A Sequential Model for Recruitment: High Volume, Random Yields, and Rigid Target

Lilun Du* Qing Li† Peiwen Yu‡

August 24, 2021

Abstract

We model a multi-phase and high-volume recruitment process as a large-scale dynamic program. The success of the process is measured by a reward, which is the total assessment score of accepted candidates minus the penalty cost of the number of accepted candidates in the end deviating from a preset hiring target. For a recruiter, two questions are important: How many offers should be made in each phase, and how many phases should there be?

We consider an upper bound, which is obtained when the information about all candidates is available at the beginning, and a lower bound, which is obtained when the recruiter sets the number of offers to make in each phase before assessing candidates. We show that when the volume (i.e., arrival rate of applicants and the target) is large, the upper bound, the lower bound, and the optimal policy all converge to the same limit. Motivated by the convergence results, we design four easily computable heuristics which are all asymptotically optimal when the volume is large. With simple yet effective heuristics in hand, we can compute the number of offers to make in each phase and examine the impact of the number of phases in the process on the reward.

We apply our modeling framework and heuristics to the recruitment process of graduate students in a business program. Our counterfactual analysis shows that the outcome of the process can be improved by up to 5.5%, if our model recommendations are adopted. Our study is the first to model a high-volume recruitment process as a dynamic program and test it in a case study.

Keywords: High-volume recruitment, dynamic programming, approximation, secretary problem, random yields

1 Introduction

High-volume recruiting involves hiring a larger number of people in a short amount of time (Hayton 2018 and Min 2019). While it can happen almost anywhere, the common high-volume

*Hong Kong University of Science and Technology, e-mail: dulilun@ust.hk

†ibid, e-mail: imqli@ust.hk

‡Chongqing University, e-mail: ypw@cqu.edu.cn

recruiters are in industries such as logistics, hospitality, accounting, consulting, universities, and global manufacturing. High-volume recruitment is needed for various reasons. For example, it is needed when organizations are growing exponentially or when they are opening new locations or offices (Hayton 2018). To create economy of scale and synchronize with the school calendar, many organizations have regular recruiting seasons during which they hire in volume.

High-volume recruiting is challenging because of the time pressure, the sheer number of candidates with diverse background and experience, and the uncertainty involved (Du and Li 2019). The traditional hiring process designed for low-volume recruiting is not scalable and simply fails in a high-volume environment. A poorly managed process may overwhelm the recruiting teams and lead to poor hiring decisions, not to mention poor candidate experience. The latter, in particular, has wide ramifications. In the short term, it will reduce candidates' interest in joining the organization. In the long term, word will spread to damage brand image, which will increase the difficulty of hiring workers and acquiring customers in the future (Hayton 2018).

One way to spread out the workload for the recruiting teams is to adopt a sequential process. That is, each recruiting season is divided into several phases, each with an application deadline. Candidates who apply in the same phase will be assessed and compared, and offers will be made before the next phase. Because of the reduced volume in each phase, such a sequential process will allow the recruiting teams to adopt a more personalized approach when communicating with candidates and reduce time to hire, an important metric for candidate experience (Moore 2018). Furthermore, in a sequential process, the recruiter determines the number of offers to make in each phase after observing the number of offers that have been accepted in the previous phases. Having multiple, rather than just one, tempts to meet the target can provide diversification over time and reduce the risk of mismatch[1]. The sequential process, however, is not without drawback. As offer decisions must be made before knowing the qualifications of the candidates applying in the future, there is no guarantee that offers are always given to the most qualified candidates. For high-volume recruiters, there are two practical questions. First, how many offers should be made in each phase? Second, how many phases should there be in a recruiting season?

To provide a rigorous framework to answer these questions, we model a high-volume recruiting process as a large-scale dynamic program. The recruiter has a preset target number of people to hire. In each phase, applications are received and assessed. The recruiter determines which and how many applicants will be given an offer, and those candidates who have received

[1]For diversification over time in lot-sizing with random yields, see Henig and Gerchak (1990), Li and Zheng (2006), or Li et al. (2007)

offers will have until the next phase to decide whether they accept their offers. The overall objective of the recruiter is to keep the number of people hired in the end in the vicinity of the target number and to make offers to the most qualified candidates. Numerically, we measure the success of the recruitment process by a total reward, which is the total assessment score of the candidates hired minus the penalty cost of the total number of candidates hired deviating from the hiring target.

Because of the large state space, it is impossible to compute the optimal policy of the dynamic program. We instead consider an upper bound, which is obtained when all the information about the candidates (i.e., their qualification and acceptance decisions) is available at the beginning, and a lower bound, which is obtained when the recruiter determines the number of offers to make in each phase before assessing candidates. We show that when the volume (i.e., arrival rate of candidates and the target) is large, the upper bound, the lower bound, and the optimal policy all converge to the same limit with known convergence rates. The convergence results motivate us to design four easily computable heuristics and the total reward under any of them converges to the maximal reward. With simple yet effective heuristics in hand for computing the number of offers to make in each phase, we can derive an intuitive closed form expression for the total reward as a function of the number of phases. We show that when both the volume is large, the total reward primarily comes from the total score of accepted candidates and it is not sensitive to the change of the number of phases.

We apply our modeling framework and heuristics to a recruitment process of graduate students in a business program. Our counterfactual analysis for the 2017-2018 and 2018-2019 academic years shows that the outcome of the recruitment process, measured by the total assessment score of the candidates enrolled minus the penalty cost of total enrollment deviating from the target number, can be improved by up to 5.5%, if our model recommendations are adopted. Our analysis also clears doubts about the implication of the number of phases, which is currently set three for said business program. We show that, as the administrators suspected, they should indeed increase the number of phases. However, the total reward changes only slightly as a result.

Due to the importance of high-volume recruiting, the topic has been intensively discussed in industries. Interestingly, most of the organizations contributing to the field are analytics-related start-ups or technology giants (e.g., Google, LinkedIn, Kiran Analytics, Ideal, Appcast, and Harver). Their clients from diverse industries have benefited from their products and services. While their focus and strength might be on different aspects of high-volume recruiting, they seem to share one view in common - data analytics and technology are crucial for an efficient and effective process (Hayton 2018, Min 2019, and Moore 2018). Our research reconfirms that view. We are the first to provide an analytical tool for computing the number of offers to make in each phase and for understanding the impact of the number of phases in a sequential process. The tool can be immediately implemented. Without such tools, recruiters would have to rely on their gut feeling, which can be time consuming and could easily lead to poor and inconsistent decisions.

We model the sequential process as a large-scale dynamic program. Dynamic programs with large state spaces are notoriously intractable and researchers have been looking for good approximations (Powell 2007). In constructing approximations, it is important to have theoretical justifications. In our problem, the reward in each phase and the state transition function can be written as additively separable functions of the state variables. From the law of large numbers, when the number of state variables is large, observing the realization of the state variables in each phase is no longer important, and as such, we can simply ignore the available information about the state variables when computing the value function and actions and such an approximation is asymptotically optimal. To our best knowledge, the way we construct approximations is new[2] and the similar idea can be applied to other applications with similar reward and transition functions. In summary, our research not only contributes to the theory and practice of high-volume recruiting, but also adds to the arsenal of approximate dynamic programming.

Our study is located at the intersection of decision theory, operations management, and human resources management (see Boudreau et al. 2003 for an early review). In statistical psychology, one important research topic is how to select candidates through multistage selection tests. The outcomes of tests are correlated, albeit not perfectly, with the selection criteria. The first formulation of sequential selection came from Cronbach and Gleser (1965). Their framework has been used in and extended by De Corte (1998) to include a probationary period and by De Corte et al. (2006) to allow a mixture of several candidate groups. De Corte et al. (2006) provide a neat summary of the decisions that must be made and their potential impact. They also provide a comprehensive literature review on personnel selection in general. The goal typically is to select personnel through sequential tests such that the utility or payoff of hiring and the cost of testing are properly balanced, or such that the probability of rejecting qualified candidates is no greater than a given percentage (Du and Li 2019). In our study, we assume that the evaluation of candidates is perfect and we focus on the question of how many offers should be made instead of who should be made an offer.

[2]Goldberg et al. (2016) and Xin and Goldberg (2016) show that a very simple constant-order policy is optimal for the lost-sales inventory models when the lead time is large. However, their way of constructing the simple policy and their proof of optimality and convergence rates are very different from ours and cannot be applied in our model.

One challenge facing the recruiter is that in each phase offer decisions must be made without knowing what qualifications of future candidates hold. This is similar to the classical prophet inequalities or the secretary problems (see, for example, Freeman 1983 or DeGroot 2004 for early reviews of the secretary problems and Gallego and Wang 2020 for recent progress on prophet inequalities problem). Our model contributes to the literature in three ways. First, in our model candidates arrive and are assessed in batches. Therefore, we are dealing with a large scale dynamic program and our focus is on approximation. Second, the candidates who are given offers may not accept them. As such the outcome of the recruitment is determined by both the quantity and quality of the candidates hired. Finally, we discuss the optimal number of phases the recruiting process should have and conduct a case study to illustrate the applicability of our model, both new to the literature.

In high-volume recruitment, acceptance yield is random. The literature on lot sizing under random yields is rich (see Henig and Gerchak 1990, Li and Zheng 2006, Li et al. 2007, or Yano and Lee 1995 for a review). Particularly relevant to our study is the strand of literature on multiple lot-sizing production to order under random yields (Anily et al. 2002 and Grosfeld-Nir et al. 2000). In this literature, a contractor receives an order that it commits to satisfying in full. Production lots require costly setup and products may be defective. The contractor's problem is to find an optimal policy that minimizes the total cost of filling the order. In this literature, non-defective products are identical. However, in our study candidates have different qualifications and the recruiter not only needs to meet the hiring target, but indeed to find the most qualified candidates, which is challenging because candidates do not apply all at once. As such, having more phases in the process does not always lead to a better outcome, even if there is no costly setup in each phase.

The rest of the paper is organized as follows. In Section 2, we formulate the problem as a large-scale dynamic program. The determination of the cutoff in each phase, assuming the number of phases is given, and the determination of the number of phases for the recruitment process, are discussed in Sections 3 and Section 4, respectively. We conduct simulation studies in Section 5. We apply our modeling framework and heuristic policies to the recruitment process of a postgraduate business program and conduct counter-factual studies in Section 6. The paper is concluded in Section 7.

2 Formulations

The process of recruiting a total of d workers before time D can take up to T phases, which correspond to T application deadlines. The phases are of equal length, which is D/T. Let q_t

be the total number of candidates who have accepted their offers up to phase t. The value to the recruiter of hiring candidate i who applies in phase t is represented by an evaluation score, a positive random variable S_i^t which has a continuous density function f_t and cumulative distribution function F_t. The scores in the same phase are independent and identically distributed. Let s_i^t be the realization of S_i^t. Random variable δ_i^t represents whether candidate i accepts his/her offer. It is equal to 1 if he/she accepts the offer and zero otherwise. The random variable δ_i^t may be dependent of s_i^t. Each candidate accepts his/her offer with probability p_i^t. The candidates arrive stochastically over time according to a Poisson process with rate λ. Let the number of candidates in the t-th phase be n_t and its mean is given by $\lambda(D/T)$. Let $\boldsymbol{\delta}^t = (\delta_1^t, \delta_2^t, ..., \delta_{n_t}^t)$, $\mathbf{s}^t = (s_1^t, s_2^t, ..., s_{n_t}^t)$ and $\mathbf{S}^t = (S_1^t, S_2^t, ..., S_{n_t}^t)$. We assume that given the scores at the current phase, the scores in each future phase are independent of each other, and offer acceptance decisions in each future phase are also independent of each other. Our main results continue to hold when the phases have different lengths and the arrival process is nonstationary, provided that the arrival rate in each phase goes to infinity and the length of each phase is positive.

For a given T, the timing of events in each phase t is as follows. First, the recruiter evaluates each of the n_t candidates and the results of the evaluation is represented by $\mathbf{s}^t$. Second, the recruiter determines the number of offers to make. Third, offers are either accepted or rejected before the recruiter determines the number of offers to make again in the next phase. At the end of the recruiting season, there is an overage cost, c_o, and a marginal underage cost, u, if the total number of candidates recruited deviates from d. We assume that c_o is an increasing convex function and $c_o(0) = 0$. In each phase, the recruiter needs to decide how many candidates to give offers to. The success of the recruiting process is measured by both the *quantity* and *quality* of the candidates eventually hired. That is, the recruiter should keep the number of candidates recruited in the end as close to d as possible and at the same time, maximize the overall quality of the candidates recruited. We measure the quality of a group of candidates by the sum of their scores. The recruiter needs to determine whether each candidate i in phase t should be extended an offer to or not. The decision is represented by a binary vector $\mathbf{a}^t = (a_1^t, a_2^t, ..., a_{n_t}^t)$. In phase t, $\mathsf{E}[\sum_{i=1}^{n_t} s_i^t a_i^t I\{\delta_i^t = 1\}]$ represents the expected quality of recruited candidates and $\sum_{i=1}^{n_t} a_i^t I\{\delta_i^t = 1\}$ represents the number of candidates who have received offers and also accepted them.

For a given T, the recruiter's objective is to find a policy such that the total expected score of the candidates hired minus the penalty cost of deviating from the target d is maximized. Let $V_t'(q_t, \mathbf{s}^t)$ denote the optimal value function in phase t. The Bellman equation for the above problem is as follows:

$$V_t'(q_t, \mathbf{s}^t) = \max_{\mathbf{a}^t} J_t'(q_t, \mathbf{a}^t, \mathbf{s}^t),$$

$$J_t'(q_t, \mathbf{a}^t, \mathbf{s}^t) = \mathsf{E}\Big[\sum_{i=1}^{n_t} s_i^t a_i^t I\{\delta_i^t = 1\} + V_{t+1}'(q_t + \sum_{i=1}^{n_t} a_i^t I\{\delta_i^t = 1\}, \mathbf{S}^{t+1})\Big].$$

Here the expectation is with respect to $\boldsymbol{\delta}^t$ and $\mathbf{S}^{t+1}$ conditioning on $\mathbf{S}^t = \mathbf{s}^t$. At the end of the recruiting season, the terminal condition is given by

$$V_{T+1}'(q_{T+1}) = -c_o(q_{T+1} - d)^+ - u(d - q_{T+1})^+.$$

Here the overage and underage costs should be properly scaled so that a fine balance is struck between the quality and quantity of candidates.

In the above formulation, we have used a simple way to combine into a single objective, which we call the total *reward*, the desire to hire qualified candidates and the need to avoid a mismatch between the quantity of hires and the target. There are certainly many other ways to quantify the objective of the recruiter and which one is the most appropriate is likely to be context specific. In addition, the policy that offers are given to candidates from those with the highest score to those with the lowest score (i.e., a cutoff policy) is not optimal for the recruiter in general because acceptance yields may be score dependent[3]. The following proposition provides a sufficient condition for the cutoff policy to be optimal.

Proposition 1 *Suppose that δ_i^t equals 1 with probability p_i and 0 with probability $1 - p_i$. If $1 > \frac{p_j}{p_i} > \frac{s_i+u}{s_j+u}$ holds for any $s_j > s_i$, then there is a cutoff score such that a candidate is given an offer if and only if his or her score is higher than the cutoff.*

The condition in Proposition 1 means that the acceptance probability of a candidate with a higher score cannot be higher or much lower than that of a candidate with a lower score. Although not optimal in general, the cutoff policy is widely used in the practice of high-volume recruiting. Giving an offer to a less qualified candidate but not a more qualified candidate creates issues about fairness, which many recruiters would object. Furthermore, not following the order from the highest score to the lowest in making offers will make the process more complicated, and even impossible to implement in a high-volume environment. As such, in this study we focus on finding the best cutoff policy.

[3]Consider the following example where there is only one phase in the process and there are three candidates with scores 100, 80, and 50. We estimate that they accept their offers with probabilities 0.4, 0.5, and 0.8, respectively. Our recruitment target is one. The overage cost is very high so that making one offer is optimal. We can see that the three candidates yield the same expected score equal to 40. The optimal solution is to give the only offer to the candidate with the lowest score because he/she is most likely to accept his/her offer, which leads to the lowest underage cost.

For a given T, the problem of finding the best cutoff policy is formulated as

$$V_t(q_t, \mathbf{s}^t) = \max_{x_t \geq x_{\min}} J_t(q_t, x_t, \mathbf{s}^t), \tag{1}$$

$$J_t(q_t, x_t, \mathbf{s}^t) = \mathsf{E}\left[\sum_{i=1}^{n_t} s_i^t I\{s_i^t \geq x_t\} I\{\delta_i^t = 1\}\right] + \mathsf{E}\left[V_{t+1}(q_t + \sum_{i=1}^{n_t} I\{s_i^t \geq x_t\} I\{\delta_i^t = 1\}, \mathbf{S}^{t+1})\right].$$

Here the first expectation is with respect to δ_i^t, the second is with respect to δ_i^t and $\mathbf{S}^{t+1}$, both conditional on $\mathbf{S}^t = \mathbf{s}^t$, and $x_{\min}$ represents the minimum requirement for a candidate to be considered. To avoid trivial cases, we assume that $x_{\min} < o$. The terminal condition is similarly defined as $V_{T+1}(q_{T+1}) = -c_o(q_{T+1} - d)^+ - u(d - q_{T+1})^+$. Let $x_t^*(q_t, \mathbf{s}^t)$ be the smallest optimal solution of (1). The state space of the above dynamic program is large. The optimal policy cannot be directly computed.

In the beginning of the hiring season, the recruiter also needs to determine the optimal number of phases. That is,

$$V_0 = \max_{T \geq 1} J_0,$$

where $J_0 = \mathsf{E}V_1(0, \mathbf{S}^1)$. Essentially, in determining the optimal number of phases, the recruiter is optimizing over a class of dynamic programs with different horizons.

