

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

Alexandre Belloni

Fuqua School of Business, Duke University, 100 Fuqua Drive, Durham, NC 27708, abn5@duke.edu,
<http://www.duke.edu/~abn5>

Mitchell J. Lovett

Simon graduate School of Business, University of Rochester, 305 Schlegel Hall, Rochester N.Y. 14627,
mitch.lovett@simon.rochester.edu
<http://www.simon.rochester.edu/faculty-research/faculty-directory/full-time-faculty-directory/mitchell-j-lovett/index.aspx>

William Boulding

Fuqua School of Business, Duke University, 100 Fuqua Drive, Durham, NC 27708, bb1@duke.edu,
<http://www.fuqua.duke.edu/faculty/alpha/bb1.htm>

Richard Staelin

Fuqua School of Business, Duke University, 100 Fuqua Drive, Durham, NC 27708, rstaelin@duke.edu,
<http://www.fuqua.duke.edu/faculty/alpha/rstaelin.htm>

Each year in the post-secondary education industry schools offers admission to nearly 3 million new students and scholarships totalling nearly \$100 billion. This problem represents a huge, previously understudied targeted marketing and price discrimination problem. This paper provides a solution tailored to this problem. This solution 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 GPA. This goal to achieve a class with a desirable set of average characteristics greatly complicates the optimization problem and doesn't allow the application of standard approaches. We develop a new approach that solves this more complex optimization problem and also accounts for the Optimizer's Curse. In order to avoid the computationally untractable task of provably solving for the optimal policy, we calculate sharp upper bounds on the value of the optimal solution, which informs a heuristic search algorithm for feasible solutions. We then use conservative confidence intervals for the value of any policy to examine whether it is statistically distinguishable from the value of the optimal solution or other solutions. We test the approach in a field study of an MBA scholarship process and implement adjusted scholarship decisions. We demonstrate the methodology can generate statistically significant improvements over current management decisions and policies that are statistically indistinguishable from the true optimal solution. By comparing our solution to what management would do on its own, we provide insight into the nature of mistakes management made in this complex decision environment.

Key words: choice sets, college choice, utility on averages, statistical approximation, non-convex optimization

1. Introduction

This paper presents a tailored solution for two major marketing problems – identifying a subset of potential customers to target and selecting individual level discounts in order to price discriminate among these targeted customers. This tailored solution is designed to be applied in the post-secondary 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 a solution 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 average characteristics greatly complicates the solution method because it renders the optimization problem to be non-separable across the different students. As a result we develop a new optimization method that handles these averages by exploiting the specific structure of the problem facing schools.

Finding a feasible solution to this problem has economic importance both because the post-secondary 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 post-secondary education industry in the 2006–2007 academic year had total revenues of over \$465 billion (U.S. Dept. of Education 2009)¹ with revenue from student tuition and fees of over \$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 were 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, since 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, neither the economics nor marketing literatures specifically address how to optimally make admission and scholarship decisions. A few papers in the marketing literature specifically model the college choice problem (Punj and Staelin 1978; Chapman 1979; Chapman and Staelin 1982; Wainer 2005), but 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 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 Manki 1983; Ehrenberg and Sherman 1984; DesJardins 2001; Epple, Romana, and Sieg 2003; Avery and Hoxby 2003; Marsh and Zellner 2004; Ehrenberg 2004; Niu and Tienda 2008; Nurnberg, Schapiro, and Zimmerman 2010) address issues that are only tangentially connected with the admission and scholarship decisions, focusing primarily on the estimation of student enrollment and its antecedents or correlates. 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 don’t 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 proposes a model that allocates a single scholarship level to all members of a pre-determined group of students. Their approach ignores uncertainty and does not provide guidance as to how to actually solve the optimization problem.

In many respects the admission and financial aid decisions are similar to problems addressed in the target marketing literature (e.g., Rossi, McCulloch and Allenby 1996; Venkatesan, Kumar, and Bohling 2007; Kahn, Lewis, and Singh 2007). While this literature is applicable to many different industries and discusses many related issues including, for example, empirical models of customer actions and reactions (e.g., Rossi, McCulloch, and Allenby 1996; Allenby, Leone, and Jen 1999; Reinartz and Kumar 2003), only a subset of this literature is concerned with making individual level offer decisions (e.g., Venkatesan and Kumar 2004). Our approach to admissions and financial aid decisions has many similarities with this subset. For example, like Venkatesan, Kumar, and Bohling (2007), we use an individual customer response model that conditions on covariates and

¹All statistics cited in this paragraph come from the U.S. Dept. of Education, National Center for Education Statistics, Integrated Post-secondary Educational Data System, downloaded on 1/23/2011 from http://nces.ed.gov/programs/digest/2009menu_tables.asp.

is calibrated with Bayesian estimation techniques. Also in order to solve our optimization problem and customize offers at the individual level, like them, we develop an approximate solution method.

However, the objective function in our problem differs 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, or customer life-time 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. Normally, these objectives imply multiple criteria and are affected by the average characteristics of the obtained students. For instance, in order 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.

It is this desire to optimize over multiple averages that distinguishes and complicates the admissions and scholarship problem. In particular, it makes the benefit of giving an offer to any one applicant dependent on not only all other offers made, but also the benefit associated with all the other offers. That is, unlike most targeted marketing applications (e.g., Rossi, McCulloch and Allenby 1996; Venkatesan, Kumar, and Bohling 2007; Kahn, Lewis, and Singh 2009), we cannot use standard methods to separate the targeting decision for each individual from those of the rest of the individuals. This non-separability, along with the large scale of the problem, rules out both simple optimization approaches, such as greedy algorithms and approaches used to solve the knapsack problem, and generic optimization methods developed in the linear 0-1 fractional programming literature (Wu 1997). Consequently, we develop a tailored optimization method that differs from existing approaches and exploits the structure of the problem facing schools. We then complement this optimization approach with an empirical system that provides the algorithm with the necessary inputs, these being 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.

We illustrate our approach with a specific application to graduate business education. As is true in most college choice situations, our greatest empirical challenge is to predict the student’s enrollment decision, since 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 one such approach to making these predictions. This approach, while based on standard Bayesian estimation methods, differs from previous targeted marketing efforts (e.g., Rossi, McCulloch, and Allenby 1996) due to the institutional fact that administrators at the time of the admission and scholarship decisions do not know whether or not students have been admitted to the competing schools nor whether and how much scholarship they will receive. As a result, we construct a two-stage estimation process and prediction model that is tailored to this setting.

We test our empirical solution via a field study 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 adjust the director’s decisions based on our predictions about the prospective students’ enrollment choices (but not our full optimization method). We show an increase in both the yield and the quality of the entering class. These results provide evidence that the predictions are accurate enough to apply our optimization method. We then use this field test data to demonstrate via policy simulations how our optimization method would be able to further improve on the (already improved) policy that benefited from the more accurate enrollment predictions. We conclude the application section by giving some initial insight into the ways in which managerial heuristics may fail in these complex decision problems.

The flow of the paper is as follows. In Section 2 we first provide an overview of our tailored optimization approach and how our approach extends prior work. We then specify the model,

our proposed general solution, its statistical properties, and its relationship to the optimization literature. In Section 3 we discuss a field test and policy simulations for the admission director’s problem of making scholarship offers to a set of admitted students. Section 4 discusses extensions to the method.

2. Problem Definition and Optimization Methodology

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). While 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 individual such as test scores, grade point average (GPA), gender, race, activities (sports, music, etc.), and other schools they 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 uncertain. As a result, the solution needs to account for this uncertainty and the admissions director needs to predict the students’ enrollment choices.

A solution to this problem requires two critical inputs. First is the institution’s utility function—a function that translates a class profile (e.g., number of enrolling students, average GPA, average SAT, %female, %minority, etc.) into a value. While typically admissions directors are 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 and admissions director) using conjoint analysis. Second is the enrollment predictions which translate the information from the application into a probability of enrolling given any particular scholarship level. Again, many approaches can be used to attain these predictions—from very simple subjective estimates of experienced admissions staff to more complex statistical approaches. In our application, we used statistical modeling to obtain these predictions. 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, which lead 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 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 post-secondary education industry where (a) the institution’s utility is a function not only of the total number (or total revenue, profit, or CLV) of acquired customers, but also of the average of several observable characteristics associated with these customers, (b) the institution faces (linear) 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/rejection of the offer, our approach uses a stochastic programming structure. 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.

2.1. Model development

Let I denote the set of potential students who have applied for admission to the school and J denote the set of different possible scholarship offers (discounts) that the school can assign 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 k th 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 one if the offer level j is assigned to individual i , and zero otherwise. The random variable for whether individual i accepts the school's offer j is denoted by a_{ij} . The random variable equals one if the individual accepts the offer and zero 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 $u_k : \mathbb{R} \rightarrow \mathbb{R}$ is evaluated on the average value of the k^{th} feature for the pool of individuals who matriculate. This school admission and scholarship problem can be cast as the following mathematical programming

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

where the matrix A and vector b define (generic) linear constraints, and, without loss of generality, we use the convention $0/0 = 0$ (see Appendix B 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 while u_k represents the utility for the average value for the acquired objects on the k^{th} 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 $Ax \leq b$ 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. For notational convenience, we denote the feasible region of (1) by \mathcal{R} .

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

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

The above formulation is related to the literature on product line design. Notably, the model studied in Chen and Hausman (2000) had 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 had

additional fixed cost decisions that linked the different terms. She also showed how to build upon Chen and Hausman (2000) to derive an efficient implementation. In contrast to Chen and Hausman (2000), 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 not able to build upon the advances in these papers. Further, 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) that also accounts for its statistical properties.

2.2. Solution Strategy

Our approach to the problem in (2) provides a solution to three related methodological challenges—high computational costs of optimization, potential for overfitting the approximation errors, and identification of feasible solutions. These three issues are closely interconnected and we explain each as we describe our solution strategy. Finding an optimal solution to (2) entails maximizing the sum of the true objective function in (1) and the 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). Hence, the optimal solution to (2) is unlikely to be the optimal solution to (1). This overfitting is referred to by Smith and Winkler (2006) as the Optimizer’s Curse since the expected value of optimal value of (2) is always larger than the optimal value of (1). This Optimizer’s Curse fundamentally shapes how we solve the problem.

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. For large scale instances of (2), however, these schemes are still computationally demanding. More importantly, the Optimizer’s Curse implies this solution is not as valuable as it seems. We cannot statistically distinguish the true expected utility associated with any feasible solution that has a value within the approximation error of the optimal value of (2) from that of the provably computed solution of (2). Consequently, our approach avoids using enumerating schemes.

Our solution strategy involves three connected elements—efficient computation of the upper bound on the optimal solution value, an efficient heuristic search strategy, and conservative confidence intervals on the value of any solution. This combination of approaches efficiently identifies solutions while guaranteeing the quality of the solutions obtained by the heuristic search without relying on computationally demanding enumeration schemes. Further, using the conservative confidence intervals we can demonstrate that the obtained solution is statistically indistinguishable from the true optimal value. In this way, we avoid the steep computational costs² of provably solving for the optimal policy and still provide the best solution.

2.3. 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. While the details are presented in Appendix A, 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

² We emphasize that sharp upper bounds are valuable even when we found the optimal solution because typically we cannot recognize its optimality. To prove its optimality is computation intractable in general and very computationally demanding even for small instances via enumeration methods.

