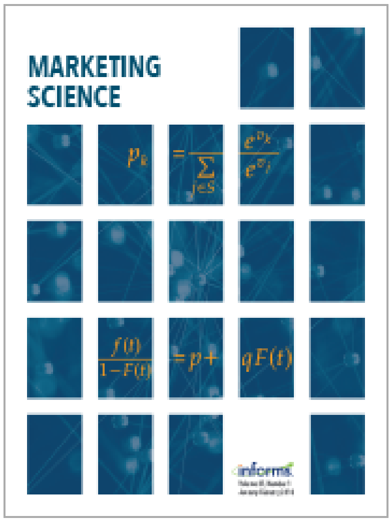


Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Optimal Admission and Scholarship Decisions: Choosing Customized Marketing Offers to Attract a Desirable Mix of Customers

Alexandre Belloni, Mitchell J. Lovett, William Boulding, Richard Staelin,

MARKETING SCIENCE
Vol. 31, No. 4, July–August 2012, pp. 621–636
ISSN 0732-2399 (print) | ISSN 1526-548X (online)

http://dx.doi.org/10.1287/mksc.1120.0707

Optimal Admission and Scholarship Decisions: Choosing Customized Marketing Offers to Attract a Desirable Mix of Customers

Alexandre Belloni
Fuqua School of Business, Duke University, Durham, North Carolina 27708,
abn5@duke.edu

Mitchell J. Lovett
Simon Graduate School of Business, University of Rochester, Rochester, New York 14627,
mitch.lovett@simon.rochester.edu

William Boulding, Richard Staelin
Fuqua School of Business, Duke University, Durham, North Carolina 27708
{bb1@duke.edu, rstaelin@duke.edu}

Each year in the postsecondary education industry, schools offer admission to nearly 3 million new students and scholarships totaling nearly $100 billion. This is a large, understudied targeted marketing and price discrimination problem. This problem falls into a broader class of configuration utility problems (CUPs), which typically require an approach tailored to exploit the particular setting. This paper provides such an approach for the admission and scholarship decisions problem. The approach accounts for the key distinguishing feature of this industry—schools value the average features of the matriculating students such as percent female, percent from different regions of the world, average test scores, and average grade point average. Thus, as in any CUP, the value of one object (i.e., student) cannot be separated from the composition of all of the objects (other students in the enrolling class). This goal of achieving a class with a desirable set of average characteristics greatly complicates the optimization problem and does not allow the application of standard approaches. We develop a new approach that solves this more complex optimization problem using an empirical system to estimate each student's choice and the focal school's utility function. We test the approach in a field study of an MBA scholarship process and implement adjusted scholarship decisions. Using a holdout sample, we provide evidence that the methodology can lead to improvements over current management decisions. Finally, by comparing our solution to what management would do on its own, we provide insight into how to improve management decisions in this setting.

Key words: choice sets; college choice; utility on averages; statistical approximation; nonconvex optimization
History: Received: March 7, 2011; accepted: December 22, 2011; Gary Lilien served as the guest editor-in-chief and Duncan Simester served as associate editor for this article. Published online in *Articles in Advance* April 10, 2012.

1. Introduction

This paper presents a tailored approach for solving two core marketing problems—identifying a subset of potential customers to target and selecting individual-level discounts to price discriminate among these targeted customers. This tailored approach is designed to be applied in the postsecondary education industry where firms (schools) use detailed information about prospective customers (students) to choose a customized offer for each applicant from a finite set of possible offers. We design an approach to capture the institutional details of that industry—most importantly, the school's objective is to achieve not only a given level of revenue (i.e., a desired enrollment level), subject to a scholarship budget constraint but also to attract a student body with desirable average characteristics. This latter objective of achieving a set of students with promising average characteristics greatly complicates the solution method because the addition of one student to the admission set affects the value of every other applicant. More technically, it renders the optimization problem to be nonseparable across the different students. As a result, we develop an optimization method that handles this issue by exploiting the specific structure of the problem facing schools.

In many respects, these admission and financial aid decisions have many similarities to problems addressed in the target marketing literature (e.g., Rossi et al. 1996, Venkatesan et al. 2007, Khan et al. 2009). That literature addresses problems arising in a range of different industries and discusses a number of related issues including, for example, empirical models of customer actions and reactions (e.g., Rossi et al. 1996, Allenby et al. 1999, Reinartz and Kumar 2003). However, only a subset of that literature is concerned with making individual-level offer decisions (e.g., Venkatesan and Kumar 2004). Our approach to admission and financial aid decisions is most similar to this subset. For example, like Venkatesan et al. (2007), we use an individual customer response model that conditions on covariates and is estimated with Bayesian techniques. Also, like them, to solve our optimization problem and customize offers at the individual level, we develop an approximate solution method.

However, our problem has an objective function that differs fundamentally from this targeting marketing literature. In the typical problem posed in the prior literature, the firm values each customer along a single financial metric (e.g., customer sales, previous-period customer revenue, past customer value, customer lifetime duration, customer lifetime value) that is not affected by the characteristics of other obtained customers (see, for example, Venkatesan and Kumar 2004). In our problem, the school cares not only about such a financial metric but also other objectives such as gift giving, school rankings, attractiveness of graduates to employers, and school culture. These objectives imply multiple criteria based on the average characteristics of the obtained students. For instance, to meet all these objectives, a university may desire an entering class that has a high average SAT score, a certain proportion of women, dean's admits, and scholar–athletes, as well as proportions of students from different regions of the world.

Because of this objective function, our problem belongs to a broader class of problems in which the decision maker values configurations of objects rather than a simple aggregation of separate objects. We refer to this broad class of problems as configuration utility problems (CUPs). In CUPs, decisions are made at the level of the individual, but the value of each individual is intrinsically linked through the obtained configuration. Examples from this class of problems include a variety of settings such as deciding a set of strategic alliances to serve multiple purposes, allocating resources to a portfolio of interrelated products, and hedging decisions like group insurance, mortgage-backed securities, and gambling spread decisions. In all of these cases, the problem can be defined by a utility function on characteristics of the attained configuration.

In general, CUPs are extremely difficult to solve. The configuration utility implies that the benefit of giving an offer to any one individual is dependent on all other offers made, because all offers affect the configuration. That is, unlike most targeted marketing applications (e.g., Rossi et al. 1996, Venkatesan et al. 2007, Khan et al. 2009), CUPs cannot use standard methods to separate the targeting decision for each individual from those of the rest of the individuals. This nonseparability, along with the large scale of the problem, typically rules out both simple optimization approaches, such as greedy algorithms and generic optimization methods. In fact, the computational difficulty is sufficiently large that the solution typically needs to be tailored closely to the setting to exploit the particular structure of the problem.

The overarching contribution of this paper is to formulate the admission and scholarship decisions problem as a large-scale CUP and to provide a highly tailored approach to solving this problem. This tailored solution has economic importance both because the postsecondary education industry is very large and because it relies heavily on price discrimination and targeted marketing under the guise of scholarships and selective admission. In terms of industry size, the postsecondary education industry in the 2006–2007 academic year had total revenues of more than $465 billion (U.S. Department of Education 2009)[1] with revenue from student tuition and fees of more than $100 billion. Total enrollments in all degree-granting institutions for the fall of 2007 were 18.2 million students, and first-time freshmen at undergraduate institutions numbered nearly 3 million, with many of these facing selective admission. Selecting the correct set of students to admit represents a large targeted marketing decision. The opportunity to price discriminate is also large, because most students receive a sizable tuition discount. Thus, although the average student tuition at higher-education institutions was approximately $16,000 in 2007–2008, with private schools closer to $30,000, 64% of these undergraduates received some grant aid with an average grant award of $7,100. This suggests an average discount of 44%, or nearly $100 billion in discounts a year.

Interestingly, although these issues are economically important, how to optimally make admission and scholarship decisions has been understudied. A few papers in the marketing literature model the college choice problem (Punj and Staelin 1978, Chapman 1979, Chapman and Staelin 1982, Wainer 2005), but they do not address how to optimally allocate a scholarship budget. For example, Punj and Staelin (1978) run policy simulations to determine the value of increasing the scholarship to a given student, but they do not embed this simulation within the school's decision problem of trading off across multiple average characteristics. Other papers in the broader economics and education literature (see, for example, Manski and Wise 1983, DesJardins 2001, Epple et al. 2003, Avery

[1] All statistics cited in this paragraph come from State Higher Education Officers (2008) and the U.S. Department of Education (2009).

and Hoxby 2004, Niu and Tienda 2008, Nurnberg et al. 2012) address issues that are tangentially connected with the admission and scholarship decisions, focusing primarily on the estimation of student enrollment and its antecedents or correlates, or they consider the role of admission criteria on academic success in a subsequent program (Carver and King 1994, Deckro and Woundenberg 1977). However, two papers are more directly relevant to our problem. Marsh and Zellner (2004) consider admission decisions where the institution only cares about the number of students enrolling (i.e., the characteristics of the students do not enter the objective) and apply a Bayesian decision framework with a loss function to account for uncertainty. Ehrenberg and Sherman (1984) consider cases where the institution cares about a single, objective (observed) quality index (e.g., SAT scores) and propose a model that allocates a single scholarship level to all members of a predetermined group of students. Their approach neither accounts for uncertainty nor provides guidance for solving the optimization problem. In contrast, our approach makes offers at the individual level, accounts for multiple averages and uncertainty, and specifies an optimization method.