3 The Optimal Cutoff in Each Phase

For a given number of phases, how should we determine the cutoff in each phase? In this section, we first provide upper bound and lower bound formulations. We then show that the value function in the dynamic program formulation of the problem, as well as those in its lower and upper bounds all converge to the same limit when the volume is large enough. This motivates us to design four heuristic policies, two related to the lower bound and two to the asymptotic limit. They are all easily computable and their performance (total reward) all converges to the optimal performance.

3.1 A Lower Bound and an Upper Bound

To reduce the state space, we may ignore the score information when determining the cutoff in each phase and this gives a lower bound of the optimal value function. Let

$$G_t(x, \mathbf{S}^t, \boldsymbol{\delta}^t) = \frac{1}{n_t} \sum_{i=1}^{n_t} S_i^t I(S_i^t \geq x) I(\delta_i^t = 1)$$

and

$$H_t(x, \mathbf{S}^t, \boldsymbol{\delta}^t) = \frac{1}{n_t} \sum_{i=1}^{n_t} I(S_i^t \geq x) I(\delta_i^t = 1)$$

be the average score function and the incremental proportion of the candidates recruited at phase t, respectively. We consider the following approximation (or heuristic):

$$V_t^L(q_t) = \max_{x \geq x_{\min}} J_t^L(q_t, x), \tag{2}$$

$$J_t^L(q_t, x) = \mathsf{E}[n_t G_t(x, \mathbf{S}^t, \boldsymbol{\delta}^t) + V_{t+1}^L(q_t + n_t H_t(x, \mathbf{S}^t, \boldsymbol{\delta}^t))].$$

Here both expectations are with respect to $\mathbf{S}^t$ and $\boldsymbol{\delta}^t$. We use a continuous function, $p_t(s_i^t)$, to represent the conditional probability of offer acceptance. For instance, the conditional probability of acceptance can be modeled by a logistic regression as

$$p_t(s_i^t) = \frac{\exp(\beta_0^t + \beta_1^t s_i^t)}{1 + \exp(\beta_0^t + \beta_1^t s_i^t)}$$

for $i = 1, \ldots, n_t$, where β_0^t and β_1^t are the regression coefficients. Define the expectation of G_t and H_t respectively as

$$g_t(x) = \int_x^\infty s p_t(s) f_t(s) ds \quad \text{and} \quad h_t(x) = \int_x^\infty p_t(s) f_t(s) ds.$$

Notice that $h_t(x) = \Pr(S_i^t \geq x, \delta_i^t = 1)$ and the number of candidates who receive and accept their offers given n_t, $\sum_{i=1}^{n_t} I\{S_i^t \geq x\} I\{\delta_i^t = 1\} \mid n_t$, is a binomial random variable with success probability $h_t(x)$. As n_t is a Poisson random variable with mean $\lambda(D/T)$, $\sum_{i=1}^{n_t} I\{S_i^t \geq x\} I\{\delta_i^t = 1\}$ is also a Poisson random variable with rate $\lambda(D/T)h_t(x)$. As a result, the objective function in the lower bound formulation at phase t given a cutoff x can be explicitly written down as

$$J_t^L(q_t, x) = \lambda(D/T) g_t(x) + \sum_{k=0}^{\infty} V_{t+1}^L(q_t + k) \exp\{-\lambda(D/T) h_t(x)\} \frac{\{\lambda(D/T) h_t(x)\}^k}{k!}.$$

We define $x_t^L(q_t)$ as the smallest optimal solution of (2).

The gap between the performance of the lower bound and that of the optimal policy is difficult to compute because the optimal policy is unknown. To evaluate the performance of the lower bound, we consider the following information relaxation as an upper bound (Brown et al. 2010). The idea is that if the gap between the lower and upper bounds is small, then the lower bound must be close to the optimal reward and hence the heuristic is satisfactory.

Suppose that the scores of all candidates from phase t to the end of the planning horizon and their offer acceptance decisions are available at the beginning of phase t. Then the score vector is no longer needed as state variables in the dynamic program. Let $\mathbf{s}^{t:T} = (\mathbf{s}^t, \mathbf{s}^{t+1}, ..., \mathbf{s}^T)$, $\mathbf{S}^{t:T} = (\mathbf{S}^t, \mathbf{S}^{t+1}, ..., \mathbf{S}^T)$, and $\boldsymbol{\delta}^{t:T} = (\boldsymbol{\delta}^t, \boldsymbol{\delta}^{t+1}, ..., \boldsymbol{\delta}^T)$, where $\boldsymbol{\delta}^t = (\delta_1^t, \delta_2^t, ..., \delta_{n_t}^t)$. Then the dynamic programming formulation of the upper bound is

$$V_t^U(q_t, \mathbf{s}^{t:T}, \boldsymbol{\delta}^{t:T}) = \max_{x \geq x_{\min}} J_t^U(q_t, x, \mathbf{s}^{t:T}, \boldsymbol{\delta}^{t:T}), \tag{3}$$

where

$$J_t^U(q_t, x, \mathbf{s}^{t:T}, \boldsymbol{\delta}^{t:T}) = n_t G_t(x, \mathbf{s}^t, \boldsymbol{\delta}^t) + V_{t+1}^U(q_t + n_t H_t(x, \mathbf{s}^t, \boldsymbol{\delta}^t), \mathbf{s}^{(t+1):T}, \boldsymbol{\delta}^{(t+1):T}).$$

The score information set is not updated over time. The optimal policy at phase t is a function of the score and acceptance yield information from phase t onwards. Let $x_t^U(q_t, \mathbf{S}^{t:T}, \boldsymbol{\delta}^{t:T})$ be the smallest optimal cutoff in the above dynamic program. The following proposition shows the existence of optimal solutions in all the three formulations and the relations between their corresponding value functions.

Proposition 2

(i) In each of the formulations (1), (2) and (3), there exists at least one optimal solution.

(ii) $\mathsf{E}[V_t^U(q_t, \mathbf{S}^{t:T}, \boldsymbol{\delta}^{t:T})] \geq \mathsf{E}[V_t(q_t, \mathbf{S}^t)] \geq V_t^L(q_t)$.

To compute $\mathsf{E}[V_t^U(q_t, \mathbf{S}^{t:T}, \boldsymbol{\delta}^{t:T})]$, we first create a large number of samples. For each sample, we rank the candidates who accept their offers based on their scores in descending order and accept candidates greedily until the target d is met. For the candidates remaining, we accept one if and only if his or her score is greater than the marginal overage cost incurred. We then compute the total reward for each sample and the average reward over all samples. Because it requires no optimization, the computation is easy.

3.2 Convergence Properties

In this section, we show that the value function in (1), the value function in the lower bound formulation (2), and the value function in the upper bound formulation (3) all converge to the same function when the rate λ is large. In all cases, we can establish the convergence rates.

Let $J_t^0(q_t, x_t)$ and $V_t^0(q_t)$ be the objective function and value function of the following dynamic program:

$$V_t^0(q_t) = \max_{x_t \geq x_{\min}} J_t^0(q_t, x_t), \tag{4}$$

where

$$J_t^0(q_t, x_t) = \lambda(D/T) g_t(x_t) + V_{t+1}^0(q_t + \lambda(D/T) h_t(x_t)),$$

and $V_{T+1}^0(q_{T+1}) = V_{T+1}(q_{T+1})$. Let $x_t^0(q_t) = \arg\max_{x_t \geq x_{\min}} J_t^0(q_t, x_t)$. In the lower bound formulation (2), at each phase the number of offers that will be accepted is random. In the formulation (4), we replace that number by its mean.

The formulation (4) is a deterministic dynamic program and can be solved by standard backward induction or with Karush-Kuhn-Tucker (KKT) conditions. Its optimal policy has the following closed form expression.

Theorem 1 *Assume that $x_{\min} < c_o'(0)$. The optimal solutions of (4) are*

$$x_t^0 = ... = x_T^0 = \begin{cases} x_{\min}, \ \textit{when } q_t < d - \lambda(D/T)\sum_{i=t}^T h_i(x_{\min}), \\ \overline{x}_t^0, \ \textit{when } d - \lambda(D/T)\sum_{i=t}^T h_i(x_{\min}) \leq q_t \leq d - \sum_{i=t}^T \lambda(D/T) h_i(c_o'(0)), \\ \overline{\overline{x}}_t^0, \ \textit{when } q_t > d - \lambda(D/T)\sum_{i=t}^T h_i(c_o'(0)), \end{cases}$$

where $\overline{x}_t^0$ is the solution to equation $\lambda(D/T)\sum_{i=t}^T h_i(x) = d - q_t$ and $\overline{\overline{x}}_t^0$ is the solution to equation $x = c_o'(q_t + \lambda(D/T)\sum_{i=t}^T h_i(x) - d)$.

The condition $x_{\min} < c_o'(0)$ means that it is not worthy to recruit one more candidate who barely meets the minimum requirement when the total number of accepted candidates is already greater than d. In Theorem 1, the cutoffs can take three different values, and they correspond to the cases where underage, no mismatch, and overage occur, respectively. The optimal cutoff does not depend on the underage cost. The theorem also implies that (a) under the optimal solution of (4), the cutoff score is the same across phases, i.e., $x_t^0(q_t) = x_{t+1}^0(q_{t+1})$; (b) $x_t^0(q_t)$ is increasing in q_t; and hence (c) $x_t^0(q_t) = x_{t+1}^0(q_{t+1}) \geq x_{t+1}^0(q_t)$; (d) $x_t^0(q_t)$ may diverge to infinity. For example, let $d = \sqrt{\lambda}$, $q_t = 0$, $c_o(x) = ox + ax^b$ for some $o > x_{\min}, a > 0, b > 1$, and assume that the scores are normally distributed and the acceptance probability is independent of score. In this example, as $\lambda \to \infty$, the optimal cutoff diverges to infinity and the total number of hires in the end converges to the target (i.e., $q_{T+1}/d \to 1$).

When the score is identically distributed over time (i.e., $F_t = F$ for all t), and the acceptance probability is also identically distributed and independent of score (i.e., $p_t(s) = p$ for all t and s), $\overline{x}_t^0$ is simply given by

$$\overline{x}_t^0 = \overline{F}^{-1}(\frac{(d-q_t)T}{p(T-t+1)\lambda D})$$

for any $q_t \leq d$. Equivalently, the number of offers in phase t is given by

$$d_t = \frac{d - q_t}{p(T-t+1)}.$$

For example, if the process consists of three phases, then the number of offers to make in the first phase (with $q_1 = 0$) is $d/(3p)$. In other words, the number of offers to make in each phase is the unmet recruiting target divided by the number of remaining phases and then inflated by the acceptance yield rate. We call this simple policy a *linear inflation heuristic*.

In probability theory, a series of estimators X_n is stochastically bounded if for any $\varepsilon > 0$, there exists a finite $C > 0$ such that $\Pr(|X_n| > C) \leq \varepsilon$ for all n. Let $X_n = O_p(a_n)$ denote that X_n/a_n is stochastically bounded. Similarly, a series $x_n = O(a_n)$ if $x_n/a_n \leq C$ for some bounded positive constant C for all n. Denote $a_n \asymp b_n$ if $a_n = O(b_n)$ and $b_n = O(a_n)$. In what follows, we show that the value functions in the upper bound formulation, the original formulation and lower bound formulation converge to the same limit with different rates.

Theorem 2 *Assume that a)* $\lambda(D/T)\sum_{i=t}^{T} h_i(x_{\min}) > d - q_t$*; b)* $\sup_{t=1,\ldots,T} \mathsf{E}|S_i^t|^4 < \infty$*. Then as* $\lambda \to \infty$*, for* $0 \le q_t < d$*,*

$$\left|V_t^U(q_t, \mathbf{S}^{t:T}, \boldsymbol{\delta}^{t:T})/V_t^0(q_t) - 1\right| = O_p(\max\{u, \sup_{x \le (1+\epsilon)\lambda D} |c_o'(x)|\}\sqrt{\frac{\log(\lambda)}{\lambda}} \times \frac{\lambda}{d - q_t}),$$
$$\left|V_t^L(q_t)/V_t^0(q_t) - 1\right| = O(\max\{u, \sup_{x \le (1+\epsilon)\lambda D} |c_o'(x)|\}\frac{1}{\sqrt{\lambda}} \times \frac{\lambda}{d - q_t}),$$
$$\left|V_t(q_t, \mathbf{S}^t)/V_t^0(q_t) - 1\right| = O_p(\max\{u, \sup_{x \le (1+\epsilon)\lambda D} |c_o'(x)|\}\sqrt{\frac{\log(\lambda)}{\lambda}} \times \frac{\lambda}{d - q_t}),$$

for some $\epsilon > 0$*.*

The above condition a) means that the expected total number of candidates hired in the end is greater than the hiring target d if all candidates who meet the minimum requirement $x_{\min}$ are made offers to. This condition is met easily in practice and is needed to avoid the trivial situation when V_t^0 is zero.

In the upper bound and the original formulations, the fastest convergence rate is $\sqrt{\log(\lambda)/\lambda}$, and in the lower bound formulation, the fastest convergence rate is $1/\sqrt{\lambda}$. For all cases, the fastest convergence rates are achieved when $d - q_t \asymp \lambda$ and $\sup_{x \le (1+\epsilon)\lambda D} |c_o'(x)| < \infty$. The fastest convergence rates, however, cannot always be achieved. For example, when $d - q_t \asymp \lambda^\zeta$ for some $1/2 < \zeta < 1$ and $c_o(x) = ox + ax^b$ for some $a > 0, b \ge 1$, the convergence rates are of order $O_p(\sqrt{\log(\lambda)}/\lambda^{\zeta - 0.5 - (b-1)})$ for the upper bound formulation and the original formulation, which is slower than $\sqrt{\log(\lambda)/\lambda}$. The theorem also implies that in this example, the value functions for upper bound, lower bound and original formulation converge to the deterministic one if $\zeta - 0.5 - (b-1) > 0$.

In a revenue management problem where a firm dynamically controls the arrival rate through pricing to sell n units of inventory, Gallego and van Ryzin (1994) show that fixed price policies are asymptotically optimal and the convergence rate is $1/\sqrt{n}$. This convergence rate is similar to the fastest convergence rate for the lower bound formulation. As the value functions in the original formulation and the upper bound formulation depend on the score data $\mathbf{S}^t$ in the current phase, they converge more slowly than in the lower bound formulation. Also because the optimal policy in the asymptotic formulation (4) does not depend on the underage cost, and when the volume is large enough, the optimal policy in the original formulation converges to their corresponding asymptotic limits derived from (4), we can conclude that estimating the underage cost is not critically important in a high-volume environment.

3.3 Heuristic Policies

As we have shown in the previous section, V_t and V_t^L both converge to V_t^0 when the recruitment volume is large. We now have four heuristic policies with varying computational requirements: 1) Lower bound heuristic (LB). We use $x_t^L(q_t)$ as the cutoff score for the current phase, regardless of the scores in the current phase. 2) Asymptotic limit (AL). We use $x_t^0(q_t)$ as the cutoff score for the current phase, regardless of the scores in the current phase. 3) Improved lower bound heuristic (ILB). We use V_{t+1}^L to approximate the value-to-go function in (1); that is, the number of offers to make in phase t is given by

$$x_t^{ILB}(q_t, \mathbf{s}^t) = \arg\max_{x_t \geq x_{\min}} \mathsf{E}[n_t G_t(x, \mathbf{s}^t, \boldsymbol{\delta}^t) + V_{t+1}^L(q_t + n_t H_t(x, \mathbf{s}^t, \boldsymbol{\delta}^t))],$$

where both expectations are with respect to $\boldsymbol{\delta}^t$. 4) Improved asymptotic limit heuristic (IAL). We use V_{t+1}^0 to approximate the value-to-go function in (1); that is, the number of offers to make in phase t is given by

$$x_t^{IAL}(q_t, \mathbf{s}^t) = \arg\max_{x_t \geq x_{\min}} \mathsf{E}[n_t G_t(x, \mathbf{s}^t, \boldsymbol{\delta}^t) + V_{t+1}^0(q_t + n_t H_t(x, \mathbf{s}^t, \boldsymbol{\delta}^t))],$$

where both expectations are with respect to $\boldsymbol{\delta}^t$. Computationally, the best cutoffs for the ILB and IAL are obtained by the golden-section search after the approximation of the value-to-go functions (i.e., $V_j^L(q_j)$ and $V_j^0(q_j)$).

The last two approximations allow us to utilize the score information in the current phase without adding computational burden. They may give better outcomes than the first two heuristics. Let $R_t(\ell, q_t, S^t)$ be the total reward from phase t onward when heuristic ℓ is used, where $\ell = LB, ILB, AL$, and IAL. Then,

$$\begin{aligned} R_t(\ell, q_t, \mathbf{s}^t) \quad = \quad & \mathsf{E}\Big[n_t G_t(x_t^\ell, \mathbf{s}^t, \boldsymbol{\delta}^t) + \sum_{j=t+1}^{T} n_j G_j(x_j^\ell, \mathbf{s}^j, \boldsymbol{\delta}^j) \\ & + V_{T+1}(q_t + H_t(x_t^\ell, \mathbf{s}^t, \boldsymbol{\delta}^t) + \sum_{j=t+1}^{T} n_j H_j(x_j^\ell, \boldsymbol{S}^j, \boldsymbol{\delta}^j)) \Big]. \end{aligned}$$

The following theorem shows that all the four heuristics are asymptotically optimal with a known rate when the volume is large.

Theorem 3 *Assume that the conditions in Theorem 2 hold. When λ is large and $0 \leq q_t < d$,*

$$\left| \frac{R_t(\ell, q_t, \mathbf{S}^t)}{V_t(q_t, \mathbf{S}^t)} - 1 \right| = O_p(\max\{u, \sup_{x \leq (1+\epsilon)\lambda D} |c_o'(x)|\} \sqrt{\frac{\log(\lambda)}{\lambda}} \times \frac{\lambda}{d - q_t}),$$

for $\ell = LB, ILB, AL,$ and IAL, $t = 1, \cdots, T$.

