The Optimal Selection of a Subset and the Roles of Testing

Lilun Du* Chenyin Gong† Qing Li‡

A firm needs to select applicants from an applicant pool to fill a number of identical job positions. Each applicant in the pool has an initial score that is related, albeit imperfectly, to their true qualification. The firm can select applicants based on these initial scores, but can also conduct tests to learn more about them. Tests are costly but produce additional signals about the applicants; however, the signals are also imperfectly related to their qualifications. How many and which applicants should the firm accept based solely on the initial scores, and how many and which applicants should the firm short-list for additional testing? Among those short-listed, how many and which applicants should the firm accept? How do the answers to these questions depend on the informativeness of the signals generated by the tests?

We develop a model framework for answering these important questions. We show that there are two cutoffs for the initial scores such that the applicants with initial scores higher than the upper cutoff should be accepted, those with initial scores lower than the lower cutoff should be rejected, and those with initial scores in between the cutoffs should be short-listed for additional testing. As the tests become more informative, the upper cutoff increases and the lower cutoff decreases; in other words, the firm accepts fewer applicants based solely on their initial scores, short-lists more for additional testing, and rejects a smaller number based on their initial scores. The short-listed applicants are ranked based on their initial and test scores, and how the scores affect the ranking depends on how they are related to the applicant qualifications. If one short-listed applicant is accepted, then all those who rank strictly higher must also be accepted.

We compare the optimal policy with two commonly used policies in practice: *screen-to-hire*, where no additional testing is conducted, and *test-to-hire*, where all accepted applicants must go through additional testing. Both are easier to compute than the optimal policy. Although they are suboptimal in general, they provide useful bounds, which facilitates the computation of the optimal policy. Our model involves a large state space with both integer and continuous variables, making the problem technically challenging. To compute the optimal policy, we use an approximation of the objective function that has theoretical performance guarantees. Overall, our model framework is general and is applicable to contexts such as high-volume recruitment, loan approvals, medical triage, and startup funding and investments.

Key words: sequential testing; structure of optimal policy; order statistics; optimal selection

---

* College of Business, City University of Hong Kong, Kowloon Tong, Hong Kong. Email: lilundu@cityu.edu.hk

† The HKUST Business School, Clear Water Bay, Kowloon, Hong Kong. Email: cgongad@connect.ust.hk

‡ ibid. Email: imqli@ust.hk

1. Introduction

Hiring people with the right qualifications is important to any organization. The process of hiring the right people, however, can be complex, costly and time consuming. For example, applications to master's programs accredited by the Association to Advance Collegiate Schools of Business (AACSB) increased by 48% between 2018 and 2024, rising from an average of 2,040 applications per institution to 3,013 (AACSB 2024), and according to data from Internet Collaborative Information Management Systems (iCIMS) in April 2024, there were an average of 30 applications per job opening across industries, with tech-related jobs experiencing a 45% increase in job openings, manufacturing a 31% increase, and healthcare a 26% increase over the past year (iCIMS 2024). The Society for Human Resource Management (SHRM) reports that the average hiring cost rose from \$4,129 in 2019 to \$4,700 in 2023, a 14% increase, with executive hiring costs averaging \$28,329 (Prokopets 2024). In addition, filling a position typically takes about two months (Navarra 2022).

In spite of considerable effort, hiring processes often lead to poor hiring outcomes. One reason for this is that the true qualifications of applicants are not known for sure at the time of hiring, and firms involved in hiring must rely on observable signals to infer their qualifications. The quantity and quality of the signals depend on how the process is managed and how many resources the firms are willing to invest. To streamline the process, firms often rely on sequential processes in which applicants are evaluated in multiple stages, and applicants who meet certain criteria proceed from one stage to the next. In a sequential process, the assessment in some stages may be automated, while that in other stages relies on human judgment. For example, earlier stages may use artificial intelligence tools to save costs and time; they are sometimes called pre-filtering or screening stages in practice. Screening can use features identified in application materials to generate screening scores, which may correlate with qualification. In later stages, applicants are interviewed by hiring managers or relevant personnel who can leverage their experience and personal judgment.

In this study, we consider a sequential process with two stages, which we call screening and testing. The firm needs to select applicants from an applicant pool to fill a number of identical job positions. After an initial screening process, each applicant has an initial score, which reflects, albeit imperfectly, the applicant's true qualification. The firm can accept applicants based on the score, but it can also conduct tests to learn more about the applicants before accept/reject decisions are made. These tests, however, are costly and the additional signals they yield are also imperfectly correlated with qualifications. In this environment, several questions are crucial. First, how many and which applicants should the firm accept based solely on the initial scores, and how many and which applicants should the firm short-list for further testing? Second, for those short-listed, how many and which applicants should the firm accept? Third, how does the informativeness of the signals generated by the tests affect the optimal policy?

Depending on how the first question is answered, there can be three policies in practice. The first, which we call *screen-to-hire*, is to base all reject/accept decisions solely on the initial scores and do no testing at all. The argument for this policy is that the screening scores are sufficiently informative, and perhaps more importantly, the size of the applicant pool is so large that testing is too costly. The second, which we call *test-to-hire*, is the opposite. That is, no one can be accepted unless they have been tested in the second stage. The argument for this policy is that testing, costly as it is, produces quality signals and hence is worth conducting. Among the many master's level programs offered by major business schools, some are adopting the first policy,[1] and others the second (Du and Li 2020). The third policy is somewhere between the first and second policy. That is, some applicants are accepted based only on their initial scores and some proceed to the second stage for further testing before accept/reject decisions are made. In late September 2014, LinkedIn's talent acquisition team needed to hire 100 employees within 60 days to align with their sales strategies. They developed a scoring system that allowed high-scoring applicants to bypass certain interview stages and medium-scoring applicants to go on to a second phone interview (Jedeikin 2015). This suggests that they are using a variation of the third policy.

The sequential process, as well as the possible policies that we described above, are important in recruitment. However, there are other business scenarios that share the same characteristics. For example, a population of start-ups approaches an investor such as Y Combinator or TechStars for funding. The investor has limited financial resources and can only fund a subset of start-ups. The investor first screens the start-ups by evaluating their business plans, market potential, and founding teams based on pitch decks or initial proposals. The results are summarized in initial scores. Based on the initial scores, the investor can immediately invest in some start-ups, reject others, and require the remaining start-ups to proceed to the next round, which involves a more detailed due diligence process. Other examples that share the same characteristics include bank loan approvals and medical triage.

Despite the broad applicability of the problem described above, there is little work done about it in the academic literature. We suspect that in practice, these important decisions are often based on gut feelings or elementary statistical analyses, as seen with LinkedIn's approach, and there are no rigorous decision tools to support. One possible reason is the complexity of the problem. To find the optimal subset in each stage, one must solve an optimization problem that involves both integer (number of applicants) and continuous (initial scores and signals from the tests) variables. In addition, when the population size is large, the problem has a large state space, which adds to the computational challenge.

[1] For example, some master's programs make selections primarily or entirely based on application materials. See, e.g., https://www.cityu.edu.hk/pg/taught-postgraduate-programmes/apply-now.

The objective of this study is to provide a model framework for a firm to answer the important questions that arise in the problem described above. We show that the optimal policy is characterized by a two-cutoff structure: applicants with initial scores above the upper cutoff are hired directly, those below the lower cutoff are rejected, and those in between the cutoffs are short-listed for additional testing. Therefore, the screen-to-hire and test-to-hire policies are in general suboptimal. As the test in the second stage becomes more informative, the firm should decrease the number of applicants accepted solely based on initial scores, increase the number short-listed for further testing, and reduce the number rejected before testing. These monotone comparative statics results are established under a general prediction model of applicant qualification.

We then analytically compare the optimal policy with the two commonly used test-to-hire and screen-to-hire policies, which are easier to compute than the optimal policy. The comparison leads to interesting bounds that can help us better understand why the two policies are suboptimal. These bounds are also useful in reducing the search region when computing the optimal policy. Finally, we compare the optimal policy with the two suboptimal policies numerically. As the optimal policy is not easily computable, we approximate the objective function. The approximation is accurate when the number of short-listed applicants is large, which is exactly the situation when approximation is needed the most. Our numerical results show that the optimal policy considerably outperforms the test-to-hire policy when the informativeness of the test is low, and it also demonstrates significant superiority over the screen-to-hire policy when the informativeness is high. These performance gaps persist even as the applicant pool increases in size.

The rest of the paper is organized as follows. In Section 2, we provide a discussion of the related literature. We formulate the problem in Section 3, and present the two-cutoff structure of the optimal policy in Section 4. In Section 5, we investigate the impact of the informativeness of the test on the optimal policy, and in Section 6, we compare the optimal policy with the suboptimal screen-to-hire and test-to-hire policies. Finally, approximations and numerical studies are presented in Section 7. All of the proofs are relegated to Appendix B.

2. Related Literature

Our study falls within the area of sequential assignment in the operations research/management literature. A primary focus in this field is the allocation of a fixed amount of resources to applicants who arrive randomly over time. Relevant studies include Vera and Banerjee (2021) on online resource allocation; Arlotto and Gurvich (2019) and Arnosti and Ma (2023) on secretary problems; Prastacos (1983) and Ahn et al. (2021) on opportunity and asset investment; and Vulcano et al. (2002) on dynamic auctions, among many others. These studies typically assume that applicants arrive one at a time and the decision maker must reject or accept them on the spot. In Li and

Yu (2021), however, applicants are assessed only at pre-determined fixed time intervals, and hence applicants are assessed in batches. Building on Li and Yu (2021), Du et al. (2024) extend the model to allow random yields. They propose asymptotically optimal algorithms and test them in a case study. In Gong and Li (2024), applicants depart the system after a random amount of time, and the firm must determine when to assess the applicants in the system and to whom to make offers. In the current study, all of the applicants are available at the beginning and have initial scores, and the firm has the option of conducting costly testing to obtain additional information about the applicants before making accept/reject decisions.

While our analysis is relevant to a broad range of business scenarios, it is specifically rooted in the context of job market hiring. Several studies examine the role of sequential testing in personnel selection. The earliest formulation of sequential testing was developed by Cronbach and Gleser (1965). Subsequent research explores various settings; for example, De Corte (1998) extends the framework to include a probationary period, and De Corte et al. (2006) consider a mixture of applicant groups and provide a comprehensive review of the literature on personnel selection. In addition, Du and Li (2020) present a statistical procedure for estimating the probability of false rejections; their goal is to keep this probability below a specified level. Multistage personnel selection problems share similarities with industrial inspection problems, where products, unlike applicants who have different qualifications, are either good or defective, and firms seek to balance inspection-repair costs with the costs associated with allowing defective units to reach end markets (see, e.g., Yao and Zheng 2002). Our study diverges from the existing research on sequential testing in personnel selection in two significant ways. First, most studies focus on determining to whom to extend offers after all of the tests have been completed, whereas we are also interested in how many and which applicants should be accepted immediately based on their initial scores, and how many and which applicants should be short-listed for further testing. Second, we investigate how the informativeness of the signals generated by tests influences the optimal policy.

Finally, our study is related to ranking and selection problems in the statistics and econometrics literature. For example, Gu and Koenker (2023) develop a nonparametric empirical Bayes approach in a compound decision framework to select a proportion of the best or worst populations, and Klein et al. (2020) and Mogstad et al. (2024) measure the uncertainty involved in estimating the ranks of true qualifications by constructing confidence sets for these ranks. For a detailed review of recent developments in this field, readers can refer to Mogstad et al. (2023). While the goal of these studies is usually to control the probability of correct or incorrect selection, ours is to maximize the aggregated value (or effect size, in statistical terms). In addition, in our study, the selection is made sequentially.

3. Problem Formulation

In the following, we use bold letters to represent vectors. For $\mathbf{s} = (s_1, s_2, \ldots, s_n)$, define $\mathbf{s}_{-j} = (s_1, \ldots, s_{j-1}, s_{j+1}, \ldots, s_n)$ and $(s_i)_{i\in I}$ the same as $\mathbf{s}$, but keep only the coordinates in $I \subset \{1, 2, \ldots, n\}$. For example, $(s_i)_{i\in\{1,n\}} = (s_1, s_n)$. For random vector $\mathbf{S} = (S_1, S_2, \ldots, S_n)$, let $S_{[i]}$, $i = 1, 2, \ldots, n$, be the ith order statistic of $\mathbf{S}$, which is defined as the ith largest element in $\mathbf{S}$.[2] Denote the size of a finite set $\mathcal{A}$ by $|\mathcal{A}|$. Let $x^+ = \max\{x, 0\}$ and $x^- = \max\{-x, 0\}$. Comparative adjectives such as "higher" and "lower" are used in a weak sense.

A firm needs to select applicants to fill d identical job positions. There are n applicants in the applicant pool. All of the applicants in the pool have gone through an initial assessment, and for applicant $i = 1, 2, \ldots, n$, let X_{i1} represent applicant i's initial score. The firm has the option of conducting an additional test, which costs c per applicant. For applicant $i = 1, 2, \ldots, n$, let X_{i2} be the test scores. The qualification of applicant i, which represents their value to the firm, is denoted by Y_i, and it is unknown but related to both the initial score and the test score. The firm can predict the qualification Y_i based on the initial score and the test score.[3] Given a random triplet $(X_1, X_2, Y) \in \mathbb{R}^3$, consider the following two regression functions:

$$f(X_1) = \mathbb{E}[Y \mid X_1] \quad \text{and} \quad g(X_1, X_2) = \mathbb{E}[Y \mid X_1, X_2]. \tag{1}$$

Assume that (X_{i1}, X_{i2}, Y_i) are independent and identically distributed (i.i.d.) copies of (X_1, X_2, Y). Then, X_{i1}, X_{i2}, and Y_i are associated as the following prediction model:

$$Y_i = f(X_{i1}) + \epsilon_i \quad \text{and} \quad Y_i = g(X_{i1}, X_{i2}) + \epsilon_i', \tag{2}$$

where $\mathbb{E}[\epsilon_i \mid X_{i1}] = \mathbb{E}[\epsilon_i' \mid X_{i1}, X_{i2}] = 0$. Because the errors ϵ_i and ϵ_i' are unpredictable, the regression functions $f(\cdot)$ and $g(\cdot, \cdot)$ are the optimal predictors (under squared error loss) of the qualification Y_i (Györfi et al. 2002), without and with further testing, respectively.[4] After observing the initial score X_{i1}, the firm can either use $f(X_{i1})$ to predict Y_i or conduct an additional test for more information. In the latter case, because X_{i1} has been observed, the predictor is $g(X_{i1}, X_{i2})$ conditionally on X_{i1}, denoted by $g(X_{i1}, X_{i2}) \mid X_{i1}$. The prediction model (2) is intimately connected to sequential or Type I analysis of variance (ANOVA), which provides a statistical test for the stepwise inclusion

[2] In the literature, the ith order statistic is normally defined as the ith smallest element.

