual case, their real job is to admit the best group of applicants. From this perspective, some decision rules that produce mistakes in judging individual students can produce a higher average quality of the group admitted. (Technically, the reason is that the standard error of an average score is smaller than the one based on the individual, because it incorporates more information—all the students admitted.) For example, if a department discriminates between students on the basis of GRE score differences of only 50 points, they will make some individual errors but they are still likely to be right more often than wrong. Hence, since there is at least some information in these scores, using them as one among many factors in making decisions will usually result in a group of admittees with higher average quality.

Do Admissions Committees Choose Well?

We began our study by asking whether admissions committees in previous years did their job well. Since most faculty have, at one time or another, served on the admissions committee, and all students made it through the admissions process successfully, everyone at Harvard thinks the committee does a terrific job. We thought we would check a bit more systematically and found that, indeed, past admissions committees have relied on measurable aspects of the students' records in making selections, and that they do a very good job at predicting success in graduate school. Moreover, we could not identify quantifiable factors that would improve this performance.

We made these studies by first selecting all potential admittees in a recent entering class, and coded many features from their admissions files to see if these variables predicted the score that the admissions committee gave each applicant. The variables used in this analysis include the undergraduate grade point average (GPA) within the major field, the overall GPA, the quantitative GRE score, two summaries of the quantitative indicators on the recommendation form, and a ranking of the quality of their undergraduate institution.⁷

One needs to be careful in devising a predictive model of this sort from a long list of possible explanatory variables. To avoid mapping the idiosyncratic features of these data, rather than the systematic parts that persist from year to year, we fit our model to one entering class and, only when finished, evaluated it with a second (see King 1991). In our data, we found the coefficients and predictions for the two separately fit samples to be very similar.

The results (which we omit here to protect the confidentiality of our students and the admissions process) showed that indeed the admissions committee *was* using these variables systematically in assigning its admissions scores to applicants. We demonstrated this by regressing the admissions scores on the variables, and replicating it in our second sample of data. All the variables had meaningfully large coefficients with

sufficiently small standard errors to indicate that the admissions committee was not arbitrarily assigning admissions scores to applicants.

We also wanted to determine whether the admissions committee was doing a good job of predicting students' performance in graduate school. In other words, having determined that the admissions score was systematically derived from information available in the admissions files, we then wanted to see if those admissions scores predicted success in graduate school. Ideally, one would code as dependent variables many measures of success in and after graduate school. These might include graduate course grades, measures of grades on the comprehensive examinations, the time to degree, the status of their first job, and perhaps even measures of success at various stages of their career. However, for legal reasons, the Harvard Graduate School destroys most application files after only four years. As such, the only measure of success we could compute was grades in graduate school. Graduate grades at Harvard, like elsewhere, are high and relatively compressed, but we found statistically that there is enough variation to meaningfully distinguish between students. Subjective judgments of admissions committee members confirm this decision. Although this is the best available measure, graduate grades are not equivalent to success in graduate school or professional success more generally.

We first used the same variables to predict the admissions score in order to predict students' graduate grades, controlling for selection bias in the data—that is, controlling for the fact that only the better students were admitted to the department in the first place. In our example, the selection bias correction was very easy, not requiring any complicated statistical analysis. Usually, statistical models for selection bias require one to estimate the probability of admission and to include this as an additional control variable in the regression, a procedure that leans heavily on some very stringent assumptions about the error process (see Achen 1986; King 1989, Ch. 9), resulting in estimates that are sensitive to small changes in model assumptions. How-

ever, in the present case, the admissions committee's 14-point score is a directly measurable surrogate for the probability of admission. We therefore simply included this variable in the regression to correct for selection bias.⁸ A better method would be to design an experiment where marginal admits are randomly admitted in certain categories.

In our analyses (which as before we omit to protect the confidentiality of our students and our admissions process), once the students' admissions scores were taken into account all other quantitative variables provided *no* additional power in predicting graduate grades. This indicates that prior admissions committees, operating without any formal statistical analyses, have been successful at incorporating *all* of the information in these quantitative measures in their qualitative judgments of candidates for admission.

In other words, we found that the admissions score is itself a very good predictor of graduate grades, even after controlling for all of our other measurable characteristics of the applicants' files. Thus, not only does the admission committee's score reflect the factors we could measure, but it is also composed of other factors we were unable to quantify (such as the written statement of purpose and letters of recommendation). These other qualitative factors significantly improve our ability to predict how well admittees will fare in classes at Harvard.⁹

Reducing the Work-Load of Faculty and Improving the Quality of Decisions

We have demonstrated that prior committees were successful in making admissions decisions, at least in large part, on the basis of expected grades in graduate school. Because we were also unable to improve their decision process with quantifiable information, we focused on reducing the workload without sacrificing the quality of their decisions.