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 functions³ 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

$$\begin{aligned} \phi(\lambda, \sigma, \alpha) = \max_{x, x^s} \sum_{s \in S} p_s & \left(u_0 \left(\sum_{i \in I} \sum_{j \in J} a_{ij}^s x_{ij}^s \right) + \frac{\sum_{i \in I} \sum_{j \in J} \left(\sum_{k \in K} \sum_{\ell=1}^{\ell_k} \alpha_{k,\ell}^s r_{k,\ell} w_{ik} \right) a_{ij}^s x_{ij}^s}{\sum_{i \in I} \sum_{j \in J} a_{ij}^s x_{ij}^s} + \sum_{k \in K} \sum_{\ell=1}^{\ell_k} \alpha_{k,\ell}^s d_{k,\ell} \right) - \\ & - \sum_{s \in S} \lambda'_s (x^s - x) - \sigma' (Ax - b) \\ \text{s.t.} & \\ & \sum_{j \in J} x_{ij}^s = 1 \quad \text{for } i \in I, s \in S \\ & x_{ij}^s \in \{0, 1\} \quad \text{for } s \in S \end{aligned} \tag{3}$$

In 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 the literature on max-min 0-1 (linear) knapsack problems of 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 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. Generating Feasible Solutions

There is a variety of heuristics that could be applied to the problem (2). We focus on the following simple heuristic:

1. Generate a feasible solution based on solution of the dual problem (3).
2. Try to make local improvements on neighborhoods constructed with dual information.

After an initial feasible solution is obtained, we proceed to make a sequence of “two-opt” moves (switch one variable from 1 to 0, and another from 0 to 1) and “three-opt” moves (similarly but using three variables), always maintaining feasibility. Such a local search is standard on combinatorial problems and were first proposed for the Traveling Salesman Problem (see Jünger, Reinelt

³ We write u_k as the minimum of ℓ_k linear pieces and each linear piece is given by a slope $r_{k,\ell}$ and an intercept $d_{k,\ell}$, $\ell = 1, \dots, \ell_k$, $k \in K$.

and Rinaldi 1994). However, in order to improve efficiency we select a small subset of decisions to try the local searches on (despite being polynomial time they can be computationally demanding). In selecting these subsets the use of dual variables led to better solutions as the quality of the dual bounds improved.

2.5. Conservative Confidence Intervals

In order to obtain valid statistical guarantees, we need to account for the overall “flexibility” of the model. Similar to Kleywegt et al. (2001), we invoke finite sample uniform laws of large numbers (Ledoux and Talagrand 1991, van der Vaart and Wellner 1996, and Dudley 2000) to derive statistical bounds on the expected value of policies that are chosen through a scenario-based procedure. That is, we provide confidence intervals that account for possible overfitting of the approximation errors. The length of these confidence intervals make explicit the relationship between the number of scenarios and the number of decision variables. The technical details of the derivation are provided in Appendices C and E and we report the main result used in the application.

COROLLARY 1. *Let $a^s, s \in S$ be independent scenarios, $x \in \mathcal{R}$, and let $f(x, a^s)$ the value of the criterion function (2) in scenario a^s under decisions x . Let $\sigma^2 = \max_{x \in \mathcal{R}} \text{var}_a(f(x, a))$ and fix $\delta \in (0, 1)$. For $0 < \gamma < \ln 2$ and*

$$\varepsilon := \frac{\sigma}{\sqrt{|S|}} \cdot \sqrt{\frac{2|I|}{2 - \exp(\gamma)} \log\left(\frac{2|J|}{\delta}\right)},$$

assume that $|f(x, a) - E[f(x, a)]| \leq \gamma \sqrt{|S|} \sigma^2$ for every a and $x \in \mathcal{R}$. Then, we have that for any feasible solution $\hat{x} \in \mathcal{R}$ selected based on the given scenarios

$$E[f(\hat{x}, a)] \in \left[\frac{1}{|S|} \sum_{s=1}^{|S|} f(\hat{x}, a^s) - \varepsilon, \frac{1}{|S|} \sum_{s=1}^{|S|} f(\hat{x}, a^s) + \varepsilon \right],$$

with probability at least $1 - \delta$.

In words, Corollary 1 provides a confidence interval to the true expected value of any feasible policy selected based on the given scenarios. The length of the interval accounts for the number of feasible decisions in \mathcal{R} which is bounded by $|J|^I$ as well as the number of scenarios $|S|$.

3. Scholarship and Admissions Decisions of MBA Applicants

In this section, we illustrate the value of our optimization method and demonstrate how this method can be implemented using actual school admissions and scholarship allocation data from our focal school. In our case the focus is on allocating scholarship dollars to a set of the focal school’s admitted M.B.A. applicants, since this aspect of the problem was the focal school’s primary concern. We start by describing how we determine two of the inputs needed for our optimization approach and provide evidence on the veracity of these inputs. We then conduct 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 predictions⁴. 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.

As shown in Figure 1 our optimization approach requires four inputs; (1) the budget constraint, (2) the applicants’ characteristics, (3) the institution’s utility function, and (3) the probability of

⁴At 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.

enrollment conditional on the scholarship level and yields one output per student, i.e. the scholarship amount, which could be zero. The first input comes from the Dean’s office and the second from the student’s application form. However, the latter two inputs are not typically available to admission directors. Consequently, we discuss them in more detail in the next two subsections.

3.1. Estimating the Institution’s Utility Function

We start with 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 M.B.A. 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 of GMAT score, the class average of the admission director’s assessment of the student’s potential (from EXCELLENT=1 to POOR=5), the percentage of the class that had an overall assessment with score four or below, the average interview score (from EXCELLENT=1 to POOR=5), the percentage of the class that had interview grades four or below, percentage of female, and percentage of foreigners.

We then asked key faculty and administrators, including the admissions director, to complete an ACA computer-based interview (e.g., Gustafsson, Herrmann, and Huber 2003) as implemented in Sawtooth Software (see www.SawtoothSoftware.com). We used their responses to estimate partworth-utilities (utilities for each level within each attribute) associated with these eight attributes for each individual completing the task. We determined that there was general agreement in terms of weights and shape of the individual partworths across our respondents, suggesting faculty, administration, and the admissions director had a common view of what was sought from the admissions process. Hence, we combined the individual estimates to form a piecewise linear function of the characteristics of the enrolling class. We then 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.

Note that while obtaining these partworth estimates, we made no attempt to ensure that each of the individual functions were concave over the total region⁵. Not surprisingly, a few of the estimated functions violated the concavity assumption as stated in (4). Consequently, we tailored a projection method to our setting as described in Section 4.3 and applied it to obtain consistent concave estimators. Figure 2 illustrates the impact of this methodology on the utility of total enrolment. Not only does this figure illustrate the worst violation of concavity in our estimates, it also points out the school’s strong disutility for under-enrollment as well as its disutility for bringing in a class size that exceeds its capacity constraint.

3.2. Predicting the Probability of Enrollment Conditional on Scholarship Offer

A second set of inputs needed for our optimization model is the scenarios (realizations) of 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. Further, interviews with the admissions director indicated such a task would involve substantial guesswork.

Coming up with an estimate a_{ij} requires knowing both the student’s choice set and choice rule. 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

⁵ There are alternative methods which enforce such constraints directly in the estimation, e.g., see Allenby, Arora, and Ginter (1995).

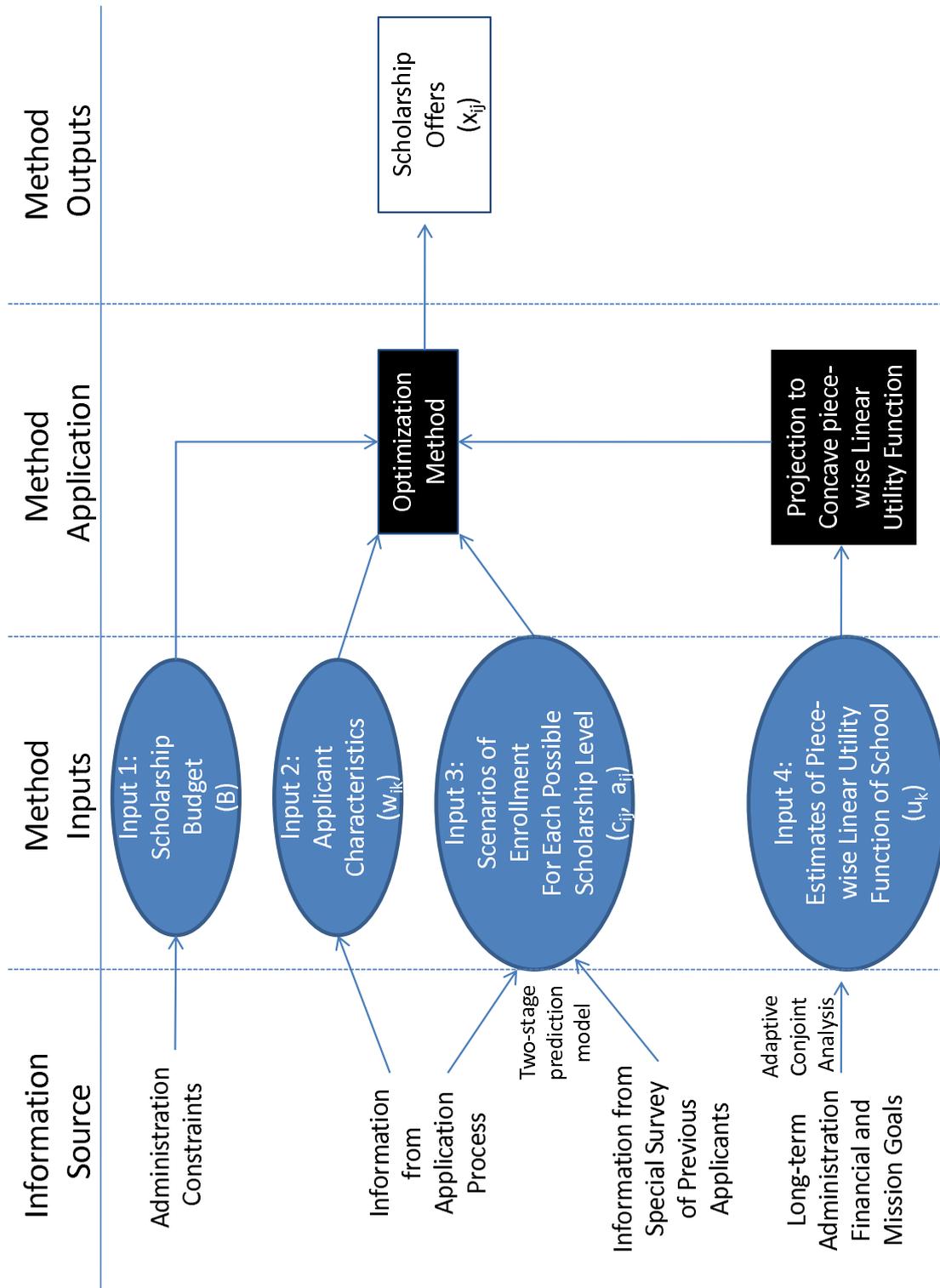


Figure 1 The figure shows the process flow associated with the admissions director's application.