According to Theorem 3, the total reward from using any of the four heuristic formulations converges to the total reward of the original formulation at the same rate as that appeared in Theorem 2. That is, the fastest rate is $\sqrt{\log(\lambda)/\lambda}$, which can be achieved, for example, when $d - q_t \asymp \lambda$ and $\sup_{x \leq (1+\epsilon)\lambda D} |c_o'(x)| < \infty$. For the same q_t, we can show that the total expected reward the recruiter can gain from using the lower bound heuristic for all t is simply $V_t^L(q_t)$, the maximal reward in the lower bound formulation. In other words, $V_t^L(q_t) = \mathsf{E}R_t(LB, q_t, \mathbf{S}^t)$.

4 The Optimal Number of Phases

A multi-phase process can help reduce time to hire and the risk stemming from uncertain offer acceptance, two major challenges facing high-volume recruiters. However, with a multi-phase process, the recruiter must determine to whom to extend offers and how many offers to extend when more qualified candidates may apply in the future. In this section, we investigate the impact of changing the number of phases in the process on the total expected reward.

In this section, we assume that the acceptance probability is independent of score and denote the probability as p. Let $S_{[i]}^{n_t}$ be the i^{th} highest score in phase t and $\delta_{[i]}^t$ the corresponding offer acceptance decision. Let d_t be the number of offers in phase t. We first rewrite the objective function at the beginning of the planning horizon in the following non-recursive form:

$$J_0 = r(T) - c(T),$$

where

$$r(T) = \mathsf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{d_t} S_{[i]}^{n_t} I\{\delta_{[i]}^t = 1\}\right],$$

and

$$c(T) = \mathsf{E}[c_o(q_{T+1} - d)^+ + u(d - q_{T+1})^+],$$

which represent the total expected score and penalty cost, respectively. A simple sample-path arguments can show that for a given policy on how the cutoffs (or d_t) are chosen, $r(T)$ is decreasing in T. How a change in T affects $c(T)$ is more involved. However, under some mild conditions, $c(T)$ has a very simple and intuitive expression.

Theorem 4 *Assume that the probability of acceptance is independent of the score and the offer acceptance decisions are independent across the phases. If the number of offers in phase t is set using the linear inflation heuristic and $c_o(x) = ox + ax^b$, then,*

$$\frac{c(T)}{(o+u)\sqrt{\frac{(1-p)d}{T}}\frac{1}{\sqrt{2\pi}} + a(\sqrt{\frac{(1-p)d}{T}})^b \frac{1}{\sqrt{2\pi}}(\sqrt{2})^{b-1}\Gamma(\frac{b+1}{2})} \longrightarrow 1$$

as $d \to \infty$, where $\Gamma(.)$ is the gamma function.

When d is large, $c(T)$ is decreasing in T. As a special case, when $a = 0$, $c(T)$ approaches a simple linear function of $(o+u)$, the sum of the marginal overage and underage costs. In Theorem 5, we discuss the asymptotic behavior of $r(T)$ and $c(T)$ when d and λ approach infinity at different rates.

Theorem 5 *Assume that the conditions in Theorem 4 hold. If scores follow a Normal distribution with mean μ and standard deviation σ and $d/\sqrt{\lambda} \to \infty$, then*

(i)

$$\frac{\sum_{t=1}^{T}\sum_{i=1}^{d_t} S_{[i]}^{n_t} I\{\delta_{[i]}^t = 1\}}{\mu p\lambda D/T \sum_{t=1}^{T} \frac{d_t}{n_t} + \sigma p\lambda D/T \sum_{t=1}^{T} \int_0^{\frac{d_t}{n_t}} \Phi^{-1}(1-z)dz} \longrightarrow 1,$$

in probability, where Φ is the standard Normal cumulative distribution function.

(ii) Suppose that $d = \gamma\lambda D$, where $0 < \gamma < p$. Then

$$\frac{r(T)}{d\mu + \sigma p\lambda D\Psi(\gamma/p)} \to 1, \text{ as } \lambda \to \infty,$$

where $\Psi(x) = \int_0^x \Phi^{-1}(1-y)dy$.

(iii) Suppose that $d = C\lambda^\gamma$ for any $1/2 < \gamma < 1$ and $C > 0$. Then

$$\frac{r(T)}{\sigma d\sqrt{2(1-\gamma)\log(\lambda)}} \to 1, \text{ as } \lambda \to \infty.$$

Furthermore, in both cases, $c(T)/r(T) \to 0$ as $\lambda \to \infty$.

Part (i) of Theorem 5 is related to the following approximation for a Normally distributed score (Royston 1982):

$$\mathsf{E}S_{[i]}^{n_t} \approx \mu + \sigma\Phi^{-1}\left(\frac{n_t + 1 - i - \alpha}{n_t + 1 - 2\alpha}\right).$$

For a general score distribution, Arnold and Groeneveld (1979) and Bertsimas et al. (2006) provide bounds on the expected value of order statistics. To make the statement more rigorous, we employ the uniform convergence results on the gap between sample quantile and its theoretical one under the condition that d is at least as large as $O(\sqrt{\lambda})$.

The proof of the convergence results in Theorem 4 and Theorem 5 relies on the observation that the number of offers over time, d_t, form a martingale. In both cases (ii) and (iii), when the candidate arrival rate and the hiring target approach infinity, $r(T)$ is asymptotically the same as a limit that is independent of T. In addition, the total reward predominantly comes from $r(T)$. One implication of this theorem is that when the volume is large, the total reward is not sensitive to the number of phases. This will be confirmed in the simulation study in Section 5.4 and in the case study in Section 6.

From the theorem, we can also see that when both d and λ approach infinity, but d does so in a slower rate than λ (i.e., part (iii)), the limit of $r(T)$ is independent of mean score. In such a high-volume environment, it is the quality variability, not the mean, that matters to the total score of accepted candidates. In addition, when a small number of top candidates is selected from a large pool of candidates, the quality of the selected candidates is similar. As such, the limit is also independent of acceptance probability.

Theorems 4 and 5 provide us interesting insight into the impact of the number of phases on the recruitment process. The objective function of the recruiter consists of two terms, $r(T)$ and $c(T)$. When T increases, both decrease. In determining the optimal T, the recruiter must balance the aggregate quality of the candidates hired and the mismatch cost. We can also see that $r(T)$ is submodular in (σ, T) and $c(T)$ in $(o+u, T)$. Therefore, when variability in score increases, a smaller T should be chosen so that more candidates are ranked in each phase. When the sum of the overage and underage costs increases, a large T should be chosen to reduce the risk of mismatch.

5 Simulation Studies

In this section, to numerically confirm the convergence results, we first test the gaps between the upper and lower bounds (both LB and ILB are lower bounds). We then test the performance of the four heuristic policies discussed in Section 3.3. The effect of having an increasing and convex overage cost on the total number of hires is then examined. Finally, with effective heuristic policies in hand, we investigate the impact of the number of phases on the outcome of recruitment, which is measured by the total reward. In this section, the sample size n_t is generated from Poisson distribution with mean $\lambda(D/T)$, for all $t = 1, 2, ..., T$.

5.1 Gaps between Bounds

We assume that the score follows a normal distribution and the offer acceptance probability depends on the score according to a logistic link. The parameters we have chosen are listed in Table 1. These parameters are estimated by using a dataset from a postgraduate business program that we have collaborated with. Other parameters are: $T = 3$, $D = 6$, $d = 60$, $u = 105$ and $c_o(x) = ox + ax^b$ with $o = 125$ and $a = 0$. For each phase, we use the normal and logistic models with the parameters in Table 1 to generate n_t pairs of scores and acceptance decisions, that is $\{(S_i^t, \delta_i^t), i = 1, \cdots, n_t, t = 1, 2, 3\}$. We then compute the optimal cutoffs and

t	μ_t	σ_t	β_0^t	β_1^t
1	118.4545	12.4154	-0.7882	0.0125
2	113.5373	10.0440	6.4463	-0.0496
3	116.1900	12.0830	-5.4837	0.0421

Table 1: *Mean and standard deviation of the score distribution in each phase, as well as the regression coefficient for the offer acceptance decision, estimated from academic year 2016-2017.*

the corresponding value functions under different formulations.

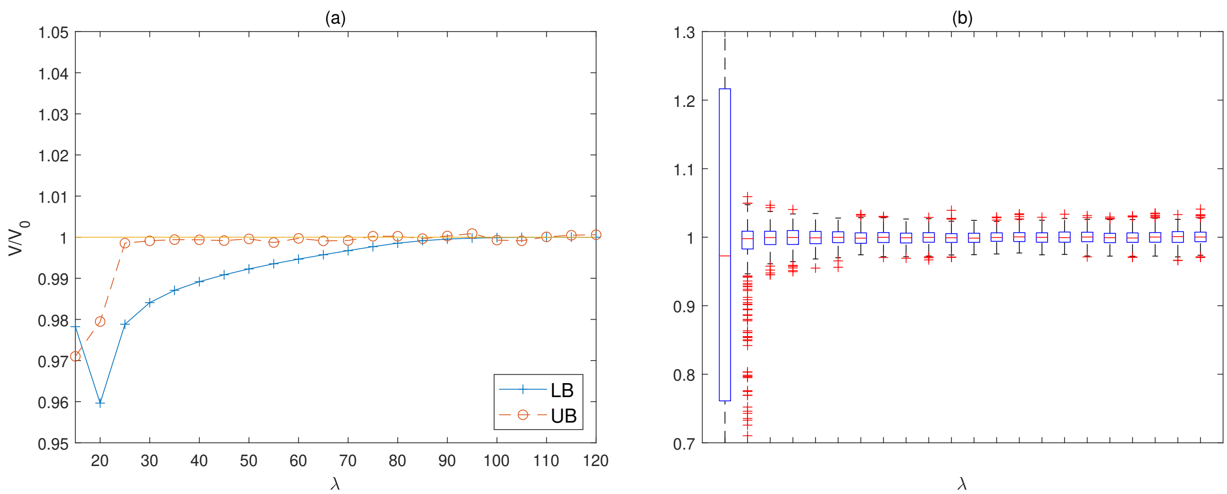

Figure 1: *The ratios of the value functions under lower bound and upper bound formulations to their asymptotic limit as functions of the arrival rate λ; we set $q_1 = 0, d = 60, o = 125$, $a = 0$, and $u = 105$; the results are based on* 500 *experimental replications.*

The results are summarized in Figure 1. In Figure 1(a), we report the value function in the lower bound formulation and the mean of the value function in the upper bound formulation when $\lambda = 15, 20, \cdots, 120$. In Figure 1(b), to show how the variance changes, we also report the value function in the upper bound formulation using box plots. In all cases, when λ increases from 15 to 120, the value functions move closer to the asymptotic limit V_t^0, and when $\lambda = 120$, they are very close to V_t^0, consistent with Theorem 2.

5.2 Performance of the Heuristics

We next evaluate the performance of the four heuristic policies. For each heuristic and a given value of q_t, we compute the cutoff, x_t^i, where $i = \text{AL}, \text{LB}$, IAL or ILB, and update q_{t+1} by $q_{t+1} = q_t + n_t H_t(x_t^i, \mathbf{S}^t, \boldsymbol{\delta}^t)$. Then the total expected reward when heuristic i is implemented is

$$\mathsf{E}[R_1(i, q_1, \mathbf{S}^1)] = \mathsf{E}\left[\sum_{t=1}^{T} n_t G_t(x_t^i, \mathbf{S}^t, \boldsymbol{\delta}^t) + V_{T+1}(q_{T+1})\right].$$

In the simulation, we keep the same setup as that in Section 5.1 and λ is either 50 or 15. We also vary the value of q_1, which is equivalent to varying the value of d, the recruiting target. For each heuristic, we create 500 random samples and compute the average reward. The computational results are summarized in Table 2. Both when $\lambda = 50$ and $\lambda = 15$, the performance of ILB is the best among all heuristics. This heuristic is also fairly close to the upper bound and therefore is obviously, even closer to the optimal solution. Also in both cases, IAL is better than AL, and ILB is better than LB; that is, using the score information in the current phase can improve performance. Finally, when λ increases from 15 to 50, the performance of all heuristics improves, percentage wise, which again confirms our earlier convergence results.

	Total reward				
	$\lambda = 50$				
q_1	AL	IAL	LB	ILB	UB
0	7565.2	7681.4	7687.3	7699.2	7742.7
6	6832.2	6954.5	6963.9	6969.8	7007
12	6142	6241.9	6247.8	6252.5	6280.1
18	5428.2	5510.3	5519.1	5521.4	5536.4
24	4726.9	4778.9	4785.6	4785.5	4793.1
30	4017.9	4038.6	4040.9	4039.5	4043.2
	$\lambda = 15$				
q_1	AL	IAL	LB	ILB	UB
0	4998.1	5006.1	5005.5	5007.6	5019.6
6	5242.6	5252.3	5251.4	5256.8	5293
12	5184.7	5266	5275.7	5287.9	5369.3
18	4772.7	4860.5	4850.9	4887.2	4982.1
24	4201.4	4281.5	4277.3	4307.2	4394
30	3553.9	3613.1	3628.2	3648.2	3718.4

Table 2: *Total reward when one of the four heuristic policies is implemented and the upper bound under different values of q_1 and λ and a low score variability.*

In Table 3, we increase the standard deviation of score in each phase to 30 while keeping other parameters the same as in Table 2. Since in the examples, only the candidates with higher than average scores are given offers, the recruiter benefits from a higher variability. However, the performance of the heuristics does not change substantially relative to the upper bound.

Finally, in Table 4, we compare the results when the underage cost u is set to 125 and 85. Our earlier theoretical results show that in a high-volume environment, the total reward obtained from using any of the four heuristics and the total reward in the upper bound formulation are

	Total reward				
	$\lambda = 50$				
q_1	AL	IAL	LB	ILB	UB
0	8933.8	8949.9	8932.1	8949.2	8963.2
6	8182.8	8192.4	8184.5	8192.4	8198.1
12	7448.5	7456.1	7454.1	7456.1	7459.3
18	6699.5	6705	6706.6	6704.9	6707.9
24	5947.2	5953.4	5955.8	5953.4	5956.2
30	5213.9	5220.3	5222.9	5220.2	5223.2
	$\lambda = 15$				
q_1	AL	IAL	LB	ILB	UB
0	4884.8	4901.6	4897.5	4901.5	4925
6	5137.3	5238.5	5217.1	5240.7	5304.2
12	4972.9	5307.1	5243.9	5344.9	5479.9
18	4603.3	4992.8	4896	5068	5224.9
24	4189.5	4533.5	4435.2	4593.4	4730.6
30	3799.2	3946.5	3907.6	3987.2	4086.6

Table 3: *Total reward from implementing each of the four heuristic policies and the upper bound under different values of* q_1 *and* λ *and a high score variability* ($\sigma_t = 30$)*.*

not sensitive to the underage cost. The simulation results here confirm the results.

5.3 The Effect of the Overage Cost

We evaluate the effect of having an increasing and convex overage cost on the total number of hires. The setup is the same as that in Section 5.1 except that $b = 1.1$ and $a = 0$ or 10. We compute the total number of hires under four heuristics and the upper bound formulation. The results are shown in Figure 2. In each case, when $a = 10$, the total number of hires only sightly exceeds the target 60 even when the arrival rate is as large as $\lambda = 150$. In contrast, when $a = 0$, the total number of hires increases linearly with the arrival rate. Therefore, when the arrival rate increases much faster than the target, choosing a linear overage cost will lead to a total number of hires exceeding the target substantially.

5.4 The Optimal Number of Phases

In this section, we test how the optimal number of recruiting phases varies with various model parameters. We keep $\lambda = 50$, $D = 6$, $a = 0$, and the underage cost $u = 150$ unchanged throughout. We assume that the score S is a normal random variable with mean and standard

q_1	Total reward $u = 125$ AL	IAL	LB	ILB	UB
0	7551.7	7679.4	7685.9	7698.1	7742.7
6	6818.7	6952.8	6963.7	6969.8	7007
12	6131.4	6241.3	6247.8	6253.7	6280.1
18	5419.9	5509.5	5519.2	5521.3	5536.4
24	4722.1	4779.7	4785.5	4785.4	4793.1
30	4016.6	4038.5	4040.9	4039.5	4043.2
	u=85				
q_1	AL	IAL	LB	ILB	UB
0	7578.6	7683.3	7689	7700.1	7742.7
6	6845.7	6955.5	6965.6	6970.8	7007
12	6152.6	6242.2	6249	6253.2	6280.1
18	5436.5	5511.3	5519.2	5521.7	5536.4
24	4731.7	4779.5	4785	4785.6	4793.1
30	4019.2	4037.8	4041	4039.5	4043.2

Table 4: *The total reward when one of the four heuristic policies is used and the upper bound for different values of q_1 and u (here $\lambda = 50$ and $o = 125$).*

deviation (μ, σ). We further assume that the acceptance probability is independent of score and let p denote the acceptance probability. We numerically investigate the following questions: How does the optimal number of recruiting phases depend on the mean quality of candidates μ, the variance of their quality σ^2, the acceptance probability p, the target number of hires d and the overage cost o?

We compute the number of offers in each phase by the linear inflation heuristic; that is, $d_t = \frac{d-q_t}{p(T-t+1)}$. The results are reported in Table 5. We have also used the lower bound formulation to compute the number of offers in each phase. The results are similar.

The numerical results show that the optimal number of phases decreases when the score of candidates becomes more variable, or when the overage cost decreases. These findings confirm the analysis in Section 4. Having fewer phases means that more candidates are evaluated and compared in each phase before offers are made, which is important when the quality of candidates is more variable. When the overage or underage cost decreases, the need for many phases to hedge against acceptance yield uncertainty is reduced.