[3] As an example, in the context of graduate student recruitment, X_{i1} can be a summary score computed based on relevant features identified in the application materials, X_{i2} interview score, and Y_i the GPA in the program (Du and Li 2020).

[4] Regression functions are often used to predict arm rewards in the literature on linear and contextual bandits (see, e.g., Goldenshluger and Zeevi 2013, Bastni and Bayati 2020). In the context of job market hiring, X_{i1} and X_{i2} can represent observable characteristics, and Y_i can be the applicants' skill levels. Each group of applicants (e.g., majority and minority groups) is linked to a specific linear regression function, and the firm predicts applicants' skills using these predictors. For a detailed discussion of this bandit setting, see Komiyama and Noda (2024).

of predictors in a regression model. This method allows for the assessment of the incremental contribution of each factor to the response.

We impose the following assumptions on the prediction model (2).

ASSUMPTION 1. *(i) $f(\cdot)$ is increasing. (ii) The error difference $\epsilon_i - \epsilon_i'$ is independent of X_{i1}.*

The first condition ensures a positive correlation between initial scores and qualifications. The requirement for independence between the error difference $\epsilon_i - \epsilon_i'$ and the initial score X_{i1} can be satisfied in many contexts, as illustrated in the examples below. Assumption 1 implies the more familiar concept of $g(X_{i1}, X_{i2})$ being *positively regression dependent* on X_{i1}; that is, the conditional probability $\mathbb{P}(g(X_{i1}, X_{i2}) \le y | X_{i1} = x)$ is decreasing in x for all y (Lehmann 1966). Positive regression dependence suggests that large (small) values of X_{i1} tend to be associated with large (small) values of $g(X_{i1}, X_{i2})$, which is relevant in our context. Lehmann (1966) provides several examples of distributions that satisfy this dependence, including bivariate normal, multinomial, and multiple hypergeometric distributions. Positive regression dependence is also frequently used in multiple hypothesis testing, such as in false discovery rate control where p-values or e-values satisfy positive regression dependence properties (see, e.g., Benjamini and Yekutieli 2001, Wang and Ramdas 2022).

Next, we introduce two specific examples and show how the assumptions hold in these examples.

EXAMPLE 1 (MULTIVARIATE NORMAL MODEL). Suppose that the random triplet (X_1, X_2, Y) follows a *multivariate normal model* with mean $\boldsymbol{\mu} = (\mu_{x_1}, \mu_{x_2}, \mu_y)$ and the covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{x_1}^2 & \sigma_{x_1,x_2} & \sigma_{x_1,y} \\ \sigma_{x_2,x_1} & \sigma_{x_2}^2 & \sigma_{x_2,y} \\ \sigma_{y,x_1} & \sigma_{y,x_2} & \sigma_y^2 \end{pmatrix}.$$

Using the recruitment data from a postgraduate business program, Du and Li (2020) confirm that a multivariate normal distribution is appropriate for their context. Assume that $\sigma_{x_1,y} \in \mathbb{R}_+$. In this case, the regression functions (1) can be decomposed into the following linear forms (Johnson and Wichern 2014, Result 4.6):

$$f(X_{i1}) = \mu_y + \frac{\sigma_{x_1,y}}{\sigma_{x_1}^2}(X_{i1} - \mu_{x_1}), \tag{3}$$

and

$$g(X_{i1}, X_{i2}) = \mu_y + \gamma_1(X_{i1} - \mu_{x_1}) + \gamma_2(X_{i2} - \mu_{x_2}), \tag{4}$$

where $\gamma_1 = (\sigma_{x_2}^2 \sigma_{x_1,y} - \sigma_{x_2,y}\sigma_{x_1,x_2})/(\sigma_{x_1}^2\sigma_{x_2}^2 - \sigma_{x_1,x_2}^2)$ and $\gamma_2 = (\sigma_{x_1}^2 \sigma_{x_2,y} - \sigma_{x_1,y}\sigma_{x_1,x_2})/(\sigma_{x_1}^2\sigma_{x_2}^2 - \sigma_{x_1,x_2}^2)$. In addition, X_{i2} is regressed on X_{i1} with the following linear form:

$$X_{i2} = \mathbb{E}(X_{i2}|X_{i1}) + \epsilon_i'' = \mu_{x_2} + \frac{\sigma_{x_1,x_2}}{\sigma_{x_1}^2}(X_{i1} - \mu_{x_1}) + \epsilon_i'', \tag{5}$$

where ϵ_i'' is normally distributed with a mean of zero and a variance of $\sigma_{x_2}^2 - \sigma_{x_1,x_2}^2/\sigma_{x_1}^2$. By (3)-(5) and some simple algebra, we have

$$g(X_{i1}, X_{i2}) = f(X_{i1}) + \gamma_2 \epsilon_i''.$$

It is clear that $f(\cdot)$ is increasing, both errors ϵ_i and ϵ_i' are independent of X_{i1}, and hence so is their difference $\gamma_2 \epsilon_i''$. □

Example 2 (Linear Probability Model). Consider a binary response of Y_i. For example, $Y_i = 1$ if employee i is skilled and $Y_i = 0$ if i is unskilled; or $Y_i = 1$ if applicant i eventually graduates and $Y_i = 0$ if i fails to graduate. In this case, the primary interest lies in the response probabilities $f(X_{i1})$ and $g(X_{i1}, X_{i2})$, which can be specified as the *linear probability model* (Wooldridge 2010):

$$f(X_{i1}) = \mathbb{P}(Y_i = 1 \mid X_{i1}) = \alpha_0 + \alpha_1 X_{i1},$$

and

$$g(X_{i1}, X_{i2}) = \mathbb{P}(Y_i = 1 \mid X_{i1}, X_{i2}) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2},$$

where $\alpha_1 \in \mathbb{R}_+$, and $\alpha_0, \beta_j \in \mathbb{R}$, $j = 0, 1, 2$. If $Y_i = 1$, then $\epsilon_i = 1 - f(X_{i1})$ with probability $f(X_{i1})$, and if $Y_i = 0$, then $\epsilon_i = -f(X_{i1})$ with probability $1 - f(X_{i1})$. Thus, given X_{i1}, ϵ_i has a mean of zero and a variance equal to $f(X_{i1})(1 - f(X_{i1}))$. The conditional distribution of ϵ_i' can be similarly described. The two random variables X_{i1} and X_{i2} are associated through the following regression

$$X_{i2} = \mathbb{E}[X_{i2} \mid X_{i1}] + \epsilon_i'',$$

where $\mathbb{E}[\epsilon_i'' \mid X_{i1}] = 0$. It can be shown that $g(X_{i1}, X_{i2}) = f(X_{i1}) + \beta_2 \epsilon_i''$. If ϵ_i'' and X_{i1} are independent, the linear probability model satisfies Assumption 1. While the error difference $\epsilon_i - \epsilon_i' = \beta_2 \epsilon_i''$ is independent of X_{i1}, unlike Example 1, here both errors ϵ_i and ϵ_i' are dependent on X_{i1}. □

From these examples, it can be seen that the prediction model (2) with Assumption 1 encompasses all of the linear regression models in which the regression function $f(\cdot)$ is linearly increasing, the regression function $g(\cdot, \cdot)$ is linear, and the test score X_{i2} can be expressed as an independent combination of the initial score X_{i1} and an error term. A discussion of the implications of relaxing Assumption 1 can be found in Appendix A.

According to Assumption 1(*ii*), the functional form of the conditional variance $\mathrm{Var}(g(X_{i1}, X_{i2}) \mid X_{i1})$ is a constant. We denote the variance as σ^2. In addition, by the law of iterated expectations, the conditional expectation is given by $\mathbb{E}[g(X_{i1}, X_{i2}) \mid X_{i1}] = f(X_{i1})$. Let $\mathbf{X}_1 = (X_{11}, X_{21}, \ldots, X_{n1})$ and its realization $\mathbf{x}_1 = (x_{11}, x_{21}, \ldots, x_{n1})$. Given $\mathbf{X}_1 = \mathbf{x}_1$, we define the following random variables:

$$S_i'(x_{i1}) = g(X_{i1}, X_{i2}) \mid X_{i1} = x_{i1} \sim (f(x_{i1}), \sigma^2), \ i = 1, 2, \ldots, n,$$

which are independent across i. Here, $S \sim (\mu, \sigma^2)$ means that random variable S has a mean of μ and a variance of σ^2. Let the realization of $S'_i(x_{i1})$ be $s'_i(x_{i1})$.

The timing of events is as follows. (1) The firm observes the initial scores of all of the applicants in the pool $\mathbf{X}_1 = \mathbf{x}_1$. With this information, the firm determines how many applicants to accept based solely on their initial scores, how many to short-list for additional testing, and who to accept and short-list. The expected value of accepting applicant i to the firm at this point is $f(x_{i1})$. (2) For those short-listed, the firm conducts testing to obtain the test results X_{i2}. The firm then determines again how many and whom to accept based on both the initial scores and test scores. The expected value of accepting applicant i to the firm at this point is $s'_i(x_{i1})$. (3) In the end, there is a penalty cost, measured by a convex function $G(\cdot)$, if the total number of applicants accepted, denoted by q, deviates from the hiring target d. The penalty cost can be defined as, for example, $G(q-d) = c_u(q-d)^- + c_o(q-d)^+$ for some positive marginal underage and overage costs, c_u and c_o, respectively. In practice, the volume of the applicant pool is typically much larger than the target. When the volume of the applicant pool is large enough, if we choose $c_u = c_o = \infty$, then the firm will always hire exactly d applicants.

Let $\mathcal{U} \subset \{1, 2, \ldots, n\}$ be the set of applicants accepted based sorely on the initial scores, and $\mathcal{Z} \subset \{1, 2, \ldots, n\} \setminus \mathcal{U}$ be the set of applicants short-listed for additional testing. After additional testing, let $\mathcal{H} \subset \mathcal{Z}$ be the set of applicants who are accepted. The objective of the firm is to maximize the total expected value of all of the accepted applicants, minus the cost of testing and the penalty cost in the end. The problem formulation is as follows:

$$
\begin{aligned}
(\mathcal{P}_0) \quad \max \quad & \sum_{i \in \mathcal{U}} f(x_{i1}) - c|\mathcal{Z}| + \mathbb{E}\left[\max_{\mathcal{H} \subset \mathcal{Z}} \left\{ \sum_{i \in \mathcal{H}} S'_i(x_{i1}) - G(|\mathcal{U}| + |\mathcal{H}| - d) \right\} \right] \\
\text{s.t.} \quad & \mathcal{U} \cup \mathcal{Z} \subset \{1, 2, \ldots, n\}, \\
& \mathcal{U} \cap \mathcal{Z} = \emptyset,
\end{aligned}
$$

where the expectation is taken with respect to the random test results. The maximization problem inside the expectation refers to the problem of selecting applicants to place on the offer list after the test is conducted.

It can be shown that, after additional testing, the optimal policy for determining the offer list $\mathcal{H}$ is a cutoff type. That is, there exists a cutoff such that any applicant i is on the offer list if and only if their expected value $s'_i(x_{i1})$ to the firm is no less than this cutoff. Then, the inner maximization problem can be written as

$$
\max_{\mathcal{H} \subset \mathcal{Z}} \left\{ \sum_{i \in \mathcal{H}} S'_i(x_{i1}) - G(|\mathcal{U}| + |\mathcal{H}| - d) \right\} = \max_{0 \le h \le |\mathcal{Z}|} \left\{ \sum_{i=1}^{h} S'_{[i]}(\mathbf{x}_1, \mathcal{Z}) - G(|\mathcal{U}| + h - d) \right\}, \tag{6}
$$

where $S'_{[i]}(\mathbf{x}_1, \mathcal{Z})$ is ith order statistic of $(S'_i(x_{i1}))_{i \in \mathcal{Z}}$.

4. The Two-Cutoff Policy

We first establish the following properties of the second-stage optimization problem.

LEMMA 1. *For any $i \in \mathcal{Z}$, we have*

$$0 \le \nabla_{x_{i1}} \mathbb{E}\left[\max_{0\le h\le|\mathcal{Z}|}\left\{\sum_{j=1}^{h} S'_{[j]}(\mathbf{x}_1,\mathcal{Z}) - G(|\mathcal{U}|+h-d)\right\}\right] \le \nabla f(x_{i1}).$$

For the first inequality, the total expected value to the firm is increasing in x_{i1}, because the qualification of applicant i is positively correlated to his or her initial score. The second inequality holds because only some short-listed applicants are accepted eventually.

With Lemma 1 as a building block, we can show how the firm should determine how many and which applicants to accept, and how many and which applicants to short-list for further testing, based on the initial scores. In particular, the optimal policy of determining $\mathcal{U}$ and $\mathcal{Z}$ is a two-cutoff policy.

THEOREM 1. *For Problem $(\mathcal{P}_0)$, before further testing is conducted, there exist two cutoffs such that the initial scores of all of the hired applicants must be higher than the upper cutoff, and the initial scores of all of the rejected applicants must be lower than the lower cutoff.*

As a special case, when several applicants have the same initial scores, under the optimal policy, it is possible that among the applicants with the same initial score, some are accepted and others are short-listed or some are short-listed and others are rejected. In the following, we let $u = |\mathcal{U}|$ and $z = |\mathcal{Z}|$. In a slight abuse of notation, let $\mathbf{S}(\mathbf{x}_1, u, z) = (S_{u+1}, S_{u+2}, \ldots, S_{u+z})$ be the random vector of the values of the short-listed applicants. Here, the ith element corresponds to the value of the applicant who has the $(u+i)$th highest initial score. For example, if u applicants have been accepted without further testing, and applicant j's initial score is the $(u+i)$th $(1 \le i \le z)$ highest, then he or she will be short-listed for further testing with the value $S'_j(x_{j1}) = S_{u+i}$. The mean of S_{u+i} is $f(x_{[u+i]1})$ and the variance is σ^2, and S_{u+i} are independent across i. We let $\mu_i = f(x_{[i]1})$ for notational convenience. Because $f(\cdot)$ is increasing, μ_i decreases with i. In addition, let the realization of $\mathbf{S}(\mathbf{x}_1, u, z)$ be $\mathbf{s}(\mathbf{x}_1, u, z) = (s_{u+1}, s_{u+2}, \ldots, s_{u+z})$.