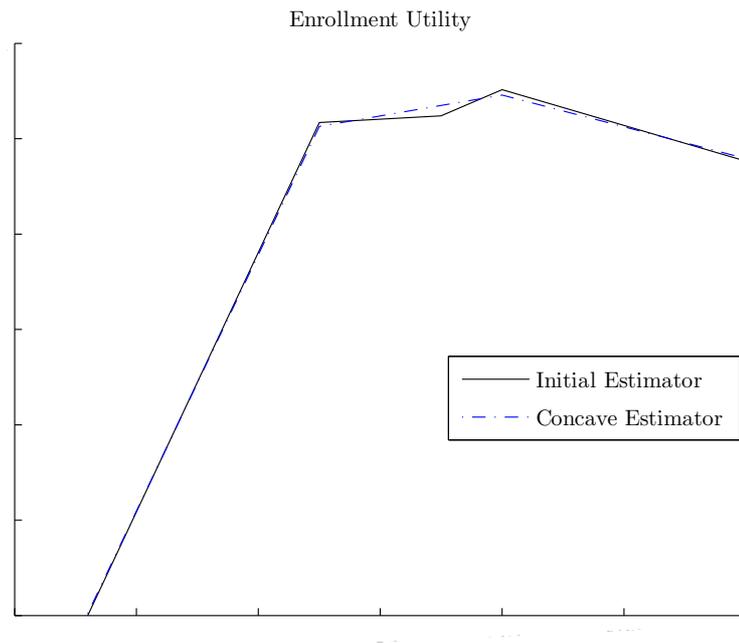


Figure 2 The figure shows the initial and concave estimators for the utility for the enrollment feature. Scales have been removed at the request of the institution.

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 applicants⁶. 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 latter 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 years⁷. While this sample is not a sample of all applicants to M.B.A. programs, it is not biased for our purposes (estimating competitor actions and applicant choices about applicants to the focal school), since it includes all the relevant groups associated with the focal school. More specifically we obtained responses from 1191 responses of which 1139 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 type of student admitted to and possibly given scholarships to 20 competing MBA programs and an "others" program that was used to represent a collection of schools that individually had only a small number of joint applications with the focal school. With this data, we estimate two related Bayesian statistical models.

⁶ Note that throughout we will loosely refer to individuals as students, though some do not ultimately matriculate to any school.

⁷ The overall response rate was 43% although the rates were slightly higher for admitted than denied populations. However, no clear systematic bias was present in terms of the observable characteristics of those responding. Perhaps more importantly, non-response becomes an issue in our case only if the model processes depends on whether the person responds or not to the survey. We are unable to come up with a reasoned answer to why such a difference might occur.

3.2.1. Estimating Competing Schools' 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 admissions and scholarship decisions, thereby allowing us to predict each new applicant's choice set and scholarship offers. Because our sample is made up of only applicants to these competitor schools that also applied to the focal school, the parameter estimates may not be accurate for the total sample of applicants to the competitor schools. However, we only care about the decision rules relevant for prospective students that also applied to the focal school. As long as the application set and decision rules for the prediction period do not differ markedly from the estimation sample, this bias should not affect the accuracy of our predictions.

More formally, 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} ,

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

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 a combination of linear and quadratic terms. 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 $\alpha_{0k} \leq sc \leq \alpha_{1k}$ (indicating admission with scholarship). The values α_{0k} and α_{1k} are observed while α_{-1k} and the σ_k are to be estimated. We translate the underlying index into a decision D_{ik} via the following rule:

$$D_{ik} = -1 \cdot 1(u_{ik} < \alpha_{-1k}) + u_{ik} \cdot 1(\alpha_{0k} \leq u_{ik} \leq \alpha_{1k}) + \alpha_{1k} 1(u_{ik} > \alpha_{1k}) \quad (4)$$

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

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

$$p(D_{ik} | \theta_k^1) = \Phi(\alpha_{-1k}; v_{ik}, \sigma_k)^{1(D_{ik}=-1)} \cdot (\Phi(\alpha_{0k}; v_{ik}, \sigma_k) - \Phi(\alpha_{-1k}; v_{ik}, \sigma_k))^{1(D_{ik}=0)} \\ \cdot \phi(D_{ik}; v_{ik}, \sigma_k)^{1(\alpha_{0k} < D_{ik} < \alpha_{1k})} \cdot (1 - \Phi(\alpha_{1k}; v_{ik}, \sigma_k))^{1(D_{ik}=\alpha_{1k})}$$

where $\Phi(x; \mu, v)$ and $\phi(x; \mu, v)$ are the c.d.f. and p.d.f. of the normal distribution with mean μ and variance v at x and $\theta_k^1 = (\alpha_{-1k}, \sigma_k, \beta_k)$ the parameters to be estimated. There are n individuals and J schools, but not every individual applies to every school. Instead, we denote the number of applications for individual i as J_i^a . Hence, the likelihood is a product of $\sum_{i=1}^n \sum_{k=1}^{J_i^a} p(D_{ik} | \theta_k^1)$. This model can be viewed as joint ordered probit (for offer admission and the minimum scholarship) and two-sided censored tobit model (for the scholarship offer), all with a single underlying index. We assume diffuse conjugate priors and employ standard MCMC methods with data augmentation to estimate the model (see Dunson et al 2003 and Chib 1992).

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

More formally, 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 $E_{ik} = \{0 \text{ if not chosen, } 1 \text{ if chosen}\}$. 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., $E_{ik} = 1(z_{ik} > \max(\{z_{ih}\}_{h \neq k}))$

and $\{\cdot\}_m$ is a set with the subscript m defining the members. 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 on enrollment choice, γ_{1i} , where the individual-level heterogeneity comes from the observed demographics of the student 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.⁸

For estimation, we use a multinomial probit with absent dimensions (Zeithammer and Lenk 2006). This approach employs data augmentation for the unobserved utility vector z_i , of length J . Note that while individuals only have $|A_i| \leq J$ options available, the approach simplifies the sampling procedure by drawing utilities for all J options. The utilities for the unavailable options are drawn without constraints while the utilities for the available options are constrained. Thus, the model can loosely be written as

$$\begin{aligned} E_i &= (E_{i1}, E_{i2}, \dots, E_{iJ}) \\ E_{ik} &= \begin{cases} 1(z_{ik} > \max(\{z_{ih}\}_{h \neq k})) & k \in A_i \\ 0 & k \notin A_i \end{cases} \\ z_i &\sim N_J(V_i, \Sigma) \end{aligned}$$

where $N_J(\cdot, \cdot)$ is the multivariate normal distribution of dimension J , V_i is the vector of means containing the V_{ik} 's, and Σ is the covariance matrix. The indicator function induces truncation for the available options. The chosen school's distribution is truncated below by the maximum utility of the competing options and the unchosen schools are truncated above by the utility of the chosen option. Thus, the parameters of interest are $\theta^2 = (\gamma_0, \Psi, \Sigma)$, where γ_0 is the vector of school intercepts. We assume standard, diffuse conjugate priors and follow the standard sampling procedure.

3.2.3. 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 knows only the applicants' characteristics and the set of schools the prospective student applied to. Using 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 use the characteristics gained from the applications X_i^p and the set of considered schools (applications) in combination with a sample from the posterior distribution of the first model to predict which schools will admit and give scholarship offers to the applicant. We predict these offers for all competing schools for which the student applied except the focal school producing a vector D_i^p of length $J_i^a - 1$. These prediction vectors have two sources of randomness – the uncertainty about the parameters $\theta^1 = \{\theta_k^1\}_{k=1}^J$ 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 $p_{J_i^a}(D_i^p | X_i)$ we draw θ^1 and a vector of the stochastic elements, ϵ_i . Jointly these determine the vector of underlying indices,

⁸Note that although the scholarship term is linear, the nature of the probit model leads to non-linear responses to scholarship. Further non-linearities (e.g., polynomial functions of sc_{ik}) can easily be accommodated, but in our application the data did not support it.

u_i , which directly maps to the vector D_i^p as expressed in equation (4). 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 s (admitted with scholarship s). The J_{i-f}^a subscript denotes that the random vector has variables for each school individual i applied to other than the focal school. Thus, for each individual we have a set of mc 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 mc 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 new applicant characteristics and use the estimates from the second model. We have uncertainty about both the parameters θ^2 and unobserved utility components. Thus, we wish to sample from the predictive distribution of utilities, $p_{|A_i^p|}(z_i)$, in order to identify whether for that sample, the focal school is chosen. To do so, we draw mz samples of θ^2 and the stochastic elements for each of the mc 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 $z_{if} > \max(\{z_{ik}\}_{k \neq f})$) in which case $a_{ij}^s = 1$ otherwise $a_{ij}^s = 0$. Thus, for each individual we have a sample of $mz * mc$ 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-scholarship level, resulting in 11,655,000 values of a_{ij}^s .

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 decisions.

3.2.4. Predictive validity. We demonstrate the predictive power of our models by calculating the percent of time we correctly predict the student's school choice using a hold-out sample of 434 individuals that come from a different year than the calibration sample. The average hit rate is 78% for one stage predictions (i.e., the choice set is known) and 73% for two stage predictions (i.e., when the choice set is also predicted). These accuracies compare favorably to previous models in which a one stage prediction garnered 74% accuracy (e.g., Punj and Staelin 1974). This is strong evidence 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, something the admissions director felt she was not able to do on her own. 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 be applied⁹. In the next section we use our field experiment to demonstrate that the accuracy and precision of these estimates enabled us to adjust decisions with a stochastic search method and improve actual outcomes.

⁹ 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 admissions and scholarship decisions) may be endogenous, introducing bias in the coefficients we estimate (e.g., Manchanda, Rossi, and Chintagunta 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.

3.3. 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 director¹⁰. 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 that benefited from the improved enrollment predictions but not the better heuristic and optimization approach explained above¹¹. Hence, any improvement in this experimental group should represent a lower bound for the benefit of our proposed methodology, since our 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 we would 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 amounts¹². The only control 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 that we identified to give an increase in scholarship. 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.

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 since most of the scholarships that we awarded were in this range. As can be seen from the results as displayed in Table 1, the group of students awarded scholarships based on our approach had a higher yield and better quality and yet “cost” the school less per student compared to 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 can make significant gains over current practice. However, whether our full methodology can garner additional improvements is still an open question that we address in the next subsection.

¹⁰ This director had been making such decisions for over 10 years and in a normal year supervised the application process for over 4000 students.

¹¹ 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 Section 2.5 and Appendices A.1 and A.2.

¹² 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.

College Admission Results: Policy Utilities

Decisions	Average		Average Characteristics		
	Yield	Scholarship	Female (%)	GMAT	Interview Score
Adjusted	70%	\$6,400	43%	726	1.9
Unaltered	54%	\$9,800	42%	698	2.1

Table 1 Field test results.

3.4. Methodological Study

In this section, we analyze whether, holding fixed the improved enrollment predictions, our full methodology can produce significant improvements over the initial stochastic search approach. We use the same set of 111 students from the experimental group and compare the expected utility under the two scholarship policies, i.e., derived from our full methodology or the initial stochastic search approach. To evaluate whether the difference is statistically significant, we use the theoretical bounds derived in Section 2.5.

The theoretical bounds suggest that the minimum number of scenarios needed to obtain uniform convergence is of the order of $|I| \ln |J|$ which is roughly a few hundred scenarios since $|I| = 111$ and $|J| = 21$. In Table 2 we show the results of our analyses for a number of different sized samples from the full set of scenarios. For each sample, we use the method outlined in Section 2.4 to identify a policy and the Lagrangian relaxation method of Appendix A.1 to estimate the bound on what an optimal policy could achieve. We calculate the values of the implemented policy (the policy that was obtained by our initial stochastic search method), the best policy found by applying the full methodology to the approximation problem, and the optimality bound. To make these values easy to interpret, we subtract the utility obtained from a “No Fellowship” policy. Thus, zero utility would mean no improvement over this benchmark policy.

College Admission Results: Policy Utilities

Number of Scenarios $ S $	Utility above “No Fellowship” Policy		
	Implemented Policy	Best Policy Found	Optimality Bound
5	0.548	4.400	5.203
10	0.687	3.592	4.275
50	0.671	2.562	3.298
100	0.732	2.372	3.029
500	0.741	2.158	2.858
1,000	0.709	2.043	2.727
5,000	0.675	1.955	2.691

Table 2 Utility above the “No Fellowship” policy benchmark for the scholarship and admissions decisions of MBAs Applicants using real data, 111 applicants, 21 fellowship levels, and 8 features.