In most of the cases we have tested, the optimal number of phases decreases when the acceptance probability increases. However, the optimal number of phases is not monotonic in

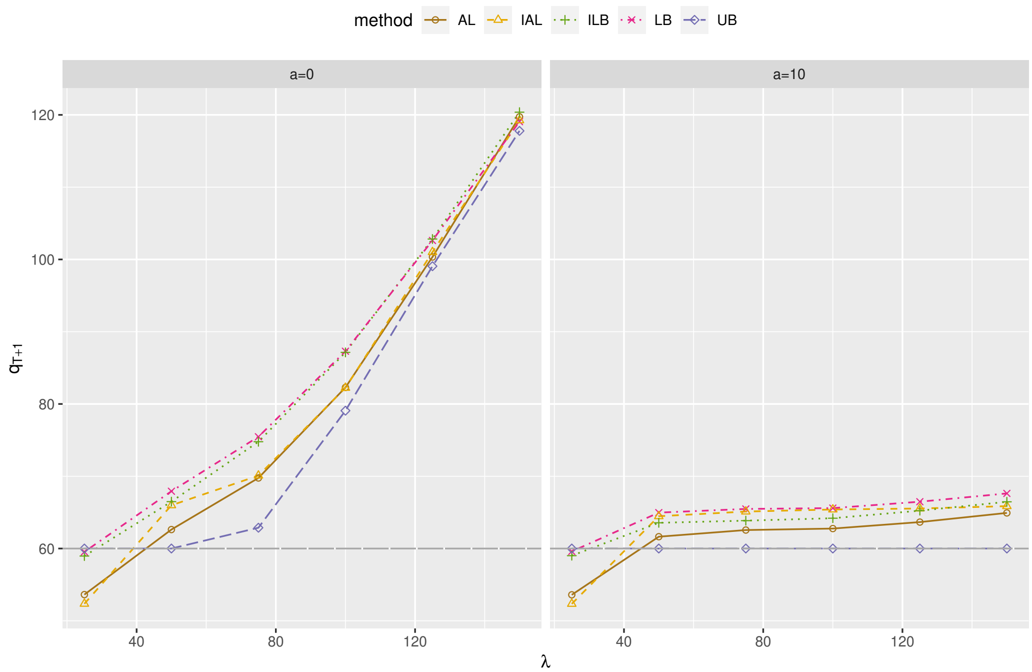

Figure 2: *The total number of hires under various methods as a function of the arrival rate λ; we set $q_1 = 0, d = 60, o = 125$, $u = 105$, $b = 1.1$; the left penal corresponds to the case when $a = 0$ while the right one to the case when $a = 10$; the results are based on 500 experimental replications.*

the target number of hires or the mean quality of candidates.

In Table 6, we show the total reward for different choices of T while keeping other model parameters fixed. For this particular instance, the optimal number of phases is 6. However, the total reward is not sensitive to the number of phases. The reduction in total reward is less than 0.8% if T deviates from the optimal by up to 3 phases. This confirms our theoretical findings in Section 4.

6 Applications

In this section, we apply our modeling framework to the admission process of a postgraduate program that we have collaborated with. The program was first launched in 2013 and they have since adopted a sequential recruitment process similar to what we described earlier. The online application system is open for a duration of six months (i.e., $D = 6$) every year and there are three application deadlines (i.e., $T = 3$). The enrollment target, d, has been set to 60 since 2014. It is considered undesirable if the actual enrollment deviates substantially from the target. If it is below the target, valuable resources are wasted. But if it is above the target, scheduling classes and other activities could prove difficult and learning experience

			$\mu = 150$			$\mu = 200$		
			$\sigma = 10$	$\sigma = 25$	$\sigma = 40$	$\sigma = 10$	$\sigma = 25$	$\sigma = 40$
$d = 60$	$p = 0.5$	$o = 150$	16	10	7	15	11	9
		$o = 200$	18	11	9	17	11	10
		$o = 250$	19	12	9	20	12	10
	$p = 0.9$	$o = 150$	6	6	6	6	6	6
		$o = 200$	6	6	6	6	6	6
		$o = 250$	7	6	6	6	6	6
$d = 90$	$p = 0.5$	$o = 150$	9	9	8	9	9	7
		$o = 200$	14	9	8	11	9	9
		$o = 250$	15	9	9	12	10	9
	$p = 0.9$	$o = 150$	9	9	5	9	9	5
		$o = 200$	10	9	5	10	9	9
		$o = 250$	10	9	9	10	10	9

Table 5: The optimal number of recruiting phases

T	1	2	3	4	5	6	7	8	9
Total reward	9379	9514	9568	9543	9581	**9621**	9615	9592	9572

Table 6: *Total reward for different choices of T under the linear inflation heuristic ($d = 60$, $p = 0.9$, $\sigma = 10$, $\mu = 150$, $o = 200$, $a = 0$).*

could be negatively impacted. In each phase, candidates are first evaluated based on their academic background, work experience, and standard test results and each is given a screening score based on a formula. The shortlisted candidates are then interviewed by a panel. Each shortlisted candidate's total score is the sum of his/her screening score and the interview scores.

Over the years, the program has adopted a cut-off policy in issuing offers, although the cutoff values as well as the number of phases were decided without much deliberation. However, there were one or two exceptions where a candidate with a lower score was given an offer while one with a higher score was not in the same phase. According to the program administrators, although they have fine-tuned the scoring scheme over the years, it remains impossible to consider all relevant factors. For example, they believe that diversity in student body can enhance learning experience, but diversity has not been explicitly incorporated in the scoring scheme. In addition, they recognize that the scores are positively correlated with, but not perfect indicators of, academic and career success, the ultimate goal of the program. They therefore adopt an "exploitation versus exploration" approach by paying special attention to candidates with unusual background and experience. Finally, currently the total score is the sum of the screening and interview scores. However, these two scores are not viewed equally. In particular, the candidates who receive a very low interview score would not be given offers even if their total scores are high. In spite of these complications, the number of exceptions is very small. It is their goal to further fine-tune the scoring scheme and minimize the need for human intervention.

To put our model into practice, we need to determine the appropriate overage and underage costs. There are three parameters to determine in the overage cost function. As we discussed earlier, on the one hand, b has to be strictly greater than 1; otherwise, the total number of hires will exceed the target substantially. On the other hand, the larger the value of b, the slower the convergence rate. To have a good balance, we choose $b = 1.1$. For o and a, we ask the following two questions. Suppose that we have two classes of students, one's class size is d and the other's $d + 1$. In order for the two classes to be considered equivalent, how much higher does the total score from the larger class have to be? Then we change the class sizes to $d + 1$ and $d + 2$, respectively, and ask the question again. The answers to the two questions give us two equations, which we can use to find o and a. Similarly, for the underage cost, suppose that we have two classes of students, one's class size is d and the other's $d - 1$. In order for the two classes to be considered equivalent, how much higher does the total score from the smaller class have to be? The answer is the underage cost.

Our retrospective study proceeds in four steps. First, we derive the population models from the data for the academic year 2016-2017. The parameters are presented in Table 1. The arrival rate is set as the total number of arrivals in that year divided by the length of the recruiting process D. We plot the quantile-quantile (Q-Q) plot for the scores in Figure 3 and confirm that they follow a normal distribution. From the data for the same academic year and based on our discussion with the program administrators, we also estimate the overage and underage costs to be $o = 125$, which is slightly higher than the average score of all the candidates who were given offers to, and $u = 105$. Other parameters are $a = 10$ and $b = 1.1$. We assume that the score distribution, underage and overage costs, acceptance yield rate, and the arrival rate in the 2017-2018 academic year are the same as those in the 2016-2017 academic year. Second, we then use ILB, the best heuristic among the four, to compute the number of offers that should be given in each of the three phases in the 2017-2018 academic year. The scores in this year are indeed normally distributed, according to Figure 4. In Table 7, we report the actual cutoffs, the numbers of offers made, the numbers of offers accepted, and the total reward in the 2017-2018 academic year and their corresponding quantities if ILB is implemented[4]. Third,

[4]Because acceptance yields are random, under ILB, the numbers of offers accepted and the total reward are random. The numbers reported in the table are their expected values.

we examine how the total reward changes if we vary the number of phases, one question the program administrators were always curious about but unable to answer because they did not have the tools. Finally, we repeat the above for the 2018-2019 academic year by using the score distributions and the number of candidates in each phase in both 2016-2017, 2017-2018 academic years.

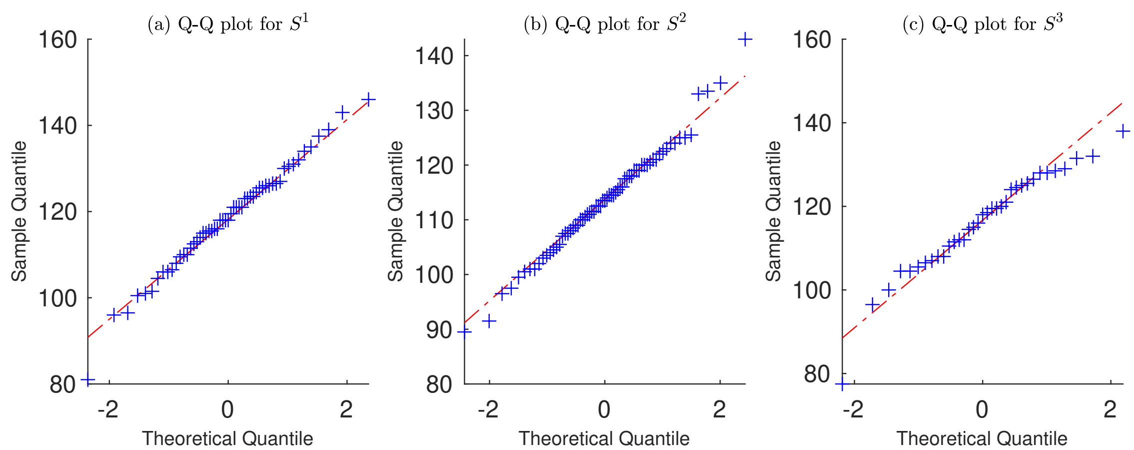

Figure 3: *Quantile-quantile plot for the score S_i^t in each phase from academic year 2016-2017. Here $n_1 = 55$, $n_2 = 67$, and $n_3 = 35$.*

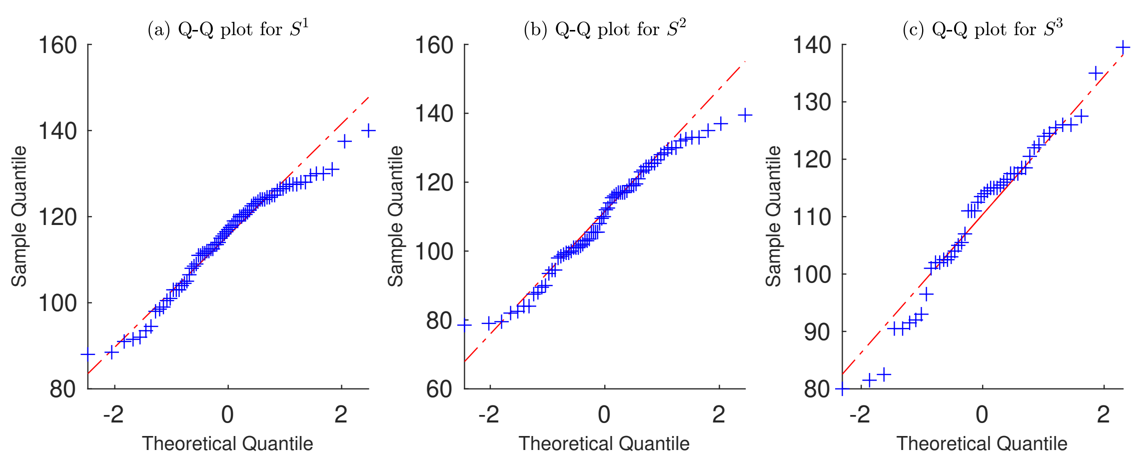

Figure 4: *Quantile-quantile plot for the score S_i^t in each phase from academic year 2017-2018. Here $n_1 = 75$, $n_2 = 70$, and $n_3 = 48$.*

Because the number of offers made in each phase is small, to estimate the offer acceptance probability, we aggregate all the data from the three phases. We run the logistic model by regressing the offer acceptance indicator on the score. The p-value of the slope is as high as 0.94484. We also substitute the logistic link function by *probit*, *comploglog*, and *loglog* and the outcomes are insignificant. To exclude the possibility of any nonlinear effects on the score, we expand the design matrix by including up to the fifth order of polynomials, and run the logistic regression again, but all of the p-values are insignificant[5]. We conclude that the offer acceptance probability does not depend on the score in this particular case. We therefore model this probability as a constant, which is equal to 0.60241. When the acceptance probability is independent of scores, the value function for ILB can be calculated exactly because the number of offers accepted given the scores and the cutoff is a binomial variable.

Because of the small number of exceptions in which a strict cut-off policy was not followed, to ensure a fair comparison between the actual outcome and what our model recommends, we need to carefully define the cutoffs that were actually set. In the first phase of the admission exercise in year 2017-2018, for example, out of $n_1 = 75$ candidates, 48 offers were made and 23 of them were accepted. We set the cutoff to the 48^{th} highest score, which is 112. Because of one exception, the lowest score among the 48 candidates who received offers is actually 111, which is slightly lower than 112. We do the same for other phases. Based on our model and ILB, the cutoff score should have been set to 111.5, or 51 offers should have been given, resulting in 24.4 expected accepted offers. The results for the whole recruiting season are summarized in Table 7. Overall, by using our modeling framework, we can raise the total reward from 6850.5 to 7224.8, an increase of about 5.5%. Interestingly, the cutoffs in three phases which were set by the program administrators are all higher than what our model suggests. The biggest difference is in the third phase when just 24 offers were made when our model suggests 28.9.

One technical point is worth mentioning. In the third phase, because the program administrators set a higher cutoff than that suggested by the model, some candidates who actually qualified for an offer according to our model did not receive one. For these candidates, we don't know whether they would have accepted their offers or not, had they been extended one. We assume that they would have accepted their offers with probability 0.60714, the average yield rate in year 2016-2017.

To test the robustness of ILB, we conduct another retrospective study for the 2018-2019 academic year. Q-Q plots, which are not reported here, confirm that the scores are approximately normally distributed. We estimate the score distribution and acceptance probability for the model by aggregating the data for the academic years 2016-2017 and 2017-2018. The parameters for the academic year 2017-2018 are summarized in Table 8. We compare the model's

[5]We originally expected that those with high scores might have lower acceptance probabilities because they might have more outside options. However, there are other forces at play, according to the program administrators. First, the recruiting team have probably made more effort on the top candidates in persuading them to accept their offers than on the average candidates. Second, the top candidates tend to apply for fewer universities and have better ideas about what they want.

	ILB			Actual			
t	Cutoff	No. of Offers	Accepted	n_t	Cutoff	No. of Offers	Accepted
1	111.5	51	24.4	75	112	48	23
2	108.7	37.5	21.3	70	109.5	37	21
3	109.6	28.9	17.1	48	114	24	14
Total		117.4	62.8	193		109	58
Reward			7224.8				6850.5

Table 7: Comparison of ILB with the actual practice for academic year 2017-2018.

results and the actual results in Table 9. According to ILB, the program administrators should have made more offers in the first two phases and about seven less in the last phase, which leads to an increase of about 1.1% in the total reward.

t	n_t	μ_t	σ_t	p
1	75	115.03	12.066	0.53211
2	70	109.96	16.217	0.53211
3	48	110.32	13.642	0.53211

Table 8: *Mean and standard deviation of the score distribution in each phase, as well as probability of offer acceptance, estimated from academic year 2017-2018.*

	ILB			Actual			
t	Cutoff	No. of Offers	Accepted	n_t	Cutoff	No. of Offers	Accepted
1	115	60	36.4	103	118	48	30
2	120.1	34.1	20.9	83	122	29	18
3	129.4	8.8	6.3	30	124.5	16	11
Total		102.9	63.6	216		93	59
Reward			7437				7356.5

Table 9: The comparison of the ILB heuristic with the actual practice for academic year 2018-2019.

The recruiting process of the program currently consists of three phases. The number was set years ago without much deliberation. The program administrators have been wondering whether this is the right choice and what would happen if the number is changed. In a past year, the yields were unusually low and the actual enrollment was only 51, significantly lower than the target of 60. The administrators were contemplating increasing the number of phases to four and were particularly interested in the potential impact. We compute the total reward by using LB for various choices of number of phases T. We assume that the score is normally distributed with mean $\mu_t = 114.97$ and standard deviation $\sigma_t = 10.96$, and offer acceptance probability is $p_t = 0.60241$ for all phases. These parameters are all estimated from the combined data in the year 2016-2017. Other parameters are $d = 60$, $o = 125, a = 10, b = 1.1, u = 105$ and $\lambda D = 158$. We then use the data to run LB. Table 10 shows the total rewards for various choices of T.

T	1	2	3	4	5	6	7	8
$V_1^L(0)$	6775	6930	6972	7004	7010	7011	**7015**	7007
T	9	10	11	12				
$V_1^L(0)$	6996	6990	6986	6980				

Table 10: *Total reward for different choices of T under LB.*

From the table, we can see that the maximum reward is achieved at $T = 7$, which is much greater than 3, the number of phases in the current practice. Over the years, the number of accepted candidates has varied, and at times deviated substantially from the target 60. The administrators have been thinking about increasing the number of phases to reduce the mismatch. Our analysis confirms that is the right thing to do. However, the numerical studies also show that the total reward is not sensitive to T, which confirms the theoretical results in Theorem 5. Given our findings, the program administrators are interested in the idea of increasing T to 4. They are unlikely to increase it further, although the optimal T is 7 according to our model. Given that the difference in reward of choosing a different T is small, other factors such as fixed costs associated with each phase, competition in the market, and impact on the candidates' behavior, which have not been modeled, will take precedence. Without our rigorous analysis, this level of clarity is next to impossible.

In our case study, the hiring target is in the dozens and the number of total applicants in the hundreds. In practice, there is no exact number that constitutes high volume and there are situations where the volume is in the ten-thousands (Hayton 2018 and Min 2019). Our results show that for a problem whose size is similar to that of the case study, the approximation approach we develop already works quite well. As the approximation converges to the optimal solution when the volume is large, it works even better for larger problems. In addition, when the volume goes up, it is harder for recruiters to manage the process relying only on their gut feeling and experience and there is a greater need for analytics.