Now, let $S_{[i]}(\mathbf{x}_1, u, z)$ represent the ith order statistic of $\mathbf{S}(\mathbf{x}_1, u, z)$. In the following, we suppress $\mathbf{x}_1$ in the notation when there is no risk of confusion. According to Theorem 1, we rewrite Problem $(\mathcal{P}_0)$ as follows:

$$\begin{aligned}(\mathcal{P}_1)\quad \max\quad & \sum_{i=1}^{u}\mu_i - cz + \mathbb{E}\left[\max_{0\le h\le z}\left\{\sum_{i=1}^{h} S_{[i]}(u,z) - G(u+h-d)\right\}\right]\\ \text{s.t.}\quad & u \in \{0,1,\ldots,n\},\\ & z \in \{0,1,\ldots,n-u\},\end{aligned}$$

where the expectation is taken with respect to $\mathbf{S}(u,z)$. Let (u^*,z^*) be the optimal solution. When there are multiple maximizers, (u^*,z^*) is defined as the largest in lexicographical order. For convenience, we also denote $k=u+z$ ($k^*=u^*+z^*$) as the total (optimal) number of applicants who are either accepted based on initial scores or short-listed for further testing. Furthermore, define

$$F_z(u,\mathbf{s})=\max_{0\le h\le z}\left\{\sum_{i=1}^{h}s_{[i]}-G(u+h-d)\right\},\ \forall \mathbf{s}=(s_1,s_2,\ldots,s_z)\in\mathbb{R}^z, \tag{7}$$

where the subscript z in $F_z(u,\mathbf{s})$ means that the second argument $\mathbf{s}$ has z dimensions, and $s_{[i]}$ is the ith largest element in $\mathbf{s}$. Denote by h^* the largest maximizer. The goal of (7) is to find a subset of short-listed applicants to accept after the test results are observed.

To proceed, we first analyze (7). Recall that $\mathbf{s}_{-i}(u,z)=(s_{u+1},\ldots,s_{u+i-1},s_{u+i+1},\ldots,s_{u+z})$. For any $i\in\{1,2,\ldots,z\}$, we concatenate the vector $\mathbf{s}_{-i}(u,z)$ and the vector of z marginal values of $G(\cdot)$ to obtain a new vector $\hat{\mathbf{s}}(u,z,i)=(\hat{s}_1,\hat{s}_2,\ldots,\hat{s}_{2z-1})=(\mathbf{s}_{-i}(u,z),\Delta G(u+1-d),\Delta G(u+2-d),\ldots,\Delta G(u+z-d))$, where $\Delta G(h-d)=G(h-d)-G(h-1-d)$. We show in the following lemma that $F_z(\cdot,\cdot)$ in (7) has a special functional form.

Lemma 2. *The function $F_z(u,\mathbf{s}(u,z))$, which is defined in (7), can be written as follows:*

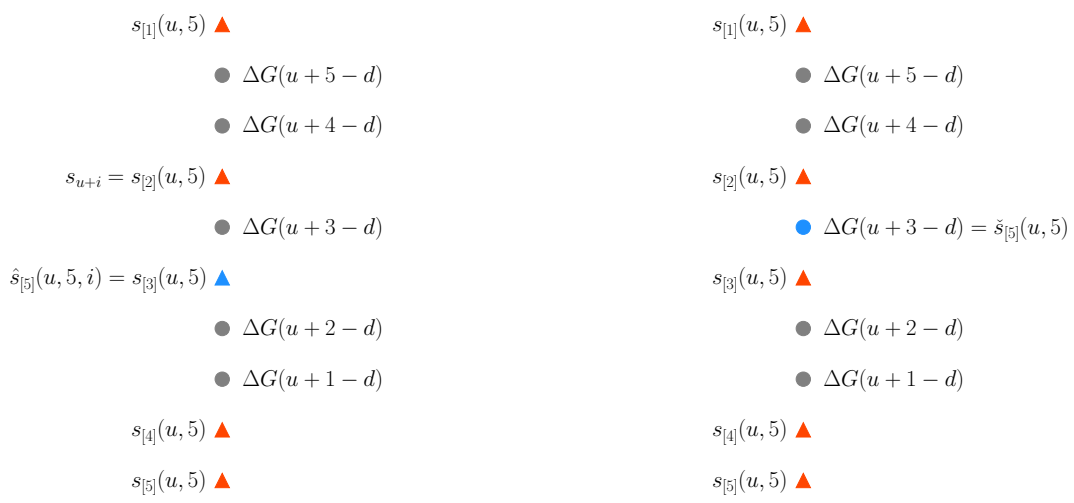
$$F_z(u,\mathbf{s}(u,z))=(s_{u+i}-\hat{s}_{[z]}(u,z,i))^+ +F_{z-1}(u,\mathbf{s}_{-i}(u,z)), \tag{8}$$

where $\hat{s}_{[z]}(u,z,i)$ is the zth largest element in $\hat{\mathbf{s}}(u,z,i)$ or the median of $\hat{\mathbf{s}}(u,z,i)$, and $F_0(u,\cdot)=-G(u-d)$.

The function $F_z(u,\mathbf{s}(u,z))$ can be iteratively expressed as the sum of a piecewise linear function of s_{u+i} and a term that is independent of s_{u+i}. The score s_{u+i} is included in the optimal score set (i.e., the set of scores of the applicants in the offer list) if and only if $F_z(u,\mathbf{s}(u,z))$ depends on s_{u+i}. This means that s_{u+i} is in the optimal score set if and only if it is larger than $\hat{s}_{[z]}(u,z,i)$. For example, in Figure 1(a), with z set to 5, the vector $\hat{\mathbf{s}}(u,5,i)=(\mathbf{s}_{-i}(u,5),\Delta G(u+1-d),\Delta G(u+2-d),\ldots,\Delta G(u+5-d))$ is arranged in descending order from top to bottom. Here, s_{u+i} is assumed to be the second largest element in $\mathbf{s}(u,5)$. Because s_{u+i} exceeds the fifth largest element in $\hat{\mathbf{s}}(u,5,i)$, which is $\hat{s}_{[5]}(u,5,i)=s_{[3]}(u,5)$, it follows that s_{u+i} should be included in the optimal score set. In summary, Lemma 2 provides us an efficient way to determine whether s_{u+i} should be included in the optimal score set.

We can also apply a similar procedure to find all s_{u+i}'s that should be included in the optimal score set. (1) We sort the concatenated vector $\check{\mathbf{s}}(u,z)=(\check{s}_1,\check{s}_2,\ldots,\check{s}_{2z})=(\mathbf{s}(u,z),\Delta G(u+1-d),\Delta G(u+2-d),\ldots,\Delta G(u+z-d))$ in descending order and (2) pick all s_{u+i} with $s_{u+i}\ge\check{s}_{[z]}(u,z)$, where $\check{s}_{[z]}(u,z)$ is the zth largest element in $\check{\mathbf{s}}(u,z)$. These picked elements, and these picked elements only, should be in the optimal score set. Indeed, because $\check{\mathbf{s}}(u,z)$ has one more element,

Figure 1 Determining the Optimal Score Set



(a) Whether a score is in the set (b) Determining the whole set

Note. We set $z = 5$. The elements are arranged in descending order from top to bottom.

s_{u+i}, than $\hat{\mathbf{s}}(u,z,i)$, we have $\check{s}_{[z]}(u,z) \geq \hat{s}_{[z]}(u,z,i)$. Then, according to Lemma 2, $s_{u+i} \geq \check{s}_{[z]}(u,z) \geq \hat{s}_{[z]}(u,z,i)$ implies that s_{u+i} should be included in the optimal score set. In turn, if there exist some s_{u+i} included in the optimal score set such that $s_{u+i} < \check{s}_{[z]}(u,z)$, then we must have $\check{s}_{[z]}(u,z) = \hat{s}_{[z]}(u,z,i)$ because the largest z elements in $\check{\mathbf{s}}(u,z)$ and $\hat{\mathbf{s}}(u,z,i)$ are the same. This implies that $s_{u+i} < \hat{s}_{[z]}(u,z,i)$, which contradicts Lemma 2. For example, in Figure 1(b), $\Delta G(u+3-d)$ is the cutoff $\check{s}_{[5]}(u,5)$. Therefore, the two largest elements, $s_{[1]}(u,5)$ and $s_{[2]}(u,5)$, in $\mathbf{s}(u,5)$ are included in the optimal score set.

As our discussion unfolds, it will become clear that Lemma 2 is very useful in analyzing our problem. Recall that $k = u + z$ refers to the total number of applicants who are either accepted based on initial scores or short-listed for further testing. By replacing z with $k-u$ in Problem $(\mathcal{P}_1)$, the new constraint set $\{0,1,\ldots,n\} \times \{u, u+1, \ldots, n\}$ of (u,k) is a lattice. As an immediate consequence of Lemma 2, the objective function of Problem $(\mathcal{P}_1)$ is shown to be submodular in (u,k), which is needed to prove Theorems 2 and 3.

LEMMA 3. *$F_{k-u}(u, \mathbf{s}(u,k-u))$ is submodular in (u,k).*

Because submodularity is preserved under expectation, it is easy to see that the objective function of Problem $(\mathcal{P}_1)$ is also submodular in (u,k). The submodularity implies that the number of applicants hired based solely on their initial scores, u, and the total number of those who are either accepted based on initial scores or short-listed for further testing, k, are economic substitutes, regardless of score distributions. In Section 5, we show that u and the number of applicants short-listed for additional testing, z, are also economic substitutes.

5. The Informativeness of the Test

In this section, we examine how the informativeness of the test affects the optimal policy. The informativeness of the test is measured by variance. We show how the optimal policy changes when the test becomes more informative. The proof, which has strong intuitive appeal, is outlined at the end of this section.

To investigate how the informativeness of the test affects the optimal policy, we need to first define what is meant by informativeness. A natural option is to use the variance σ^2 of S_{u+i}. To see this, we first expand the conditional variance:

$$\mathrm{Var}(g(X_{i1}, X_{i2}) \mid X_{i1}) = \mathbb{E}\left[\epsilon_i^2 \mid X_{i1}\right] - \mathbb{E}\left[\epsilon_i'^2 \mid X_{i1}\right]. \tag{9}$$

The derivation of (9) is provided in Appendix C. By (9) and the law of total expectations, we have $\sigma^2 = \mathbb{E}\left[\mathrm{Var}(g(X_{i1}, X_{i2}) \mid X_{i1})\right] = \mathrm{Var}(\epsilon_i) - \mathrm{Var}(\epsilon_i')$. If the test score becomes more informative, then the regression function $g(X_{i1}, X_{i2})$ has more power for predicting the qualification Y_i. This implies that the error variance $\mathrm{Var}(\epsilon_i')$ should be smaller, and therefore σ^2 should be larger. The conditional variance (9) can also be interpreted as the reduction in the mean squared error (MSE) in sequential ANOVA. This reduction in MSE quantifies the degree to which the test score explains the variability of the qualifications, or alternatively the reduction in error achieved by testing. Motivated by these observations, we say that *the test becomes more informative when σ^2 is larger*.

To better understand how the variance can change, let us consider the multivariate normal model in Example 1. We have derived the regression $g(X_{i1}, X_{i2}) = f(X_{i1}) + \gamma_2\epsilon_i''$. The variance of $\gamma_2\epsilon_i''$ is $\gamma_2^2(1-\rho_{x_1,x_2}^2)\sigma_{x_2}^2$, where $\rho_{x_1,x_2} = \sigma_{x_1,x_2}/(\sigma_{x_1}\sigma_{x_2})$ is the Pearson correlation coefficient of X_{i1} and X_{i2}. The Pearson correlation coefficient of X_{i2} and Y_i is $\rho_{x_2,y} = \sigma_{x_2,y}/(\sigma_{x_2}\sigma_y)$. We keep the diagonal of the covariance matrix $\mathbf{\Sigma}$ unchanged and assume that $g(\cdot,\cdot)$ is also increasing for the sake of simplicity. The partial derivatives of the variance of $\gamma_2\epsilon_i''$ with respect to $\sigma_{x_2,y}$ and σ_{x_1,x_2} are

$$\frac{\partial \mathrm{Var}(\gamma_2\epsilon_i'')}{\partial \sigma_{x_2,y}} = 2\gamma_2 \geq 0 \quad \text{and} \quad \frac{\partial \mathrm{Var}(\gamma_2\epsilon_i'')}{\partial \sigma_{x_1,x_2}} = -2\gamma_1\gamma_2 \leq 0.$$

If the test score and the qualification are more correlated, i.e., $\rho_{x_2,y}$ is closer to one, $\sigma^2 = \mathrm{Var}(g(X_{i1}, X_{i2}) \mid X_{i1}) = \mathrm{Var}(\gamma_2\epsilon_i'')$ is larger; similarly if the test score and the initial score are less correlated, i.e., ρ_{x_1,x_2} is closer to zero, σ^2 is larger. In summary, in the multivariate normal model, the test is more informative if the test score is more correlated to the qualification or if it is less correlated to the initial score. The former indicates a boost in the predictive accuracy of the test score on its own, and the latter indicates an increase in the amount of new information provided by the test.

The random variable S_{u+i} can be expressed as a function of its mean and variance:

$$S_{u+i} = \mu_{u+i} + \sigma\varepsilon_{u+i},\ i = 1, 2, \ldots, z, \tag{10}$$

where $\varepsilon_i \sim (0,1)$ are i.i.d. and independent of $\mathbf{X}_1$. Let $\phi_i(\cdot)$ be the probability density function (PDF) of the error ε_i. We present the main result in this section.

THEOREM 2. *Suppose that $\phi_i(\varepsilon_i) = \phi_i(-\varepsilon_i)$ for all $\varepsilon_i \geq 0$. For Problem $(\mathcal{P}_1)$, when the test result becomes more informative, u^* is smaller, and both z^* and k^* are larger.*

The condition in the theorem imposes a symmetry assumption on the error distribution. Symmetric errors, such as Gaussian errors, are widely adopted in the literature (see, e.g., Bickel 1982, Chae et al. 2019) due to their substantial computational and theoretical benefits. The multivariate normal model in Example 1 satisfies this symmetry condition. For many non-symmetric models, such as linear regressions with log-normal responses, power transform techniques like Box–Cox transformation (Box and Cox 1964) can be applied to achieve symmetry.

To show Theorem 2, a standard approach is to show that the objective function of Problem $(\mathcal{P}_1)$ is supermodular in $(\sigma, -u, k)$ and the constraint set is a lattice. This is a challenging undertaking because the constraint set is not a lattice, the prediction model (2) is general, and the parameter σ^2 is associated with the distributions of *all* z non-identical random variables S_{u+i}. There are two important ideas in our proof.