Whereas we provide a holistic approach that solves each aspect of the admission and scholarship decision problem, the main contributions of our tailored approach are specializing the utility function and related optimization procedure and accurately predicting prospective students' enrollment choices. By specializing the utility function to this setting, we are able to develop a new optimization approach that exploits the specific structure of the problem facing schools. We complement this optimization approach with an empirical system that provides the optimization algorithm with the necessary inputs: (1) demographic information for the prospective students, (2) the school's budget constraint, (3) the utility function for the school, and (4) predictions about the prospective students' enrollment choices conditional on the school's offers. The most critical element of this system is the last one—providing accurate predictions of prospective students' enrollment choices. Without accurate predictions, the optimization approach cannot hope to improve actual decisions. Our complete system includes solutions to each of these empirical problems of estimation and optimization and accounts for prediction uncertainty. More importantly, in a field test, we provide initial evidence that this system can improve on existing decision processes.

As is true in most college choice situations, the most difficult empirical task is to predict the student's enrollment decision, as this decision is not only conditional on the level of the focal school's scholarship offer but also on what other offers are available to the student. We put forward an approach to making these predictions. This approach, although based on standard Bayesian estimation methods, differs from previous targeted marketing efforts (e.g., Rossi et al. 1996) because of the institutional fact that administrators at the time of the admission and scholarship decisions do not know whether a particular student has been admitted to competing schools nor how much of a scholarship the individual will receive, if any. To account for this uncertainty, we predict not only enrollments but also the competing offers that students choose among. As a result, we construct a two-stage prediction model that first predicts a student's set of offers (choice set) and then predicts the student's choice of schools. We provide evidence that these predictions improve over those from a simpler model.

We test our tailored solution via both a field study and a methodological study in the context of graduate business education. The field study is composed of control and experimental sets of admitted students. In the control group, students were offered scholarships by the school's admissions director. In the experimental group, we adjusted the director's decisions based on our predictions about the prospective students' enrollment choices (but not our full optimization method). We provide evidence of an increase in both the yield and the quality of the entering class. These results suggest that the predictions are accurate enough to apply our optimization method. We then use these field test data in a methodological study that holds constant the sample of students and thus the potential benefit from our empirical system. We provide evidence via policy simulations for how our optimization method would be able to further improve on the (already improved) implemented policy. Using these results, we conclude the application section by giving some initial insight into the ways in which existing managerial heuristics may not perform as well as our approach.

2. Scholarship and Admission Decisions for MBA Applicants

In this section, we present our tailored approach to solving the specific CUP of scholarship and admission decisions. The basic problem is how to target (i.e., admit) a set of customers (students) from a larger set of potential applicants and then attract these selected students by offering them individualized prices (i.e., tuition minus scholarships).[2] Although the exact title differs from school to school, we will refer to the person charged with this problem as the admissions director. This admissions director takes in applications that include a host of information about the

[2] Note that we will refer to individuals as "students," though some do not ultimately matriculate to any school.

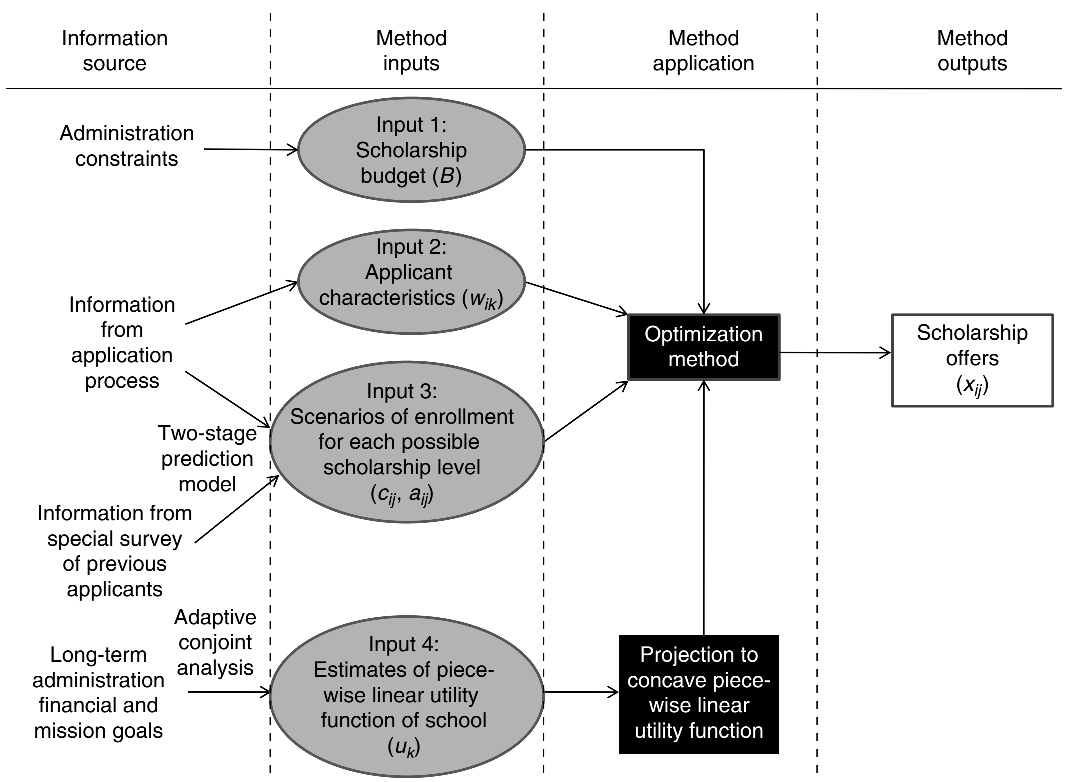
individual such as test scores, grade point average (GPA), gender, race, activities (sports, music, etc.), and other schools the prospective student has applied to. Based on this information, and under a budget constraint, the admissions director would like to make decisions that result in the best expected class profile as measured against some objective function. At the time of making offers, the enrollment decisions of the prospective students are not known. As a result, the solution needs to account for this uncertainty, and the admissions director needs to predict the students' enrollment choices. In §2.1, we present the mathematical formulation of this decision problem.

Figure 1 presents an overview of our approach to solving the admission director's problem. To begin, our optimization approach requires four inputs: (1) the budget constraint, (2) the applicants' characteristics, (3) the institution's utility function, and (4) the probability of enrollment conditional on the scholarship level. This approach yields one output per student, i.e., the admission decision and scholarship amount, which could be zero. The first input comes from the dean's office and the second from the student's application form. The last two inputs are not typically available to admission directors and form critical inputs to the decision process.

To begin, we discuss the available information on the applicant, which falls into four categories: demographics, test scores, admissions ratings, and the schools that students applied to. The available demographics include age, gender, ethnicity, marital status, and country of origin. The test scores include GMAT for all applicants and TOEFL for nonnative English-speaking applicants. The admissions ratings are provided by a set of admissions staff who rate each application on a small number of key dimensions that reflect the perceived quality of the applicant. These dimensions include scores for an essay, recommendations, leadership, work experience, and an overall score that incorporates all of these ratings. For each dimension, two admissions staff provide ratings on a five-point scale (1 = excellent, 5 = poor). We use the average of these two staff members' ratings. In addition, a combination of alumni and admissions staff interview each applicant in person and provide an interview score. Finally, the students provide a list of the schools to which they applied.

Although this application information is essential to solving our problem, the most critical inputs are the enrollment predictions and the institution's utility function. The enrollment predictions translate the information from the application into a probability of enrolling given any particular scholarship level. Many approaches can be used to attain these predictions—from very simple subjective estimates of experienced admissions staff to more complex statistical approaches like the one we use. In our application, we used statistical modeling to obtain these predictions

Figure 1 **The Process Flow Associated with the Admissions Director's Problem**



to improve accuracy. Because the competitive offers and enrollment choices are uncertain at the time of decisions, we model both the admission and scholarship decision rules of competing schools. This leads to the set of predicted offers an applicant receives and the enrollment decision rules of the applicants. We predict across these two stages to form the enrollment predictions. The details of the estimation, prediction strategy, and predictive validity are presented in §2.2.