Table 2 presents the obtained values for the implemented and best policies as well as the optimal bound as the number of scenarios increases. Note that the implemented policy does not depend on the number of scenarios, since it is fixed. Consequently, as expected, the obtained value of the implemented policy neither systematically increases nor decreases with the number of scenarios (i.e., it simply fluctuates randomly). In contrast, the value of the best policy obtained decreases with the number of scenarios. This is due to the fact that the obtained best policy may overfit when there are too few scenarios, i.e., it may suffer from the Optimizer’s Curse (Smith and Winkler 2006). Thus, we find in-sample performance of the computed solution for a small number of scenarios being

much higher than the performance achieved with a larger number of scenarios. More importantly the performance of the proposed solutions seems to be converging after about 1000 scenarios, suggesting that there is no longer substantial overfitting.

Another important aspect of these solutions is the confidence interval associated with each policy. Using Corollary 1 it is possible to determine these intervals. As can be seen in Table 3 the 95% confidence intervals for both policies decrease with the number of scenarios. Interestingly, these intervals no longer overlap when we used 5000 scenarios. Consequently, we can conclude that the best solution found based on these 5000 scenarios is a statistically significant improvement over the implemented policy. In addition we note that the obtained result (i.e., 1.955) associated with this 5000 scenario best policy is statistically indistinguishable from the unknown true optimal policy. This is due to the fact that we know a) the true optimum must be less than or equal to the upper bound (i.e., 2.691) and b) the lower end of the confidence interval for the true optimum is less than or equal to $1.507 (= 2.691 - (3.139 - .770)/2)$ and c) our best policy is above that.

In summary we note the tradeoff between the number of scenarios, the potential for overfitting, the ability to distinguish our best policy from the implemented policy and the ability to say that our best policy is indistinguishable from the upper bound. Clearly the larger the number of scenarios, the less likely we overfit. However, as discussed in Section 2.5, more scenarios are also more computationally costly. Using 5000 scenarios we were able to conclude a) the best solution found was a statistically significant improvement over the implemented policy and b) the best policy found is statistically indistinguishable from the optimality bounds of Table 3. This latter finding indicates that even if we could provably compute the optimal solution by complete enumeration, we would not be able to statistically distinguish this optimal solution from the best solution found.

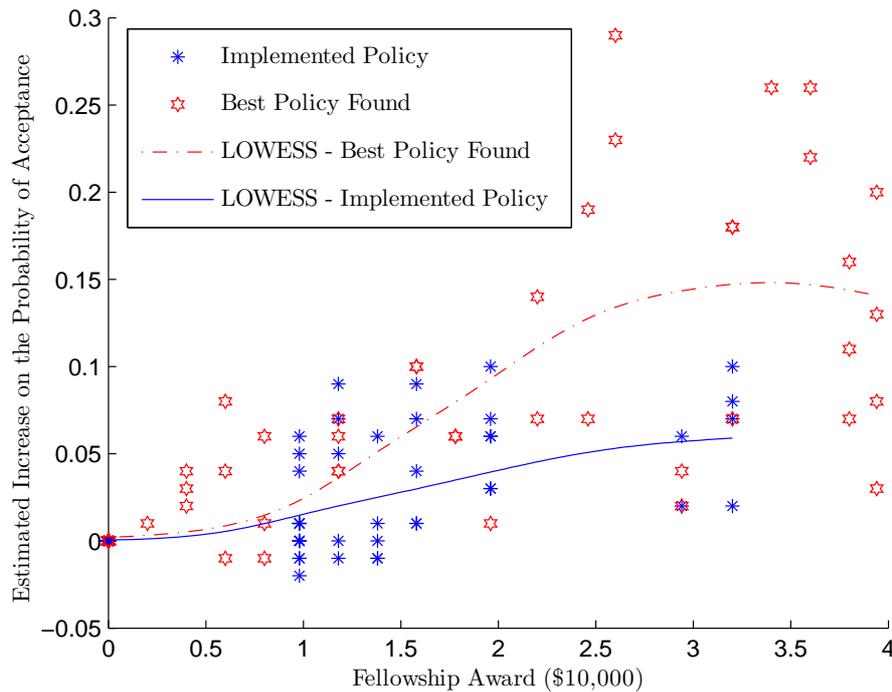
College Admission Results: Confidence Intervals

Number of scenarios $ S $	Utility above “No Fellowship” Policy	
	Implemented Policy	Best Policy Found
5	[-2.279, 3.377]	-
10	[-1.312, 2.687]	-
50	[-0.223, 1.566]	-
100	[0.100, 1.366]	[-6.983, 11.728]
500	[0.458, 1.025]	[-1.719, 6.036]
1,000	[0.509, 0.910]	[-0.656, 4.743]
5,000	[0.585, 0.765]	[0.770, 3.139]

Table 3 The table provides confidence intervals for the improvement over the “No Fellowship” policy benchmark. For the implemented policy a standard 95% confidence interval is provided while for the proposed policy a conservative 95% confidence interval based on Lemma 1 is reported. (Lemma 1 is invoked with $\delta = 0.05$.)

3.5. 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 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 over-estimated 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 scholarship under the original decisions did



not under our decisions and vice versa. 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.

In essence, our method produces much higher increases in enrollment probabilities than does the admission’s director. In Figure 3.4 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 much 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. 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, as compared to the admission 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.

4. Extensions

This section discusses four extensions that allow the method to be used in a broader set of applications. The first extension incorporates dynamic decisions in which there is additional uncertainty about future incoming individuals. The second concerns improving statistical properties when we deny applicants. The third extension, which was used for the application of Section 3, handles noisy estimates of the institution’s utility function.

4.1. Dynamic Decisions

In this section we discuss how to model a dynamic decision environment in the framework above. Fundamentally, all the methodology can still be applied. Consider that we have a set of “past”

decisions associated with applicants in a set P , “current” decisions associated with applicants in a set C , and “future” decisions associated with applicants in a set F .

The decision maker is at the current time. He already committed with offers for some applicants in P for which uncertainty was resolved. He observes the applicants that are in C for which he needs to make offers now, or for whom he made offers that are not yet accepted or rejected. However, he also needs to balance with the future applicants in F which have not yet arrived. That is, the set of applicants F is random. The applicants in F are drawn accordingly to some distribution (that is, the features of the applicants in F are random).

The model in the paper can accommodate this setting by properly enlarging the number of scenarios. The objective function will have the form

$$E \left[u_0 \left(\sum_{i \in P, j \in J} a_{ij} x_{ij} + \sum_{i \in C, j \in J} a_{ij} x_{ij} + \sum_{i \in F, j \in J} a_{ij} x_{ij} \right) + \sum_{k \in K} u_k \left(\frac{\sum_{i \in P, j \in J} w_{ik} a_{ij} x_{ij} + \sum_{i \in C, j \in J} w_{ik} a_{ij} x_{ij} + \sum_{i \in F, j \in J} w_{ik} a_{ij} x_{ij}}{\sum_{i \in P, j \in J} a_{ij} x_{ij} + \sum_{i \in C, j \in J} a_{ij} x_{ij} + \sum_{i \in F, j \in J} a_{ij} x_{ij}} \right) \right].$$

The information structure at the current time has the following aspects:

- (i) for $i \in P$, x_{ij} is fixed, $x_{ij} a_{ij}$ is observed, w_{ik} is known;
- (ii) for $i \in C$, x_{ij} is a decision variable, a_{ij} is a random variable, w_{ik} is known;
- (iii) for $i \in F$, x_{ij} is a decision variable, a_{ij} is a random variable, w_{ik} is a random variable;

Thus, the expectation above is over the random variables a_{ij} for $i \in C \cup F$, and the (random) features w_{ik} , $i \in F$. Note that the draws of a_{ij} can be conditional on the draws of w_{ik} for $i \in F$. Further recall that outstanding offers that have not been accepted or rejected are considered in C and the linear constraints fix the corresponding decision variables.

We emphasize that the non-separability of the objective function requires one to consider all three sets of applicants “simultaneously” in our decision. This is in contrast to most dynamic models in which one can define a (future) value function that enters the objective function additively (Bertsekas 2000 and Birge and Louveaux 1997). As expected the number of scenarios will be larger but all the results and methods can be directly applied. If the number of required scenarios becomes too large we discuss in Appendix D an alternative approach to evaluate solutions that should reduce the computational burden.

We have the scenarios S_C for the current uncertainty, and each of these branch out into scenarios for the future uncertainty S_{F^s} (which could be conditional on $s \in S_C$). Letting $p_s = 1/|S_C|$ and $p_{\tilde{s}} = 1/|S_{F^s}|$, the scenario based approximation is given by

$$\sum_{s \in S_C} p_s \sum_{\tilde{s} \in S_{F^s}} p_{\tilde{s}} \left[u_0 \left(\sum_{i \in P, j \in J} a_{ij} x_{ij} + \sum_{i \in C, j \in J} a_{ij}^s x_{ij} + \sum_{i \in F^s, j \in J} a_{ij}^{\tilde{s}} x_{ij}^s \right) + \sum_{k \in K} u_k \left(\frac{\sum_{i \in P, j \in J} w_{ik} a_{ij} x_{ij} + \sum_{i \in C, j \in J} w_{ik} a_{ij}^s x_{ij} + \sum_{i \in F^s, j \in J} w_{ik}^s a_{ij}^{\tilde{s}} x_{ij}^s}{\sum_{i \in P, j \in J} a_{ij} x_{ij} + \sum_{i \in C, j \in J} a_{ij}^s x_{ij} + \sum_{i \in F^s, j \in J} a_{ij}^{\tilde{s}} x_{ij}^s} \right) \right].$$

where $p_{\tilde{s}}$ is the probability associated with $\tilde{s} \in S_{F^s}$. Note that the linear constraints $Ax \leq b$ can be modified accordingly.

Waiting lists can also be considered in this dynamic setting. To model these waiting list decisions we add binary variables and additional linear constraints, both of which still allow us to apply the

methodology described above. Specifically, each current applicant, $i \in C$, with variables x_{ij} , will also be included in the future applicant pool creating variables x_{ij}^F . To model the waiting list, we create a new variable z_i to denote the decision of including the i th applicant into the waiting list. The constraint

$$\sum_{i \in C} z_i \leq K$$

models that we can have at most K students in the waiting list, and the additional constraints

$$x_{ij} \leq 1 - z_i \quad \text{and} \quad x_{ij}^F \leq z_i$$

ensure that applicant i can only be selected at the current time if he is not in the waiting list, and can only be selected at the future time if he is in the waiting list.

4.2. Denying Applicants and Sparsity Property

As noted earlier, it is possible to use the model (1) to exclude specific applicants by creating an “exclusion offer” that no applicant would accept. Consequently, all the probabilistic analysis and bounds developed earlier remain valid. However, in such a situation the total number of potential applicants might be substantially larger than the number of non-excluded applicants. This would be the case of many firms with capacity constraints. For example, in the school’s admission and scholarship problem, class sizes restrict the number of students than can be accepted from a much larger application pool.

This aspect of the business situation makes the optimal solution (to both (1) and (2)) to have the “exclusion offer” assigned to a large proportion of applicants. Consequently, those applicants are excluded and do not contribute with their “noise” in the further evaluation of the objective function. Hence, we can focus on “sparse solutions” that select at most \bar{C} applicants out of the total initial pool of \bar{T} applicants.