First, we change the decision variable z to $k-u$, where k is the total number of applicants who are either accepted solely based on initial scores or short-listed for additional testing. We then relax the range of k from $\{u, u+1, \ldots, n\}$ to $\{0, 1, \ldots, n\}$, and replace z in the objective function by $(k-u)^+$. The redefined problem is

$$\begin{aligned} \max \quad & \sum_{i=1}^{u} \mu_i - c(k-u)^+ + \mathbb{E}\left[F_{(k-u)^+}(u, \mathbf{S}(u, (k-u)^+))\right] \\ \text{s.t.} \quad & -u \in \{-n, -n+1, \ldots, 0\}, \\ & k \in \{0, 1, \ldots, n\}. \end{aligned}$$

It can be shown that the new formulation is equivalent to the original one and the feasible region is a lattice.

To show the theorem, we prove that the new objective function is pairwise supermodular in $(\sigma, -u, k)$. To illustrate the second important idea in our proof, we use the proof of the supermodularity in (σ, k) as an example. Based on Lemma 2, the marginal value of k has the following form (for ease of exposition, we only consider the case $k > u$ here):

$$\mathbb{E}\left[F_{k-u}(u, \mathbf{S}(u, k-u))\right] - \mathbb{E}\left[F_{k-1-u}(u, \mathbf{S}(u, k-1-u))\right] = \mathbb{E}\left[(S_k - \hat{S}_{[k-u]}(u, k-u, k-u))^+\right], \quad (11)$$

where $\hat{S}_{[k-u]}(u, k-u, k-u)$ is the $(k-u)$th order statistic of the random vector $\hat{\mathbf{S}}(u, k-u, k-u) = (S_{u+1}, S_{u+2}, \ldots, S_{k-1}, \Delta G(u+1-d), \Delta G(u+2-d), \ldots, \Delta G(k-d))$, which is defined in Section 4.

That is, we transform the difference of two maximization problems into a much simpler convex function of the random variable S_k subtracting an order statistic. Because only the mean-variance information for each random variable is available, we further replace each S_{u+i} with the linear form $\mu_{u+i}+\sigma\varepsilon_{u+i}$ (see (10)) and obtain

$$\mathbb{E}\left[(S_k-\hat{S}_{[k-u]}(u,k-u,k-u))^+\right]=\mathbb{E}\left[(\mu_k+\sigma\varepsilon_k-\hat{S}_{[k-u]}(u,k-u,k-u))^+\right]. \tag{12}$$

We then show that (12) is increasing in σ.

The proof of the supermodularity in $(\sigma,-u)$ is similar. The marginal value of $-u$ is

$$\begin{aligned}&\mathbb{E}\left[F_{k-u}(u,\mathbf{S}(u,k-u))\right]-\mathbb{E}\left[F_{k-u-1}(u+1,\mathbf{S}(u+1,k-u-1))\right]\\ =\;&\mathbb{E}\left[\max\left\{S_{u+1},\hat{S}_{[k-u]}(u,k-u,u+1)\right\}\right],\end{aligned} \tag{13}$$

where $\hat{S}_{[k-u]}(u,k-u,u+1)$ is the $(k-u)$th order statistic of the random vector $\hat{\mathbf{S}}(u,k-u,u+1)=(S_{u+2},S_{u+3},\ldots,S_k,\Delta G(u+1-d),\Delta G(u+2-d),\ldots,\Delta G(k-d))$. The derivation of (13) also relies on Lemma 2.

Theorem 2 shows that if the test result is more informative, the firm should short-list more applicants for additional testing and accept fewer based solely on their initial scores. In addition, the total number of applicants who are either accepted based on initial scores or short-listed for further testing is larger (i.e., the number of applicants rejected based on initial scores is smaller). This, interestingly, means that when $z^*>0$, the optimal number of short-listed applicants is more sensitive to the changes in the informativeness of the test than the optimal number of applicants accepted in the first stage .

6. Two Suboptimal Policies

In this section, we consider two formulations that correspond to two commonly adopted policies in practice and are special cases of Problem $(\mathcal{P}_1)$. In the first formulation, the firm must make accept/reject decisions solely based on the initial scores, which corresponds to the screen-to-hire policy; that is,

$$\begin{aligned}(\mathcal{P}_0^u)\quad \max\quad &\sum_{i\in\mathcal{U}}f(x_{i1})-G(|\mathcal{U}|-d)\\ \text{s.t.}\quad &\mathcal{U}\subset\{1,2,\ldots,n\}.\end{aligned}$$

In the second formulation, the firm must conduct further testing before an applicant can be accepted, which corresponds to the test-to-hire policy; that is,

$$\begin{aligned}(\mathcal{P}_0^z)\quad \max\quad &-c|\mathcal{Z}|+\mathbb{E}\left[\max_{\mathcal{H}\subset\mathcal{Z}}\left\{\sum_{i\in\mathcal{H}}S_i'(x_{i1})-G(|\mathcal{H}|-d)\right\}\right]\\ \text{s.t.}\quad &\mathcal{Z}\subset\{1,2,\ldots,n\}.\end{aligned}$$

We summarize the optimal policies of determining $\mathcal{U}$ for Problem $(\mathcal{P}_0^u)$ and $\mathcal{Z}$ for Problem $(\mathcal{P}_0^z)$ in the following proposition.

PROPOSITION 1. *(i) For Problem $(\mathcal{P}_0^u)$, there exists a cutoff such that the initial scores of all of the rejected applicants must be lower than the cutoff; (ii) For Problem $(\mathcal{P}_0^z)$, there exists a cutoff such that the initial scores of all of the applicants not short-listed must be lower than the cutoff.*

For Problem $(\mathcal{P}_0^z)$, Proposition 1 requires Assumption 1 to hold. The fact that cutoff policies may not be optimal in general has been discussed in the literature. In the literature on screening, for example, the monotone likelihood ratio (MLR) property has been identified as a sufficient condition for the optimality of a cutoff policy (see, e.g., Lagziel and Lehrer 2019, Koren 2024). The MLR property, like Assumption 1, implies positive regression dependence. Whether optimal or not, cutoff policies serve as a useful tool for applicant selection, as they are straightforward to implement and promote a level playing field for transparent selection procedures.

Based on Proposition 1, we respectively reformulate Problems $(\mathcal{P}_0^u)$ and $(\mathcal{P}_0^z)$ as follows:

$$\begin{aligned} (\mathcal{P}_1^u) \quad \max \quad & \sum_{i=1}^{u} \mu_i - G(u-d) \\ \text{s.t.} \quad & u \in \{0,1,\ldots,n\}, \end{aligned}$$

and

$$\begin{aligned} (\mathcal{P}_1^z) \quad \max \quad & -cz + \mathbb{E}\left[F_z(0,\mathbf{S}(0,z))\right] \\ \text{s.t.} \quad & z \in \{0,1,\ldots,n\}, \end{aligned}$$

where $\mu_i = f(x_{[i]1})$, and $F_z(\cdot,\cdot)$ is as defined in (7). Let u' and z' denote the largest optimal solutions of Problems $(\mathcal{P}_1^u)$ and $(\mathcal{P}_1^z)$, respectively. Problem $(\mathcal{P}_1^u)$ is easy to solve because the objective function is discrete concave in u. In Section 7, we provide an explicit formula for u'. We can also show that Problem $(\mathcal{P}_1^z)$ is a discrete concave optimization. Specifically, we show that $\mathbb{E}\left[F_z(u,\mathbf{S}(u,z))\right]$ is discrete concave in z under the usual stochastic ordering condition. For two random variables V and W, if $\mathbb{P}(V>x) \geq \mathbb{P}(W>x)$ for all $x \in \mathbb{R}$, then V is said to be *larger than W in the usual stochastic order*, denoted by $V \geq_{\text{st}} W$ (Shaked and Shanthikumar 2007).

LEMMA 4. *If $S_{u+1} \geq_{\text{st}} S_{u+2} \geq_{\text{st}} \cdots \geq_{\text{st}} S_n$, then $\mathbb{E}\left[F_z(u,\mathbf{S}(u,z))\right]$ is discrete concave in z and has decreasing differences in (u,z).*

To see that the usual stochastic ordering condition holds in our setting, recall that $S_i = \mu_i + \sigma\varepsilon_i$, where μ_i decreases with i and ε_i are i.i.d. across i. This implies that $S_i - (\mu_i - \mu_j)$ and S_j are equal in distribution, and thus, $\mathbb{P}(S_i \geq x) = \mathbb{P}(S_i - (\mu_i - \mu_j) \geq x - (\mu_i - \mu_j)) = \mathbb{P}(S_j \geq x - (\mu_i - \mu_j)) \geq$

$\mathbb{P}(S_j \geq x)$ for any $x \in \mathbb{R}$ and $j \geq i$. The concavity property is useful for computing the optimal policy, but it is also needed to prove Theorem 3. The property of decreasing differences implies that the number of applicants hired based solely on their initial scores, u, and the number of applicants short-listed for further testing, z, are economic substitutes. It can also be shown that when the test becomes more informative, the firm adopting the test-to-hire policy will short-list more applicants for further testing.

We now present the main result in this section, which establishes the connections among the optimal policy, the screen-to-hire policy, and the test-to-hire policy.

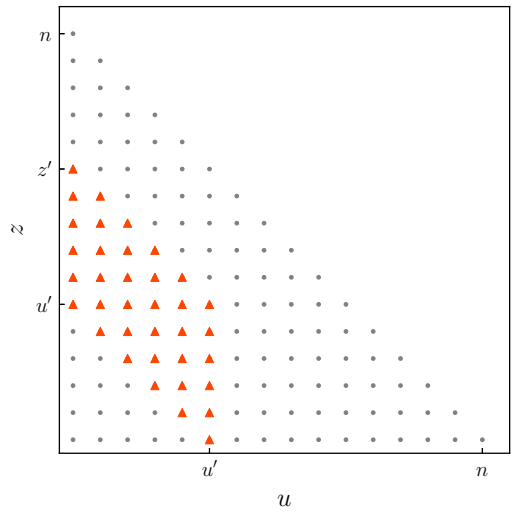
THEOREM 3. *(i)* $u^* \leq u' \leq u^* + z^*$*; (ii) If* $z^* > 0$*, then* $u^* + z^* \leq z'$.

Relative to the screen-to-hire policy, the firm adopting the optimal policy would accept fewer applicants in the first stage (i.e., $u^* \leq u'$), because it also accepts applicants in the second stage. Furthermore, because it plans to accept only some of the short-listed applicants, we naturally expect $u' \leq u^* + z^*$. Part (ii) is obvious only in hindsight. Under the test-to-hire policy, the firm short-lists more applicants than $u^* + z^*$ because even some of the top u^* applicants based on their initial scores may be rejected in the second stage, while under the optimal policy, the top u^* applicants receive offers. The firm needs to increase the number of short-listed applicants to account for that possibility. Parts (i) and (ii) of the theorem together imply that if $z^* > 0$, then $z' \geq u^* + z^* \geq u'$, or if $z' < u'$, then $z^* = 0$. This means that if $z' < u'$, then the screen-to-hire policy is actually optimal.

Theorems 3 can be viewed in terms of the informativeness of the test. Specifically, when the informativeness of the test is extremely low, the firm will not conduct any testing, and so the screen-to-hire policy is optimal. In this case, we have $u^* = u'$, $z^* = 0$, and $k^* = u^* + z^* = u'$. As the informativeness increases, according to Theorem 2, u^* decreases and k^* increases. Therefore, we have $u^* \leq u' \leq u^* + z^*$. When the cost of testing is sufficiently low, the firm will short-list all of the applicants for the second stage, and the test-to-hire policy is optimal. In this scenario, we have $u^* = 0$, $z^* = z'$, and $k^* = z'$. Again based on Theorem 2, as testing becomes less informative, u^* increases, and k^* and z' decrease. Under the optimal policy, u^* is positive, and under the test-to-hire policy, no applicants are accepted in the first stage. Suppose that we optimize k and u sequentially. The quantity $u^* + z^*$ is the optimal k when $u = u^*$ and z' is the optimal k when $u = 0$. We have $u^* + z^* \leq z'$ because k and u are substitutes (Lemma 3).

We also visualize the result for Theorem 3 in Figure 2. The optimal policy (u^*, z^*) is located within the parallelogram marked by triangles. This reduced search region can be much smaller than the feasible region represented by the triangular area marked with dots, which is useful computationally. Take AACSB-accredited master's programs. For the 2018 to 2024 period, the

Figure 2 **Search Region of the Optimal Policy** (u^*, z^*)



Note. The parallelogram region marked by triangles denotes the reduced search region based on Theorem 3, and the triangular region marked by dots denotes the whole feasible region.

average acceptance rate (i.e., the percentage of applicants who are offered admission by a school) is approximately 42% (AACSB 2024). Assuming that most programs adopt the screen-to-hire policy, we have $u' \approx 0.42n$. If we set the testing rate to be 60%, i.e., $z' = 0.6n$, the reduced search region only accounts for 15% of the feasible region. In Section 7, we conduct approximations within this reduced search region.

7. Approximations

In this section, we develop methods for computing the optimal policy (u^*, z^*) approximately. We start by proposing an efficient approximation for the objective function of Problem $(\mathcal{P}_1)$. We then reduce the search region using the results established in Section 6. Finally, we evaluate the performance of our method and compare the performance of the two suboptimal policies against that of the optimal policy numerically.

7.1. Approximating the Objective Function

To approximate the objective function of Problem $(\mathcal{P}_1)$, we focus on the expected value of the second-stage problem, $\mathbb{E}[F_z(u, \mathbf{S}(u, z))]$. According to Lemma 2, this expected value can be iteratively expressed as the sum of z convex functions involving order statistics. One possible approach is to directly approximate these order statistics. For example, we can apply the bounds on the expected value of order statistics provided by Arnold and Groeneveld (1979) and Bertsimas et al. (2006), which only require the mean-variance information for general distributions. In this section, however, we propose an alternative method based on concentration bounds to approximate the second-stage problem, which is shown to be efficient and accurate in our context.

As alluded to earlier, we iteratively expand $F_z(u, \mathbf{S}(u,z))$ from the zth element to the first in $\mathbf{S}(u,z)$ as follows:

$$\begin{aligned} F_z(u,\mathbf{S}(u,z)) &= (S_{u+z} - \hat{S}_{[z]}(u,z,z))^+ + F_{z-1}(u,\mathbf{S}(u,z-1)) \\ &= (S_{u+z} - \hat{S}_{[z]}(u,z,z))^+ + (S_{u+z-1} - \hat{S}_{[z-1]}(u,z-1,z-1))^+ + F_{z-2}(u,\mathbf{S}(u,z-2)) \\ &\cdots \\ &= \sum_{i=1}^{z} (S_{u+i} - \hat{S}_{[i]}(u,i,i))^+ - G(u-d), \end{aligned} \tag{14}$$

where $\hat{S}_{[i]}(u,i,i)$ is the ith order statistic or the median of the $(2i-1)$-dimensional random vector $(S_{u+1}, S_{u+2}, \ldots, S_{u+i-1}, \Delta G(u+1-d), \Delta G(u+2-d), \ldots, \Delta G(u+i-d))$.