The institution's utility function translates a class profile (e.g., number of enrolling students, average GPA, average SAT, percent female, percent minority) into a value. Although admissions directors are typically given a set of goals or measures to manage, this function may not be completely characterized in explicit form. Many approaches can be used to specify this function; in our application, it was obtained from the management team (i.e., the dean's office, a faculty committee, and the admissions director) using conjoint analysis, which was postprocessed to accommodate noisy estimates. More detail on how we obtain the objective function is presented in §2.3.

The enrollment predictions and school utility function along with the application information and budget constraint form the necessary inputs to the optimization procedure. The optimization procedure is tailored to the postsecondary education industry where (a) the institution's utility is a function not only of the total number (or total revenue, profit, or customer lifetime value) of acquired customers but also of the average of several observable characteristics associated with these customers, (b) the institution faces constraints on the offers it can make, and (c) the institution is uncertain about acquiring the customers after making offers. The institution chooses for each customer an offer from a discrete set of potential offers (which can include a non-offer, i.e., denied admission) in order to maximize the institution's expected utility.

Given the uncertainty surrounding the acceptance or rejection of the offer, as is the case with many stochastic programming problems (Birge and Louveaux 1997), exact computation of the expected utility is computationally infeasible. Consequently, we approximate the expectation with an empirical average of (i.i.d.) scenarios. However, it is well known that a good (uniform) approximation for the objective function requires a suitably large number of scenarios. Such a large number of scenarios and the different nonconvexities in the problem make computing a solution very challenging. We provide an approach that does so relatively efficiently. The details of the optimization procedure are presented in §2.4.

2.1. Model Development

Let I denote the set of potential students who have applied for admission to the school and let J denote the set of different possible scholarship offers (discounts) that the school can assign to each individual, including options for no admission and admission with no scholarship offer. For each individual in I, there is a set of observable features such as SAT score, gender, etc., denoted by K. The level of the kth feature, $k \in K$, for individual $i \in I$ is denoted by w_{ik}. The school's decision variables are denoted by binary variables x_{ij}, $i \in I$, $j \in J$, which assume the value of 1 if the offer level j is assigned to individual i and 0 otherwise. The random variable for whether individual i accepts the school's offer j is denoted by a_{ij}. This random variable equals 1 if the individual accepts the offer and 0 otherwise.

The school's true objective function is based on the expectation of the sum of utility functions u_0 and u_k, $k \in K$. The utility u_0 is evaluated on the total number of matriculating students. For $k \in K$, each function $u_k: \mathbb{R} \to \mathbb{R}$ is evaluated on the average value of the kth feature for the pool of individuals who matriculate. Thus, the value of a particular matriculated student depends on other students. This is an example of a CUP in which the value of each student (individual) cannot be computed separately from other students.

This school admission and scholarship problem can be cast as follows:

$$\begin{gathered}\max_x E\left[u_0\left(\sum_{i\in I,\, j\in J} a_{ij}x_{ij}\right) + \sum_{k\in K} u_k\left(\frac{\sum_{i\in I,\, j\in J} w_{ik}a_{ij}x_{ij}}{\sum_{i\in I,\, j\in J} a_{ij}x_{ij}}\right)\right],\\ Ax \le b, \sum_{j\in J} x_{ij} = 1, \quad i \in I,\ x_{ij} \in \{0,1\}\ i \in I,\ j \in J,\end{gathered} \tag{1}$$

where the matrix A, the vector x (the stacked vector of x_{ij}'s), and vector b define (generic) linear constraints, and, without loss of generality, we use the convention $0/0 = 0$ (see Technical Appendix B at http://dx.doi.org/10.1287/mksc.1120.0707 for details). We call attention to three features of (1). First, the function u_0 represents the utility associated with the number of acquired objects, and u_k represents the utility for the average value for the acquired objects on the kth characteristic. Second, the assignment constraint, $\sum_{j\in J} x_{ij} = 1$, restricts each individual to be given one and only one offer. Note that if the school wishes to target only a subset of potential students, the formulation accommodates this by constructing one offer type that the individual always rejects. Third, the linear constraints $Ax \le b$ can capture many different aspects of the school's problem such as a budget constraint in expectation and fixed decision variables to represent outstanding offers. In our application, for example, the row of A corresponding to the budget constraint has elements equal to $c_j E(a_{ij})$. For notational convenience, we denote the set of policies that satisfy the constraints in (1) by $\mathcal{R}$.

In practice, the calculation of the exact expectation is usually not possible for the problem of interest. Consequently, we approximate the expectation by using scenarios that are generated randomly and independently. We denote by S the set of scenarios used to approximate the expectation. For any scenario $s \in S$, a_{ij}^s is 1 if individual i is "acquired" when the choice j is assigned under scenario s (otherwise, the value of a_{ij}^s is 0). We denote the probability of scenario s as $p_s = 1/|S|$. We then look to find the policy $x \in \mathcal{R}$ that maximizes the schools's utility averaged over all the generated scenarios in S. This leads to the following empirical formulation, which we used to approximate (1):

$$\max_{x \in \mathcal{R}} \sum_{s \in S} p_s \left[u_0 \left(\sum_{i \in I,\, j \in J} a_{ij}^s x_{ij} \right) + \sum_{k \in K} u_k \left(\frac{\sum_{i \in I,\, j \in J} w_{ik} a_{ij}^s x_{ij}}{\sum_{i \in I,\, j \in J} a_{ij}^s x_{ij}} \right) \right]. \tag{2}$$

The optimization problem (2) that arises from this specific CUP is related to the literature on product line design. Notably, the model studied in Chen and Hausman (2000) has one linear fractional term (and no utility function). The authors exploit this additional structure to invoke unimodularity results that guarantee the existence of an integer optimal solution for their linear relaxation. Recently, Schön (2010) studies a model that maximizes the sum of ratios that do not have variables in common. Her model has additional fixed cost decisions that linked the different terms. She also shows how to build on Chen and Hausman (2000) to derive an efficient implementation. In contrast to Chen and Hausman, our model has the sum of many fractional terms, and unlike Schön (2010), these terms involve the same decision variables in different terms. As a result, the unimodularity property is lost, and we are unable to build on the advances in these papers. Furthermore, unlike either paper, we allow for an arbitrary utility function on the total number of acquired objects (e.g., which could be used to accommodate scale efficiencies or network effects). These issues led us to develop a tailored computational method for solving (2).

2.2. Predicting the Probability of Enrollment Conditional on Scholarship Offer

The first set of critical inputs needed for our optimization model is the scenarios, a_{ij}, i.e., the random variables for whether each student will enroll at the focal school given any possible scholarship amount. In theory, it is possible to simply ask an expert to create the scenarios. However, the task of creating these estimates would be onerous without some algorithm to relate scholarship amounts to characteristics of students. Furthermore, interviews with the admissions director indicated such a task would involve substantial guesswork.

A number of alternative statistical models could be used in estimating the a_{ij}. The leading example comes from the marketing literature (Punj and Staelin 1978), but this model was not designed to accommodate uncertainty in the choice set (i.e., the choice set was known). The decision maker, however, is generally uncertain about the choice sets (i.e., the actual offers its competitors will make to the applicants) until after it has made decisions about admissions and scholarships. Two generic approaches exist for dealing with this uncertainty: (1) ignore it, predicting student enrollment decisions in a single stage based on observable factors at the time of the decision, and (2) estimate both the student's choice set and choice rule and use a two-stage prediction. Our tests on holdout data suggested that the latter option performed better than the former.

This led us to develop two models—one that relates the competing schools' choices about prospective students to the characteristics of those students and one that relates the students' enrollment choice to the set of admission and scholarship offers they received. The data used to estimate these models come from the students' application form and from a special survey of past applicants. These application and survey data include for each student (a) the student's characteristics; (b) the schools where the student applied; (c) the schools that admitted the student and the scholarships, if any, that the student received from these schools; and (d) the student's ultimate choice, i.e., the identity of the school, if any, where the student matriculated.

We collected the last three sets of data via a Web survey that was sent to a large number of applicants to the focal school. These potential respondents included both admitted and denied students from the previous two years.[3] Although this sample is not a sample for the population of all applicants to MBA programs, it is exactly the sample desired for our purposes (estimating competitor actions and applicant choices about applicants to the focal school).[4] More specifically, we obtained 1,191 responses, of which 1,139 contained complete information. Of the total set, 558 matriculated at the focal school, 494 turned down the focal school's offer, and 139 applied but were not admitted. We used the responses to determine the types of students admitted to, and possibly receiving scholarships from, 20 competing MBA programs and an "others" program that was used to represent a collection of schools that individually had only a small

[3] The overall response rate was 43%, although the rates were slightly higher for admitted than denied populations. However, conditional on admission status, no clear nonresponse bias was present in terms of the observable characteristics of those responding versus those not responding.