7 Discussion and Concluding Remarks

It has been well recognized in the industries that high-volume recruitment is complex and data is crucial. In this study, we have formulated a high-volume recruitment problem as a large scale dynamic program. Our model captures three important features of high-volume recruiting: multiple phases, random yields, and a preset hiring target. We have provided a decision tool to answer practical questions about how many offers should be made in each phase of a recruitment season, and how many phases the season should have. To solve the dynamic program, we rely on approximations and the approximations are asymptotically optimal when the volume is large. Our simulation studies confirm the convergence results. We illustrate how our modeling framework can be put into practice in a case study, which shows that our decision tool can improve the recruiting outcome.

In our earlier analysis, we assumed that the value to the recruiter of hiring a candidate is perfectly measured by the candidate's assessment score. An extension of this is the case where the score is positively correlated with the value, but not perfectly so. In that case, we would first need to estimate the relationship between value and score and use the relationship to predict the value. We would then replace the sum of scores in the dynamic programing formulation with the sum of predicted values. We have used the total reward, which is the total assessment score of the candidates hired minus the penalty cost of the number of candidates hired deviating from the preset target, to measure the outcome of the recruiting process. There are other ways to measure performance. For example, in Kao and Rowan (1959), assessment is imperfect. Hiring each candidate incurs a cost and each bad hire incurs another cost. The recruiter determines a strategy regarding the number of candidates to be recruited and the cutoff score, which will yield a minimum cost subject to a given probability that at least a fixed number of good employees will be hired.

Another assumption we have made is that there is no recall - the recruiter does not go back to previous phases to try to accept candidates who were previously not accepted (or put on a waiting list). This assumption is valid in our case study because time-to-hire is an important performance metric there. In addition, the market is very competitive and qualified candidates won't be available unless they are notified in the same phase they submit their applications. A more general model should consider recall and each candidate's offer acceptance probability is lower if the offer is made in later phases. Interestingly, with recall, our earlier convergence results continue to hold true. They hold true no matter how likely the candidates on waiting list will accept offers later extended to them.

Our upper bound formulation corresponds to the case when we have full information in the beginning about all candidates' scores and their acceptance decisions, and it continues to be an upper bound when the recruiter has the option of recall, irrespective of how likely the candidates on waiting list will accept offers later extended to them. At the same time, having the option of recall can only benefit the recruiter. In other words, our current lower bound for the case without recall is also a lower bound for the case with recall. Because the upper bound and lower bound converge to the same limit asymptotically, the formulation with the option of recall will also converge to the same limit. This has the interesting implication that the option value of recall is negligible in a high volume environment.

One important application of the classical secretary problem is dynamic auctions for revenue management (see, for example, Vulcano et al. 2002). Although our study is in a different business context, the model in our study is similar to the model in Vulcano et al. (2002). In their model, the score vector in each phase represents bid prices and the hiring target represents the number of items to auction. Like in our study, because of the curse of dimensionality, computing the optimal number of items to auction in each phase is challenging. Since they do not consider random yields, there is no mismatch in quantity and their model is a special case of ours. Our heuristics can be directly applied to and the asymptotic properties continue to hold in their setting. Our central idea in terms of approximation is that we do not need to keep track of a large number of state variables when the reward function and the state transition function have a simple additively separable form. We believe that this angle is new to the literature of approximate dynamic programming and that our approach may have applications elsewhere. We leave that to future research.

References

Anily S, Beja A, Mendel A (2002) Optimal lot sizes with geometric production yield and rigid demand. *Operations Research* 50(3):424-432.

Arnold BC, Groeneveld RA (1979) Bounds on expectations of linear systematic statistics based on dependent samples. *The Annals of Statistics* 7(1):220-223.

Barr DR, Sherrill ET (1999) Mean and variance of truncated normal distributions. *The American Statistician* 53(4):357-361.

Bertsimas D, Natarajan K, Teo CP (2006) Tight bounds on expected order statistics. *Probability in the Engineering and Informational Sciences* 20(4):667-686.

Boudreau J, Hopp W, McClain J, Thomas L (2003) On the interface between operations and human resources management. *Manufacturing & Service Operations Management* 5(3):179-202.

Brown DB, Smith JE, Sun P (2010) Information relaxations and duality in stochastic dynamic programs. *Operations Research* 58(4): 785-801.

Cronbach LJ, Gleser GC (1965) *Psychological Tests and Personnel Decisions* (University of Illinois Press, Urbana, Ill.).

De Corte W (1998) Estimating and maximizing the utility of sequential selection decisions with a probationary period. *British Journal of Mathematical and Statistical Psychology* 51(1):101-121.

De Corte W, Lievens F, Sackett PR (2006) Predicting adverse impact and mean criterion performance in multistage selection. *Journal of Applied Psychology* 91(3):523-537.

DeGroot MH (2004) *Optimal Statistical Decision.* Wiley Interscience. Hoboken, New Jersey.

Du L, Li Q (2020) A Data-driven approach to high-volume recruitment: Application to student admission. *Manufacturing & Service Operations Management* 22(5):942-957.

Freeman PR (1983) The secretary problem and its extensions: A review. *International Statistical Review* 51(2):189-206.

Gallego G, van Ryzin G (1994) Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Science* 40(8):999-1020.

Gallego G, Wang R (2020) On the soft k-prophet problem. Working paper, Hong Kong University of Science and Technology, HKSAR.

Goldberg DA, Katz-Rogozhnikov DA, Lu Y, Sharma M, Squillane MS (2016) Asymptotic optimality of constant-order policies for lost sales inventory models with large lead times. *Mathematics of Operations Research* 41(3):898-913.

Grosfeld-Nir A, Gerchak Y, He Q (2000) Manufacturing to order with random yield and costly inspection. *Operations Research* 48(5):761-767.

Hayton E (2018) 5 ways technology can help with high volume recruiting. Accessed July 21, 2020, https://harver.com/blog/high-volume-recruiting/.

Henig M, Gerchak Y (1990) The structure of periodic review policies in the presence of random yield. *Operations Research* 38(4):634-643.

Kao, RC, Rowan TC (1959) A model for personnel recruiting and selection. *Management Science* 5(2):192-203.

Li Q, Xu H, Zheng S (2008) Periodic-review inventory systems with random yield and demand: bounds and heuristics. *IIE Transactions* 40(4): 434-444.

Li Q, Zheng S (2006) Joint inventory replenishment and pricing control for systems with uncertain yield and demand. *Operations Research* 54(4): 696-705.

Min J (2019) 4 Innovative Strategies For High Volume Hiring. Accessed July 21, 2020, https://ideal.com/high-volume-hiring/.

Moore J (2018) A Modern Take on Data-Driven Recruiting from a Google Recruiting Leader. Accessed July 21, 2020, https://hire.google.com/articles/a-modern-take-on-data-driven-recruiting/.

Powell WB (2007) *Approximate Dynamic Programming* (John Wiley and Sons, New York).

Royston JP (1982) Expected normal order statistics (exact and approximate): algorithm AS 177. *Applied Statistics* 31(2):161-165.

Shao J (2003) *Mathematical Statistics* (Springer Science & Business Media).

Vulcano G, van Ryzin G, Maglaras C (2002) Optimal Dynamic Auctions for Revenue Management. *Management Science* 48(11):1388-1407.

Xin L, Goldberg DA (2016) Optimality gap of constant-order policies decays exponentially in the lead time for lost sales models. *Operations Research* 64(6):1556-1565.

Yano C, Lee HL (1995) Lot sizing with random yields: a review. *Operations Research* 43(2):311-334.

Appendix

Proof of Proposition 1: Let $\overline{a}_i^t$ denote the optimal decision for candidate i. The proof is by contradiction. Suppose that there exist i and j such that $\overline{a}_i^t = 1$, $\overline{a}_j^t = 0$ and $s_j > s_i$. Then the revenue-to-go under the optimal solution is given by (To simplify notation, we omit the score state vector in V'_{t+1}.)

$$\sum_{k\neq i,j} s_k p_k \overline{a}_k^t + s_i p_i + p_i \mathsf{E}\Big[V'_{t+1}(q+1+\sum_{k\neq i,j} \overline{a}_k^t I\{\delta_k^t = 1\})\Big] + (1-p_i)\mathsf{E}\Big[V'_{t+1}(q+\sum_{k\neq i,j} \overline{a}_k^t I\{\delta_k^t = 1\})\Big],$$

which is greater than the revenue-to-go if we swap the offer decisions for candidates i and j,

$$\sum_{k\neq i,j} s_k p_k \overline{a}_k^t + s_j p_j + p_j \mathsf{E}\Big[V'_{t+1}(q+1+\sum_{k\neq i,j} \overline{a}_k^t I\{\delta_k^t = 1\})\Big] + (1-p_j)\mathsf{E}\Big[V'_{t+1}(q+\sum_{k\neq i,j} \overline{a}_k^t I\{\delta_k^t = 1\})\Big].$$

This means

$$p_i(s_i + \Delta) > p_j(s_j + \Delta), \tag{A.1}$$

where $\Delta = \mathsf{E}\Big[V'_{t+1}(q+1+\sum_{k\neq i,j} \overline{a}_k^t I\{\delta_k^t = 1\}) - V'_{t+1}(q+\sum_{k\neq i,j} \overline{a}_k^t I\{\delta_k^t = 1\})\Big]$. Here, Δ measures the expected marginal value of having an additional position filled. At state $q + \sum_{k\neq i,j} \overline{a}_k^t I\{\delta_k^t = 1\}$, if we always follow the optimal policy under the state $q + 1 + \sum_{k\neq i,j} \overline{a}_k^t I\{\delta_k^t = 1\}$, then for every sample path of candidates' scores and acceptance decisions, the total scores under the two states are the same, so the marginal value is upper bounded by the reduction in mismatch cost at the end of the recruiting season, which can not be higher than u. Rearranging the terms in (A.1) and noting that we have assumed $p_j < p_i$, then

$$\Delta > \frac{p_j s_j - p_i s_i}{p_i - p_j}.$$

However, we know $\Delta < u$ and hence the above inequality is a contradiction with the assumption $\frac{p_j}{p_i} > \frac{s_i+u}{s_j+u}$. ■

Proof of Proposition 2: (i) Formulation (1): $J_t(q_t, x_t, \mathbf{s}^t)$ is upper bounded by the total expected score of all candidates from phase t to phase T, and the total expected score is given by:

$$\sum_{i=1}^{n_t} s_i^t + \sum_{j=t+1}^{T} \lambda \frac{D}{T} \mathsf{E}[S_i^j],$$

which is bounded. Furthermore, $J_t(q_t, x_t, \mathbf{s}^t)$ is a step function of x_t with at most n_t jumps. Therefore, formulation (1) has at least one optimal solution.

Formulation (2): The conclusion holds if we can show that the objective function in the lower bound formulation is bounded and continuous of x_t. The objective function in the lower bound formulation is upper bounded by $\mathsf{E}[V_t(q_t, \mathbf{S}^t)]$, which is finite. To show that it is also continuous, we first explicitly write down the objective function as

$$J_t^L(q_t, x) = \lambda(D/T)g_t(x) + \sum_{k=0}^{\infty} V_{t+1}^L(q_t + k)\exp\{-\lambda(D/T)h_t(x)\}\frac{\{\lambda(D/T)h_t(x)\}^k}{k!}.$$

The second term on the right-hand side is equal to $\mathsf{E}[B(n_t, x)]$, where

$$B(n_t, x) = \sum_{i=1}^{n_t} \binom{n_t}{k} h_t(x)^k \{1 - h_t(x)\}^{n_t - k} V_{t+1}^L(q_t + k)$$

and n_t is sampled from a Poisson distribution with rate λ. The expectation of $B(n_t, x)$ can be decomposed into two parts:

$$\sum_{n_t=0}^{N} B(n_t, x)\exp(-\lambda)\frac{\lambda^{n_t}}{n_t!} + \sum_{n_t=N+1}^{\infty} B(n_t, x)\exp(-\lambda)\frac{\lambda^{n_t}}{n_t!} \tag{A.2}$$

Because $V_t^L(q_t)$ is upper bounded by $\mathsf{E}[V_t(q_t, \mathbf{S}^t)]$ and $V_t(q_t, \mathbf{S}^t)$ is bounded, $V_t^L(q_t)$ is bounded. Therefore, we have $B(n_t, x) \le C$ for some $C > 0$, and the second summation in (A.2) is upper bounded by

$$C\sum_{n_t=N+1}^{\infty} \exp(-\lambda)\frac{\lambda^{n_t}}{n_t!} \le C\frac{\lambda}{(N+1)}\sum_{n_t=N+2}^{\infty} \exp(-\lambda)\frac{\lambda^{n_t-1}}{(n_t-1)!} \le C\frac{\lambda}{N+1}.$$

Note that for a given n_t, $B(n_t, x)$ is continuous in x. Therefore, for any $\epsilon > 0$, there exists a $\delta > 0$ such that when $|x - x_0| \le \delta$, we have $|B(n_t, x) - B(n_t, x_0)| \le \epsilon$, for $n_t = 0, \cdots, N$. For $\mathsf{E}[B(n_t, x)]$, we have

$$\Big|\mathsf{E}[B(n_t, x)] - \mathsf{E}[B(n_t, x_0)]\Big| \le \epsilon + C\frac{\lambda}{N+1}$$

Let's choose $N + 1 = \frac{1}{\epsilon}$. Then the right hand side is simply $(1 + C\lambda)\epsilon \to 0$.

Formulation (3): It is easy to show that $J_t^U(q_t, x_t, \mathbf{S}^{t:T}, \boldsymbol{\delta}^{t:T})$ is upper bounded by $\sum_{j=t}^{T}\sum_{i}^{n_t} s_i^t$. Furthermore, it is a step function of x_t with at most $\sum_{i=t}^{T} n_i$ jumps. Therefore, there exists at least one optimal solution.

(ii) The proof is standard and hence omitted. ■

Proof of Theorem 1: Let $w = \lambda(D/T)$. The dynamic program (4) can also be formulated in the following non-recursive form:

$$\max_{x_t, \ldots, x_T \ge x_{\min}} \sum_{i=t}^{T} w g_i(x_i) - c_o(q_t + \sum_{i=t}^{T} w h_i(x_i) - d)^+ - u(d - q_t - \sum_{i=t}^{T} w h_i(x_i))^+.$$

If $q_t + \sum_{i=t}^{T} w h_i(x_{\min}) - d \le 0$, because g_i and h_i are decreasing functions for all $i = t, \ldots, T$, it is obvious that the optimal solution is $x_t^0 = \ldots = x_T^0 = x_{\min}$.

Suppose that $q_t+\sum_{i=t}^T wh_i(x_{\min})-d>0$. Then the above optimization problem is equivalent to

$$\max \sum_{i=t}^{T} wg_i(x_i) - c_o(q_t + \sum_{i=t}^{T} wh_i(x_i) - d), \tag{A.3}$$

subject to constraints $q_t + \sum_{i=t}^T wh_i(x_i) - d \geq 0$ and $x_t, ..., x_T \geq x_{\min}$. The condition $q_t + \sum_{i=t}^T wh_i(x_{\min}) - d \geq 0$ ensures that the feasible set is nonempty and there is always an optimal solution. The KKT conditions are:

1) $x_t^0, ..., x_T^0$ are feasible; that is, $q_t + \sum_{i=t}^T wh_i(x_i^0) - d \geq 0$ and $x_t^0, ..., x_T^0 \geq x_{\min}$;

2) There exists $\lambda^0, \lambda_t^0, ..., \lambda_T^0 \geq 0$ such that

$$\lambda^0(q_t + \sum_{i=t}^{T} wh_i(x_i^0) - d) = 0$$

and for all $i = t, ..., T$, $\lambda_i^0(x_i^0 - x_{\min}) = 0$;

3) and for $i = t, ..., T$

$$w[g_i'(x_i^0) - c_o'(q_t + \sum_{i=t}^{T} wh_i(x_i^0) - d)h_i'(x_i^0) + \lambda^0 h_i'(x_i^0)] + \lambda_i^0 = 0.$$

We first show that the optimal cutoffs are either all equal to $x_{\min}$ or all strictly greater than it by contradiction. Suppose there exist $x_m^0 = x_{\min}$ and $x_n^0 > x_{\min}$, where $m \neq n$. Then for $i = n$, from 2) and 3) above and the fact that $g_t'(x) = xh_t'(x)$, we have $\lambda_n^0 = 0$ and $x_n^0 = c_o'(q_t + \sum_{i=t}^T wh_i(x_i^0) - d) - \lambda^0 > x_{\min}$. For $i = m$, by 3) above, we have

$$w[x_{\min} - (c_o'(q_t + \sum_{i=t}^{T} wh_i(x_i^0) - d) - \lambda^0)]h_m'(x_{\min}) + \lambda_m^0 = 0,$$

which is a contradiction because $c_o'(q_t + \sum_{i=t}^T wh_i(x_i^0) - d) - \lambda^0 > x_{\min}$, $h_m'(x_{\min}) < 0$, and $\lambda_m^0 \geq 0$. The KKT conditions also imply that when the cutoffs are strictly greater than $x_{\min}$, they are all the same.

The optimal solution can only be one of the following two possibilities.

1) If

$$q_t + \sum_{i=t}^{T} wh_i(c_o'(0)) - d \geq 0,$$

there must exist a unique fixed point $\overline{\overline{x}}_t^0$ that solves

$$x = c_o'(q_t + \sum_{i=t}^{T} wh_i(x) - d).$$

The reason is as follows. First notice that the right hand side of the equation is decreasing in x. At $x = x_{\min}$, $x_{\min} < c_o'(0) < c_o'(q_t + \sum_{i=t}^T wh_i(x_{\min}) - d)$. Consider two sub-cases. 1a). If $q_t > d$, then when x is large enough, the left hand side must be greater than the RHS, and hence a unique fixed point must exist. 1b). If $q_t < d$, then $\overline{x}_t^0$, the solution to the equation $q_t + \sum_{i=t}^T wh_i(x) - d = 0$ must satisfy $\overline{x}_t^0 \geq c_o'(0)$. This means that at $x = \overline{x}_t^0$, we have

$$\overline{x}_t^0 \geq c_o'(0) = c_o'(q_t + \sum_{i=t}^T wh_i(\overline{x}_t^0) - d),$$

and a unique fixed point must exist between $[x_{\min}, \overline{x}_t^0]$.