Next, we let $\tilde{\mathbf{S}}(u,z) = (\tilde{S}_{u+1}, \tilde{S}_{u+2}, \ldots, \tilde{S}_{u+z}) = (g(X_{i1}, X_{i2}))_{i\in\mathcal{Z}}$, where $g(X_{i1}, X_{i2})$ is the regression function as defined in the prediction model (2). The random vector $\tilde{\mathbf{S}}(u,z)$ represents the unsorted version of $\mathbf{S}(u,z)$ with the initial scores being random and $X_{[u]1} \geq X_{i1} \geq X_{[u+z]1}$ for all $i \in \mathcal{Z}$. In a slight abuse of notation, we continue to use $\hat{S}_{[i]}(u,i,i)$ to denote the median of the random vector $(\tilde{S}_{u+1}, \tilde{S}_{u+2}, \ldots, \tilde{S}_{u+i-1}, \Delta G(u+1-d), \Delta G(u+2-d), \ldots, \Delta G(u+i-d))$. According to (14), the objective function of Problem $(\mathcal{P}_1)$ can be expressed as follows:

$$\sum_{i=1}^{u} \mu_i - cz + \mathbb{E}\left[F_z(u,\mathbf{S}(u,z))\right] = \sum_{i=1}^{u} \mu_i - cz - G(u-d) + \mathbb{E}\left[\mathcal{G}_z(u,\tilde{\mathbf{S}}(u,z)) \mid \mathbf{X}_1 = \mathbf{x}_1\right], \tag{15}$$

where

$$\mathcal{G}_z(u,\tilde{\mathbf{S}}(u,z)) = \sum_{i=1}^{z} (\tilde{S}_{u+i} - \hat{S}_{[i]}(u,i,i))^+. \tag{16}$$

In addition, for two sequences of positive real numbers $\{a_N\}$ and $\{b_N\}$, we say that a_N and b_N have the same order of growth as N changes, denoted by $a_N \asymp b_N$, if there exist two positive constants $\tau_1 < \tau_2$ such that $\tau_1 \leq a_N/b_N \leq \tau_2$ for all N. With these notations, the following theorem shows that $\mathcal{G}_z(u,\tilde{\mathbf{S}}(u,z))$ is close to its conditional mean $\mathbb{E}[\mathcal{G}_z(u,\tilde{\mathbf{S}}(u,z)) \mid \mathbf{X}_1]$ with high probability.

THEOREM 4. *Assume that $\Delta G(\cdot)$ is bounded. For $d \asymp n$ and all $z \in \{1,\ldots,n-u\}$,*

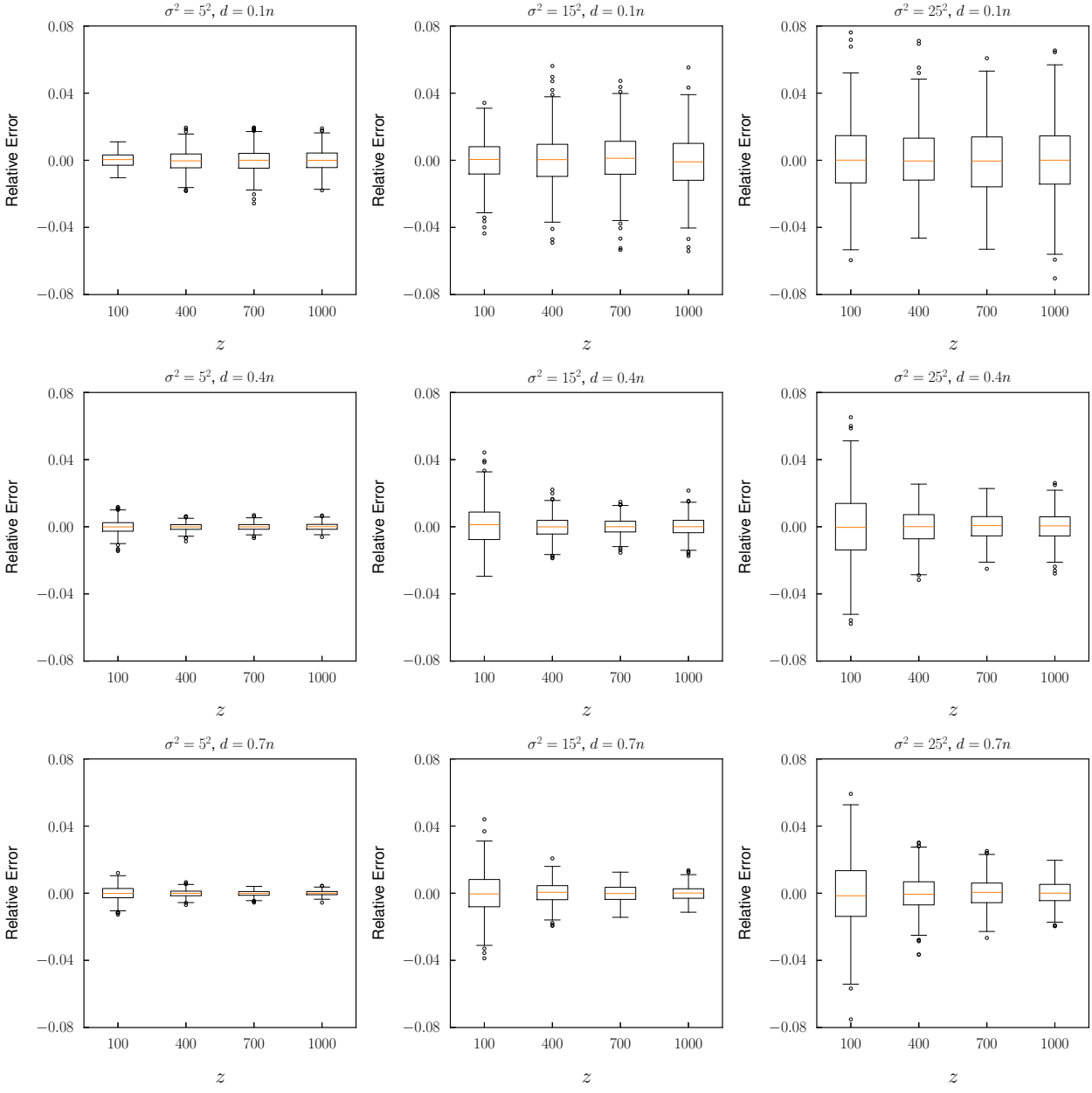
$$\mathbb{P}\left(\left|\frac{\mathcal{G}_z(u,\tilde{\mathbf{S}}(u,z))}{\mathbb{E}\left[\mathcal{G}_z(u,\tilde{\mathbf{S}}(u,z)) \mid \mathbf{X}_1\right]} - 1\right| \geq \frac{\log(z)}{\sqrt{z}}\right) \leq \frac{C_1}{\log(z)}, \tag{17}$$

where C_1 is a positive constant that does not depend on z, n, and d.

The boundedness condition on $\Delta G(\cdot)$ means that the marginal cost of hiring an additional applicant is bounded. This condition is easily met in practice; for example, the penalty cost with linear marginal underage and overage costs, given by $G(q-d) = c_u(q-d)^- + c_o(q-d)^+$, clearly satisfies this condition. Given an input data $\mathbf{X}_1 = \mathbf{x}_1$ and for any sample of test scores

$(X_{(u+1)2}, X_{(u+2)2}, \ldots, X_{(u+z)2})$, we use $\mathcal{G}_z(u, \tilde{\mathbf{S}}(u,z)(\omega))$ to approximate $\mathbb{E}[\mathcal{G}_z(u, \tilde{\mathbf{S}}(u,z)) \mid \mathbf{X}_1 = \mathbf{x}_1]$ in the objective function (15). The approximation significantly simplifies the computation, and according to Theorem 4, it does so without incurring too much error with high probability. In addition, by sorting $\tilde{\mathbf{S}}(u,z)(\omega)$ in descending order, Lemma 4 shows that the approximated objective function is discrete concave in z. This allows us to efficiently find the optimal solution to the approximated problem.

Figure 3 The Single-Sample Approximation Relative to the Monte Carlo Simulations



Note. We set $u = 0$ and $n = 1,000$. The results are based on 600 samples of initial scores.

The side-by-side boxplots in Figure 3 evaluate the approximations of $\mathbb{E}[\mathcal{G}_z(u, \tilde{\mathbf{S}}(u,z)) \mid \mathbf{X}_1 = \mathbf{x}_1]$ across different levels of σ^2, short-list sizes z, and hiring targets d. Let $V^{\mathrm{SS}}(\mathbf{x}_1)$ represent the value

obtained from the single-sample approximation $\mathcal{G}_z(u, \tilde{\mathbf{S}}(u,z)(\omega))$, and let $V^{\mathrm{MC}}(\mathbf{x}_1)$ denote the value obtained from the Monte Carlo (MC) simulations. The y-axis is the relative difference between the two approximations, calculated as $(V^{\mathrm{SS}}(\mathbf{x}_1) - V^{\mathrm{MC}}(\mathbf{x}_1))/V^{\mathrm{MC}}(\mathbf{x}_1)$. From the figure, we can see that the single-sample approximation is close to the MC benchmark in most cases.

7.2. Approximating under the Reduced Search Region

We now propose a simple algorithm to compute the optimal policy approximately, which is based on the theoretical results obtained in Sections 6 and 7.1. First, according to Theorem 3, the optimal policy (u^*, z^*) is in the reduced search region

$$\mathcal{R} = \{(u,z) : u \le u' \le u + z \le z'\} \bigcup \{(u', 0)\},$$

where the optimal solution u' of Problem $(\mathcal{P}_1^u)$ (screen-to-hire) can be found by the following simple formula:

$$u' = \begin{cases} \max\{i : \mu_i \ge \Delta G(i-d)\} & \text{if } \mu_1 \ge \Delta G(1-d), \\ 0 & \text{otherwise}. \end{cases} \tag{18}$$

As it is challenging to compute $\mathbb{E}[F_z(u, \mathbf{S}(u,z))]$ exactly, we also resolve to approximation when computing the test-to-hire policy. In particular, we use the single-sample approximation in Section 7.1 to approximate $\mathbb{E}[\mathcal{G}_z(0, \tilde{\mathbf{S}}(0,z)) \mid \mathbf{X}_1 = \mathbf{x}_1] = \mathbb{E}[F_z(0, \mathbf{S}(0,z))] + G(-d)$ in the objective function of Problem $(\mathcal{P}_1^z)$, and z' is replaced by the optimal solution of the approximation for Problem $(\mathcal{P}_1^z)$, denoted by $\hat{z}'$. Finally, we approximate the objective function (15) of Problem $(\mathcal{P}_1)$ using the single-sample approximation and search for the optimal policy within the approximated search region $\widehat{\mathcal{R}} = \{(u,z) : u \le u' \le u + z \le \hat{z}'\} \bigcup \{(u', 0)\}$. Algorithm 1 summarizes the procedure.

Algorithm 1

Input: Initial scores $\mathbf{x}_1$

Step 1. Determine the search region

1. Use (18) to compute u'.
2. Approximate $\mathbb{E}[\mathcal{G}_z(0, \tilde{\mathbf{S}}(0,z)) \mid \mathbf{X}_1 = \mathbf{x}_1]$ in the objective function of Problem $(\mathcal{P}_1^z)$ using the single-sample approximation in Section 7.1. Compute the optimal solution $\hat{z}'$.
3. Set the search region $\widehat{\mathcal{R}} = \{(u,z) : u \le u' \le u + z \le \hat{z}'\} \bigcup \{(u', 0)\}$.

Step 2. Approximate Problem $(\mathcal{P}_1)$

1. Approximate $\mathbb{E}[\mathcal{G}_z(u, \tilde{\mathbf{S}}(u,z)) \mid \mathbf{X}_1 = \mathbf{x}_1]$ in the objective function of Problem $(\mathcal{P}_1)$ using the single-sample approximation in Section 7.1.
2. Search for the optimal solution $(\hat{u}^*, \hat{z}^*)$ within the region $\widehat{\mathcal{R}}$.

Output: The approximated policy $(\hat{u}^*, \hat{z}^*)$

7.3. Numerical Studies

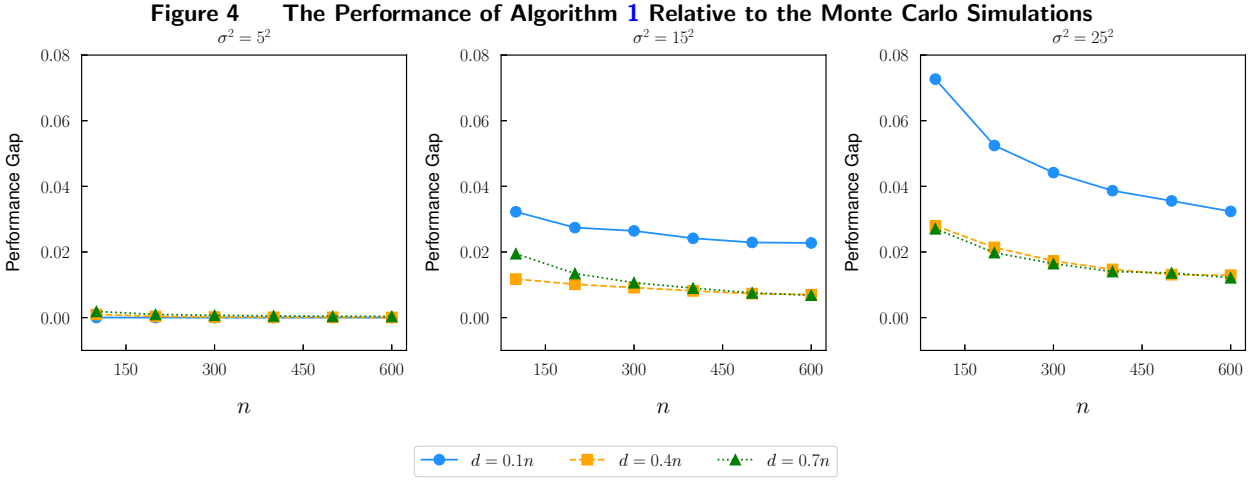
In this section, we evaluate the performance of Algorithm 1 and compare the performance of the optimal policy with those of the screen-to-hire and test-to-hire policies.