As a result, the probabilistic bounds developed in Section 2.5, which rely on bounding the number of different functions induced by the number of different feasible solutions by $|\mathcal{F}| \leq J^I$, can be further specialized. Instead of bounding $|\mathcal{F}| \leq (|J| + 1)^{\bar{T}}$, we can use the sparsity property and bound

$$|\mathcal{F}| \leq |J|^{\bar{C}} \cdot \binom{\bar{T}}{\bar{C}}.$$

The latter is substantially smaller than the first when $\bar{C} \ll \bar{T}$. In particular, we have $\log |\mathcal{F}| \leq \bar{C} \log |J| + \bar{C} \log \bar{T}$. The following corollary summarizes the new result which shows that the precision depends only logarithmically on the total number of applicants \bar{T} .

COROLLARY 2. *Under assumptions of Corollary 1, and the condition that at most \bar{C} applicants can be selected from a pool of \bar{T} , the result in Corollary 1 holds with*

$$\varepsilon := \frac{\sigma}{\sqrt{|S|}} \cdot \sqrt{\frac{2\bar{C}}{2 - \exp(\gamma)} \log \left(\frac{2|J|\bar{T}}{\delta} \right)}.$$

4.3. Noisy Estimates of Concave Utility Functions

The methodologies presented above require (assume) that each utility function u_k is concave for $k \in K$. There are many situations where the utility functions of interest are unknown a priori and must be estimated (see Section 3 for an example). In such cases it is not unusual to obtain consistent estimates that violate the assumed concavity property in a particular region of the utility function. While such violations are likely due to “noise,” they have serious implications on the computational tractability of the problem. So that our method can allow utility function estimates to contain such noise, we specialize a method to correct for such violations.

Different approaches have been proposed for correcting these violations. One set of approaches impose shape constraints on the (original) estimation of functions. Some examples include Bayesian estimation with constraints in the prior (Allenby, Arora, and Ginter 1995) and nonparametric estimation based on entropy (Allon et al 2007). A second approach post-processes consistent estimates to produce revised (consistent) estimates that do not violate concavity. Some post processing approaches include rearrangement to isotone the derivative (Birke and Dette 2007) and projections using Reproducing Kernels (Delecroix, Simioni and Thomas-Agnan 1996).

We specialize the projection idea of Delecroix, Simioni and Thomas-Agnan (1996) to our context. An important theoretical property of the projection of an initial estimate $\hat{u} : \mathcal{D} \rightarrow \mathbb{R}$ of the utility function $u : \mathcal{D} \rightarrow \mathbb{R}$, is that it will produce a new concave estimate $\tilde{u} : \mathcal{D} \rightarrow \mathbb{R}$ such that

$$\int_{\mathcal{D}} |u(t) - \tilde{u}(t)|^2 dt \leq \int_{\mathcal{D}} |u(t) - \hat{u}(t)|^2 dt. \quad (5)$$

That is, the new estimator \tilde{u} is closer to the true (unknown) utility function with respect to the metric induced by (5). This allows us to circumvent the estimation errors that disturb the concavity property. Note that (5) implies that the new estimator will inherit the consistency properties of the original estimator.

We are interested in the particular case of the domain $\mathcal{D} = [0, 1]$ (without loss of generality for our application), and that the initial estimate \hat{u} is piecewise linear. Let $\{(t_0, \hat{u}_0), (t_1, \hat{u}_1), \dots, (t_N, \hat{u}_N)\} \subset [0, 1] \times \mathbb{R}$ denote the “breakpoints” and the associated values of the piecewise linear estimate. That is, $\hat{u}(t_i) = \hat{u}_i$ and $\hat{u}(t)$ is linear for $t \in [t_i, t_{i+1}]$, $i = 1, \dots, N - 1$. The following convex quadratic problem

$$\begin{aligned} \min_{\tilde{u}} \sum_{i=1}^N \gamma_{i,1}(\hat{u}_i - \tilde{u}_i)^2 + \gamma_{i,2}(\hat{u}_{i-1} - \tilde{u}_{i-1})^2 + 2\gamma_{i,3}(\hat{u}_i - \tilde{u}_i)(\hat{u}_{i-1} - \tilde{u}_{i-1}) \\ \tilde{u}_i \geq \frac{t_i - t_{i-1}}{t_{i+1} - t_{i-1}} \tilde{u}_{i+1} + \frac{t_{i+1} - t_i}{t_{i+1} - t_{i-1}} \tilde{u}_{i-1} \quad \text{for } i = 1, \dots, N - 1 \end{aligned} \quad (6)$$

defines our new function estimate \tilde{u} , where $\gamma_{i,1} = (t_i - t_{i-1}) \int_0^1 (1-t)^2 dt$, $\gamma_{i,2} = (t_i - t_{i-1}) \int_0^1 t^2 dt$, $\gamma_{i,3} = (t_i - t_{i-1}) \int_0^1 t(1-t) dt$ implies convexity of the objective function of (6). Note that additional (artificial) breakpoints on \hat{u}_k could be introduced if one desires to achieve a more refined approximation for the projection. However, this projection approach will never produce a smooth function since the projection is always piece-wise linear (see Vapnik 1998).

5. Conclusions

In this paper, we provide a tailored solution to the admission and scholarship decisions of institutions in the post-secondary industry. In this industry, nearly 3 million admission offers and \$100 billion of scholarships are targeted to individual applicants each year. Our tailored solution accounts for the key distinguishing feature of this industry—schools value the average characteristics of the matriculating students such as percent female, percent from different regions of the world, average test scores, and average GPA. This goal to achieve a class with a desirable set of average characteristics greatly complicates the optimization problem and doesn’t allow the application of standard approaches. We develop a new approach that solves this more complex optimization problem and also accounts for the Optimizer’s Curse. Rather than provably solving for the optimal policy, we calculate sharp upper bounds on the value of the optimal solution, which informs a heuristic search algorithm for feasible solutions. We then use conservative confidence intervals for the value of any policy to examine whether it is statistically distinguishable from the value of the optimal solution or other solutions. We test the approach in a field study of an MBA scholarship process and implement adjusted scholarship decisions. We show how to develop the necessary inputs for the optimization method in this setting. We demonstrate the methodology can generate statistically

significant improvements over current management decisions and that, in the field test, the solution we identify is statistically indistinguishable from the true optimal solution. By comparing our solution to what management would do on its own, we provide insight into the nature of mistakes management made in this complex decision environment. Finally, we present extensions to handle decisions made on a rolling rather than single-shot basis, to further reduce the computational burden by taking advantage of sparsity properties when denying applicants, and to accommodate noisy estimates of the utility functions via a projection method that ensures concavity of each utility component.

Appendix A: Optimality Bounds

In this section we explore the structure of (2) in order to obtain bounds on the optimal value. In particular we address the lack of global concavity of the objective function due to the average characteristics. We explicitly use the concave piecewise linear structure (Hiriart-Urruty and Lemaréchal 1993) of each utility function

$$u_k(t) = \min_{\ell \leq \ell_k} t r_{k,\ell} + d_{k,\ell} = \min_{\substack{\sum_{\ell=1}^{\ell_k} \alpha_{k,\ell} = 1 \\ \alpha_{k,\ell} \geq 0}} \sum_{\ell=1}^{\ell_k} \alpha_{k,\ell} (t r_{k,\ell} + d_{k,\ell}) \quad (7)$$

where ℓ_k denotes the number of linear pieces of u_k , $k \in K$, and each linear piece is given by a slope $r_{k,\ell}$ and an intercept $d_{k,\ell}$. We propose a new Lagrangian relaxation (Appendix A.1) which is compared with a more standard relaxation based on linear programming for generic 0–1 fractional programming proposed by Wu (1997) (Appendix A.2).

Although the linear programming based relaxation is more standard, we note that the Lagrangian relaxation provides us with two technical advantages over the linear programming relaxation. First, there are no required assumptions on the function u_0 . Second, it allows us to consider the case of having a zero in the denominator (i.e., no object is acquired in a particular scenario, see Appendix B). Moreover, in computational experiments, the bounds obtained with the Lagrangian relaxation were substantially tighter, see A.3.

A.1. Lagrangian Relaxation

In this section we discuss a Lagrangian relaxation method designed to exploit the particular structure of the problem as stated in (2). This methodology will lead to bounds on the optimal value of (2). 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. By setting the penalty appropriately, we improve the bound and bring the policies closer.

Starting from the formulation (2) and using (7) we can write

$$\begin{aligned} \max_x \quad & \sum_{s \in S} p_s \left(u_0 \left(\sum_{i \in I, j \in J} a_{ij}^s x_{ij}^s \right) + \sum_{k \in K} \min_{\substack{\sum_{\ell=1}^{\ell_k} \alpha_{k,\ell} = 1 \\ \alpha_{k,\ell} \geq 0}} \sum_{\ell=1}^{\ell_k} \alpha_{k,\ell} \left(\frac{\sum_{i \in I, j \in J} w_{ik} a_{ij}^s x_{ij}^s}{\sum_{i \in I, j \in J} a_{ij}^s x_{ij}^s} r_{k,\ell} + d_{k,\ell} \right) \right) \\ \text{s.t.} \quad & Ax \leq b \\ & \sum_{j \in J} x_{ij} = 1, \quad i \in I \\ & x_{ij} \in \{0, 1\} \end{aligned}$$

Then, for each scenario $s \in S$, we create additional variables x^s with the constraint $x^s = x$. The Lagrangian relaxation will consist of relaxing the following three types of constraints: (i) the constraints just introduced (introducing Lagrangian multipliers λ_s for each scenario); (ii) the generic linear constraints (introducing the

multiplier σ); and (iii) the piecewise linear objective functions u_k (introducing the multipliers α). Therefore, for each fixed value of $(\lambda, \sigma, \alpha)$ we have the so-called dual function

$$\begin{aligned} \phi(\lambda, \sigma, \alpha) = \max_{x, x^s} \sum_{s \in S} p_s & \left(u_0 \left(\sum_{i \in I} \sum_{j \in J} a_{ij}^s x_{ij}^s \right) + \frac{\sum_{i \in I} \sum_{j \in J} \left(\sum_{k \in K} \sum_{\ell=1}^{\ell_k} \alpha_{k, \ell}^s r_{k, \ell} w_{ik} \right) a_{ij}^s x_{ij}^s}{\sum_{i \in I} \sum_{j \in J} a_{ij}^s x_{ij}^s} + \sum_{k \in K} \sum_{\ell=1}^{\ell_k} \alpha_{k, \ell}^s d_{k, \ell} \right) - \\ & - \sum_{s \in S} \lambda'_s (x^s - x) - \sigma' (Ax - b) \\ \text{s.t.} & \\ & \sum_{j \in J} x_{ij}^s = 1 \quad \text{for } i \in I, s \in S \\ & x_{ij}^s \in \{0, 1\} \quad \text{for } s \in S \end{aligned} \quad (8)$$

As shown in Appendix B, the dual function $\phi(\cdot, \cdot, \cdot)$ can be efficiently evaluated for any fixed value of $(\lambda, \sigma, \alpha)$. Moreover, by well-known results of weak duality (see Hiriart-Urruty and Lemaréchal 1993), for any $(\lambda, \sigma, \alpha)$ we have that

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

Therefore the best upper bound on the optimal value is obtained by minimizing the dual function $\phi(\lambda, \sigma, \alpha)$. Although $\phi(\lambda, \sigma, \alpha)$ is nondifferentiable, this dual function is known to be convex and many algorithms like subgradient, cutting planes, or bundle methods can be used to achieve this bound (see Bonnans et al 2006 for details).

A.2. Linear Programming Relaxation

The bilinear structure with binary variables will allow us to re-formulate (2) in such a way that nonlinearities are completely removed. This enables us to obtain global bounds on the problem based on linear programming (LP) techniques. This approach can be traced back to Wu (1997).