In this case, $\lambda^0 = \lambda_t^0 = ... = \lambda_T^0 = 0$, and $x_t^0 = ... = x_T^0 = \overline{\overline{x}}_t^0$ satisfy all the KKT conditions. The solution is optimal because the objective function in (A.3) reaches its global maximum at $x_i = \overline{\overline{x}}_t^0$ for all i.

2) If

$$q_t + \sum_{i=t}^T wh_i(c_o'(0)) - d < 0,$$

then the solution $\overline{x}_t^0$ to the equation $q_t + \sum_{i=t}^T wh_i(x) - d = 0$ must satisfy $\overline{x}_t^0 < c_o'(0)$. In this case, $x < c_o'(q_t + \sum_{i=t}^T wh_i(x) - d)$ for all $x \in [x_{\min}, \overline{x}_t^0]$. We can show that $\lambda^0 > 0$, $\lambda_i^0 = 0$ for all $i = t, ..., T$ and $x_t^0 = ... = x_T^0 = \overline{x}_t^0$ satisfy the all the KKT conditions. The solution is optimal because the objective function in (A.3) is increasing in $x_i \in [x_{\min}, \overline{x}_t^0]$ for all i. ■

To facilitate the derivation of the proofs of Theorems 2, we first provide some lemmas.

Lemma 1 *Suppose $\{X_i, i = 1, \cdots, n\}$ are i.i.d. random variables and $\mathsf{E}[|X_i|^4] < \infty$. Then for any $\ell > 0$ and $\tau_n = \sqrt{4\mathsf{E}[X_i^2](1-\delta)\log(n)} \to \infty$,*

$$\Pr\left(\left|n^{-1}\sum_{i=1}^n (X_i - E[X_i])\right| > \frac{\tau_n}{\sqrt{n}}\right) = O(n^{-(1-\delta)}),$$

for some small $\delta > 0$.

Proof of Lemma 1: The proof relies on Bernstein inequalities which we quote here: *Let Y_i be independent zero-mean random variables. Suppose that $|Y_i| \leq M$ almost surely, for all i. Then for all positive t,*

$$\Pr\left(\sum_{i=1}^n Y_i > t\right) \leq \exp\left(-\frac{\frac{1}{2}t^2}{\sum_{i=1}^n \mathsf{E}[Y_i^2] + \frac{1}{3}Mt}\right).$$

To apply this inequality, define $Y_i = X_i I(|X_i| \leq \sqrt{n}/\log(n))$, for $i = 1, \cdots, n$. Then, by Chebyshev's inequality,

$$\Pr\left(\sum_{i=1}^n (X_i - \mathsf{E}[X_i]) \neq \sum_{i=1}^n (Y_i - \mathsf{E}[Y_i])\right) \leq n\Pr(|X_i| > \sqrt{n}/\log(n)) \leq n\{\log(n)\}^4 \times \frac{\mathsf{E}[|X_i|^4]}{n^2} = o(n^{-(1-\delta)}).$$

Thus, without loss of generality, assume $|X_i| \leq \sqrt{n}/\log(n)$. According to Bernstein's inequality, we obtain

$$\Pr\left(\left|n^{-1}\sum_{i=1}^{n}(X_i - E[X_i])\right| > \frac{\tau_n}{\sqrt{n}}\right) \leq 2\exp\left(-\frac{\frac{1}{2}\tau_n^2 n}{n\mathsf{E}[X_i^2] + \frac{1}{3}\tau_n n/\log(n)}\right) \leq 2\exp(-\frac{\tau_n^2}{4\mathsf{E}[X_i^2]}) = O(n^{-(1-\delta)}).$$

■

With Lemma 1, we can obtain a uniform convergence rate for $G_t(x, \mathbf{S}^t, \boldsymbol{\delta}^t)$ and $H_t(x, \mathbf{S}^t, \boldsymbol{\delta}^t)$.

Lemma 2 *Suppose* $\sup_{t=1,\cdots,T}\mathsf{E}[|S_i^t|^4] < \infty$. *Then, given a large* n_t,

$$\sup_{x\geq 0}\left|G_t(x, \mathbf{S}^t, \boldsymbol{\delta}^t) - g_t(x)\right| = O_p(\sqrt{\frac{\log(n_t)}{n_t}}), \quad \sup_{x\geq 0}\left|H_t(x, \mathbf{S}^t, \boldsymbol{\delta}^t) - h_t(x)\right| = O_p(\sqrt{\frac{\log(n_t)}{n_t}}),$$

for $t = 1, \cdots, T$.

Proof of Lemma 2 Denote $C = \mathsf{E}[|S_i^t|] < \infty$. To show the uniform convergence results, we partition the range of x by a grid of points $\{x_j, j = 0, \cdots, k\}$, such that $g_t(x_j) = C \times \frac{j}{k}$. This implies that $0 = x_k < x_{k-1} < ... < x_1 < x_0 = \infty$. Becasue $G_t(x, \mathbf{S}^t, \boldsymbol{\delta}^t)$ is non-increasing in x, the supremum of $\left|G_t(x, \mathbf{S}^t, \boldsymbol{\delta}^t) - g_t(x)\right|$ over x can be upper bounded by

$$\max_{j=0,\cdots,k}\left|G_t(x_j, \mathbf{S}^t, \boldsymbol{\delta}^t) - g_t(x_j)\right| + \max_{j=0,\cdots,k}|g_t(x_j) - g_t(x_{j-1})| \leq \max_{j=0,\cdots,k}\left|G_t(x_j, \mathbf{S}^t, \boldsymbol{\delta}^t) - g_t(x_j)\right| + \frac{C}{k}$$

By Lemma 1,

$$\Pr\left(\max_{j=0,\cdots,k}\left|G_t(x_j, \mathbf{S}^t, \boldsymbol{\delta}^t) - g_t(x_j)\right| > \sqrt{\frac{4\mathsf{E}[|S_i^t|^2](1-\delta)\log(n_t)}{n_t}}\right) \leq \frac{k}{n_t^{1-\delta}}.$$

By choosing k such that $k\sqrt{\log(n_t)/n_t} \to \infty$ and $k = o(n_t^{1-\delta})$, the bias due to grid-approximation is negligible. Combining these results, we arrive at

$$\sup_{x\geq 0}\left|G_t(x, \mathbf{S}^t, \boldsymbol{\delta}^t) - g_t(x)\right| = O_p(\sqrt{\frac{\log(n_t)}{n_t}}).$$

The derivation for $H_t(x, \mathbf{S}^t, \boldsymbol{\delta}^t)$ is simpler than that of $G_t(x, \mathbf{S}^t, \boldsymbol{\delta}^t)$, as $H_t(x, \mathbf{S}^t, \boldsymbol{\delta}^t)$ is the sum of i.i.d. bounded random variables. ■

When n_t is sampled from a Poisson distribution with mean $\lambda\frac{D}{T}$ diverging to infinity, λ acts as the role of sample size, as demonstrated in Lemma 3.

Lemma 3 *As the rate* $\lambda \to \infty$,

$$\left|\frac{n_t}{\lambda} - \frac{D}{T}\right| = O_p(\frac{1}{\sqrt{\lambda}}).$$

for $t = 1, \ldots, T$.

Proof of Lemma 3: The conclusion follows by Chebyshev's inequality and the fact that $Var(n_t) = \lambda \frac{D}{T}$. ■

Lemma 4 *Let q_{t+1} be the total number of candidates hired at the end of t^{th} phase. Then, as $\lambda \to \infty$,*

$$\Pr\left(q_{t+1} > (1+\epsilon)\lambda(D/T)t\right) \leq \frac{1}{\epsilon^2 \lambda(D/T)t},$$

for any $\epsilon > 0$.

Proof of Lemma 4: The conclusion follows by the fact that $q_{t+1} \leq \sum_{i=1}^{t} n_i$, the results in Lemma 3, and Chebyshev's inequality. ■

Lemma 5 *For any $t = 1, \cdots, T$, the value function for the deterministic formulation satisfies Lipschitz continuity:*

$$|V_t^0(q_t') - V_t^0(q_t'')| \leq \max\{u, \sup_{x \leq (1+\epsilon)\lambda D} |c_o'(x)|\}|q_t' - q_t''|,$$

for any $0 \leq q_t' \leq (1+\epsilon)\lambda(D/T)(t-1)$ and $0 \leq q_t'' \leq (1+\epsilon)\lambda(D/T)(t-1)$.

Proof of Lemma 5: From Theorem 1, we have

$$V_t^0(q_t) = \begin{cases} \sum_{i=t}^{T} \lambda \frac{D}{T} g_i(x_{\min}) - u(d - q_t - \sum_{i=t}^{T} \lambda \frac{D}{T} h_i(x_{\min})), \text{ when } q_t < d - \sum_{i=t}^{T} \lambda \frac{D}{T} h_i(x_{\min}), \\ \sum_{i=t}^{T} \lambda \frac{D}{T} g_i(\overline{x}_t^0(q_t)), \text{ when } d - \sum_{i=t}^{T} \lambda \frac{D}{T} h_i(x_{\min}) \leq q_t \leq d - \sum_{i=t}^{T} \lambda \frac{D}{T} h_i(c_o'(0)), \\ \sum_{i=t}^{T} \lambda \frac{D}{T} g_i(\overline{\overline{x}}_t^0) - c_o(q_t + \sum_{i=t}^{T} \lambda \frac{D}{T} h_i(\overline{\overline{x}}_t^0) - d), \text{ when } q_t > d - \sum_{i=t}^{T} \lambda \frac{D}{T} h_i(c_o'(0)). \end{cases}$$

In the case that $V_t^0(q_t') > V_t^0(q_t'')$, we can relax $V_t^0(q_t'')$ by $J_t^0(q_t'', x_t(q_t'))$, where $x_t(q_t')$ is the cutoff from Theorem 1. Thus, by mean value theorem,

$$V_t^0(q_t') - V_t^0(q_t'') \leq V_t^0(q_t') - J_t^0(q_t'', x_t(q_t')) \leq \max\{u, \sup_{x \leq (1+\epsilon)\lambda D} |c_o'(x)|\}|q_t' - q_t''|.$$

The case that $V_t^0(q_t') < V_t^0(q_t'')$ can be analysed similarly. This completes the proof. ■

Proof of Theorem 2: We first prove the result for the upper bound method. Recall that $V_t^0(q_t)$ can be expressed in the following non-recursive form:

$$\max_{x_t, \ldots, x_T \geq x_{\min}} \sum_{i=t}^{T} \lambda \frac{D}{T} g_i(x_i) - c_o(q_t + \sum_{i=t}^{T} \lambda \frac{D}{T} h_i(x_i) - d)^+ - u(d - q_t - \sum_{i=t}^{T} \lambda \frac{D}{T} h_i(x_i))^+.$$

From Theorem 1, we have $x_i \geq x_{\min}$ for $i = t, \ldots, T$. This together with condition a) implies that $V_t^0(q_t) \geq \sum_{i=t}^T \lambda D/T g_i(\overline{x}_t^0) \geq \sum_{i=t}^T \lambda D/T h_i(\overline{x}_t^0) = d - q_t$. As a result, it suffices to show that, for $0 \leq q_t < d$,

$$\Big|\frac{1}{d-q_t}V_t^U(q_t, \mathbf{S}^{t:T}, \boldsymbol{\delta}^{t:T}) - \frac{1}{d-q_t}V_t^0(q_t)\Big| = O_p(\max\{u, \sup_{x\leq(1+\epsilon)\lambda D}|c_o'(x)|\}\sqrt{\frac{\log(\lambda)}{\lambda}} \times \frac{\lambda}{d-q_t}),$$

for $t = 1, \cdots, T$. To this end, we need

$$\Big|\frac{1}{\lambda}V_t^U(q_t, \mathbf{S}^{t:T}, \boldsymbol{\delta}^{t:T}) - \frac{1}{\lambda}V_t^0(q_t)\Big| = O_p(\max\{u, \sup_{x\leq(1+\epsilon)\lambda D}|c_o'(x)|\}\sqrt{\frac{\log(\lambda)}{\lambda}}),$$

for $0 \leq q_t < d$. For notational convenience, let $\mathrm{Con}(\lambda) = \max\{u, \sup_{x\leq(1+\epsilon)\lambda D}|c_o'(x)|\}\sqrt{\frac{\log(\lambda)}{\lambda}}$.

The following inequality is the key to the solution:

$$\begin{aligned}
&\Big|\frac{1}{\lambda}V_t^U(q_t, \mathbf{S}^{t:T}, \boldsymbol{\delta}^{t:T}) - \frac{1}{\lambda}V_t^0(q_t)\Big| \\
\leq\ & \max\Big\{\Big|\frac{1}{\lambda}J_t^U(q_t, x_t^U(q_t, \mathbf{S}^{t:T}, \boldsymbol{\delta}^{t:T}), \mathbf{S}^{t:T}, \boldsymbol{\delta}^{t:T}) - \frac{1}{\lambda}J_t^0(q_t, x_t^U(q_t, \mathbf{S}^{t:T}, \boldsymbol{\delta}^{t:T}))\Big|, \\
&\Big|\frac{1}{\lambda}J_t^U(q_t, x_t^0(q_t), \mathbf{S}^{t:T}, \boldsymbol{\delta}^{t:T}) - \frac{1}{\lambda}J_t^0(q_t, x_t^0(q_t))\Big|\Big\} \\
\leq\ & \sup_{x_t\geq x_{\min}}\Big|\frac{1}{\lambda}J_t^U(q_t, x_t, \mathbf{S}^{t:T}, \boldsymbol{\delta}^{t:T}) - \frac{1}{\lambda}J_t^0(q_t, x_t)\Big|.
\end{aligned} \tag{A.4}$$

By Lemma 4, assume without loss of generality that $q_t \leq (1+\epsilon)\lambda(D/T)(t-1)$, for $t = 2, \cdots, T+1$. It then suffices to show that

$$\sup_{x_t\geq x_{\min}}\ \sup_{0\leq q_t\leq(1+\epsilon)\lambda(D/T)(t-1)}\Big|\frac{1}{\lambda}J_t^U(q_t, x_t, \mathbf{S}^{t:T}, \boldsymbol{\delta}^{t:T}) - \frac{1}{\lambda}J_t^0(q_t, x_t)\Big| = O_p(\mathrm{Con}(\lambda)). \tag{A.5}$$

We use backward induction to prove (A.5).

Step I: We show the uniform consistency results at phase T:

$$\sup_{x_T\geq x_{\min}}\ \sup_{0\leq q_T\leq\lambda(D/T)(T-1)}\Big|\frac{1}{\lambda}J_T^U(q_T, x_T, \mathbf{S}^T, \boldsymbol{\delta}^T) - \frac{1}{\lambda}J_T^0(q_T, x_T)\Big| = O_p(\mathrm{Con}(\lambda)). \tag{A.6}$$

By definition,

$$\begin{aligned}
&\sup_{x_T\geq x_{\min}}\ \sup_{0\leq q_T\leq\lambda(D/T)(T-1)}\Big|\frac{1}{\lambda}J_T^U(q_T, x_T, \mathbf{S}^T, \boldsymbol{\delta}^T) - \frac{1}{\lambda}J_T^0(q_T, x_T)\Big| \\
\leq\ & \sup_{x_T\geq x_{\min}}\Big|\frac{n_t}{\lambda}G_T(x_T, \mathbf{S}^T, \boldsymbol{\delta}^T) - \frac{D}{T}g_T(x_T)\Big| \\
&+ \sup_{x_T\geq x_{\min}}\ \sup_{0\leq q_T\leq\lambda(D/T)(T-1)}\Big|\frac{1}{\lambda}V_{T+1}(q_T + n_T H_T(x_T, \mathbf{S}^T, \boldsymbol{\delta}^T)) - \frac{1}{\lambda}V_{T+1}(q_T + \lambda\frac{D}{T}h_T(x_T))\Big|
\end{aligned} \tag{A.7}$$

Note that the penalty function $V_{T+1}(q) = -c_o(q-d)^+ - u(d-q)^+$ is Lipschitz continuous because it consists of two functions with bounded slope, that is u and $\sup_{x\le(1+\epsilon)\lambda D}|c_o'(x)|$. Consequently, the second term on the right hand side of (A.7) is bounded by

$$\sup_{x_T\ge x_{\min}} \max(u, \sup_{x\le(1+\epsilon)\lambda D}|c_o'(x)|)\Big|(n_T/\lambda)H_T(x_T,\mathbf{S}^T,\boldsymbol{\delta}^T) - (D/T)h_T(x_T)\Big|,$$

which is independent of q_T. By Lemma 2 and Lemma 3, (A.6) holds.