Our first policy change was to adjust the physical procedures by which files were read and scores and comments were recorded. Previously, faculty would write their scores and

comments on the small piece of paper accompanying each file. This made it difficult to make comments (and more difficult to read them), but it also caused a problem when the second faculty member evaluated a file, since he or she would see the comments of the first member. No matter how diligent committee members are, it is almost impossible to avoid being influenced by the person who previously judged the file. This is especially true after having read 200-300 files. Statistically, this means that the two scores are dependent and therefore contain less information than two independent readings would (the equivalent process to autocorrelation in time series data).

We improved on this process by developing a computer program that quickly displayed all relevant codable information about the applicant on the faculty member's computer screen, such as the applicant's name, our explanatory variables, their undergraduate school, etc. We also provided room for faculty to provide comments and a score. This made it easier to find information and much easier to enter detailed comments. Each faculty member's computer program was tied together so that they could easily write notes to the committee staff (such as asking them to call for additional references or to be notified when the undergraduate transcripts arrive), and the staff could respond. We also wrote the program in such a way as to prevent any faculty member from seeing the scores of any other member until the process was over. This alone substantially increased the amount of information elicited from the faculty readers.

We also attempted to be somewhat more race-neutral, at least during the ranking process. To do this, we had the graduate school remove the green sticker and pink sheet identifying minority applicants, and, although we had this information in our computer program, we did not display it for faculty until after the scorings were complete. One can still read the files carefully and often figure out whether an applicant was a member of a minority group, but at least the graders were not focused on it as much. We also asked committee members to judge solely based on

The Profession

merit and, where possible, to ignore race, gender, intended field of study, and other such criteria. In principle, this would produce more meaningful rankings based more on merit than individual faculty beliefs about Affirmative Action or preferred fields of study. Moreover, since Harvard requires us to use a lower admissions threshold for Affirmative Action after we were finished, the procedure would not affect whether minorities would be admitted. Indeed, it should give us better evaluations of all applications.

One problem with the existing procedures was the scale used to judge applicants. The 1-14 scale was not anchored anywhere; committee members were merely told to give higher scores to better applicants. As a result, the scores were not really interpersonally comparable across committee members. Although each member had about the same variance in scores, the mean of some members was a lot higher than for some other members. Although one can fix this problem statistically, after the fact, solutions based on better data are almost always preferable to statistical fixes.

Our goal was to provide a scale with some direct meaning and, especially, to help distinguish among those near the top of the list, where our hardest decisions will have to be made. One possibility was to require committee members to judge applicants by giving their predicted grade point average if admitted. This provided a lot of meaning with such an explicit reference, but it would not have distinguished sufficiently among our top candidates. After some experimentation and compromise, we changed the scale to 0-100, where every 10 points were labeled (i.e., “anchored”) as follows:

- 100 best applicant
- 90 among top 10% of students we admit
- 80 very strong, almost certainly should be admitted
- 70 strong; probably admit; equals average Harvard graduate
- 60 maybe; significant weaknesses in some areas
- 50 marginal for admittance
- 40 too many weaknesses to admit
- 30 very weak
- 20 extremely weak
- 0 do not admit under any foreseeable circumstances

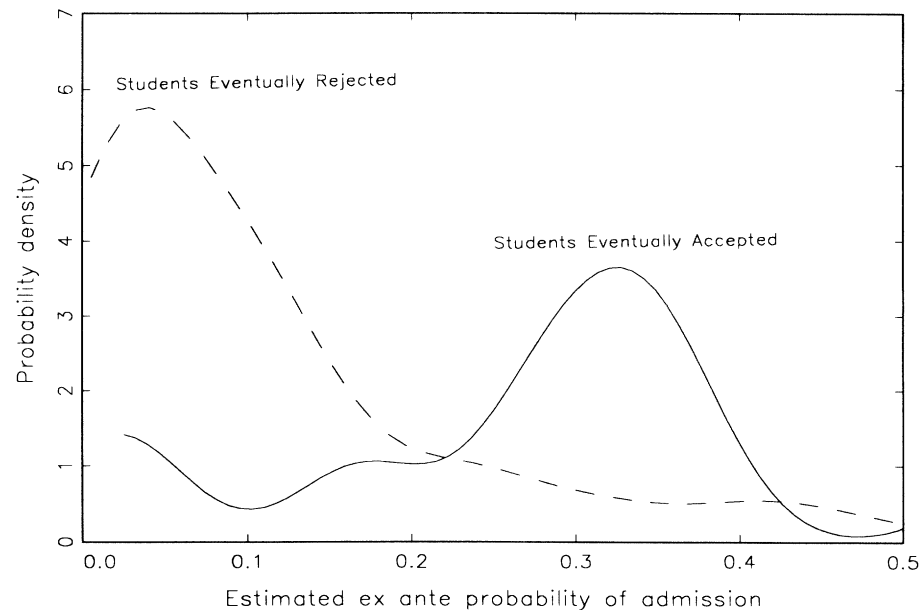
Our computer program reminds committee members of these anchors by displaying the appropriate label as they enter their evaluation. With a few exceptions, our analyses of the data from the new system indicated that different committee members had roughly the same mean and variance across students; in addition, different committees tend to give very similar scores to the same applications. This implies that they are interpreting the scores in approximately the same fashion.