[4] It is important that the application set and decision rules for the prediction period do not differ markedly from the sample period.

number of joint applications with the focal school. With these data, we estimate two related Bayesian statistical models.

2.2.1. Estimating Competing Schools's Decision Rules. The first of these models estimates school-specific admission and scholarship decision rules for the competing 21 MBA programs. This model provides a mapping between applicant characteristics and each school's admission and scholarship decisions, thereby allowing us to predict each new applicant's choice set and scholarship offers.

Each individual i applying to school k has an underlying index $u_{ik} = v_{ik} + \epsilon_{ik}$, and ϵ_{ik} is i.i.d. normal with mean zero and variance σ_k. The deterministic component v_{ik} is defined as

$$v_{ik} = \beta_k X_i,$$

where X_i is the vector of applicant characteristics that schools value and β_k is the vector of parameters. Any functional form of the original characteristics can be included in the data vector X_i. In practice, we included linear terms for all variables and both linear and quadratic terms for GMAT. Each school also has its own cutoffs for admission, α_{-1k}, the minimum scholarship offer, α_{0k}, and the maximum scholarship offer, α_{1k}. The school k decision about individual i is observed in the survey data and is denoted as D_{ik}. It takes a value -1 (indicating denied), 0 (indicating admission without scholarship), or some positive value $\alpha_{0k} \le sc \le \alpha_{1k}$ (indicating admission with scholarship). The values α_{0k} and α_{1k} are observed, and α_{-1k} and the σ_k are to be estimated. We translate the underlying index into a decision D_{ik} via the following rule:

$$\begin{aligned} D_{ik} = {} & -1 \cdot 1(u_{ik} < \alpha_{-1k}) + u_{ik} \cdot 1(\alpha_{0k} \le u_{ik} \le \alpha_{1k}) \\ & + \alpha_{1k} 1(u_{ik} > \alpha_{1k}), \end{aligned} \tag{3}$$

where $1(\cdot)$ denotes the indicator function.

Hence, we have the following likelihood function for a single decision D_{ik}:

$$\begin{aligned} p(D_{ik} \mid \theta_{1,k}) = {} & \Phi(\alpha_{-1k}; v_{ik}, \sigma_k)^{1(D_{ik}=-1)} \\ & \cdot (\Phi(\alpha_{0k}; v_{ik}, \sigma_k) - \Phi(\alpha_{-1k}; v_{ik}, \sigma_k))^{1(D_{ik}=0)} \\ & \cdot \phi(D_{ik}; v_{ik}, \sigma_k)^{1(\alpha_{0k} < D_{ik} < \alpha_{1k})} \\ & \cdot (1 - \Phi(\alpha_{1k}; v_{ik}, \sigma_k))^{1(D_{ik}=\alpha_{1k})}, \end{aligned}$$

where $\Phi(x; \mu, v)$ and $\phi(x; \mu, v)$ are the cumulative distribution function and probability density function, respectively, of the normal distribution with mean μ and variance v at x, and $\theta_{1,k} = (\alpha_{-1}, \sigma_k, \beta_k)$ are the parameters to be estimated. There are n individuals and J total schools with individual i applying to only J_i^a of these. Hence, the likelihood is the product of $\prod_{i=1}^{n} \prod_{k=1}^{J_i^a} p(D_{ik} \mid \theta_{1,k})$. We provide details of the complete model and sampling chain in Technical Appendix D.

2.2.2. Estimating Students's Enrollment Choices. The second model estimates the utility function of the applicants to make predictions about student matriculation decisions conditional on the focal school's scholarship offer and predicted admission and scholarship offers of competitors. It is estimated using the above-referenced survey data.

We assume individual i has utility z_{ik} for available option k. The observed enrollment decision is formulated as a vector, E_i, with components $E_{ik} = \{0$ if not chosen, 1 if chosen$\}$. Individuals choose from the set of options to which they applied and were admitted, A_i. The individual chooses the option in A_i offering the maximum utility; i.e., $E_{ik} = 1(z_{ik} > \max_{h \neq k} z_{ih})$. The utility for an option has both a stochastic and a deterministic component. The deterministic component, V_{ik}, is composed of a college intercept (i.e., the college brand effect), γ_{0k}, and the effect of the scholarship offer, γ_{1i}, where the individual-level heterogeneity comes from the observed demographics, X_i. The deterministic component is represented as

$$\begin{aligned} V_{ik} &= \gamma_{0k} + \gamma_{1i} sc_{ik}, \\ \gamma_{1i} &= \Psi X_i, \end{aligned}$$

where sc_{ik} is the scholarship offer from option k and Ψ is a vector of linear parameters.[5]

For estimation, we use a multinomial probit with absent dimensions (Zeithammer and Lenk 2006). We assume standard, diffuse conjugate priors and follow the standard sampling procedure (Zeithammer and Lenk 2006). We provide more detail in Technical Appendix D.

2.2.3. Predicting a_{ij}. We use the estimates obtained from applying these two models to make predictions of whether a given student will matriculate at the focal school, i.e., realizations from the random variables a_{ij}. At the time of decisions (and these predictions), the admissions director has information on only the applicants' characteristics and the set of schools the prospective student applied to. Using only this information, we wish to predict the probability of enrolling given each scholarship level. In our problem, we have 21 scholarship levels. Because we wish to account for uncertainty, we make many predictions for each prospective student and scholarship level.

We make the predictions as follows. First, we wish to predict which schools will admit and give scholarships to the applicant. We use the characteristics obtained from the applications X_i^p and the set of considered schools (applications) in combination with a

[5] Note that although the scholarship term is linear, the nature of the probit model leads to nonlinear responses to scholarship. Further, nonlinearities (e.g., polynomial functions of sc_{ik}) can easily be accommodated, but in our application, the data did not support it.

sample from the posterior distribution of the first model. With these inputs, we predict the offers for each individual in the applicant set. Specifically, we predict all competing schools for which the student applied (i.e., not the focal school), producing a vector D_i^p of length $J_i^a - 1$.

These prediction vectors account for two sources of randomness—uncertainty about the parameters $\theta_1 = \{\theta_{1,k}\}_{k=1}^J$ contained in the posterior distribution and uncertainty about the school decisions arising from the stochastic component of the underlying index. Thus, to sample from the predictive distribution $p_{J_i^a}(D_i^p \mid X_i)$, we draw θ_1 and a vector of the stochastic elements, ϵ_i. Jointly, these determine the vector of underlying indices, u_i, which directly maps to the vector D_i^p as expressed in Equation (3). Thus, we have a predictive distribution, $p_{J_{i-f}^a}(D_i^p)$, where D_i^p is a vector taking values -1 (denied admission), 0 (admitted, no scholarship), or $\tilde{sc}$ (admitted with scholarship $\tilde{sc}$). The J_{i-f}^a subscript denotes that the random vector has variables for each school the individual i applied to other than the focal school. Thus, for each individual we have a set of m_c such vectors sampled from $p_{J_{i-f}^a}(D_i^p)$ containing indicators of which schools admitted the student and offered scholarships.

Second, we use these vectors as inputs to determine which school each individual will attend. For each of the m_c samples, we make predictions about the applicant's utility, z_{ik}, for the set of available competitive options A_i^p (i.e., options with $D_i^p \geq 0$). To make these utility predictions, we incorporate the applicant characteristics and use the posterior samples from the second model. We have uncertainty about both the parameters θ_2 and the unobserved utility components. Thus, we wish to sample from the predictive distribution of utilities, $p_{|A_i^p|}(z_i)$, to identify whether for that sample the focal school is chosen. To do so, we draw m_z samples of θ_2 and the stochastic elements for each of the m_c samples of D_i^p. We then calculate the utility effect of scholarship offers by the focal school at each scholarship level. We use the draw of Ψ and multiply the scholarship amount by ΨX_i. This allows us to calculate for the sample s whether the focal school has the highest utility (i.e., whether $z_{if} > \max_{k \neq f} z_{ik}$), in which case $a_{ij}^s = 1$; otherwise, $a_{ij}^s = 0$. Thus, for each individual we have a sample of $m_z m_c$ realizations of a_{ij} drawn from the predictive distribution $p(a_{ij})$. These predictions are conditional on the scholarship offer from the focal school (hence, the j subscript) and the application characteristics of the individuals X_i. We also use these a_{ij} values to calculate the expected costs $c_{ij} = \sum_{s \in S} a_{ij}^s c_j$, where c_j is the value of scholarship level j and S is the set of scenarios used. In our application, the largest number of scenarios we use is $|S| = 5{,}000$ samples per individual (111) and scholarship level (21), resulting in $11{,}655{,}000$ values of a_{ij}^s. Note that this is 1,000 scenarios for optimization and 4,000 scenarios for the holdout sample.