As expected there is a cost associated with such a reformulation. It will be necessary to introduce several (continuous and discrete) additional variables. That is, we lift the problem to a higher dimensional space where it can be cast as a LP. Nonetheless, the total number of additional variables increases with the number of scenarios turning the problem into a large-scale LP. In this formulation we will assume that at least one object is always selected for any scenario $s \in S$ and any feasible solution $x \in \mathcal{R}$, such that we have

$$\sum_{i \in I} \sum_{j \in J} a_{ij}^s x_{ij} > 0 \quad (9)$$

which ensures that the problem is well defined.

For any object $i \in I$ and scenario $s \in S$ consider the following binary quantity

$$y_{is} = \sum_{j \in J} x_{ij} a_{ij}^s \in \{0, 1\} \quad (10)$$

which equals 1 if object i is acquired under scenario s given the choice of x . Next we define the (continuous) quantity

$$v_{sk} = \frac{\sum_{i \in I} w_{ik} y_{is}}{\sum_{i \in I} y_{is}} \quad (11)$$

which is the average value for the feature k under scenario s . The quantity $v_{isk} = v_{sk} y_{is}$ is the “contribution” of student i to the average of feature k . Thus we can rewrite

$$v_{sk} \sum_{i \in I} y_{is} = \sum_{i \in I} v_{isk} = \sum_{i \in I} w_{ik} y_{is}. \quad (12)$$

The second equality constraint is linear on the variables of interest and can be included in our formulation. Next we need to ensure the first equality in (12). By adding the following inequalities we properly model the average

$$0 \leq v_{isk} \leq v_{sk}, \quad v_{isk} \leq M_k y_{is}, \quad v_{sk} - M_k (1 - y_{is}) \leq v_{isk}$$

where M_k is a feature dependent big- M constant (could be simply set to $M_k = \max_{i \in I} w_{ik}$).

Therefore, by also using the piecewise linear representation for each u_k (7), the reformulated problem can be cast as

$$\left\{ \begin{array}{l} \max_{x,y,v,t} \sum_{s \in S} p_s \left(t_{s0} + \sum_{k \in K} t_{sk} \right) \\ x \in \mathcal{R} \\ t_{sk} \leq v_{sk} r_{k,\ell} + d_{k,\ell} \quad \ell \leq \ell_k, k \in K \\ t_{s0} \leq \sum_{i \in I} y_{is} r_{0,\ell} + d_{0,\ell} \quad \ell \leq \ell_0 \\ y_{is} = \sum_{j \in J} x_{ij} a_{ijs} \quad i \in I, s \in S \\ \sum_{i \in I} v_{isk} = \sum_{i \in I} w_{ik} y_{is} \quad s \in S, k \in K \\ 0 \leq v_{isk} \leq v_{sk} \quad i \in I, s \in S, k \in K \\ v_{isk} \leq M_k y_{is} \quad i \in I, s \in S, k \in K \\ v_{sk} - M_k(1 - y_{is}) \leq v_{isk} \quad i \in I, s \in S, k \in K \end{array} \right. \quad (13)$$

The linear relaxation of (13) yields an upper bound on the optimal value of (2).

A.3. Preliminary Comparison with Random Data

In this section we provide the reader with a feel for the appropriateness of our proposed optimality bounds. We do this by solving different instances of (2). Table 4 summarizes solutions for four different problem sizes. We present the value and gap obtained from the optimality bound based on the Lagrangian relaxation (LR), the generic linear programming relaxation (LP), and also display the values of the optimal solutions obtained from an enumerating scheme based on the Lagrangian relaxation. For these small to medium problem sizes the heuristic also found the optimal solutions.

Computational Results for Random Data

Instance	Instance Parameters				Optimality Bounds			Optimality Gaps	
	I	J	K	S	Optimal Value	LP	LR	Gap-LP(%)	Gap-LR (%)
Rnd-01	10	5	4	2	8.8944	11.6463	8.8967	30.93	0.02
Rnd-02	20	10	4	10	9.6556	13.0812	10.3194	35.47	6.90
Rnd-03	221	5	4	2	7.9792	13.0985	8.0718	64.16	1.16
Rnd-04	200	20	10	100	15.7799	-	15.8008	-	0.13

Table 4 Computational results for randomly generated instances. The columns with LP correspond to the bound obtained by the linear programming relaxation described in Appendix A.2. The columns with LR correspond to the bound obtained by Lagrangian relaxation described in Section A.1. The optimal value was obtained with an enumerating scheme implemented on top of the Lagrangian relaxation.

The linear relaxation used CPLEX version 10.0. However, the memory requirements were too demanding to run the largest examples. The Lagrangian relaxation called for substantial coding but was much less demanding in terms of memory and produced much tighter bounds. Tight bounds were also produced by the Lagrangian relaxation in the work of Yu (1996) and Iida (1999) for the related max-min 0-1 (linear) knapsack problem. Hence, the Lagrangian relaxation appears to perform better for larger numbers of scenarios (as we used in our application). Based on these results we used the Lagrangian relaxation approach in Section 3.

Appendix B: Lagrangian Relaxation Details: evaluating $\phi(\lambda, \sigma, \alpha)$

Without loss of generality we use the convention that $0/0 = 0$. Indeed, if $\sum_{i \in I} \sum_{j \in J} a_{ij} x_{ij} = 0$, we also have that $\sum_{i \in I} \sum_{j \in J} w_{ij} a_{ij} x_{ij} = 0$ for every $k \in K$ (so that we cannot have a non-zero number divided by zero). Moreover, since all the arguments of u_k are fixed in the case of $\sum_{i \in I} \sum_{j \in J} a_{ij} x_{ij} = 0$, we can redefine $u_0(0)$ to the overall utility of having zero individuals minus the sum of all $u_k(0)$ over $k \in K$. Therefore there is no loss of generality in using the convention $0/0 = 0$.

For notational convenience define $g_i^s = g_i(\alpha^s) = \left(\sum_{k \in K} \sum_{\ell=1}^{\ell_k} \alpha_{k,\ell}^s r_{k,\ell} w_{ik} \right)$, and A^{ij} denote the column of A associated with x_{ij} . The dual function defined in (8) decomposes into smaller problems as follows

$$\phi(\lambda, \sigma, \alpha) = \sigma' b + \sum_{s \in S} p_s \sum_{k \in K} \sum_{\ell=1}^{\ell_k} \alpha_{k,\ell}^s d_{k,\ell} + \sum_{i \in I} \max_{x_i} \sum_{j \in J} x_{ij} \left(\sum_{s \in S} \lambda_{ij}^s - \sigma' A^{ij} \right) + \sum_{s \in S} p_s \phi_s(\lambda, \sigma, \alpha^s)$$

$$\sum_{j \in J} x_{ij} = 1$$

$$x_{ij} \in \{0, 1\}$$

where for each $s \in S$

$$\phi_s(\lambda, \sigma, \alpha^s) = \max_{x^s} u_0 \left(\sum_{i \in I} \sum_{j \in J} a_{ij}^s x_{ij}^s \right) + \frac{\sum_{i \in I} \sum_{j \in J} g_i^s a_{ij}^s x_{ij}^s}{\sum_{i \in I} \sum_{j \in J} a_{ij}^s x_{ij}^s} - \sum_{i \in I} \sum_{j \in J} \lambda_{ij}^s x_{ij}^s$$

$$\sum_{j \in J} x_{ij}^s = 1 \quad i \in I$$

$$x_{ij}^s \in \{0, 1\}$$

The first set of maximization problems were decomposed for each $i \in I$ and a simple greedy procedure with running time of $O(|J|)$ can solve each one of them. The last set of maximization problems were decomposed for each scenario $s \in S$. For each scenario this is equivalent to solving for $m = 0, 1, 2, \dots, |I|$

$$\phi_s(\lambda, \sigma, \alpha^s) = \max_{0 \leq m \leq |I|} \max_{x^s} u_0(m) + \frac{\sum_{i \in I} \sum_{j \in J} g_i^s a_{ij}^s x_{ij}^s}{m} - \sum_{i \in I} \sum_{j \in J} \lambda_{ij}^s x_{ij}^s$$

$$\sum_{i,j} a_{ij}^s x_{ij}^s = m$$

$$\sum_{j \in J} x_{ij}^s = 1 \quad i \in I$$

$$x_{ij}^s \in \{0, 1\}, \quad i \in I, j \in J.$$
(14)

(Recall our convention that $0/0 = 0$, so for $m = 0$ the middle term is simply zero. All dual bounds are still valid.) For each particular scenario s and value of m , we define for each object $i \in I$ the sets $J_0^s(i) = \{j : a_{ij}^s = 0\}$ and $J_1^s(i) = \{j : a_{ij}^s = 1\}$, which are respectively the set of policies that object i would not be acquired and the set of policies for which object i would be acquired. Associated with these sets we define

$$d_i^s = \max \left\{ \frac{g_i^s}{m} - \lambda_{ij}^s : j \in J_1^s(i) \right\} - \max \{ -\lambda_{ij}^s : j \in J_0^s(i) \}.$$

Next consider the “new” decision variable $z_i^s = 1$ if the we choose from $J_1^s(i)$, $z_i^s = 0$ otherwise. Then the problem (14) can be recast as

$$\phi_s(\lambda, \sigma, \alpha^s) = \max_{0 \leq m \leq |I|} u_0(m) + \sum_{i \in I} \max \{ -\lambda_{ij}^s : j \in J_0^s \} + \max_{z^s} \sum_{i \in I} d_i^s z_i^s$$

$$\sum_{i \in I} z_i^s = m$$

$$z_i^s \in \{0, 1\}, \quad i \in I$$

which can be solved by simply setting to one the components of z^s associated with the m largest values of d_i^s . The overall cost of solving the second set of problems is $O(|S||I||J| + |S||I|^2 \ln |I|)$.

Appendix C: Probabilistic Approximation Guarantees

The severity of the Optimizer’s Curse and the achievable probabilistic guarantees are linked to the trade-off between the flexibility of the model and the number of scenarios used. If the model is extremely flexible relatively to the number of scenarios, the model will take “advantage” of random fluctuations to generate what appears to be a better solution. Such solutions have been well documented to be unstable in the sense that if you drew a different sample, the value of the new solution could be dramatically different from the previous value. We would expect that such a solution would not perform as well in practice.

In order to avoid this pitfall, two broad approaches have been suggested in the literature. The first is to recognize that such solutions are “irregular.” Regularization methods limit the solver from selecting these

solutions by biasing the solution towards a pre-specified pattern (see Engl, Hanke and Neubauer 1996). The effect of such biases, however, is very difficult to predict in a complex nonlinear system like ours. As a result, we prefer the second approach. In this approach, we generate a large number of scenarios to avoid over-fitting. In particular, we need a large number of scenarios relative to the flexibility in the model, which in our case is represented by the number of different possible solutions. With enough scenarios, it is possible to invoke an uniform law of large numbers (ULLNs), see van der Vaart and Wellner (1996) and Dudley (2000), for the set of feasible solutions. We use the size of the set of feasible solutions to link the trade-off between approximation errors and number of scenarios needed. This approach does not completely eliminate the Optimizer’s curse, but it does provide bounds on how large the over-fitting can be across all the policies, based on a given level of confidence.

ULLNs are generalizations of the law of large numbers. They have been central in the empirical process literature during the past decades and their applications can be found in statistics (e.g., van der Vaart and Wellner 1996), machine learning (e.g., Vapnik 1998) and other fields (e.g., Andrews 1994). ULLNs answer a similar question to the law of large numbers, but now the convergence must be obtained uniformly (i.e., simultaneously) across many functions. Intuitively, the “larger” the class of functions for which one wants to achieve convergence, the harder it will be. While it is possible to establish uniform convergence over an infinite number of functions (e.g., the first result being the Glivenko-Cantelli Theorem in 1933), uniform convergence simply cannot happen for some classes of functions no matter how large the number of scenarios (i.e. sample size, see Dudley 2000).