In order to reduce the work of the committee, we conducted a second study to see how well we could predict admissions committee scores with easily measurable characteristics such as GRE scores, grades, and the other variables we used above, but without a committee member’s time to do the reading.¹⁰ Our goal was to reduce the number of faculty who read each folder without sacrificing the quality of the decision. We ran two types of analyses on four separate years of data, using all 600+ applicants for each year. We first estimated a logistic regression of the dichotomous variable admit/reject as a function of our predictor variables. Both analyses showed clearly that we are able to provide some information about what the committee will do, with information easily gathered and processed—and requiring virtually no faculty time.¹¹

Our logistic regression gives the predicted or “ex ante” probability of admission for each applicant. We then checked to see which applicants were admitted and compared those eventually admitted to those eventually rejected. Figure 1 gives an idea of how successful our predictions were. This figure gives two estimated probability distributions (a smooth version of a histogram called a density estimate) representing the probability of admission estimated by our logistic regression. The distribution in Figure 1 represented by the solid line is the ex ante probability of admission for students who were eventually admitted; the dashed line represents the probability distribution of students eventually rejected.¹²

This figure helps us make several key points. First, the distribution for students eventually admitted is clearly distinguishable and to the right of the other: that is, those eventually admitted had on average a higher estimated ex ante probability of being admitted. Second, extremely few students had very high ex ante probabilities of admission into our graduate program (so few that we truncated the figure at 0.5 probability). Third, the prediction is far from perfect. The committee admits some students with very low ex ante probabilities of admission and turns down some with high probabilities. In part, this reflects the goal of diversity

FIGURE 1
Predicting the Admissions Committee’s Decisions



across fields in the department, but it also reflects variables that are omitted from our measurable characteristics but obviously taken into account by the committee.

Given these results, and those of the last section, we certainly would not propose any arbitrary cutoffs, either using standardized tests, grades, or even a statistical prediction. But this does not mean that we should ignore the fact that one can predict likely admissions committee decisions with some degree of accuracy, using information that can be much more easily gathered by staff. We therefore viewed the committee's task as a search process: Given a finite amount of time the committee could devote to its task, we attempted to design a policy that would focus the committee's search where good candidates are more likely to be found—on the right side of Figure 1. Since some very good candidates will be found on the left side of the figure, we still do some searching on that end, but just not as much.

We therefore assigned more faculty to read folders with higher estimated ex ante probabilities of admission, but still made sure that every individual folder was read by at least one committee member. We operationalized this policy by reducing the number of faculty who read each admissions file on the first round from two to one. Applicants made it to the second round if any one of three conditions held: (1) the applicant received a high score by the faculty member reading the application; or (2) our estimated ex ante probability of admission was high; or (3) if the faculty member pushed a button on the computer which indicated that even though they did not like the applicant, another committee member should have a look at the file. These procedures, in addition to eliminating the dependence in faculty scores with the use of our computer program, caused us to lose very little, if any, information, as compared to the old procedures, which required considerably more faculty time.

Concluding Remarks

We have found that the admissions committee to the Harvard depart-

ment of government is successful at picking applicants who are likely to succeed in graduate school (at least as measured by class grades). We have not been able to improve on this performance using quantitative measures. However, the procedures we implemented have reduced the workload of the committee considerably without sacrificing the quality of its decisions.

Since we have shown that admissions committees are quite successful at choosing and ranking students according to their eventual success in graduate school, it may make sense to allocate financial aid more on the basis of perceived merit, to ensure that we attract the best candidates. However, other factors are relevant here, too, of course, such as building comradery among a group of entering students by treating them all equally at the start.

Finally, the process has a number of beneficial side effects. We now have better records on our admissions process, better continuity in our data for different years, and admissions files that are easier to work with in generating statistics of interest in the future.