It is important to recognize that the estimation of $p(a_{ij})$ can be accomplished with any number of methods without altering the optimization problem. For the optimization, the important outcome is that the a_{ij}^s scenarios form a reasonably accurate link between the focal school's decision variables x_{ij} and the true random variable a_{ij}, i.e., whether the student ultimately matriculates at the focal school. We next provide evidence that the predictive distribution has a reasonable relationship to actual student decisions.

2.2.4. Predictive Validity. We investigate the predictive power by calculating the percentage of time we correctly predict the student's school choice using a holdout sample of 434 individuals that come from a year different from the calibration sample. We use the hit-rate measure because it most closely reflects what is required by the optimization procedure. To evaluate the predictive validity, we evaluate both the single-stage and two-stage predictions by comparing them against simpler alternatives.

First, we consider the case when the choice set is known (i.e., we know more information than the decision maker actually has) and refer to this as a single-stage prediction. Our model has an average hit rate of 78% for this one-stage prediction. This accuracy compares favorably to previous models in which a single-stage prediction garnered 74% accuracy (e.g., Punj and Staelin 1978).

Second, we discuss the two-stage prediction accuracy of our model. In this case, we predict both the choice set (decisions of competitors) and the enrollment decisions. The average hit rate was 73% across the two stages. As expected, we lost accuracy relative to the single-stage predictions, but this loss is not very dramatic. To compare this against a simple option, we created a model in the spirit of Punj and Staelin (1978), but ours uses a probit form and operates on the set of schools to which the candidate applied (i.e., matching the information available at the time of the decision). This model assumes that the unobserved index for each school to which the candidate applied includes an intercept for the school and an interaction of the school intercept and applicant demographics. This comparison model is much simpler to estimate and the predictions are much less complicated to simulate, but it ignores the fact that the choice set might differ from the application set. For this model, we find a much lower hit rate of 59%. The intuition behind this decrease is that the process that turns application sets into admission and scholarship offers appears to be difficult to approximate in the single-stage estimation model.[6] Furthermore, the simpler model does not

[6] In a preliminary examination, we found that our first-stage predictions of individuals' choice sets are accurate enough to dramatically

appear to effectively incorporate historical information on competing schools' decisions.

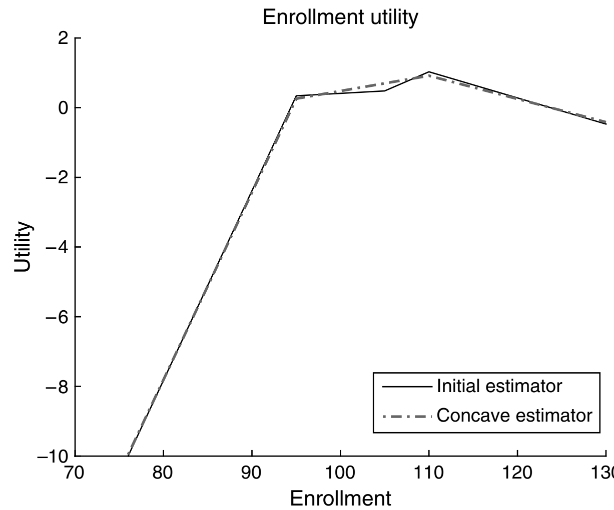
This evidence suggests that we are able to predict, to a reasonable degree of accuracy, an individual's choice conditional on only application information and the focal school's scholarship offer. Thus, although the estimates contain uncertainty, they create a reasonably accurate link to actual enrollment choices. This is the necessary requirement for the optimization method to yield meaningful results.[7]

2.3. Estimating the Institution's Utility Function

We now discuss the fourth input shown in Figure 1—namely, an estimate of the institution's utility function. Getting such an estimate requires the institution to translate its long-term mission and financial goals into a tangible, actionable short-term utility function stated in terms of the desired characteristics of an entering class of MBA students. We obtained this translation by first interviewing the dean's office to identify the key mission and financial goals of the school, which resulted in identifying the following eight incoming class characteristics used to measure those goals: total enrollment (revenues), average GMAT score, class average of the admissions director's assessment of the student's potential (from excellent = 1 to poor = 5), percentage of the class that had an overall assessment with a score of 4 or below, average interview score (from excellent = 1 to poor = 5), percentage of the class that had interview grades of 4 or below, percentage of females, and percentage of foreigners.

We then asked selected faculty and administrators, including the admissions director, to complete an adaptive conjoint analysis computer-based interview (e.g., Gustafsson et al. 2007) as implemented in Sawtooth Software (see http://www.sawtoothsoftware.com). We used the responses to estimate for each individual completing the task their part-worth utilities (utilities for each level within each attribute). We determined that there was general agreement in terms of the weights and shape of the individual part-worths across our respondents, suggesting that the faculty, administration, and admissions director had a common view of the objective of the admissions process. Hence, we combined the individual estimates to form a piecewise linear function of the characteristics of the enrolling class. We showed the results to the administration, and they agreed that this function effectively translated the school's long-term goals into the desired trade-offs for the incoming class.

Figure 2 **The Initial and Concave Estimators for the Utility for the Enrollment Feature**



While obtaining these part-worth estimates, we did not ensure that each of the individual functions were concave over the total region.[8] Not surprisingly, a few of the estimated functions violated the concavity assumption in (4). Consequently, we tailored a projection method to our setting, as described in Technical Appendix C, and applied it to obtain consistent concave estimators. Figure 2 illustrates the impact of this methodology on the utility of total enrollment. Not only does this figure illustrate the worst violation of concavity in our estimates, but it also points out the school's strong disutility for underenrollment as well as its disutility for bringing in a class size that exceeds its capacity constraint.

2.4. Solution Strategy

The approximate utility as presented in (2) contains both the true objective function (1) and an approximation error. The value of an optimal solution to (2) contains this approximation error. Because there are a large number of decision variables in our problem, the model has considerable flexibility to fit the approximate solution. Consequently, if the approximation error is too large relative to this flexibility, the maximizer is likely to overfit the approximation error in order to increase the value of the objective function in (2). This overfitting is referred to by Smith and Winkler (2006) as the "optimizer's curse" because the expected value of the optimal value of (2) is always larger than the true optimal value of (1). As the number of scenarios increases, the concern about overfitting decreases,

reduce the probability of some options, particularly for individuals who applied to more than three schools.

[7] Other techniques could also be used to model and predict enrollments. For instance, one might argue that the conditioning data in the enrollment choice (school admission and scholarship decisions) may be endogenous, introducing bias in the coefficients we estimate (e.g., Manchanda et al. 2004). We did not attempt to solve this issue with the knowledge that the focal school uses no other information in its decisions beyond what we observe.

[8] Some alternative methods enforce such constraints directly in the estimation (see Allenby et al. 1995).

but the computational demands for the optimization increase.

The problem formulation in (2) belongs to the class of nonlinear integer problems. Recently, impressive progress has been made in the field of mixed integer nonlinear programming (Bonami et al. 2008). These solvers use enumerating schemes to provably compute the optimal solution. However, these schemes are computationally demanding for the number of scenarios we require in order to keep overfitting negligible. Thus, we focus on solving approximately by developing upper bounds on the optimal solution value and an efficient heuristic search strategy that utilizes the information in the upper bound. This combination of approaches efficiently identifies solutions while providing a bound on how far the heuristic solution is from the best obtainable solution. To that end, we develop a new Lagrangian relaxation method to achieve sharp upper bounds on the optimal value and use the Lagrangian multipliers as input for the heuristic search.

2.4.1. Computing the Optimal Bounds. The efficient computation of sharp upper bounds on the optimal value of the scenario-based problem is one of our main methodological contributions. We exploit the particular structure of our problem and show how these bounds can be computed efficiently in theory and practice. The details are presented in Technical Appendix A, but we introduce the essential idea here. The motivation is to relax the requirement to use the same policy for all scenarios and introduce a penalty if different policies are used in different scenarios. We relax three constraints, introducing a multiplier for each: (i) the constraint that the policy is the same across scenarios (which we introduced in the last sentence) with the multiplier λ, (ii) the generic linear constraints with the multiplier σ, and (iii) the piecewise linear objective functions[9] u_k with the multiplier α. Ultimately, we reformulate the problem so that for each fixed value of the three multipliers (λ, σ, and α), we have the so-called dual function:

$$\begin{aligned}\phi(\lambda,\sigma,\alpha)&\\ =\max_{x,\,x^s}\quad&\sum_{s\in S}p_s\Bigg(u_0\bigg(\sum_{i\in I}\sum_{j\in J}a_{ij}^s x_{ij}^s\bigg)\\ &+\frac{\sum_{i\in I}\sum_{j\in J}(\sum_{k\in K}\sum_{l=1}^{l_k}\alpha_{k,l}^s r_{k,l}w_{ik})a_{ij}^s x_{ij}^s}{\sum_{i\in I}\sum_{j\in J}a_{ij}^s x_{ij}^s}\\ &+\sum_{k\in K}\sum_{l=1}^{l_k}\alpha_{k,l}^s d_{k,l}\Bigg)\\ &-\sum_{s\in S}\lambda_s'(x^s-x)-\sigma'(Ax-b)\\ \text{s.t.}\quad&\sum_{j\in J}x_{ij}^s=1\quad\text{for } i\in I,\ s\in S,\ x_{ij}^s\in\{0,1\}\\ &\text{for } s\in S,\ i\in I,\ j\in J.\end{aligned}\tag{4}$$

[9] We write u_k as the minimum of l_k linear pieces, and each linear piece is given by a slope $r_{k,l}$ and an intercept $d_{k,l}$, $l=1,\ldots,l_k$, $k\in K$.