In the problem as stated in (1), our interest is in the class of functions defined by each feasible solution $x \in \mathcal{R}$, namely,

$$\mathcal{F} = \left\{ f_x(a) = u_0 \left(\sum_{i \in I, j \in J} a_{ij} x_{ij} \right) + \sum_{k \in K} u_k \left(\frac{\sum_{i \in I, j \in J} w_{ik} a_{ij} x_{ij}}{\sum_{i \in I, j \in J} a_{ij} x_{ij}} \right) : x \in \mathcal{R} \right\}. \quad (15)$$

Thus, we have one function for each possible solution to the problem, where that function is a function of the scenario. Because this set of functions is finite and does not vary with the number of scenarios, it is possible to achieve a uniform convergence guarantee over the entire class of functions \mathcal{F} as the sample size increases, see van der Vaart and Wellner (1996). Such a guarantee can be used to meaningfully bound the effect of the Optimizer’s Curse.

However, the generation of an arbitrarily large number of scenarios might be costly or might lead to an untractable computational problem in (2). Therefore, we are interested in results that are valid in finite samples. To obtain non-asymptotic bounds under which ULLNs will hold one needs to properly characterize the size of the class of functions (i.e., the number of different possible solutions). This number should quantify how much flexibility the model has to solve the problem. Although there are several possible notions to measure the size of a class of functions, the most relevant notion is the covering number of \mathcal{F} (Dudley 2000). The covering number is trivially bounded by the cardinality of the set \mathcal{F} . Uniform convergence results have been investigated in many literatures, see van der Vaart and Wellner (1996) for a statistical perspective and Kleywegt et al (2001) for a stochastic programming perspective. In Appendix E we provide a simple proof under mild assumptions relying on Martingale inequalities and the union bound.

LEMMA 1. Let $|\mathcal{F}|$ denote the cardinality of \mathcal{F} and $a^s, s \in S$, be independent scenarios, and $\sigma^2 = \max_{f_x \in \mathcal{F}} \text{var}(f_x)$. Assume that for some $\varepsilon > 0$ and $\gamma > 0$, $|f_x(a) - E[f_x]| \leq \gamma \sqrt{|S| \sigma^2}$ for every a and $f_x \in \mathcal{F}$. Then, we have

$$P \left(\max_{f_x \in \mathcal{F}} \left| \frac{1}{|S|} \sum_{s=1}^{|S|} f_x(a^s) - E[f_x] \right| > \varepsilon \right) \leq 2|\mathcal{F}| e^{-\frac{\varepsilon^2 |S|}{2\sigma^2} (2 - \exp(\gamma))}.$$

A consequence of Lemma 1 is to provide the relationship between the cardinality \mathcal{F} (i.e., number of solutions) and the number of scenarios, $|S|$. In particular, to achieve a particular precision and confidence level, the number of scenarios depends only logarithmically on the number of solutions. Corollary 3 translates this dependence into a form that is directly applicable.

COROLLARY 3. Under the assumptions of Lemma 1, let \hat{x} be any data-driven selected feasible solution of (2). Then, we have that the interval

$$\left[\frac{1}{|S|} \sum_{s=1}^{|S|} f_{\hat{x}}(a^s) - \varepsilon, \frac{1}{|S|} \sum_{s=1}^{|S|} f_{\hat{x}}(a^s) + \varepsilon \right]$$

contains $E[f_{\hat{x}}]$ with probability at least $1 - \delta$ provided that

$$\varepsilon \geq \frac{\sigma}{\sqrt{|S|}} \cdot \sqrt{\frac{2}{2 - \exp(\gamma)} \log\left(\frac{2|\mathcal{F}|}{\delta}\right)}.$$

Note that Corollary 3 implies Corollary 1 since $|\mathcal{F}| \leq |J|^{|I|}$. The proof of Corollary 3 is given in Appendix E.

The relationship described in Corollary 3 provides us with a conservative estimate of the number of scenarios we need in order to achieve a desired precision with a specified probability. This relationship can be used in practice to create a valid (conservative) confidence interval on the value of any solution obtained via optimization or heuristics that rely on the given data. Thus, this interval bounds the potential impact of the Optimizer’s Curse.

Appendix D: Computationally Intractable Numbers of Scenarios

The computational complexity of solving the optimization problem (2) tends to increase as more scenarios are used. More specifically, given the combinatorial and nonlinear nature of the formulation, it is reasonable to expect that the running time of solving (2), exactly or even approximately, would increase superlinearly with the number of scenarios. If this is the case, it might be appealing to develop a less computationally intensive approach that generates an upper bound on the optimal value for a solution.

The idea is to solve many versions of the model (2) each having a relatively small number of scenarios. When the computational cost increases superlinearly with the number of scenarios, a gain is possible here. A similar idea has been proposed in Kleywegt et al (2001) using the true optimal values; instead, we focus on utilizing optimality bounds, which are usually easier to compute (as proposed in Appendix A).

The statistical inference of such a procedure is based on the fact that each replication of the model is independent and identically distributed. This will allow us to bound by how much the objective function value of a proposed solution \hat{x} can deviate from the true (unknown) optimal value. Although there are several ways of constructing the replication scheme, we focus on the following standard replication procedure for the sake of clarity (see Bayraksan and Morton 2006 for more sophisticated schemes).

1. Set confidence level $\delta \in (0, 1)$, replication size R , number of scenarios $|S|$, and a policy $\hat{x} \in \mathcal{R}$.
2. For each replica $r = 1, \dots, R$
 - 2.1 Generate i.i.d. scenarios $a^r = \{a^{r,1}, \dots, a^{r,|S|}\}$
 - 2.2 Compute an optimality bound \hat{v}^r to the optimal value of (2) with a^r
 - 2.3 Compute the r th gap $G_r = \hat{v}^r - \frac{1}{|S|} \sum_{s \in S} f_{\hat{x}}(a^{r,s}) \geq 0$, where $f_{\hat{x}} \in \mathcal{F}$, see (15)
3. Compute the sample mean and variance of the gap

$$\bar{G} = \frac{1}{R} \sum_{r=1}^R G_r \quad \text{and} \quad s_G^2 = \frac{1}{R-1} \sum_{r=1}^R (G_r - \bar{G})^2$$

4. Calculate a conservative confidence interval for the value of $\max_{x \in \mathcal{R}} E[f_x] - E[f_{\hat{x}}]$

$$\left[0, \bar{G} + \frac{t_{R-1, \delta} s_G}{\sqrt{R}} \right]$$

where $t_{R-1, \delta}$ denotes the $1 - \delta$ quantile of the Student’s t distribution with $R - 1$ degrees of freedom. Asymptotically as the number of replicas R grows to infinity we have that (see Bayraksan and Morton 2006 and the references therein)

$$P\left(\max_{x \in \mathcal{R}} E[f_x] - E[f_{\hat{x}}] \leq \bar{G} + \frac{t_{R-1, \delta} s_G}{\sqrt{R}}\right) \geq 1 - \delta. \quad (16)$$

That is, with high probability, we can bound how much larger the optimal value of (1) can be from the true expectation of the selected policy. In this way we are able to trade off having to compute bounds with one large set of scenarios and instead compute bounds using a number of sets of scenarios where we compute the bounds for each set and then use the first two moments to get the bounds.

Appendix E: Proofs

Proof of Lemma 1. By Lemma 1.6 in Ledoux and Talagrand (1991), we have that for a fixed function $f_x \in \mathcal{F}$

$$P\left(\left|\frac{1}{|S|}\sum_{s=1}^{|S|}f_x(a^s) - E[f_x]\right| > \varepsilon\right) \leq 2e^{-\frac{\varepsilon^2|S|}{2\sigma^2}(2-\exp(\gamma))}. \quad (17)$$

Therefore, using the union bound, we have

$$\begin{aligned} P\left(\max_{f_x \in \mathcal{F}} \left|\frac{1}{|S|}\sum_{s=1}^{|S|}f_x(a^s) - E[f_x]\right| > \varepsilon\right) &\leq P(\exists f_x \in \mathcal{F} : \left|\frac{1}{|S|}\sum_{s=1}^{|S|}f(a^s) - E[f(a)]\right| > \varepsilon) \\ &\leq \sum_{f \in \mathcal{F}} P\left(\left|\frac{1}{|S|}\sum_{s=1}^{|S|}f(a^s) - E[f(a)]\right| > \varepsilon\right) \\ &\leq |\mathcal{F}| \max_{f \in \mathcal{F}} P\left(\left|\frac{1}{|S|}\sum_{s=1}^{|S|}f(a^s) - E[f(a)]\right| > \varepsilon\right) \\ &\leq 2|\mathcal{F}|e^{-\frac{\varepsilon^2|S|}{2\sigma^2}(2-\exp(\gamma))} \end{aligned} \quad (18)$$

where the last step follows from (17). *Q.E.D.*

Proof of Corollary 3. Since Lemma 1 provides a bound that is uniform over all $x \in \mathcal{R}$, we have that it also applies to $\hat{x} \in \mathcal{R}$ that is selected via any data-driven procedure. The choice of ε makes the probability bound in (18) less than δ . Thus, with probability at least $1 - \delta$ we have

$$\left|\frac{1}{|S|}\sum_{s=1}^{|S|}f_{\hat{x}}(a^s) - E[f_{\hat{x}}]\right| \leq \varepsilon,$$

and the result follows. *Q.E.D.*

Acknowledgments

The authors are in debt to Pierre Bonami, Katya Scheinberg, Jon Lee, Andreas Wächer, Oktay Günlük, and Peng Sun for several discussions. We 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

- [1] G. M. ALLENBY, N. ARORA AND J. L. GINTER, *Incorporating Prior Knowledge into the Analysis of Conjoint Studies*, Journal of Marketing Research, May 1995, p. 152–162.
State Higher Education Officers, State Higher Education Finance Fiscal Year 2008 Report.
- [2] G. M. ALLENBY, R. P. LEONE AND L. JEN, *A Dynamic Model of Purchase Timing with Application to Direct Marketing.*, Journal of the American Statistical Association, Volume 94, number 446, pages 365–366, 1999.
- [3] G. ALLON, M. BEENSTOCK, S. HACKMAN, U. PASSY, AND A. SHAPIRO, *Nonparametric Estimation of Concave Production Technologies by Entropic Methods*, Journal of Applied Econometrics, 22: 795816 (2007).
- [4] D. W. K ANDREWS, *Empirical Process Methods in Econometrics*, Handbook of Econometrics, Volume IV, Edited by R.F. Engle and D.L. McFadden, Elsevier Science, 1994.
- [5] A. ANSARI AND C. F. MELA, *E-customization*, Journal of Marketing Research, Volume 40, 2, pages 131–145, 2003.
- [6] C. AVERY AND C. M. HOXBY, *Do and Should Financial Aid Packages Affect Students' College Choices?*, NBER Working Paper No. 9482, February 2003 JEL No. I2, J0, H0.
- [7] G. BAYRAKSAN AND D. P. MORTON, *Assessing Solution Quality in Stochastic Programs*, Mathematical Programming, Volume 108, Numbers 2-3, September, 2006.
- [8] A. BELLONI, R. M. FREUND, M. SELOVE, AND D. SIMESTER, *Optimal Product Line Design: Efficient Methods and Comparisons*, Management Science Vol. 54, No. 9, September 2008.
- [9] P. D. BERGER AND N. I. NASR, *Customer lifetime value: Marketing models and applications*, Journal of Interactive Marketing, Vol. 12, No. 1, pp 17–30, 1998.