Parameters. The parameter setting is the same as the one used in the numerical example in Section 7.1. Specifically, we consider the multivariate normal model presented in Example 1 with mean $\boldsymbol{\mu} = (\mu_{x_1}, \mu_{x_2}, \mu_y) = (50, -, 60)$[5] and the covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{x_1}^2 & \sigma_{x_1,x_2} & \sigma_{x_1,y} \\ \sigma_{x_2,x_1} & \sigma_{x_2}^2 & \sigma_{x_2,y} \\ \sigma_{y,x_1} & \sigma_{y,x_2} & \sigma_y^2 \end{pmatrix} = \begin{pmatrix} 35^2 & 100 & 150 \\ 100 & 25^2 & \sigma_{x_2,y} \\ 150 & \sigma_{x_2,y} & 30^2 \end{pmatrix}.$$

We adjust the parameter σ^2 by changing $\sigma_{x_2,y}$ (see the relationship of σ^2 and $\sigma_{x_2,y}$ discussed in Section 5). Other parameters are $c = 5$ and $G(q-d) = 55(q-d)^- + 60(q-d)^+$.

Figure 4 The Performance of Algorithm 1 Relative to the Monte Carlo Simulations



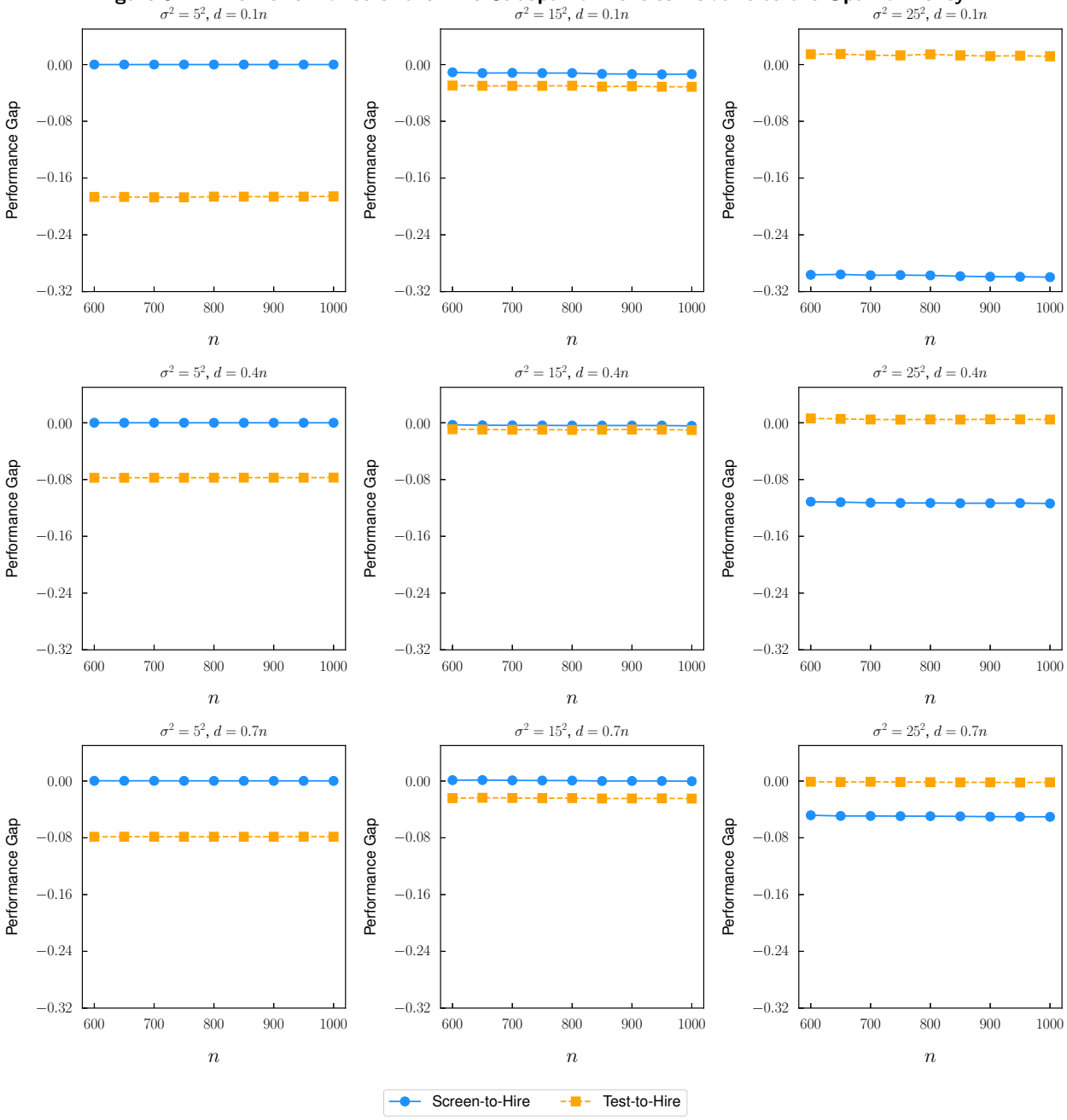
Note. The results are averaged based on 600 samples of initial scores.

We first evaluate the performance of Algorithm 1 (referred to as "Algo1") by comparing its performance gap against the MC simulation results. The expected reward for each approach i, where $i = \text{Algo1}, \text{MC}$, is denoted as $R^i(\mathbf{x}_1)$. The performance gap for Algorithm 1 relative to the MC simulation results is defined as $(R^{\text{Algo1}}(\mathbf{x}_1) - R^{\text{MC}}(\mathbf{x}_1))/R^{\text{MC}}(\mathbf{x}_1)$. The comparison is conducted across three levels of test informativeness, specifically $\sigma^2 = 5^2, 15^2, 25^2$, for varying sizes of applicant pools, $n = 100, 200, \ldots, 600$, and under three hiring targets, $d = 0.1n, 0.4n, 0.7n$. As illustrated in Figure 4, Algorithm 1 demonstrates strong performance across different parameter settings. Therefore, in the following, we present the results from Algorithm 1 as the optimal values, rather than relying on the more cumbersome MC simulations.

Next, we compare the optimal policy with the screen-to-hire and test-to-hire policies. The expected rewards and performance gaps for these policies (relative to the optimal policy) are defined

[5] Because μ_{x_2} does not play a role here, we suppress it as "$-$".

Figure 5 The Performance of the Two Suboptimal Policies Relative to the Optimal Policy



Note. The results are averaged based on 600 samples of initial scores.

similarly to the above. The comparison is again made across three levels of test informativeness, i.e., $\sigma^2 = 5^2, 15^2, 25^2$, for different sizes of applicant pools, $n = 600, 650, \ldots, 1,000$, and under three hiring targets, $d = 0.1n, 0.4n, 0.7n$. Figure 5 shows that the test-to-hire policy eventually outperforms the screen-to-hire policy as the test becomes more informative, which is expected. In our numerical studies, the performance gap between the suboptimal policies and the optimal policy can be as much as 30%. It is worth noting that the performance gaps do not diminish as n increases,

which suggests that the screen-to-hire and test-to-hire policies do not converge to the optimal policy, even as the applicant pool grows infinitely large.

8. Concluding Remarks

In this paper, we develop a model framework for business scenarios in which firms seek to select applicants from an applicant pool to fill multiple identical job positions. Firms can accept applicants based solely on their initial scores in the first stage and can also short-list them for additional costly testing to gain more information about them before accept/reject decisions are made. We show that under the optimal policy, the applicants are categorized into three groups based on their initial scores: high-scoring applicants are accepted, low-scoring ones are rejected, and those with intermediate scores are short-listed for further testing. The sizes of the groups depend on the informativeness of the signals generated by the tests: as the tests become more informative, fewer applicants are accepted and fewer are rejected based solely on their initial scores, and more proceed to the second stage for further testing. We also discuss two policies commonly observed in practice. The two policies are easier to compute than the optimal policy, and although suboptimal, they provide bounds that are useful for computing the optimal policy. The optimal policy is hard to compute exactly. We develop innovative ideas to compute it approximately, which allows us to compare the performance of the two suboptimal policies against that of the optimal policy numerically.

We believe that this research topic can motivate further studies. For example, in our analysis, we assume a variable cost of testing. There are situations where the cost of testing is a fixed cost and there is a capacity constraint on the number of applicants that can be tested. In this case, it can be shown that the optimal policy is still characterized by two cutoffs, but how the informativeness of the test affects the optimal policy is an open question. Further, our study focuses entirely on the interest of the firm but neglects the implications for applicants. A lengthy recruiting process is not only costly for firms but also for applicants. According to a recent study by Barclays, the average UK graduate attends over three interviews before receiving an offer (Barclay Simpson n.d.). Our analysis suggests that compared to the test-to-hire policy, the optimal policy benefits applicants as a whole because some applicants are accepted without going through costly testing. Compared to the screen-to-hire policy, however, the opposite is true. A relevant extension to our model would be to incorporate the strategic behavior of applicants, such as in Koren (2024). It is unclear what the optimal policies would be if the design of the recruiting process influences the number and the types of applicants the firm receives, and we consider this an important topic for future investigation.

References

AACSB (2024) Master's enrollment trends at AACSB schools: A shifting landscape. Accessed November 17, 2024, https://www.aacsb.edu/insights/reports/2024/masters-enrollment-trends-at-aacsb-schools.

Ahn H-S, Wang DD, Wu OQ (2021) Asset selling under debt obligations. *Oper. Res.* 69(4):1305-1323.

Arlotto A, Gurvich I (2019) Uniformly bounded regret in the multisecretary problem. *Stochastic Systems* 9(3):231-260.

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

Arnosti N, Ma W (2023) Tight guarantees for static threshold policies in the prophet secretary problem. *Oper. Res.* 71(5):1777-1788.

Barclay Simpson (n.d.) How much does attending an interview cost graduates? Accessed November 17, 2024, https://www.barclaysimpson.com/how-much-does-attending-an-interview-cost-graduates/.

Bastni H, Bayati M (2020) Online decision making with high-dimensional covariates. *Oper. Res.* 68(1):276-294.

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* 29(4):1165-1188.

Bertsimas D, Natarajan K, Teo CP (2006) Tight bounds on expected order statistics. *Probab. Engrg. Inform. Sci.* 20(4):667-686.

Bickel PJ (1982) On adaptive estimation. *Ann. Statist.* 10(3):647-671.

Box GEP, Cox DR (1964) An analysis of transformations. *J. Royal Statist. Soc. Series B (Statistical Methodology)* 26(2):211-243.

Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press).

Chae M, Lin L, Dunson DB (2019) Bayesian sparse linear regression with unknown symmetric error. *Information and Inference: A Journal of the IMA* 8(3):621-653.

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

De Corte W (1998) Estimating and maximizing the utility of sequential selection decisions with a probationary period. *British J. Math. Statist. Psych.* 51(1):101-121.

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

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

Du L, Li Q, Yu P (2024) A sequential model for high-volume recruitment under random yields. *Oper. Res.* 72(1):60-90.

Durrett R (2019) *Probability: Theory and Examples* (Cambridge University Press).

Goldenshluger A, Zeevi A (2013) A linear response bandit problem. *Stochastic Systems* 3(1):230-261.

Gong C, Li Q (2024) A rolling recruitment process under applicant stochastic departures. Preprint, submitted March 28, http://dx.doi.org/10.2139/ssrn.4787415.

Gu J, Koenker R (2023) Invidious comparisons: Ranking and selection as compound decisions. *Econometrica* 91(1):1-41.

Györfi L, Kohler M, Krzyżak A, Walk H (2002) *A Distribution-Free Theory of Nonparametric Regression* (Springer Series in Statistics).

Hayashi F (2011) *Econometrics* (Princeton University Press).

iCIMS (2024) Full-time jobs in the retail sector are in high demand, per new iCIMS data. Accessed November 17, 2024, https://www.icims.com/company/newsroom/aprilinsights2024/.

Jedeikin J (2015) How this recruiting team made 100 hires in 60 days. Accessed November 17, 2024, https://www.linkedin.com/business/talent/blog/talent-strategy/how-recruiting-team-made-100-hires-in-60-days.

Johnson RA, Wichern DW (2014) *Applied Multivariate Statistical Analysis* (Prentice-Hall).

Klein M, Wright T, Wieczorek J (2020) A joint confidence region for an overall ranking of populations. *J. Royal Statist. Soc. Series C (Applied Statistics)* 69(3):589-606.

Komiyama J, Noda S (2024) On statistical discrimination as a failure of social learning: A multiarmed bandit approach. *Management Sci.*, ePub ahead of print March 29, https://doi.org/10.1287/mnsc.2022.00893.

Koren M (2024) The gatekeeper effect: The implications of pre-screening, self-selection, and bias for hiring processes. *Management Sci.*, ePub ahead of print May 17, https://doi.org/10.1287/mnsc.2021.03918.

Lagziel D, Lehrer E (2019) A bias of screening. *Amer. Econom. Rev. Insights* 1(3):343-356.

Lehmann EL (1966) Some concepts of dependence. *Ann. Math. Statist.* 37(5):1137–1153.

Li Q, Yu P (2021) The secretary problem with multiple job vacancies and batch candidate arrivals. *Oper. Res. Lett.* 49(4):535-542.

Mogstad M, Romano JP, Shaikh AM, Wilhelm D (2023) A comment on: "Invidious comparisons: Ranking and selection as compound decision" by Jiaying Gu and Roger Koenker. *Econometrica* 91(1):53-60.

Mogstad M, Romano JP, Shaikh AM, Wilhelm D (2024) Inference for ranks with applications to mobility across neighbourhoods and academic achievement across countries. *Rev. Econ. Stud.* 91(1):476-518.

Navarra K (2022) The real costs of recruitment. Accessed November 17, 2024, https://www.shrm.org/mena/topics-tools/news/talent-acquisition/real-costs-recruitment.

Neumeyer N, Dette H (2007) Testing for symmetric error distribution in nonparametric regression models. *Stat. Sin.* 17(2):775-795.

Prastacos GP (1983) Optimal sequential investment decisions under conditions of uncertainty. *Management Sci.* 29(1):118-134.

Prokopets E (2024) The true cost of hiring an employee in 2024. Accessed November 17, 2024, https://toggl.com/blog/cost-of-hiring-an-employee.

Shaked M, Shanthikumar JG (2007) *Stochastic Orders* (Springer).

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

Topkis DM (1998) *Supermodularity and Complementarity* (Princeton University Press).

Vera A, Banerjee S (2021) The Bayesian prophet: A low-regret framework for online decision making. *Management Sci.* 67(3):1368-1391.

Vulcano G, van Ryzin G, Maglaras C (2002) Optimal dynamic auctions for revenue management. *Management Sci.* 48(11):1388-1407.

Wang R, Ramdas A (2022) False discovery rate control with e-values. *J. Royal Statist. Soc. Series B (Statistical Methodology)* 84(3):822-852.

Wooldridge JM (2010) *Econometric Analysis of Cross Section and Panel Data* (MIT Press).

Yao DD, Zheng S (2002) *Dynamic Control of Quality in Production-Inventory Systems: Coordination and Optimization* (Springer Science & Business Media).