In Technical Appendix §A.1, we show how to efficiently calculate this dual function and that, by well-known results of weak duality (see Hiriart-Urruty and Lemaréchal 1993), it bounds the optimal value to (2). Hence, we minimize the dual function $\phi(\lambda,\sigma,\alpha)$ to achieve the tightest bound.

This approach—by handling multiple assignment constraints, linear constraints, and piecewise linear concave utility functions of multiple 0–1 linear fractional terms—contributes to a large literature that uses the methodology of Lagrangian relaxation (see the seminal work of Held and Karp 1971; the references in Fisher 1981, 2004; and the related literature on max–min 0–1 (linear) knapsack problems—e.g., Yu 1996 and Iida 1999, who consider piecewise linear concave utility functions of univariate variables and one linear constraint). The standard alternative approach is found in the literature on 0–1 fractional programming with multiple fractions. This literature has focused on bounding the optimal value using linear programming relaxations (Wu 1997). However, when we apply this linear relaxation method to our scenario-based problem, the obtained bounds are not as tight as those obtained using our new Lagrangian relaxation method (see Technical Appendix §A.3). Our intuition is that the tighter bounds are achieved by taking advantage of the presence of assignment constraints and the particular integrality of denominators to better account for nonlinearities in the fraction terms and on the utility over the total number of acquired objects. Beyond the tighter bounds, the Lagrangian relaxation method has two additional advantages over the linear relaxation method. First, there are no required assumptions on the utility over the total number of acquired objects. Second, it allows us to consider the case of acquiring no students in a particular scenario.

2.4.2. Generating Feasible Solutions. There is a variety of heuristics that could be applied to problem (2). Our approach is to take advantage of the information generated from the Lagrangian relaxation described in §2.4.1. Broadly speaking, the heuristic has two phases:

1. Generate a feasible solution based on the solution of dual problem (4).

2. Try to make local improvements on neighborhoods constructed with dual information.

In Phase 1, we generate feasible solutions based on the current Lagrangian relaxation solution $\{x_{ij}^s\}$ as follows (recall that the Lagrangian relaxation awards one scholarship level to each student in each scenario where one level is, in effect, not being admitted). Pick an applicant randomly (denote its index by i). For this applicant, compute $p_j = \sum_{s\in S} x_{ij}^s/|S|$, which is the proportion of scenarios that the applicant was awarded scholarship level j in $\{x_{ij}^s\}$. To pick a scholarship level for this applicant, draw the candidate scholarship level from the distribution $\{p_1, \ldots, p_J\}$. Award the drawn scholarship if it does not violate the current budget constraint. Otherwise, the student does not receive a scholarship. Remove the applicant i from consideration and loop to remaining students.

After an initial feasible solution is obtained, we proceed to make a sequence of local moves to further improve the objective value while maintaining feasibility. The local search is inspired by "two-opt" and "three-opt" moves, as were originally proposed for the traveling salesman problem (see Jünger et al. 1994). However, to improve efficiency, we select a small subset of decisions on which to try the local searches (despite being polynomial time, they can be computationally demanding).

In our problem, a "move" corresponds to switching the scholarship-level choice for a student. A two-opt move picks two applicants and chooses two other scholarships for them. Similarly, a three-opt move changes the scholarships of three students simultaneously. To select a set of promising scholarship levels for each student, we select the five most frequent scholarship levels awarded by the Lagrangian relaxation.

Finally, we note that we rerun the heuristic at each iteration at which the Lagrangian relaxation achieved an improvement in the optimum bound, rather than only once at the end. This allowed us to efficiently explore the objective function.

3. Results

In this section, we apply our approach to solving the scholarship and admission problem in order to demonstrate its potential value. In this application, the focus is on allocating scholarship dollars to a set of the focal school's admitted MBA applicants, because this aspect of the problem was the focal school's primary concern. In this setting, we first conducted a field experiment to test the value of our improved enrollment predictions. We compare the allocations determined by the admissions director for one group of admitted students with a second group, where the allocations benefited from our enrollment predictions.[10] Finally, we use the same set of students to run a policy simulation using our proposed optimization methodology to determine the degree to which we can improve the school's overall utility subject to a budget constraint.

3.1. Field Experiment and Results

Our field experiment involved a single admission round. We divided the admitted students into two groups using probabilistic assignment to ensure approximately equal profiles on the characteristics included in the institution's utility function. The first group (control) had 112 admitted students and scholarships assigned by an experienced admissions director.[11] The second group (experimental) had 111 admitted students where we were responsible for assigning scholarships. We note that at this time, we had not completed the proposed methodology and instead used a stochastic search approach (Gallant and Tauchen 2010) that benefited from the improved enrollment predictions from our two-stage estimation, but not the better heuristic and optimization approach we propose.[12] Hence, any improvement that we achieve in this experimental group should represent a lower bound for improvements associated with our proposed methodology, because the stochastic search was able to improve, but not optimize, the focal school's utility function in any practical amount of time.

Using this limited stochastic search method, we identified 19 students who we believed should be awarded increased scholarship support. We predicted that by giving these students additional scholarship funding, we would not only greatly increase their probability of matriculating but also increase the average quality level of the enrolling class. Interestingly, 14 of these students had initially not been awarded any scholarship. It was decided to give these 19 students the identified scholarships. We then compared the actual matriculation results for this experimental group to the sample of 112 students who received scholarship awards (if any) as determined by the admissions director.

We compare the two different scholarship policies both (a) in terms of an increased yield rate and (b) in terms of average profiles for the two groups. We explore the differences in yield rates across the two groups after controlling for student characteristics and the scholarship amounts.[13] The only control

[10] At the time of the field study, the full methodology discussed above was not yet available, and thus we used a simpler search heuristic to determine our suggested allocation of the scholarships.

[11] This director had been making such decisions for more than 10 years and in a normal year supervised the application process for more than 4,000 students.

[12] In fact, our inability to solve this problem to our satisfaction was the prime motivation for the development of the more complex procedure described in §2.4.

[13] We estimated a simple binary logistic regression, which included student characteristics and scholarship amount along with numerous interactions. In addition, we allowed for a flexible polynomial function of scholarships.

variables that were significant were the interview score (students who interviewed better were less likely to matriculate) and the scholarship amount (higher scholarship amounts increased the matriculation rate at a decreasing rate). More interestingly, we found a significant and positive coefficient for any student who we identified to have been given an increase in scholarship funding. In effect, this coefficient shifted up the scholarship response function, indicating that the students we identified were more price sensitive, i.e., more likely to respond to the scholarship offer, than the other remaining 194 students in our sample, even after controlling for all other student characteristics.

3.2. Methodological Study

Next, we formed two comparable groups to examine average profiles. To do this, we limited each group to students who received scholarships of \$9,800 or less because most of the scholarships that we awarded were in this range. As can be seen from the results displayed in Table 1, the group of students awarded scholarships based on our approach had a higher yield and better quality, and yet they "cost" the school less per student compared with those who received scholarships based on the admissions director's allocation.

These field results are encouraging and suggest that the improved enrollment predictions along with the initial stochastic search approach have the potential to make significant gains over current practice. However, it is still an open question as to whether our full methodology can garner additional improvements. We address this in the next subsection.

In this section, we analyze whether our full methodology can produce improvements over the initial stochastic search results. Both methods take advantage of the improved enrollment predictions and use the same set of 111 students from the experimental group. In identifying solutions using our full methodology, we need to be concerned about overfitting. As a result, we present six different numbers of scenarios ranging from 5 to 1,000. As the number of scenarios increases, so does the computation cost and the confidence that we are not overfitting. We then compare these solutions against the initial stochastic search results using a holdout sample of 4,000 scenarios. We note that this analysis assumes that both the models for the

Table 1 Field Test Results

College admission results: Policy utilities					
	Average		Average characteristics		
Decisions	Yield (%)	Scholarship (\$)	Female (%)	GMAT	Interview score
Adjusted	70	6,400	43	726	1.9
Unaltered	54	9,800	42	698	2.1

objective function and the enrollment predictions are correct.