- [10] J.R. BIRGE AND F. LOUVEAUX, *Introduction to Stochastic Programming*, Springer-Verlag, New York, 1997.
- [11] D. BERTSEKAS, *Dynamic Programming and Optimal Control*, Vol. 1, Athena Scientific, Belmont, 2000.
- [12] M. BIRKE AND H. DETTE, *Estimating a Convex Function in Nonparametric Regression*, Scandinavian Journal of Statistics, Vol 34: 384404, 2007.
- [13] R. C. BLATTBERG AND J. DEIGHTON, *Manage Marketing by the Customer Equity Test*, Harvard Business Review, 74 (July-August 1996), 136-44.
- [14] P. BONAMI, L. T. BIEGLER, A. R. CONN, G. CORNUÉJOLS, I.E. GROSSMANN, C. D. LAIRD, J. LEE, A. LODI, F. MARGOT, N. SAWAYA, A. WÄCHTER, *An Algorithmic Framework for Convex Mixed Integer Nonlinear Programs*. Discrete Optimization, 5(2):186–204, 2008.
- [15] J.F. BONNANS , J.C. GILBERT, C. LEMARÉCHAL, AND C. SAGASTIZBAL, *Numerical Optimization Theoretical and Practical Aspects*, 2nd edition, 2006, Springer.
- [16] R. G. CHAPMAN, *Pricing Policy and the College Choice Process*, Research in Higher Education, 10 (1), 37-57, 1979.
- [17] R. G. CHAPMAN AND R. STAELIN, *Exploiting Rank Ordered Choice Set Data Within the Stochastic Utility Model*, Journal of Marketing Research, XIX (August), 288-301, 1982.
- [18] K. D. CHEN AND W. H. HAUSMAN, *Technical Note: Mathematical Properties of the Optimal Product Line Selection Problem Using Choice-Based Conjoint Analysis*, Management Science, Vol. 46, No. 2, February 2000, pp. 327-332.
- [19] S. CHIB, *Bayes Inference in the Tobit Censored Regression Model*. Journal of Econometrics, 51, 79-99, 1992.
- [20] W. S. CLEVELAND, *Robust Locally Weighted Regression and Smoothing Scatterplots*, Journal of the American Statistical Association, 74, 829-836, 1979.
- [21] W. S. CLEVELAND AND S. J. DEVLIN, *Locally-Weighted Regression: An Approach to Regression Analysis by Local Fitting*. Journal of the American Statistical Association 83 (403): 596–610, 1988.
- [22] M. DELECROIX, M. SIMIONI, AND C. THOMAS-AGNAN, *Functional Estimation under Shape Constraints*, Nonparametric Statistics, Vol. 6, 1996, pp. 69-89.
- [23] S. J. DESJARDINS, *Assessing the Effects of Changing Institutional Aid Policy*. Research in Higher Education, 42(6), 653-678.
- [24] S. J. DESJARDINS, H. DUNDAR AND D. D. HENDEL, *Modeling the College Application Decision Process in a Land-Grant University*. Economics of Education Review, 18, 1999, 117-132.
- [25] X. DRÈZE AND A. BONFRER, *Moving from customer lifetime value to customer equity*, Quantitative Marketing and Economics, Volume 7, number 3, pages 289–320, 2009.
- [26] R. M. DUDLEY, *Uniform Central Limit Theorems*, Cambridge Studies in advanced mathematics: 63, Cambridge Press, 1999.
- [27] D. B.DUNSON, Z. CHEN, AND J. HARRY, *A Bayesian Approach for Joint Modeling of Cluster Size and Subunit-Specific Outcomes*. Biometrics, 59(3), 521-530, 2003.
- [28] R.G.. EHRENBERG, *Econometric studies of higher education*, Journal of Econometrics, Volume 121, Issues 1-2, Higher education (Annals issue), July-August 2004, Pages 19-37.
- [29] R. G. EHRENBERG AND D. R. SHERMAN, *Optimal Financial Aid Policies for a Selective University*, Journal of Human Resources vol. 19 No. 2 (Spring 1984) 202-230.
- [30] H. W. ENGL, M. HANKE AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.
- [31] D. EPPLE, R. ROMANO AND H. SIEG, *Peer effects, financial aid and selection of students into colleges and universities: an empirical analysis*. Journal of Applied Econometrics 18, 501-525.
- [32] M. L. FISHER, *The Lagrangian Relaxation Method for Solving Integer Programming Problems*, Management Science, Vol. 27, No. 1, January 1981, pp 1–18.

- [33] M. L. FISHER, *The Lagrangian Relaxation Method for Solving Integer Programming Problems*, Management Science, Vol. 50, No. 12, Ten Most Influential Titles of “Management Science’s” First Fifty Years (Dec., 2004), pp. 1861–1871.
- [34] M. HELD AND R. M. KARP, *The traveling-salesman problem and minimum spanning trees: part II*, Mathematical Programming 1 (1971), 6–25.
- [35] A. HERRMANN AND A. GUSTAFSSON, *Conjoint Measurement: Methods and Applications*. Huber, Frank Editors: Forth Edition Berlin : Springer, 2003.
- [36] J.-B. HIRIART-URRUTY AND C. LEMARÉCHAL, *Convex Analysis and Minimization Algorithms*. Number 305-306 in Grund. der math. Wiss. Springer-Verlag, 1993. (two volumes).
- [37] C. HOMBURG, V. V. STEINER AND D. TOTZEK, *Managing dynamics in a customer portfolio*, Journal of Marketing, Volume 73, number 5, pages 70–89, 2009.
- [38] H. IIDA, *A note on the max-min 0-1 knapsack problem*, Journal of Combinatorial Optimization 3 (1999), pp 89–94.
- [39] M. D. JOHNSON AND F. SELNES, *Customer Portfolio Management: Toward a Dynamic Theory of Exchange Relationships*, The Journal of Marketing, Vol. 68, No. 2 (Apr., 2004), pp. 1–17.
- [40] M. JÜNGER, G. REINELT, AND G. RINALDI, *The Traveling Salesman Problem*, Preprint 94-12, Interdisziplinäres Zentrum für Wissenschaftliches Rechnen der Universität Heidelberg, March 1994.
- [41] R. E. KALLIO, *Factors influencing the college choice decisions of graduate students*, Research in Higher Education Vol. 36, no 1, February 1995, pp 109–124.
- [42] R. S. KAPLAN AND D. P. NORTON, *Using the Balanced Scorecard as a Strategic Management System*, Harvard Business Review, 1996, Vol. 74; Number 1, pp. 75–87.
- [43] R. KHAN, M. LEWIS AND V. SINGH, *Dynamic Customer Management and the Value of One-to-One Marketing*, Marketing Science Vol. 28, No. 6, November-December 2009, pp. 1063–1079.
- [44] A. J. KLEYWEGT, A. SHAPIRO, AND T. HOMEM-DE-MELLO, *The sample average approximation method for stochastic discrete optimization*, SIAM Journal on Optimization, Vol. 12, No. 2, 2001, pp 479–502.
- [45] M. LEDOUX AND M. TALAGRAND, *Probability in Banach Spaces (Isoperimetry and processes)*. Ergebnisse der Mathematik und ihrer Grenzgebiete, Springer-Verlag, 1991.
- [46] L. H. LITTEN, *Different Strokes in the Applicant Pool: Some Refinements in a Model of Student College Choice*, The Journal of Higher Education, Vol. 53, No. 4 (Jul. - Aug., 1982), pp 383–402.
- [47] P. MANCHANDA, P. E. ROSSI AND P. K. CHINTAGUNTA, *Response Modeling with Non-Random Marketing Mix Variables*, Journal of Marketing Research, 41 (November 2004), 467–478.
- [48] L. C. MARSH AND A. ZELLNER, *Bayesian Solutions to Graduate Admissions and Related Selection Problems*. Journal of Econometrics, 2004, 121, 405–426.
- [49] B. P. S. MURTHI AND S. SARKAR, *The role of the management sciences in research on personalization*, Management Science, Vol 49, 10, pages 1344–1362, 2003.
- [50] S. X. NIU AND M. TIENDA, *Choosing colleges: Identifying and modeling choice sets*, Social Science Research, Volume 37, Issue 2, June 2008, Pages 416–433.
- [51] PETER NURNBERG, MORTON SCHAPIRO, AND DAVID ZIMMERMAN, *Students Choosing Colleges: Understanding the Matriculation Decision at a Highly Selective Private Institution*, NBER Working Paper No. 15772, February 2010, JEL No. I21.
- [52] , S. OLAFSSON, X. LI AND S. WU, *Operations research and data mining*, European Journal of Operational Research, Volume 187, number 3, pages 1429–1448, 2008.
- [53] G. N. PUNJ AND R. STAELIN, *The Choice Process For Graduate Business Schools*, Journal of Marketing Research, XV (November), 1978, pp 588–98.
- [54] W. J. REINARTZ AND V. KUMAR, *The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration*, The Journal of Marketing, Vol. 67, No. 1 (Jan., 2003), pp. 77–99.
- [55] W. REINARTZ, J. S. THOMAS AND V. KUMAR, *Balancing acquisition and retention resources to maximize customer profitability*, Journal of Marketing, Volume 69, number 1, pages 63–79, 2005.

- [56] P. E. ROSSI, R. E. MCCULLOCH AND G. M. ALLENBY, *The Value of Purchase History Data in Target Marketing*. Marketing Science, 1996, 15(4), pp. 321–40.
- [57] , R. T. RUST, K. N. LEMON. AND V. A. ZEITHAML, *Return on marketing: Using customer equity to focus marketing strategy*, Journal of marketing, Volume 68, number 1, pages 109–127, 2004.
- [58] C. SCHÖN, *On the Optimal Product Line Selection Problem with Price Discrimination*, Management Science, Vol. 56, No. 5, May 2010, pp. 896–902.
- [59] J. E. SMITH, R. L. WINKLER, *The Optimizers Curse: Skepticism and Postdecision Surprise in Decision Analysis*, Vol. 52, No. 3, March 2006, pp. 311–322.
- [60] A. W. VAN DER VAART AND J. A. WELLNER, *Weak Convergence and Empirical Processes, with applications to statistics*, Springer Series in Statistics, 1996.
- [61] V. N. VAPNIK, *Statistical Learning Theory*, Wiley Interscience, 1998.
- [62] , R. VENKATESAN AND V. KUMAR, *A customer lifetime value framework for customer selection and resource allocation strategy*, Journal of Marketing, Volume 68, number=4, pages 106–125, 2004.
- [63] R. VENKATESAN, V. KUMAR AND T. BOHLING, T., *Optimal customer relationship management using bayesian decision theory: An application for customer selection*, Journal of Marketing Research, Volume 44, number 4, pages 579–594, 2007.
- [64] H. WAINER, *Shopping for Colleges When What We Know Ain't*. Journal of Consumer Research, 2005, 32, 337–342.
- [65] T.-H. WU, *A note on a global approach for general 01 fractional programming*, European Journal of Operational Research Volume 101, Issue 1, 16 August 1997, pp. 220–223.
- [66] G. YU, *On the Max-Min 0-1 Knapsack Problem with Robust Optimization Applications*, Operations Research, Vol. 44, No. 2, 1996, pp. 407–415.
- [67] R. ZEITHAMMER AND P. LENK, *Bayesian estimation of multivariate-normal models when dimensions are absent*, Quantitative Marketing and Economics, 4(3), 2006, 241–265.
- [68] STATE HIGHER EDUCATION EXECUTIVE OFFICERS, *State Higher Education Finance Fiscal Year 2008 Report*.
- [69] U.S. DEPARTMENT OF EDUCATION, *National Center for Education Statistics*, 200304 National Post-secondary Student Aid Study (NPSAS:04).