Appendix A: Discussion on Modeling Choice

In Assumption 1, we assume that the error difference $\epsilon_i - \epsilon_i'$ is independent of the initial score X_{i1}. In this section, we show that this assumption can be relaxed. Specifically, we consider the following heteroskedastic regression model (Neumeyer and Dette 2007):

$$g(X_{i1}, X_{i2}) \mid X_{i1} = f(X_{i1}) + (\epsilon_i - \epsilon_i') \mid X_{i1} = f(X_{i1}) + \sigma(X_{i1})\epsilon_i'', \tag{A1}$$

where ϵ_i'' is independent of X_{i1} with a mean of zero and a variance of one. The conditional variance is given by $\mathrm{Var}(g(X_{i1}, X_{i2}) \mid X_{i1}) = \sigma^2(X_{i1})$, and hence the regression (A1) does not require conditional homoskedasticity, a standard component in classical linear regression models (Hayashi 2011). For ease of exposition, we further assume that $f(\cdot)$ is strictly increasing and both $f(\cdot)$ and $\sigma(\cdot)$ are differentiable.

We first provide a sufficient condition on $\sigma(\cdot)$ for the validity of the two-cutoff optimal policy in Theorem 1.

Proposition A1. *Theorem 1 remains true if, for any $i \in \{1, 2, \ldots, n\}$ and $x_{i1} \in \mathbb{R}$, $\sigma(x_{i1})$ satisfies*

$$\sup_{x \in \mathbb{R}} \frac{1}{\mathbb{E}\left[\epsilon_i'' \mid \epsilon_i'' < x\right]} \leq -\frac{\sigma'(x_{i1})}{f'(x_{i1})} \leq \inf_{x \in \mathbb{R}} \frac{1}{\mathbb{E}\left[\epsilon_i'' \mid \epsilon_i'' \geq x\right]}.$$

The denominators $\mathbb{E}\left[\epsilon_i'' \mid \epsilon_i'' < x\right]$ in the lower bound and $\mathbb{E}\left[\epsilon_i'' \mid \epsilon_i'' \geq x\right]$ in the upper bound represent the average errors when applicant i is rejected or accepted based on a threshold $f(x_{i1}) + \sigma(x_{i1})x$, respectively. The condition in Proposition A1 requires that the rate of change of $\sigma(x_{i1})$ relative to that of $f(x_{i1})$ is bounded by the reciprocals of these average errors.

For Theorem 2, because the conditional variance $\sigma^2(X_{i1})$ varies across different realizations of X_{i1}, it is straightforward to redefine the informativeness of the test by using the expectation of $\sigma^2(X_{i1})$. More sophisticated approaches may also be possible, and we leave these for future research.

For Theorem 3, we need to impose the usual stochastic ordering condition on S_i to ensure that Lemma 4 holds. To that end, we assume that $g(X_{i1}, X_{i2})$ is positively regression dependent on X_{i1}, as discussed in Section 3. This condition is equivalent to the usual stochastic ordering condition on S_i.

Finally, for the condition in Proposition A1, the lower bound is negative while the upper bound is positive. It follows that the constant variance σ^2 under Assumption 1(*ii*) also meets this condition. Furthermore, Assumption 1 implies the positive regression dependence condition. Therefore, Assumption 1 is a special case of the heterogeneity in regression (A1).

Appendix B: Proofs for the Main Theoretical Results

Proof of Lemma 1. Define

$$J(x_{i1}) = \mathbb{E}_{S_i'(x_{i1})}\left[\max_{0 \leq h \leq |\mathcal{Z}|}\left\{\sum_{j=1}^{h} S_{[j]}'(\mathbf{x}_1, \mathcal{Z}) - G(|\mathcal{U}| + h - d)\right\} \middle| (S_j'(x_{j1}))_{j \in \mathcal{Z} \setminus \{i\}}\right],$$

where the subscript $S_i'(x_{i1})$ in the expectation operator means that the expectation is taken with respect to $S_i'(x_{i1})$. Given a realization $(s_j'(x_{j1}))_{j \in \mathcal{Z}}$, we first show that the maximization inside the conditional expectation is increasing in $s_i'(x_{i1})$ with a slope of less than one. Because

$$\sum_{j=1}^{h} s_{[j]}'(\mathbf{x}_1, \mathcal{Z}) = \max\left\{\sum_{k=1}^{h} s_{j_k}'(x_{j_k 1}) :\ j_1 < j_2 < \cdots < j_h,\ j_k \in \mathcal{Z},\ \forall k \in \{1, 2, \ldots, h\}\right\}$$

is the pointwise maximum of $|\mathcal{Z}|!/(h!(|\mathcal{Z}|-h)!)$ linear functions (Boyd and Vandenberghe 2004, Example 3.6), each increasing in $s'_i(x_{i1})$ with a slope of less than one, $\sum_{j=1}^{h} s'_{[j]}(\mathbf{x}_1,\mathcal{Z})$ is increasing in $s'_i(x_{i1})$ with a slope of less than one. Because $G(|\mathcal{U}|+h-d)$ does not depend on $s'_i(x_{i1})$, the maximization $\max_{0\le h\le|\mathcal{Z}|}\left\{\sum_{j=1}^{h} s'_{[j]}(\mathbf{x}_1,\mathcal{Z})-G(|\mathcal{U}|+h-d)\right\}$ is the maximum of $|\mathcal{Z}|+1$ increasing functions of $s'_i(x_{i1})$ with a slope of less than one. Then, by the monotonicity of conditional expectation, $J(x_{i1})$ is increasing in $s'_i(x_{i1})$ with a slope of less than one. This further implies that $\mathbb{E}_{(S'_j(x_{j1}))_{j\in\mathcal{Z}\setminus\{i\}}}[J(x_{i1})]$ has the same property. Finally, $S'_i(x_{i1})$ can be written as $S'_i(x_{i1})=f(x_{i1})+\sigma\varepsilon'_i$, where the distribution of ε'_i is not related to x_{i1}. It follows that $\mathbb{E}_{(S'_j(x_{j1}))_{j\in\mathcal{Z}\setminus\{i\}}}[J(x_{i1})]$ is increasing in x_{i1} with a slope of less than $\nabla f(x_{i1})$. □

Proof of Theorem 1. To show the two-cutoff policy, it suffices to show that if it is optimal to accept applicant i, then there exists an optimal policy in which applicant j with $x_{j1}\ge x_{i1}$ is accepted; if it is optimal to reject applicant i, then there exists an optimal policy in which applicant j with $x_{j1}\le x_{i1}$ is rejected.

We first show the first half of the statement. It is obvious that accepting j and rejecting i, ceteris paribus, generates a $f(x_{j1})-f(x_{i1})$ higher reward than accepting i and rejecting j. Therefore, we only need to verify that accepting j and short-listing i for further testing, ceteris paribus, also generates a higher reward than the other way around. Given a set $\mathcal{U}_0$ of applicants being accepted with $i,j\notin\mathcal{U}_0$, and a set $\mathcal{Z}_0$ of all applicants being short-listed for further testing with $i,j\notin\mathcal{Z}_0$, using Lemma 1, we have

$$\begin{aligned}f(x_{j1})-f(x_{i1})\ge\ &\mathbb{E}\left[\max_{0\le h\le|\mathcal{Z}_0\cup\{j\}|}\left\{\sum_{k=1}^{h}S'_{[k]}(\mathbf{x}_1,\mathcal{Z}_0\cup\{j\})-G(|\mathcal{U}_0\cup\{i\}|+h-d)\right\}\right]\\&-\mathbb{E}\left[\max_{0\le h\le|\mathcal{Z}_0\cup\{i\}|}\left\{\sum_{k=1}^{h}S'_{[k]}(\mathbf{x}_1,\mathcal{Z}_0\cup\{i\})-G(|\mathcal{U}_0\cup\{j\}|+h-d)\right\}\right].\end{aligned}$$

This implies that accepting j and short-listing i is no worse than accepting i and short-listing j. In summary, it is optimal to accept j.

We next show the second half of the statement. Because rejecting j and accepting i, ceteris paribus, generates a $f(x_{i1})-f(x_{j1})$ higher reward than rejecting i and accepting j, we only need to show that rejecting j and short-listing i for further testing, ceteris paribus, also generates a higher reward than the other way around. Using the same notations as above and Lemma 1, we have

$$\begin{aligned}&\mathbb{E}\left[\max_{0\le h\le|\mathcal{Z}_0\cup\{i\}|}\left\{\sum_{k=1}^{h}S'_{[k]}(\mathbf{x}_1,\mathcal{Z}_0\cup\{i\})-G(|\mathcal{U}_0|+h-d)\right\}\right]\\\ge\ &\mathbb{E}\left[\max_{0\le h\le|\mathcal{Z}_0\cup\{j\}|}\left\{\sum_{k=1}^{h}S'_{[k]}(\mathbf{x}_1,\mathcal{Z}_0\cup\{j\})-G(|\mathcal{U}_0|+h-d)\right\}\right].\end{aligned}$$

This implies that rejecting j and short-listing i is no worse than rejecting i and short-listing j. In summary, it is optimal to reject j. □

Proof of Lemma 2. Without loss of generality, we let $u=0$. When $z=1$, we have $F_1(0,\mathbf{s}(0,1))=\max\{s_1-G(1-d),-G(-d)\}=(s_1-\Delta G(1-d))^+-G(-d)$, which clearly satisfies (8). Therefore, in the sequel, we assume that $z>1$. Let

$$H(h)=\sum_{j=1}^{h}s_{[j]}(0,z)-G(h-d),$$

which is discrete concave in h. Recall that h^* is the largest maximizer of (7). Then, $s_{[h^*]}(0,z)$ is the smallest element in the optimal score set, and we set $s_{[h^*]}(0,z)=\Delta G(1-d)$ if $h^*=0$. The idea of the proof is to first

show that $s_i \geq s_{[h^*]}(0,z)$ if and only if $s_i \geq \hat{s}_{[z]}(0,z,i)$, and then conduct the decomposition based on the relationship between s_i and $s_{[h^*]}(0,z)$. Note that $s_{[h^*]}(0,z)$ is a function of $\mathbf{s}(0,z) \in \mathbb{R}^z$, whereas $\hat{s}_{[z]}(0,z,i)$ is a function of $\mathbf{s}_{-i}(0,z) \in \mathbb{R}^{z-1}$. We proceed in two steps.

Step 1 ($s_i \geq s_{[h^*]}(0,z)$ if and only if $s_i \geq \hat{s}_{[z]}(0,z,i)$). If $h^* = 0$, then $H(0) > H(1)$ implies that $G(1-d) - G(-d) > s_{[1]}(0,z)$. Thus, $s_{[h^*]}(0,z) = \Delta G(1-d)$ is the zth largest element in $\hat{\mathbf{s}}(0,z,i)$, i.e., $s_{[h^*]}(0,z) = \hat{s}_{[z]}(0,z,i)$. If $h^* \geq 1$, we assume that $s_i = s_{[k]}(0,z)$ for some $k \in \{1,2,\ldots,h^*\}$, and sequentially prove necessity and sufficiency.

"Only if" part: Because $s_{[k]}(0,z) \geq s_{[h^*]}(0,z)$ and $H(\cdot)$ is discrete concave, we have

$$H(k-1) \leq H(k) \leq \cdots \leq H(h^*),$$

where the first inequality implies that $s_i = s_{[k]}(0,z) \geq G(k-d) - G(k-1-d) = \Delta G(k-d)$. Therefore, $s_i \geq \Delta G(h-d)$ for $h = k, k-1, \ldots, 1$. Because we also have $s_i \geq s_{[h]}(0,z)$ for $h = k+1, k+2, \ldots, z$, at least z elements in $\hat{\mathbf{s}}(0,z,i)$ are smaller than s_i, which implies that $s_i \geq \hat{s}_{[z]}(0,z,i)$.

"If" part: Recall that $\hat{s}_j$ is the jth element in $\hat{\mathbf{s}}(0,z,i) = (\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_{2z-1}) = (\mathbf{s}_{-i}(0,z), \Delta G(1-d), \Delta G(2-d), \ldots, \Delta G(z-d))$. Let n_1 and n_2 be the numbers of $\hat{s}_j$'s with $\hat{s}_j \leq s_i$ that are from $\mathbf{s}_{-i}(0,z)$ and $(\Delta G(1-d), \Delta G(2-d), \ldots, \Delta G(z-d))$, respectively. Then, $s_i \geq \hat{s}_{[z]}(0,z,i)$ implies that $n_1 + n_2 \geq z$. By the definition of n_1, $s_{[k]}(0,z) = s_i \geq s_{[z-n_1]}(0,z)$. Therefore, $k \leq z - n_1 \leq n_2$. In addition, because $s_i \geq \Delta G(n_2 - d)$ by the definition of n_2, we obtain

$$H(k) - H(k-1) = s_i - \Delta G(k-d) \geq s_i - \Delta G(n_2 - d) \geq 0,$$

which further implies that $s_i \geq s_{[h^*]}(0,z)$ because $H(\cdot)$ is discrete concave.

Step 2 (Decomposition). Let $\tilde{H}(h) = \sum_{j=1}^{h} \tilde{s}_{[j]}(0,z) - G(h-d)$, where $\tilde{s}_{[j]}(0,z)$ denotes the jth largest element in $\mathbf{s}_{-i}(0,z)$. We consider two cases.

Case 1. If $s_i < s_{[h^*]}(0,z)$, then $h^* \in \{0,1,\ldots,z-1\}$, and we have $F_z(0,\mathbf{s}(0,z)) = H(h^*) = \tilde{H}(h^*)$. Thus, for $h = h^*+1, h^*+2, \ldots, z-1$ (if $h^* = z-1$, then no such h exists),

$$\tilde{H}(h^*) = H(h^*) > H(h) \geq \tilde{H}(h), \tag{A2}$$

where the first inequality follows from the optimality of h^*, and the second inequality follows because $\mathbf{s}(0,z)$ has one more element than $\mathbf{s}_{-i}(0,z)$. Similarly, for $h = 0,1,\ldots,h^*$,

$$\tilde{H}(h^*) = H(h^*) \geq H(h) = \tilde{H}(h), \tag{A3}$$

where the second equality follows because the largest h elements in $\mathbf{s}(0,z)$ are exactly the same as the largest h elements in $\mathbf{s}_{-i}(0,z)$ when $h \leq h^*$.

By (A2), (A3), and Step 1, we have

$$F_z(0,\mathbf{s}(0,z)) = H(h^*) = F_{z-1}(0,\mathbf{s}_{-i}(0,z)) = (s_i - \hat{s}_{[z]}(0,z,i))^+ + F_{z-1}(0,\mathbf{s}_{-i}(0,z)),$$

which satisfies (8).