In Table 2, we show the results of our analyses for the six different-sized samples from the full set of scenarios. For each sample, we use the method outlined in §2.4.2 to identify a policy and the Lagrangian relaxation method of Technical Appendix §A.1 to estimate the bound on what an optimal policy could achieve. We calculate the in-sample and out-of-sample values of the implemented policy (the policy that was obtained by our initial stochastic search method) and the best policy found by applying the full methodology to the approximation problem. We also display the optimality bound obtained by the Lagrangian for each sample. Although this optimality bound is only in-sample, standard arguments show that with high probability, it is also an upper bound on the true expectation (exactly because of the optimization overfitting). To make these values easy to interpret, we subtract the utility obtained from a "No scholarship" policy. Thus, zero utility would mean no improvement over this benchmark policy.

We draw attention to a few features highlighted by the results in Table 2. Note that the implemented policy does not depend on the number of scenarios, because it is fixed. Consequently, as expected, the obtained in-sample value of the implemented policy neither systematically increases nor decreases with the number of scenarios (i.e., it simply fluctuates randomly), and the holdout sample value is constant. In contrast, the in-sample value of the best policy obtained decreases with the number of scenarios. This is because the obtained best policy tends to overfit more with fewer scenarios. This is confirmed by the finding that in-sample performance of the computed solution for a small number of scenarios is much higher than the out-of-sample performance. Importantly, the performance of the proposed solutions in the holdout sample seems to converge with 500 scenarios, suggesting that there is no longer substantial overfitting.

Table 3 contains 95% confidence intervals based on the holdout sample for the implemented policy and the best policies. It shows that regardless of the number of scenarios, the policies computed based on our full methodology are statistically significantly better than the implemented policy according to the objective function. Also, the upper end of the confidence interval is close to the optimality bound, suggesting that there is limited room for further improvement, because we expect a nonzero duality gap in such a nonconvex combinatorial problem.

In summary, the larger the number of scenarios, the less likely we are to overfit. However, as discussed previously, more scenarios are also more computationally costly. Using 1,000 scenarios, we were able

Table 2 Utility Above the "No Scholarship" Policy Benchmark (Computed Based on the Holdout Sample) for the Scholarship and Admission Decisions of MBA Applicants Using Real Data, 111 Applicants, 21 Scholarship Levels, and 8 Features

College admission results: Policy utilities					
	Utility above "no scholarship" policy				
	Implemented policy		Best policy found		Optimality bound
No. of scenarios ($\|S\|$)	In-sample	Holdout	In-sample	Holdout	In-sample
5	1.013	0.719	5.277	1.234	5.667
10	1.812	0.719	4.967	1.563	5.344
50	0.991	0.719	3.198	1.990	3.563
100	1.010	0.719	2.883	2.045	3.252
500	0.812	0.719	2.409	2.181	2.889
1,000	0.698	0.719	2.188	2.181	2.674

to conclude (a) the improvement of the best solution found over the implemented policy was statistically significant, and (b) the best policy found is (statistically) close to the optimality bound of Table 2. This latter finding indicates that even if we could provably compute the optimal solution by complete enumeration, we would not be able to substantially improve on the best solution found.

3.3. Managerial Insights

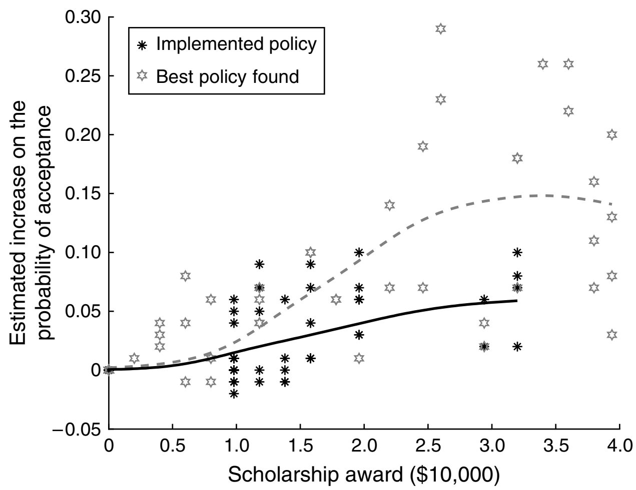
We next compare the solution from our optimization method to the original decisions of the admissions director to better understand what the method is doing to increase the utility. First, we note a difference in how close our expected use of scholarships was to the budget constraint. The admissions director used, as a heuristic, the historical yield rate on all admissions and applied it to those receiving scholarship offers. However, as it turns out, many of the people targeted for scholarships are far below the average yield rate. As a result, the admissions director overestimated the use of the scholarship budget and thus did not allocate as many scholarship offers as our methodology proposed. Second, the correlation between the original decisions and our decisions is only 0.46, indicating that many of the applicants who received scholarships under the original decisions did not under our decisions, and vice versa. For example, our method decreases scholarships for students who are likely to enroll without scholarship and increases scholarships for students who may get rejected by higher-ranking schools and for whom the scholarship could lead to choosing the focal school over similarly ranked schools. These possibilities are balanced against the best other possible uses of scholarship funds. Such complex trade-offs are much more difficult for a manager to do when assigning so many scholarships. Hence, it is not merely an increase in scholarships that leads to our improvements; it is also a large shift in who is offered scholarships.

Table 3 Confidence Intervals for the Improvement Over the "No Scholarship" Policy Benchmark Computed Based on the Holdout Sample

College admission results: Confidence intervals		
	Utility above "no scholarship" policy	
No. of scenarios ($\|S\|$)	Implemented policy	Best policy found
5	[0.621, 0.816]	[1.138, 1.330]
10	[0.621, 0.816]	[1.468, 1.659]
50	[0.621, 0.816]	[1.894, 2.086]
100	[0.621, 0.816]	[1.950, 2.141]
500	[0.621, 0.816]	[2.084, 2.277]
1,000	[0.621, 0.816]	[2.084, 2.277]

In essence, our method appears to produce larger increases in enrollment probabilities than does the admissions director for the same or better (expected) quality of students. In Figure 3, we plot scholarship offers under the two different policies against the increases in estimated matriculation probabilities. Two observations are directly evident. First, our approach produced a larger variation in the amounts given; i.e., we were more likely to give both in small and large scholarship amounts. This suggests that our ability to more finely distinguish between the effect of scholarship on different applicants' enrollment decisions translates into more evenly spread decisions. Second, our approach provided a larger average increase in forecasted yield for most levels of scholarship for similar quality students. This can be observed by the two lines in the plot. These lines smooth the points via a locally weighted regression technique, LOWESS (Cleveland 1979). We use a bandwidth of 0.5 to depict the smoothed average probability lift for the two policies. Thus, compared with the admissions director's policy, our method increases the school's expected utility function both by increasing the number and breadth of scholarship offers and by more effectively targeting those offers to increase enrollment probabilities.

Figure 3 The Estimated Increase in the Probability of Enrollment Against the Scholarship Award for Both the Original Policy of the Admissions Director and the Best Policy Found by the Proposed Approach



Note. The lines represent LOWESS estimators for the plotted relationships.

4. Discussion

4.1. Contributions

Our contribution includes both new methodology and a practical implementation. First, we provide an example of how to specialize the complex utility function in a CUP to make the problem feasible to analysis. Our particular problem of interest is the scholarship and admission decision. This involves the institution's utility function, which is composed of averages; this requires us to develop feasible, tailored optimization methods. Second, we develop a two-stage model for predicting prospective student decisions in the college choice setting in order to improve predictions. Third, we integrate these optimization and prediction methods into a holistic approach that can solve each of the empirical challenges in the scholarship and admission problem—data collection, estimation, prediction, and optimization. We then apply the components of this approach in a field test. Using the field test and methodological study, we provide evidence that the empirical system can improve on existing decisions processes. Whereas we view this work as a first step, we believe that this approach can contribute to the practice of marketing in the higher-eduction industry and serve as a model for working with CUP settings.

4.2. Extensions, Limitations, and Future Research

In many institutions, admissions are performed on a rolling or round basis (several sets of admission decision points) rather than in a single round. In this case, the admissions director has two additional challenges. First, the potential applicants in future rounds are uncertain, and a model is needed to forecast these applicants. Second, the admissions director is trading off the qualities of the current set of applicants with the qualities of the potential future set of applicants. This involves a modest extension to the existing decision model. The reason this extension is modest is because the methodology does not fundamentally change. Instead, a larger set of scenarios is required to accommodate the additional uncertainty about the future applicant qualities. In Technical Appendix E, we discuss the formal details of this extension.