Notes

*For help with the admissions process, in implementing our procedures, and allowing us to experiment on them, we thank Amanda Hurd, Mary Jane Carty, Chris Willemsen, and the faculty members of the past several admissions committees. We also thank former Dean Brendan Maher, former Government Department Chair Robert Keohane, and Graduate Admissions Officer Dee Blackman, for funding and staff support; Josh Blatt for his usual computer wizardry; and Mitchell Duneier, Claudine Gay, John Fox, Robert Keohane, David Lublin, Robert Putnam, Paul Quirk, Katherine Tate, and Celeste Wallander for many helpful comments.

1. Gary King was chair of the admissions committee; John Bruce was a fellow in the Harvard Data Center 1991-93; and Michael Gilligan was a graduate student in the program until 1992. Our email addresses are gk@isr.harvard.edu, brucej@guvax.acc.-georgetown.edu, and gilligan@acfcluster.-nyu.edu, respectively.

2. We have no terminal M.A. program in the department of government.

3. Protected minorities include self-identified African-Americans, Puerto Ricans, Chicanos, and Native Americans.

4. "Harvard University" is an extremely decentralized organization. We use the phrase "Harvard policy" to refer to a set of state-

ments in official documents from the Faculty of Arts and Sciences and federal and state legislation and court rulings, and as interpreted by various chairs of the Department of Government. These interpretations have varied widely across chairs of Government and other departments at Harvard. Thus, we use the term "Harvard policy" in this article to refer to aspects of the admissions process that are not under control of the admissions committee.

5. In some years, the department gives additional funds to students at the top of our primary list. The financial aid package for minorities is more lucrative than that for all non-minority admits. However, at the departmental level, minority admittees do not compete for funds from the same fixed account as others.

6. Calculating a predictive confidence interval is not as simple as running a regression of success in graduate school on GREs. Since admissions decisions depend on more than GREs, and presumably only the best students are admitted, the regression will be biased by the process of selection. The better job the admissions committee does in selecting good graduate students, the worse GREs will appear to do in predicting success in graduate school.

7. We took the ranking of undergraduate colleges and universities from *Barron's Guide to Undergraduate Colleges*. This ranking measures the competitiveness of admission to the undergraduate institution, not the rigor of the education offered, although the two are related. Our subjective judgment is that this score is somewhat biased against large state universities of high quality such as the University of Michigan but is a reasonable, and the best available, measure of quality.

8. We used an average of the scores assigned for as many faculty as judged each file.

9. One plausible alternative hypothesis is that students given higher admissions scores get better grades in graduate school because their better financial aid package enables students to focus on their studies more. Although this certainly operates at some level, it is unlikely to have much of an effect on our analyses. One reason is the partial "need-blind" admissions process; the other is that the department gave no extra aid to students topping our admissions list for some of the years in our analyses.

10. We had staff help in putting the many different undergraduate grading scores on a common scale. The process is not difficult, but it does take time. Fortunately, the task can be completed by a college student or other inexpensive employee, and, with our other procedures, it can save a lot of faculty time. In principle, it could also be added to the application so that applicants would do the work themselves. The tradeoff of staff time for faculty time is not always appreciated by university administrations; after all, staff time is directly measurable, and therefore expensive, but faculty time is free!

11. We tried many versions of these equations, and the vast majority of the specifications confirmed our working hypothesis. We also developed several procedures to deal with the minority applicants, because the standards used to judge these applications

The Profession

were different from non-minorities.

12. If probability distributions are not familiar, imagine laying out all application files along a straight line on the floor. Place files more to the left when they have a low ex ante probability of admission; and to the right if our statistical procedures indicated that their admission probability was high. Then imagine that folders of applicants who were admitted were sprayed white and the rest red. In Figure 1, the solid line is a smoothed outline of the piles of white folders; the dashed line is a smoothed outline of the piles of red folders.

References

- Achen, Christopher. 1986. *The Statistical Analysis of Quasi-experiments*. Berkeley: University of California Press.
- Dawes, Robyn M. 1982. "The Robust Beauty of Improper Linear Models in Decision Making." In *Judgment Under Uncertainty*, ed. Daniel Kahneman, Paul Slovic, and Amos Tversky. New York: Cambridge University Press.
- King, Gary. 1991. "Stochastic Variation." *Political Analysis* 2: 185-200.
- King, Gary. 1989. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. New York: Cambridge University Press.
- King, Gary, Robert O. Keohane, and Sidney Verba. N.d. *Designing Social Inquiry: Scientific Inference in Qualitative Research*, forthcoming.

About the Authors

Gary King is a Professor of Government, Department of Government, Harvard University. John M. Bruce is Instructor, Department of Government, Georgetown University. Michael Gilligan is Assistant Professor of Politics, Department of Politics, New York University.