Step II: Suppose the uniform consistency results at phase t hold, that is,

$$\sup_{x_t\ge x_{\min}} \sup_{0\le q_t\le(1+\epsilon)\lambda(D/T)(t-1)} \Big|\frac{1}{\lambda}J_t^U(q_t,x_t,\mathbf{S}^{t:T},\boldsymbol{\delta}^{t:T}) - \frac{1}{\lambda}J_t^0(q_t,x_t)\Big| = O_p(\mathrm{Con}(\lambda)).$$

Based on this, we intend to show that

$$\sup_{0\le q_{t-1}\le(1+\epsilon)\lambda(D/T)(t-2)} \sup_{x_{t-1}\ge x_{\min}} \Big|\frac{1}{\lambda}J_{t-1}^U(q_{t-1},x_{t-1},\mathbf{S}^{(t-1):T},\boldsymbol{\delta}^{(t-1):T}) - \frac{1}{\lambda}J_{t-1}^0(q_{t-1},x_{t-1})\Big| = O_p(\mathrm{Con}(\lambda)). \tag{A.8}$$

Equation (A.8) can be derived as

$$\begin{aligned}
&\sup_{x_{t-1}\ge x_{\min}} \sup_{0\le q_{t-1}\le(1+\epsilon)\lambda(D/T)(t-2)} \Big|\frac{1}{\lambda}J_{t-1}^U(q_{t-1},x_{t-1},\mathbf{S}^{(t-1):T},\boldsymbol{\delta}^{(t-1):T}) - \frac{1}{\lambda}J_{t-1}^0(q_{t-1},x_{t-1})\Big| \\
\le& \sup_{x_{t-1}\ge x_{\min}} \Big|\frac{n_{t-1}}{\lambda}G_{t-1}(x_{t-1},\mathbf{S}^{t-1},\boldsymbol{\delta}^{t-1}) - \frac{D}{T}g_{t-1}(x_{t-1})\Big| \\
+& \sup_{x_{t-1}\ge x_{\min}} \sup_{0\le q_{t-1}\le(1+\epsilon)\lambda(D/T)(t-2)} \Big|\frac{1}{\lambda}V_t^U(q_{t-1}+n_{t-1}H_{t-1}(x_{t-1},\mathbf{S}^{t-1},\boldsymbol{\delta}^{t-1}),\mathbf{S}^{t:T},\boldsymbol{\delta}^{t:T}) \\
&- \frac{1}{\lambda}V_t^0(q_{t-1}+\lambda\frac{D}{T}h_{t-1}(x_{t-1}))\Big|
\end{aligned}$$

By Lemma 2 and Lemma 3, the first term is bounded in probability with the order $\sqrt{\log(\lambda)/\lambda}$. The second term is then split into two terms as

$$\begin{aligned}
&\sup_{x_{t-1}\ge x_{\min}} \sup_{0\le q_{t-1}\le(1+\epsilon)\lambda(D/T)(t-2)} \Big|\frac{1}{\lambda}V_t^U(q_{t-1}+n_{t-1}H_{t-1}(x_{t-1},\mathbf{S}^{t-1},\boldsymbol{\delta}^{t-1}),\mathbf{S}^{t:T},\boldsymbol{\delta}^{t:T}) \\
&\qquad - \frac{1}{\lambda}V_t^0(q_{t-1}+n_{t-1}H_{t-1}(x_{t-1},\mathbf{S}^{t-1},\boldsymbol{\delta}^{t-1}))\Big| \\
+& \sup_{x_{t-1}\ge x_{\min}} \sup_{0\le q_{t-1}\le(1+\epsilon)\lambda(D/T)(t-2)} \Big|\frac{1}{\lambda}V_t^0(q_{t-1}+n_{t-1}H_{t-1}(x_{t-1},\mathbf{S}^{t-1},\boldsymbol{\delta}^{t-1})) - \frac{1}{\lambda}V_t^0(q_{t-1}+\lambda\frac{D}{T}h_{t-1}(x_{t-1}))\Big|
\end{aligned}$$

The first term is upper bounded by

$$\sup_{0\le q_t\le(1+\epsilon)\lambda(D/T)(t-1)} \Big|\frac{1}{\lambda}V_t^U(q_t,\mathbf{S}^{t:T},\boldsymbol{\delta}^{t:T}) - \frac{1}{\lambda}V_t^0(q_t)\Big|,$$

which is $O_p(\mathrm{Con}(\lambda))$ by (A.4) and by induction. By Lemma 5, $V_t^0(q_t)$ satisfies Lipschitz continuity for $0\le q_t\le(1+\epsilon)\lambda(D/T)(t-1)$. This together with Lemma 2 and Lemma 3 implies that the second term is of order $O_p(\mathrm{Con}(\lambda))$.

We turn to obtain the convergence rate of the lower bound method. By induction, it holds that

$$\Big|\frac{1}{\lambda}J_t^L(q_t,x_t)-\frac{1}{\lambda}J_t^0(q_t,x_t)\Big|\le C\sum_{j=t}^{T}\sup_{x_j}\mathsf{E}\Big[\Big|\lambda^{-1}\sum_{i=1}^{n_j}I(S_i^j\ge x_j)I(\delta_i^j=1)-\frac{D}{T}h_j(x_j)\Big|\Big],$$

where $C=\max\{u,\max_{x\le(1+\epsilon)\lambda D}|c_o'(x)|\}$. By Lemma 3, it suffices to show that

$$\sup_{x\ge 0}\mathsf{E}\Big[\big|H_t(x,\mathbf{S}^t,\boldsymbol{\delta}^t)-h_t(x)\big|\Big]=O(\frac{1}{\sqrt{n_t}}), \tag{A.9}$$

for $t=1,\cdots,T$. By Cauchy-Schwarz inequality, $\mathsf{E}[|X|]\le\sqrt{\mathsf{E}[|X|^2]}$, for any random variable X. We then obtain

$$\sup_x\mathsf{E}\Big[\big|H_t(x,\mathbf{S}^t,\boldsymbol{\delta}^t)-h_t(x)\big|\Big]\le\sup_x\sqrt{n_t^{-1}\mathsf{E}[I(S_i^t\ge x)I(\delta_i^t=1)]}\le\frac{1}{\sqrt{n_t}}.$$

We finally derive the convergence rate for the original formulation. For the objective function in the original formulation, we have

$$\begin{aligned}
&\Big|\frac{1}{\lambda}J_t(q_t,x_t,\mathbf{S}^t)-\frac{1}{\lambda}J_t^0(q_t,x_t)\Big|\\
\le\;&\Big|\lambda^{-1}\sum_{i=1}^{n_t}S_i^tI(S_i^t\ge x_t)p_t(S_i^t)-\frac{D}{T}g_t(x_t)\Big|\\
&+\sum_{j=t+1}^{T}\sup_{x_j}\mathsf{E}\Big[\Big|\lambda^{-1}\sum_{i=1}^{n_j}S_i^jI(S_i^j\ge x_j)p_t(S_i^j)-\frac{D}{T}g_j(x_j)\Big|\mid\mathbf{S}^t\Big]\\
&+C\sum_{j=t+1}^{T}\sup_{x_j}\mathsf{E}\Big[\Big|\lambda^{-1}\sum_{i=1}^{n_j}I(S_i^j\ge x_j)I(\delta_i^j=1)-\frac{D}{T}h_j(x_j)\Big|\mid\mathbf{S}^t\Big],
\end{aligned}$$

where $C=\max\{u,\max_{x\le(1+\epsilon)\lambda D}|c_o'(x)|\}$. Following similar steps as those in Lemma 2, we can obtain

$$\sup_{x_t\ge 0}\Big|{n_t}^{-1}\sum_{i=1}^{n_t}S_i^tI(S_i^t\ge x_t)p_t(S_i^t)-g_t(x_t)\Big|=O_p(\sqrt{\frac{\log(n_t)}{n_t}}) \tag{A.10}$$

This together with Lemma 3 implies that the first term is $O_p(\sqrt{\log(\lambda)/\lambda})$. For the second term, we need to prove

$$\sup_{x_j\ge 0}\mathsf{E}\Big[\Big|{n_j}^{-1}\sum_{i=1}^{n_j}S_i^jI(S_i^j\ge x_j)p_j(S_i^j)-g_j(x_j)\Big|\mid\mathbf{S}^t\Big]=O_p(\frac{1}{\sqrt{n_j}}), \tag{A.11}$$

for $j = t+1, \cdots, T$. This is achieved by

$$\sup_{x_j} \mathsf{E}\Big[\Big|\sum_{i=1}^{n_j} S_i^j I(S_i^j \geq x_j) p_t(S_i^j) - g_j(x_j)\Big| \mid \mathbf{S}^t\Big]$$
$$\leq \sup_{x_j} \sqrt{n_j^{-2} \sum_{i=1}^{n_j} \mathsf{E}[|S_i^j|^2 I(S_i^j \geq x_j) \mid \mathbf{S}^t]}$$
$$\leq \sqrt{n_j^{-2} \sum_{i=1}^{n_j} \mathsf{E}[|S_i^j|^2 \mid \mathbf{S}^t]},$$

where the first inequality is due to the conditional independence of $\{S_i^j, i = 1, \cdots, n_j\}$ given $\mathbf{S}^t$. Because $\mathsf{E}[|S_i^j|^2 \mid \mathbf{S}^t]$ has the mean $\mathsf{E}[|S_i^j|^2]$, which is bounded for all $i = 1, \cdots, n_j$, the above term is of order $O_p(n_j^{-1/2})$. Thus the second term is of order $O_p(\lambda^{-1/2})$. The third term is also $O_p(\max\{u, \max_{x \leq (1+\epsilon)\lambda D} |c_o'(x)|\}\lambda^{-1/2})$ by (A.9). ■

Proof of Theorem 3: We first show the convergence of the lower bound heuristic. In each phase from t onwards, the q_js are updated using the observed data, that is,

$$q_{j+1} = q_j + n_j H_j(x_j^L, \mathbf{S}^j, \boldsymbol{\delta}^j), \quad j = t, \cdots, T.$$

Then the expected reward from phase t onwards using the cutoff x_j^L can be expressed as

$$\mathsf{E}\Big[\sum_{j=t}^{T} n_j G_j(x_j^L, \mathbf{S}^t, \boldsymbol{\delta}^t) + V_{T+1}(q_{T+1})\Big] = \sum_{j=t}^{T} V_j^L(q_j) - \sum_{j=t+1}^{T} V_j^L(q_j)$$

By the results in Theorem 2, it holds that, for $0 \leq q_t < d$,

$$\Big|\frac{\mathsf{E}\big[\sum_{j=t}^{T} n_j G_j(x_j^L, \mathbf{S}^j, \boldsymbol{\delta}^j) + V_{T+1}(q_{T+1})\big]}{V_t^0(q_t)} - 1\Big| = O(\max\{u, \sup_{x \leq (1+\epsilon)\lambda D} |c_o'(x)|\}\lambda^{-1/2} \times \frac{\lambda}{d - q_t}).$$

Noting that $R_t(LB, q_t, \mathbf{S}^t) = \mathsf{E}[\sum_{j=t}^{T} n_j G_j(x_j^L, \mathbf{S}^t, \boldsymbol{\delta}^t) + V_{T+1}(q_{T+1}) \mid \mathbf{S}^t]$ and using (A.11), we conclude that, for $0 \leq q_t < d$,

$$\Big|\frac{R_t(LB, q_t, \boldsymbol{S}^t)}{V_t^0(q_t)} - 1\Big| = O_p(\mathrm{Con}(\lambda)).$$

By Theorem 2, we arrive at

$$\Big|\frac{R_t(LB, q_t, \boldsymbol{S}^t)}{V_t(q_t, \mathbf{S}^t)} - 1\Big| = O_p(\mathrm{Con}(\lambda)),$$

for $0 \leq q_t < d$.

To obtain the result for AL, we define the sequence of q_js under the asymptotic limit framework as

$$q_{j+1}^0 = q_j^0 + \lambda(D/T) h_j(x_j^0), \quad j = t, \cdots, T,$$

where x_j^0 is the cutoff under AL. Then the expected reward under AL can be formulated in the following non-recursive form:

$$\sum_{j=t}^{T} \lambda(D/T) g_j(x_j^0) + V_{T+1}(q_{T+1}^0) = \sum_{j=t}^{T} V_j^0(q_j^0) - \sum_{j=t+1}^{T} V_j^0(q_j^0).$$

Comparing the above formula with the $R_t(AL, q_t, \mathbf{S}^t)$, we derive, for $0 \le q_t < d$,

$$\begin{aligned}
&\Big|\frac{\mathsf{E} R_t(AL, q_t, \mathbf{S}^t)}{V_t^0(q_t)} - 1\Big| \\
\le& \frac{|V_{T+1}(q_{T+1}^0) - \mathsf{E} V_{T+1}(q_{T+1})|}{V_t^0(q_t)} \\
\le& \frac{\max(u, \sup_{x \le (1+\epsilon)\lambda D} c_o'(x)) \mathsf{E}|q_{T+1} - q_{T+1}^0|}{V_t^0(q_t)} \\
=& O(\max(u, \sup_{x \le (1+\epsilon)\lambda D} c_o'(x)) \times \lambda^{-1/2} \times \frac{\lambda}{d - q_t}).
\end{aligned}$$

This together with Theorem 2 and (A.11) yields that, for $0 \le q_t < d$

$$\Big|\frac{R_t(AL, q_t, \boldsymbol{S}^t)}{V_t(q_t, \mathbf{S}^t)} - 1\Big| = O_p(\mathrm{Con}(\lambda)).$$

We next show that the objective functions for the IAL and ILB formulations converge to that of the AL formulation:

$$\begin{aligned}
\sup_{0 \le q_t \le (1+\epsilon)\lambda(D/T)(t-1)} \sup_{x_t \ge 0} \Big|\frac{1}{\lambda} J_t^{IAL}(q_t, x_t, \mathbf{S}^t) - \frac{1}{\lambda} J_t^0(q_t, x_t)\Big| &= O_p(\max\{u, \sup_{x \le (1+\epsilon)\lambda D} |c_o'(x)|\} \sqrt{\frac{\log(\lambda)}{\lambda}}), \\
\sup_{0 \le q_t \le (1+\epsilon)\lambda(D/T)(t-1)} \sup_{x_t \ge 0} \Big|\frac{1}{\lambda} J_t^{ILB}(q_t, x_t, \mathbf{S}^t) - \frac{1}{\lambda} J_t^0(q_t, x_t)\Big| &= O_p(\max\{u, \sup_{x \le (1+\epsilon)\lambda D} |c_o'(x)|\} \sqrt{\frac{\log(\lambda)}{\lambda}}).
\end{aligned} \tag{A.12}$$

Similar to the proof for the upper bound method, it suffices to show that

$$\sup_{x \ge 0} \mathsf{E}\Big[\big|H_t(x, \mathbf{S}^t, \boldsymbol{\delta}^t) - h_t(x)\big| \mid \mathbf{S}^t\Big] = O_p(\sqrt{\frac{\log(\lambda)}{\lambda}})$$

The left hand side of the above equation is split into two terms as

$$\mathsf{E}\Big[\Big|n_t^{-1} \sum_{i=1}^{n_t} I(S_i^t \ge x)\{I(\delta_i^t = 1) - p_t(S_i^t)\}\Big| \mid \mathbf{S}^t\Big] + \Big|n_t^{-1} \sum_{i=1}^{n_t} I(S_i^t \ge x) p_t(S_i^t) - h_t(x)\Big|$$

By Cauchy-Schwartz inequality and law of large numbers, the first term is further bounded by

$$\frac{2\sqrt{\mathsf{E}[I(S_i^t \ge x) p_t(S_i^t)\{1 - p_t(S_i^t)\}]}}{\sqrt{n_t}}.$$

The convergence rate of the second term is given in (A.10).

With the results in (A.12), we proceed to derive the convergence rate for ILB. Define $V_j^{ILB}(q_j, \mathbf{S}^j)$ as the value function when the data-driven cutoff $x_j^{ILB}(q_j, \mathbf{S}^j)$ is plugged in the objective function $J_j^{ILB}(q_j, x_j, \mathbf{S}^j)$, for $j = t, \cdots, T$. We observe that

$$\sum_{j=t}^{T} n_j G_j(x_j^{ILB}, \mathbf{S}^t, \boldsymbol{\delta}^t) + V_{T+1}(q_{T+1}) = \sum_{j=t}^{T} V_j^{ILB}(q_j, \mathbf{S}^j) - \sum_{j=t+1}^{T} V_j^L(q_j).$$

By Theorem 2 and Equation (A.12), the right hand side can be derived as

$$\Big|\frac{\sum_{j=t}^{T} V_j^{ILB}(q_j, \mathbf{S}^j) - \sum_{j=t+1}^{T} V_j^L(q_j)}{V_t^0(q_t)} - 1\Big| = O_p(\text{Con}(\lambda)).$$

Combining the above two results, we obtain, for $0 \leq q_t < d$,

$$\Big|\frac{\sum_{j=t}^{T} n_j G_j(x_j^{ILB}, \mathbf{S}^t, \boldsymbol{\delta}^t) + V_{T+1}(q_{T+1})}{V_t^0(q_t)} - 1\Big| = O_p(\text{Con}(\lambda)).$$

By taking the conditional expectation on the total reward and using (A.11), we prove that, for $0 \leq q_t < d$,

$$\Big|\frac{R_t(ILB, q_t, \mathbf{S}^t)}{V_t^0(q_t)} - 1\Big| = O_p(\text{Con}(\lambda)).$$

The proof for the IAL formulation is analogous to that for ILB, thus is omitted. This completes the proof of Theorem 3. ■

To prove Theorem 4, we provide the following Lemma, which characterizes the asymptotic behavior of d_t, for $t = 1, \cdots, T$.

Lemma 6 *Let $d_t = (d - q_t)/\{p(T - t + 1)\}$. Then Td_t/d converges to $1/p$ in probability, as $d \to \infty$.*

Proof of Lemma 6: By definition, d_t can be recursively defined as

$$d_t = d_{t-1} + \frac{1}{T - t + 1}\{d_{t-1} - (q_t - q_{t-1})/p\}.$$

As a result, $\mathsf{E}[d_t \mid d_{t-1}] = d_{t-1}$ and $\mathsf{E}[d_t] = \mathsf{E}[d_1] = d/(pT)$. This means that d_t is a martingale. Further, the variance of $d_t - d_{t-1}$ can be derived as

$$\begin{aligned} &Var(d_t - d_{t-1}) \\ =&\mathsf{E}[Var(d_t - d_{t-1} \mid d_{t-1})] + Var\{\mathsf{E}[d_t - d_{t-1} \mid d_{t-1}]\} \\ =&\mathsf{E}[\frac{d_{t-1}p(1-p)}{p^2(T-t+1)^2}] = \frac{d(1-p)}{Tp^2} \times \frac{1}{(T-t+1)^2}. \end{aligned}$$

The covariance between $d_i - d_{i-1}$ and $d_j - d_{j-1}$ is simply zero for $i \neq j$, due to the fact that d_t is a martingale. Based on these results, we obtain that

$$Var(d_t) = \sum_{j=2}^{t} Var(d_j - d_{j-1}) = \frac{d(1-p)}{Tp^2} \times \sum_{j=2}^{t} \frac{1}{(T-j+1)^2} \leq C\frac{d(1-p)}{T^2p^2} \times \frac{t-1}{T-t+1}$$

As a result, $Var(Td_t/d) \to 0$ as $d \to \infty$. This implies that Td_t/d converges to $1/p$ in probability. ■

The following lemma is from Shao (2003) and it will be used in the proofs of Theorem 4 and Theorem 5.