In a sense, the limitations of our study suggest fruitful areas of research for a large targeted marketing industry. First, although our field study allows us to establish the feasibility of our method, its limited size does not provably demonstrate improvement over existing practice. Future research could directly apply the optimization procedure and conduct a larger-scale field test. Second, our approach involves measuring and predicting enrollments and estimating the institution's utility function. In our method, we incorporate parameter and choice uncertainty into the enrollment predictions, but we do not account for potential model misspecification, measurement error, or parameter uncertainty in our estimates of the institution's utility function. Future research could delve deeper into the influence of these types of errors on the ability to improve decisions. Third, we implemented our full empirical system, but our field evaluation did not allow us to fully determine the relative contribution of each of the pieces in our setting. Without a more formal experiment, it is hard to draw definitive conclusions about which pieces are most critical, particularly because the relative contribution is likely to differ across contexts. In our field study, we were also unable to evaluate particular aspects that could lead to improvement, such as endogenizing the set of possible financial aid offers.

We hope that admissions directors consider adopting, in part or in whole, the approach we propose here. Along these lines, we provide several suggestions. First, the survey data need to be regularly updated, and caution must be used if the school experiences marked changes in the market conditions or school ranking, as this will likely change the set of applications and the decision rules of competing institutions. Second, building the data collection into an automated part of the application and admission process can help make the process sustainable. Third, in the survey, the school must select the set of competitors most likely to affect prospective students' decisions. Ensuring that historically close competitors, reach and safety schools for the focal school's applicants, and potentially up-and-coming competitors are included in the survey is important to maintain consistency over the period necessary to estimate and validate the prediction model.

Finally, given the financial implications of this highly relevant target marketing and pricing problem, we hope that others will build on this research by either applying it to other major CUP problems or developing novel ways of extending our approach to other problems.

Electronic Companion

An electronic companion to this paper is available as part of the online version at http://dx.doi.org/10.1287/mksc.1120.0707.

Acknowledgments

The authors are in debt to Pierre Bonami, Katya Scheinberg, Jon Lee, Andreas Wächer, Oktay Günluk, and Peng Sun for several discussions. The authors also thank Liz Riley Hargrove, Abhijit Guha, and the admissions staff for assistance in implementing the field test. The first author gratefully acknowledges the support of the IBM Herman Goldstien Fellowship.

References

Allenby GM, Arora N, Ginter JL (1995) Incorporating prior knowledge into the analysis of conjoint studies. *J. Marketing Res.* 32(2):152–162.

Allenby GM, Leone RP, Jen L (1999) A dynamic model of purchase timing with application to direct marketing. *J. Amer. Statist. Assoc.* 94(446):365–366.

Avery C, Hoxby CM (2004) Do and should financial aid packages affect students' college choices? Hoxby C, ed. *College Choices: The Economics of Which College, When College, and How to Pay for It* (University of Chicago Press, Chicago).

Birge JR, Louveaux F (1997) *Introduction to Stochastic Programming* (Springer-Verlag, New York).

Bonami P, Biegler LT, Conn AR, Cornuéjols G, Grossmann IE, Laird CD, Lee J, et al. (2008) An algorithmic framework for convex mixed integer nonlinear programs. *Discrete Optim.* 5(2):186–204.

Carver MR Jr, King TE (1994) An empirical investigation of the MBA admission criteria for nontraditional programs. *J. Ed. Bus.* 70(2):94–98.

Chapman RG (1979) Pricing policy and the college choice process. *Res. Higher Ed.* 10(1):37–57.

Chapman RG, Staelin R (1982) Exploiting rank ordered choice set data within the stochastic utility model. *J. Marketing Res.* 19(3):288–301.

Chen KD, Hausman WH (2000) Mathematical properties of the optimal product line selection problem using choice-based conjoint analysis. *Management Sci.* 46(2):327–332.

Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* 74(368): 829–836.

Cominole M, Siegel P, Dudley K, Roe D, Gilligan T, Griffith J (2006) 2004 national postsecondary student aid study (NPSAS:04): Full-scale methodology report. Report, U.S. Department of Education, Washington, DC.

Deckro RF, Woundenberg HW (1977) M.B.A. admission criteria and academic success. *Decision Sci* 8(4):765–769.

DesJardins SJ (2001) Assessing the effects of changing institutional aid policy. *Res. Higher Ed.* 42(6):653–678.

Ehrenberg RG, Sherman DR (1984) Optimal financial aid policies for a selective university. *J. Human Resources* 19(2): 202–230.

Epple D, Romano R, Sieg H (2003) Peer effects, financial aid and selection of students into colleges and universities: An empirical analysis. *J. Appl. Econometrics* 18(5):501–525.

Fisher ML (1981) The Lagrangian relaxation method for solving integer programming problems. *Management Sci.* 27(1):1–18.

Fisher ML (2004) The Lagrangian relaxation method for solving integer programming problems. *Management Sci.* 50(12): 1861–1871.

Gallant AR, Tauchen G (2010) EMM: A program for efficient method of moments estimation version 2.6. Accessed March 7, 2012, http://econ.duke.edu/~get/emm.html.

Gustafsson A, Herrmann A, Huber F, eds. (2007) *Conjoint Measurement: Methods and Applications*, 4th ed. (Springer, Berlin).

Held M, Karp RM (1971) The traveling-salesman problem and minimum spanning trees: Part II. *Math. Programming* 1(1):6–25.

Hiriart-Urruty J-B, Lemaréchal C (1993) *Convex Analysis and Minimization Algorithms* (Grund. der math. Wissenschaften), Vols. 305 and 306 (Springer-Verlag, Berlin).

Iida H (1999) A note on the max-min 0-1 knapsack problem. *J. Combin. Optim.* 3(1):89–94.

Jünger M, Reinelt G, Rinaldi G (1994) The traveling salesman problem. Preprint 94-12, Interdisziplinäres Zentrum für Wissneschaftliches Rechnen der Universität Heidelberg, Heidelberg, Germany.

Khan R, Lewis M, Singh V (2009) Dynamic customer management and the value of one-to-one marketing. *Marketing Sci.* 28(6):1063–1079.

Manchanda P, Rossi PE, Chintagunta PK (2004) Response modeling with non-random marketing mix variables. *J. Marketing Res.* 41(November):467–478.

Manski C, Wise D (1983) *College Choice in America* (Harvard University Press, Cambridge, MA).

Marsh LC, Zellner A (2004) Bayesian solutions to graduate admissions and related selection problems. *J. Econometrics* 121(1–2):405–426.

Niu SX, Tienda M (2008) Choosing colleges: Identifying and modeling choice sets. *Soc. Sci. Res.* 37(2):416–433.

Nurnberg P, Schapiro M, Zimmerman D (2012) Students choosing colleges: Understanding the matriculation decision at a highly selective private institution. *Econom. Ed. Rev.* 31(1):1–8.

Punj GN, Staelin R (1978) The choice process for graduate business schools. *J. Marketing Res.* 15(4):588–598.

Reinartz WJ, Kumar V (2003) The impact of customer relationship characteristics on profitable lifetime duration. *J. Marketing* 67(1):77–99.

Rossi PE, McCulloch RE, Allenby GM (1996) The value of purchase history data in target marketing. *Marketing Sci.* 15(4):321–340.

Schön C (2010) On the optimal product line selection problem with price discrimination. *Management Sci.* 56(5):896–902.

Smith JE, Winkler RL (2006) The optimizers curse: Skepticism and postdecision surprise in decision analysis. *Management Sci.* 52(3):311–322.

State Higher Education Executive Officers (2009) State higher education finance fiscal year 2008 report. Report, SHEEO, Boulder, CO.

U.S. Department of Education (2009) Digest of education statistics—2009 digest tables directory. Retrieved January 23, 2011, http://nces.ed.gov/programs/digest/2009menu_tables.asp.

Venkatesan R, Kumar V (2004) A customer lifetime value framework for customer selection and resource allocation strategy. *J. Marketing* 68(4):106–125.

Venkatesan R, Kumar V, Bohling T (2007) Optimal customer relationship management using Bayesian decision theory: An application for customer selection. *J. Marketing Res.* 44(4):579–594.

Wainer H (2005) Shopping for colleges when what we know ain't. *J. Consumer Res.* 32(2):337–342.

Wu T-H (1997) A note on a global approach for general 0–1 fractional programming. *Eur. J. Oper. Res.* 101(1):220–223.

Yu G (1996) On the max-min 0-1 knapsack problem with robust optimization applications. *Oper. Res.* 44(2):407–415.

Zeithammer R, Lenk P (2006) Bayesian estimation of multivariate-normal models when dimensions are absent. *Quant. Marketing Econom.* 4(3):241–265.