Lemma 7 *Suppose that* $X_n \to_d X$. *Then,*

$$\lim_{n\to\infty} \mathsf{E}[|X_n|] = E[|X|] < \infty$$

if and only if $\{|X_n|\}$ *is uniformly integrable in the sense that*

$$\lim_{t\to\infty} \sup_n \mathsf{E}[|X_n| I_{\{|X_n|>t\}}] = 0.$$

Proof of Theorem 4: The proof proceeds in three steps.

(i) To derive the approximation for the penalty cost, we first show that under the linear inflation heuristic, the total number of accepted offers q_{T+1} has mean d and variance $d(1-p)/T$.

Obviously $\mathsf{E}q_2 = \frac{d}{T}$,

$$\mathsf{E}q_3 = \mathsf{E}[\mathsf{E}[q_3|q_2]] = \mathsf{E}\Big[q_2 + \frac{d-q_2}{(T-1)}\Big] = \frac{2d}{T}.$$

Similarly,

$$\mathsf{E}q_4 = \mathsf{E}[\mathsf{E}[q_4|q_3]] = \mathsf{E}\Big[q_3 + \frac{d-q_3}{(T-2)}\Big] = \frac{3d}{T}.$$

We can similarly show that for all i, we have $\mathsf{E}q_{i+1} = i\frac{d}{T}$. Therefore $\mathsf{E}q_{T+1} = d$.

The variance of q_2 is given by $Var(q_2) = \frac{d}{T}(1-p)$. According to the law of total variance,

$$Var(q_3) \quad = \quad \mathsf{E}[Var(q_3|q_2)] + Var(\mathsf{E}[q_3|q_2]).$$

And

$$\begin{aligned} Var(\mathsf{E}[q_3|q_2]) &= Var\left(q_2 + \frac{d-q_2}{(T-1)}\right) \\ &= \left(\frac{T-2}{T-1}\right)^2 Var(q_2). \end{aligned}$$

and

$$\begin{aligned} \mathsf{E}[Var(q_3|q_2)] &= \mathsf{E}\Big[Var(Bin(\frac{d-q_2}{p(T-1)}, p)|q_2)\Big] \\ &= \mathsf{E}\Big[\frac{d-q_2}{T-1}(1-p)\Big] \\ &= \frac{d}{T}(1-p). \end{aligned}$$

Therefore

$$Var(q_3) = \frac{d}{T}(1-p) + \left(\frac{T-2}{T-1}\right)^2 \frac{d}{T}(1-p).$$

Similarly,

$$\begin{aligned} Var(q_4) &= \mathsf{E}[Var(q_4|q_3)] + Var(\mathsf{E}[q_4|q_3]) \\ &= \mathsf{E}\Big[\frac{d-q_3}{T-2}(1-p)\Big] + Var\left(q_3 + \frac{d-q_3}{T-2}\right) \\ &= \frac{d}{T}(1-p) + \left(\frac{T-3}{T-2}\right)^2 Var(q_3) \end{aligned}$$

In general,

$$Var(q_{i+1}) = \frac{d}{T}(1-p) + \left(\frac{T-i}{T-i+1}\right)^2 Var(q_i)$$

So

$$Var(q_{T+1}) = \frac{d}{T}(1-p).$$

(ii) We show that $(q_{T+1}-d)/\sqrt{d(1-p)/T}$ converges to the standard normal distribution as $d \to \infty$. To this end, rewrite $q_{T+1} = \sum_{t=1}^{T}\sum_{i=1}^{d_t}\delta_i^t$, where δ_i^t's are independent Bernoulli random variables with mean p. According to Lemma 6, for any $\epsilon > 0$, we obtain

$$\lim_{d\to\infty} \Pr\left(\Big|\frac{d_t}{d} - \frac{1}{Tp}\Big| > \epsilon\right) \to 0.$$

With these results, we decompose the distribution function of q_{T+1} as

$$\begin{aligned} &\Pr\left(\frac{\sum_{t=1}^{T}\sum_{i=1}^{d_t}\delta_i^t - d}{\sqrt{d(1-p)/T}} \le x\right) \\ =&\Pr\left(\frac{\sum_{t=1}^{T}\sum_{i=1}^{d_t}\delta_i^t - d}{\sqrt{d(1-p)/T}} \le x, \cap_{t=1}^{T}\{\frac{d}{pT} - \epsilon d \le d_t \le \frac{d}{pT} + \epsilon d\}\right) \\ &+\Pr\left(\frac{\sum_{t=1}^{T}\sum_{i=1}^{d_t}\delta_i^t - d}{\sqrt{d(1-p)/T}} \le x, \cup_{t=1}^{T}\{|\frac{d_t}{d} - \frac{1}{Tp}| > \epsilon\}\right) \end{aligned}$$

The second term is bounded by $T\Pr(|d_t/d - 1/(Tp)| > \epsilon)$, which is negligible by Lemma 6. The first term is upper bounded by

$$\begin{aligned} &\Pr\left(\frac{\sum_{t=1}^{T}\sum_{i=1}^{d_t}\delta_i^t - d}{\sqrt{d(1-p)/T}} \le x, \cap_{t=1}^{T}\{\frac{d}{pT} - \epsilon d \le d_t \le \frac{d}{pT} + \epsilon d\}\right) \\ \le&\Pr\left(\frac{\sum_{t=1}^{T}\sum_{i=1}^{d/(pT)+\epsilon d}\delta_i^t - d}{\sqrt{d(1-p)/T}} \le x\right) \\ =&\Pr\left(\frac{\sum_{t=1}^{T}\sum_{i=1}^{d/(pT)}\delta_i^t - d}{\sqrt{d(1-p)/T}} + \frac{\sum_{t=1}^{T}\sum_{i=d/(pT)+1}^{d/(pT)+\epsilon d}\delta_i^t}{\sqrt{d(1-p)/T}} \le x\right) \to \Phi(x), \end{aligned}$$

as long as the second term is $o_p(1)$. To show this, we reformulate the second term as

$$\frac{\sum_{t=1}^{T}\sum_{i=d/(pT)+1}^{d/(pT)+\epsilon d}\delta_i^t}{\sqrt{\epsilon d(1-p)/T}}\sqrt{\epsilon}$$

According to Central Limit Theorem, it converges to zero in probability as $d\to\infty$ and $\epsilon\to 0$. Analogously, we can show that

$$\Pr\left(\frac{\sum_{t=1}^{T}\sum_{i=1}^{d_t}\delta_i^t - d}{\sqrt{d(1-p)/T}}\leq x\right)\geq\Phi(x),$$

as $d\to\infty$.

(iii) By (iii), we can approximate q_{T+1} by a normal distribution $N(d,(1-p)d/T)$. For a normal random variable $X\sim N(\mu,\sigma^2)$, we have

$$\mathsf{E}[X|X\leq\mu]=\mu-\sigma\sqrt{\frac{2}{\pi}}.$$

The proof can be found in Barr and Sherrill (1999). Thus, by Lemma 7 and the fact that $\mathsf{E}[|q_{T+1}-d|/\sqrt{d}]\leq\sqrt{(1-p)/T}$, the expected penalty cost converges to

$$u\sqrt{\frac{(1-p)d}{T}}\frac{1}{\sqrt{2\pi}}+\int_0^\infty c_o(\sqrt{\frac{(1-p)d}{T}}z)\phi(z)dz$$

in ratio. This together with a simple derivation using gamma function completes the proof. ■

Lemma 8 below provides an approximation of the tail probability for standard normal distribution.

Lemma 8 *For any $\delta>0$,*

$$\frac{\delta}{1+\delta^2}\exp\{-\frac{\delta^2}{2}\}\leq\int_\delta^\infty\exp\{-\frac{x^2}{2}\}dx\leq\frac{1}{\delta}\exp\{-\frac{\delta^2}{2}\}.$$

Proof of Lemma 8: For the right hand side, note that

$$\int_\delta^\infty\exp\{-\frac{x^2}{2}\}dx\leq\frac{1}{\delta}\int_\delta^\infty x\exp\{-\frac{x^2}{2}\}dx=\frac{1}{\delta}\exp\{-\frac{\delta^2}{2}\}.$$

The inequality in the left hand side can be verified by

$$\int_\delta^\infty\exp\{-\frac{x^2}{2}\}dx\geq\int_\delta^\infty\frac{x^2}{1+x^2}\exp\{-\frac{x^2}{2}\}dx\geq\frac{\delta}{1+\delta^2}\int_\delta^\infty x\exp\{-\frac{x^2}{2}\}dx=\frac{\delta}{1+\delta^2}\exp\{-\frac{\delta^2}{2}\}.$$

■

Proof of Theorem 5: Part (i): By Lemma 2 and Lemma 3, it holds that

$$\sup_{x \geq x_{\min}} \left| \frac{\sum_{t=1}^{T} \sum_{i=1}^{n_t} S_i^t I(S_i^t \geq x) I(\delta_i^t = 1)}{\sum_{t=1}^{T} \lambda(D/T) p \int_x^\infty s f_t(s) p ds} - 1 \right| = O_p(\frac{\sqrt{\log(\lambda)}}{\sqrt{\lambda}}),$$

due to the fact that $\int_x^\infty s f_t(s) ds \geq x \geq x_{\min} > 0$. As a result,

$$\left| \frac{\sum_{t=1}^{T} \sum_{i=1}^{d_t} S_{[i]}^{n_t} I(\delta_i^t = 1)}{\sum_{t=1}^{T} \lambda(D/T) p \int_{S_{[d_t]}^{n_t}}^\infty s f_t(s) p ds} - 1 \right| = O_p(\frac{\sqrt{\log(\lambda)}}{\sqrt{\lambda}}).$$

When S_i^t is normally distributed with mean μ and standard deviation σ, the integral $\int_{S_{[d_t]}^{n_t}}^\infty s f_t(s) ds$ can be simplified as

$$\int_{Z_{[d_t]}^{n_t}}^\infty (\mu + \sigma z) \phi(z) dz = \mu\{1 - \Phi(Z_{[d_t]}^{n_t})\} + \sigma \int_{Z_{[d_t]}^{n_t}}^\infty z \phi(z) dz,$$

where $Z_{[d_t]}^{n_t} = (S_{[d_t]}^{n_t} - \mu)/\sigma$. By Theorem 5.11 in Shao (2003), the gap between the sample quantile $Z_{[d_t]}^{n_t}$ and its theoretical one adopts the following asymptotic representation

$$Z_{[d_t]}^{n_t} - \Phi^{-1}(1 - \frac{d_t}{n_t}) = \frac{(1 - \frac{d_t}{n_t}) - \widehat{F}_{n_t}(\Phi^{-1}(1 - \frac{d_t}{n_t}))}{\phi(\Phi^{-1}(1 - \frac{d_t}{n_t}))} + o_p(\frac{1}{\sqrt{n_t}}), \tag{A.13}$$

where $F_{n_t}(\cdot)$ denotes the empirical distribution of the transformed scores Z_i^t. Lemma 2 and Lemma 3 again reveal that

$$\left|(1 - \frac{d_t}{n_t}) - \widehat{F}_{n_t}(\Phi^{-1}(1 - \frac{d_t}{n_t}))\right| = O_p(\frac{\sqrt{\log(\lambda)}}{\sqrt{\lambda}}).$$

Thus, to make the first term in Equation (A.13) being negligible, it is required that

$$\phi(\Phi^{-1}(1 - \frac{d_t}{n_t})) \geq C\lambda^{-\alpha}, \text{ for some } \alpha < 1/2,$$

with probability tending to one. The requirement is equivalent to

$$\Pr\left(\frac{d_t}{n_t} \geq 1 - \Phi(\sqrt{2\alpha \log(\lambda)})\right) \to 1.$$

According to Lemma 6, $Td_t/d \to_p 1/p$. Further, by Lemma 3, $n_t/\lambda \to D/T$ in probability, as $\lambda \to \infty$. Combining these two results, we arrive at

$$\frac{d_t}{n_t} \to \frac{d}{p\lambda D}$$

in probability. By Lemma 8, the above probability converges to one as long as $d \geq Dp/\sqrt{2\alpha \log(\lambda)} \times \lambda^{1-\alpha}$ with some $\alpha < 1/2$. This completes the proof by condition on the d.

To prove Part (ii) and Part (iii), we need to further derive an asymptotic behaviour of following expression under two scenarios on d, that is,

$$\mu p(\lambda D)/T \sum_{t=1}^{T} \frac{d_t}{n_t} + \sigma p(\lambda D)/T \sum_{t=1}^{T} \int_0^{\frac{d_t}{n_t}} \Phi^{-1}(1-z)dz. \tag{A.14}$$

Part (ii): We first analyze the case when $d = \gamma\lambda D$ with $0 < \gamma < p$. Recall that n_t is a Poisson random variable with mean $\lambda(D/T)$, then the assumption on d guarantees $d_t < n_t$ with probability tending to one. This implies that $d \to \infty$ as $\lambda \to \infty$. Similar to the proof of Part (i), $\frac{d_t}{n_t} \to \frac{\gamma}{p}$ in probability. Thus, the first term in the asymptotic expression (A.14) satisfies that

$$\frac{\mu(\lambda D)/T \sum_{t=1}^{T} \frac{d_t}{n_t}}{d\mu} \to 1,$$

in probability. By continuous mapping theorem (Shao 2003), we obtain

$$\Psi(\frac{d_t}{n_t}) \to_p \Psi(\frac{\gamma}{p}),$$

because $\gamma/p \leq 1$ and $\Phi^{-1}(1-\gamma/p) < \infty$. This implies that the second term in (A.14)

$$\frac{\sigma p(\lambda D)/T \sum_{t=1}^{T} \int_0^{\frac{d_t}{n_t}} \Phi^{-1}(1-z)dz}{\sigma p(\lambda D)\Psi(\gamma/p)} \to 1$$

in probability. Combing the above two results, we have

$$\frac{\mu p(\lambda D)/T \sum_{t=1}^{T} \frac{d_t}{n_t} + \sigma p(\lambda D)/T \sum_{t=1}^{T} \int_0^{\frac{d_t}{n_t}} \Phi^{-1}(1-z)dz}{d\mu + \sigma p(\lambda D)\Psi(\gamma/p)} \to 1$$

in probability. By Lemma 7,

$$\frac{r(T)}{d\mu + \sigma p(\lambda D)\Psi(\gamma/p)} \to 1.$$

Part (iii): We then derive the total expected reward when $d = C\lambda^{\gamma}$ with $1/2 < \gamma < 1$ and $C > 0$. This implies that $d_t/n_t \to_p 0$ as $\lambda \to \infty$. By Taylor expansion, we obtain that

$$\begin{aligned}
&\sum_{t=1}^{T} (\lambda D)/T\Psi(\frac{d_t}{n_t}) \\
&= \sum_{t=1}^{T} n_t \Phi^{-1}(1-\frac{d_t}{n_t})\frac{d_t}{n_t}\{1+o_p(1)\} \\
&= \sum_{t=1}^{T} d_t \Phi^{-1}(1-\frac{d_t}{n_t})\{1+o_p(1)\}
\end{aligned}$$

By Lemma 8, $\Phi^{-1}(1-\frac{d_t}{n_t})$ can be approximated by the root of the following identity

$$\frac{d_t}{n_t} = \frac{1}{\sqrt{2\pi}x}\exp\{-x^2/2\}\{1+o_p(1)\},$$

from which $\Phi(1-\frac{d_t}{n_t}) = \sqrt{2[\log(n_t)-\log(d_t)-\log\{\sqrt{2\pi}\Phi(1-\frac{d_t}{n_t})\}]}\{1+o_p(1)\}$. The term $\log\{\Phi(1-\frac{d_t}{n_t})\}$ plays a minor role on the approximation and can be ignored asymptotically. As a result, we arrive at

$$\sum_{t=1}^{T}(\lambda D)/T\Psi(\frac{d_t}{n_t}) = \sum_{t=1}^{T} d_t\sqrt{2\{\log(n_t)-\log(d_t)\}}\{1+o_p(1)\}$$

By Lemma 6 and $d = C\lambda^\gamma$, it holds that

$$\frac{d_t\sqrt{2\{\log(n_t)-\log(d_t)\}}}{d\sqrt{\log(\lambda)}} \to_p \frac{1}{pT}\sqrt{2(1-\gamma)}$$

Consequently, we have

$$\frac{\sum_{t=1}^{T}(\lambda D)/T\Psi(\frac{d_t}{n_t})}{d/p\sqrt{2(1-\gamma)\log(\lambda)}} \to_p 1.$$

Using Lemma 7 again, we obtain

$$\frac{\mathsf{E}[\sum_{t=1}^{T}(\lambda D)/T\Psi(\frac{d_t}{n_t})]}{d/p\sqrt{2(1-\gamma)\log(\lambda)}} \to 1.$$

Compared with the mean term $d\mu$ with $\mu < \infty$, we conclude

$$\frac{r(T)}{\sigma d\sqrt{2(1-\gamma)\log(\lambda)}} \to 1.$$

Finally, by Theorem 4, $c(T)$ is of order $\sqrt{d}$. As a result, in all the three scenarios, $c(T)/r(T) \to 0$ as $\lambda \to \infty$.

■