Case 2. If $s_i \geq s_{[h^*]}(0,z)$, then $h^* \in \{1,2,\ldots,z\}$. Expanding $F_z(0,\mathbf{s}(0,z))$ yields

$$\begin{aligned}
F_z(0,\mathbf{s}(0,z)) &= \sum_{j=1}^{h^*} s_{[j]}(0,z) - G(h^*-d) \\
&= s_i + \left(\sum_{j=1}^{h^*} s_{[j]}(0,z) - s_i\right) - G(h^*-d) \\
&= s_i + \sum_{j=1}^{h^*-1} \tilde{s}_{[j]}(0,z) - G(h^*-d) \\
&= (s_i - \hat{s}_{[z]}(0,z,i))^+ + \sum_{j=1}^{h^*-1} \tilde{s}_{[j]}(0,z) + \hat{s}_{[z]}(0,z,i) - G(h^*-d), \qquad \text{(A4)}
\end{aligned}$$

where the third equality follows because s_i is among the h^* largest elements in $\mathbf{s}(0,z)$, and after taking s_i out from $\mathbf{s}(0,z)$, the remaining (h^*-1) largest elements in $\mathbf{s}(0,z)$ are exactly the h^* largest elements in $\mathbf{s}_{-i}(0,z)$. The positive part in the last equality follows because $s_i \geq s_{[h^*]}(0,z)$ implies that $s_i \geq \hat{s}_{[z]}(0,z,i)$ according to Step 1. It then suffices to show that

$$F_{z-1}(0,\mathbf{s}_{-i}(0,z)) = \sum_{j=1}^{h^*-1} \tilde{s}_{[j]}(0,z) + \hat{s}_{[z]}(0,z,i) - G(h^*-d). \qquad \text{(A5)}$$

To this end, let $h' \in \{0,1,\ldots,z-1\}$ be the optimal solution of $F_{z-1}(0,\mathbf{s}_{-i}(0,z))$. We first show that $h' \in \{h^*-1, h^*\}$ by contradiction.

If $h' > h^*$, then

$$\begin{aligned}
H(h') - H(h^*) &= s_i + \sum_{j=1}^{h'-1} \tilde{s}_{[j]}(0,z) - G(h'-d) - \left(s_i + \sum_{j=1}^{h^*-1} \tilde{s}_{[j]}(0,z) - G(h^*-d)\right) \\
&\geq \sum_{j=1}^{h'} \tilde{s}_{[j]}(0,z) - G(h'-d) - \left(\sum_{j=1}^{h^*} \tilde{s}_{[j]}(0,z) - G(h^*-d)\right) \\
&= \tilde{H}(h') - \tilde{H}(h^*) \\
&\geq 0,
\end{aligned}$$

which contradicts the optimality of h^*. Here, the first equality follows from the same argument for the third equality in (A4), and the first inequality follows from the concavity of $\sum_{j=1}^{h} \tilde{s}_{[j]}(0,z)$ in h.

If $h' < h^* - 1$, then

$$\begin{aligned}
H(h'+1) - H(h^*) &= \sum_{j=1}^{h'+1} s_{[j]}(0,z) - G(h'+1-d) - \left(s_i + \sum_{j=1}^{h^*-1} \tilde{s}_{[j]}(0,z) - G(h^*-d)\right) \\
&\geq s_i + \sum_{j=1}^{h'} \tilde{s}_{[j]}(0,z) - G(h'+1-d) - \left(s_i + \sum_{j=1}^{h^*-1} \tilde{s}_{[j]}(0,z) - G(h^*-d)\right) \\
&\geq \sum_{j=1}^{h'} \tilde{s}_{[j]}(0,z) - G(h'-d) - \left(\sum_{j=1}^{h^*-1} \tilde{s}_{[j]}(0,z) - G(h^*-1-d)\right) \\
&= \tilde{H}(h') - \tilde{H}(h^*-1) \\
&> 0,
\end{aligned}$$

which contradicts the optimality of h^*. Here, the first inequality follows because $\sum_{j=1}^{h'+1} s_{[j]}(0,z) = s_i + \sum_{j=1}^{h'} \tilde{s}_{[j]}(0,z)$ if $s_i \geq s_{[h'+1]}(0,z)$ and $\sum_{j=1}^{h'+1} s_{[j]}(0,z) > s_i + \sum_{j=1}^{h'} \tilde{s}_{[j]}(0,z)$ if $s_i < s_{[h'+1]}(0,z)$, and the second inequality follows from the convexity of $G(\cdot)$.

Therefore, we only need to consider two subcases: $h' = h^* - 1$ and $h' = h^*$. Note that $H(h^*) > H(h^*+1)$ implies that $s_{[h^*+1]}(0,z) < \Delta G(h^*+1-d)$, and $H(h^*) \geq H(h^*-1)$ implies that $s_{[h^*]}(0,z) \geq \Delta G(h^*-d)$.

Subcase 2.1. If $h' = h^* - 1$, then $\tilde{s}_{[h^*-1]}(0,z) \geq s_{[h^*]}(0,z) \geq \Delta G(h^*-d)$, where we set $\tilde{s}_{[0]}(0,z) = \Delta G(1-d)$. Therefore, $\tilde{s}_{[h]}(0,z) \geq \Delta G(h^*-d)$ for $h = 1,2,\ldots,h^*-1$. If $h^* = z$, then $\tilde{s}_{[h]}(0,z) \geq \Delta G(h^*-d)$ for $h = 1,2,\ldots,z-1$, and if $h^* < z$, then $0 < \tilde{H}(h^*-1) - \tilde{H}(h^*) = -\tilde{s}_{[h^*]}(0,z) + \Delta G(h^*-d)$, i.e., $\tilde{s}_{[h]}(0,z) < \Delta G(h^*-d)$ for $h = h^*, h^*+1, \ldots, z-1$. Furthermore, we have $\Delta G(h-d) \geq \Delta G(h^*-d)$ for $h = h^*, h^*+1, \ldots, z$ and $\Delta G(h-d) \leq \Delta G(h^*-d)$ for $h = 1,2,\ldots,h^*-1$. Therefore, we conclude that $\Delta G(h^*-d) = \hat{s}_{[z]}(0,z,i)$. It follows that

$$\sum_{j=1}^{h^*-1} \tilde{s}_{[j]}(0,z) + \hat{s}_{[z]}(0,z,i) - G(h^*-d) = \sum_{j=1}^{h^*-1} \tilde{s}_{[j]}(0,z) - G(h^*-1-d) = F_{z-1}(0, \mathbf{s}_{-i}(0,z)),$$

which satisfies (A5).

Subcase 2.2. If $h' = h^*$, then $h^* < z$, and $\tilde{s}_{[h^*]}(0,z) = s_{[h^*+1]}(0,z) < \Delta G(h^*+1-d)$. In addition, $0 \leq \tilde{H}(h^*) - \tilde{H}(h^*-1) = \tilde{s}_{[h^*]}(0,z) - \Delta G(h^*-d)$. Therefore, $\tilde{s}_{[h]}(0,z)$, $h = 1,2,\ldots,h^*-1$, and $\Delta G(h-d)$, $h = h^*+1, h^*+2, \ldots, z$, are all larger than $\tilde{s}_{[h^*]}(0,z)$, and the remaining elements in $\hat{\mathbf{s}}(0,z,i)$ are less than $\tilde{s}_{[h^*]}(0,z)$. Therefore, we conclude that $\tilde{s}_{[h^*]}(0,z) = \hat{s}_{[z]}(0,z,i)$. It follows that

$$\sum_{j=1}^{h^*-1} \tilde{s}_{[j]}(0,z) + \hat{s}_{[z]}(0,z,i) - G(h^*-d) = \sum_{j=1}^{h^*} \tilde{s}_{[j]}(0,z) - \tilde{s}_{[h^*]}(0,z) + \hat{s}_{[z]}(0,z,i) - G(h^*-d) = F_{z-1}(0, \mathbf{s}_{-i}(0,z)),$$

which satisfies (A5).

By Cases 1 and 2, we complete the proof. □

For the proof of Lemma 3, we need the following auxiliary lemma. It shows that $\hat{s}_{[z]}(u,z,i)$ is increasing in z. Intuitively, as more applicants are short-listed for the second stage, the chances of those already short-listed being hired decrease.

LEMMA A1. *For any $i \in \{1,2,\ldots,z\}$, we have $\hat{s}_{[z+1]}(u,z+1,i) \geq \hat{s}_{[z]}(u,z,i)$ and $\hat{s}_{[z+1]}(u,z+1,z+1) \geq \hat{s}_{[z]}(u,z,i)$.*

Proof of Lemma A1. To show the first equality $\hat{s}_{[z+1]}(u,z+1,i) \geq \hat{s}_{[z]}(u,z,i)$, instead of directly comparing the two (sampled) order statistics, we consider a truncated version of $\hat{\mathbf{s}}(u,z+1,i)$ by deleting s_{u+z+1}: $\tilde{\mathbf{s}} = (\tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_{2z}) = (\mathbf{s}_{-i}(u,z), \Delta G(u+1-d), \Delta G(u+2-d), \ldots, \Delta G(u+z+1-d))$. Let $\tilde{s}_{[z+1]}$ be the $(z+1)$th largest element in $\tilde{\mathbf{s}}$. If we can show $\tilde{s}_{[z+1]} = \hat{s}_{[z]}(u,z,i)$, then the result follows from $\hat{s}_{[z+1]}(u,z+1,i) \geq \tilde{s}_{[z+1]}$.

We first list three observations about the two vectors $\hat{\mathbf{s}}(u,z,i)$ and $\tilde{\mathbf{s}}$: (1) the largest z elements in $\hat{\mathbf{s}}(u,z,i)$ must contain at least one element from $(\Delta G(u+1-d), \Delta G(u+2-d), \ldots, \Delta G(u+z-d))$ because there are only $(z-1)$ elements in $\mathbf{s}_{-i}(u,z)$; (2) vector $\tilde{\mathbf{s}}$ has one more element $\Delta G(u+z+1-d)$ than $\hat{\mathbf{s}}(u,z,i)$, and all of their other elements are the same; and (3) $\Delta G(u+z+1-d)$ is larger than all of the elements in $(\Delta G(u+1-d), \Delta G(u+2-d), \ldots, \Delta G(u+z-d))$.

From these observations, it is evident that the largest $(z+1)$ elements in $\tilde{\mathbf{s}}$ comprise the top z elements from $\hat{\mathbf{s}}(u,z,i)$ along with the element $\Delta G(u+z+1-d)$, which ranks at least as the zth largest in $\tilde{\mathbf{s}}$. Thus, the result follows.

For the second equality $\hat{s}_{[z+1]}(u,z+1,z+1) \geq \hat{s}_{[z]}(u,z,i)$, we note that $\hat{\mathbf{s}}(u,z+1,z+1)$ has one more element s_{u+i} than $\tilde{\mathbf{s}}$, so $\hat{s}_{[z+1]}(u,z+1,z+1) \geq \tilde{s}_{[z+1]} = \hat{s}_{[z]}(u,z,i)$. □

Proof of Lemma 3. Take any $i \in \{1,2,\ldots,z\}$. Recall (8) in Lemma 2 that

$$F_z(u,\mathbf{s}(u,z)) = (s_{u+i} - \hat{s}_{[z]}(u,z,i))^+ + F_{z-1}(u,\mathbf{s}_{-i}(u,z)). \tag{A6}$$

Assign a large enough value to s_{u+i} such that $s_{u+i} \geq \hat{s}_{[z]}(u,z,i)$. Then, by Step 1 in the proof of Lemma 2, $s_{u+i} \geq s_{[h^*]}(u,z)$, where $h^* \in \{1,2,\ldots,z\}$ is the largest maximizer of $F_z(u,\mathbf{s}(u,z))$. Let $\tilde{s}_{[j]}(u,z)$ denote the jth largest element in $\mathbf{s}_{-i}(u,z)$. Expanding $F_z(u,\mathbf{s}(u,z))$ yields

$$\begin{aligned} F_z(u,\mathbf{s}(u,z)) &= \max_{0\leq h\leq z-1} \left\{ s_{u+i} + \sum_{j=1}^{h} s_{[j+1]}(u,z) - G(u+1+h-d) \right\} \\ &= s_{u+i} + \max_{0\leq h\leq z-1} \left\{ \sum_{j=1}^{h} \tilde{s}_{[j]}(u,z) - G(u+1+h-d) \right\} \\ &= s_{u+i} + F_{z-1}(u+1,\mathbf{s}_{-i}(u,z)), \end{aligned} \tag{A7}$$

where the first equality follows because $s_{u+i} \geq s_{[h^*]}(u,z)$, and then the problem is equivalent to maximizing over the remaining $(z-1)$ elements after s_{u+i} is included in the optimal score set.

By (A6) and (A7), we have

$$F_{z-1}(u,\mathbf{s}_{-i}(u,z)) = F_{z-1}(u+1,\mathbf{s}_{-i}(u,z)) + \hat{s}_{[z]}(u,z,i). \tag{A8}$$

It is important to note that (A8) *does not* depend on the value of s_{u+i}. Plugging (A8) into (A6), we have

$$F_z(u,\mathbf{s}(u,z)) - F_{z-1}(u+1,\mathbf{s}_{-i}(u,z)) = (s_{u+i} - \hat{s}_{[z]}(u,z,i))^+ + \hat{s}_{[z]}(u,z,i) = \max\left\{s_{u+i}, \hat{s}_{[z]}(u,z,i)\right\}, \tag{A9}$$

which holds for all $s_{u+i} \in \mathbb{R}$.

Now, to show the submodularity, it suffices to show that $F_{k-u}(u,\mathbf{s}(u,k-u)) - F_{k-u-1}(u+1,\mathbf{s}(u+1,k-u-1))$ is increasing in k. By (A9), we have

$$F_{k-u}(u,\mathbf{s}(u,k-u)) - F_{k-u-1}(u+1,\mathbf{s}(u+1,k-u-1)) = \max\left\{s_{u+1}, \hat{s}_{[k-u]}(u,k-u,1)\right\},$$

where $\hat{s}_{[k-u]}(u,k-u,1)$ is the $(k-u)$th largest element in $\hat{\mathbf{s}}(u,k-u,1) = (\mathbf{s}_{-1}(u,k-u), \Delta G(u+1-d), \Delta G(u+2-d), \ldots, \Delta G(k-d))$. Similarly,

$$F_{k+1-u}(u,\mathbf{s}(u,k+1-u)) - F_{k-u}(u+1,\mathbf{s}(u+1,k-u)) = \max\left\{s_{u+1}, \hat{s}_{[k+1-u]}(u,k+1-u,1)\right\},$$

where $\hat{s}_{[k+1-u]}(u,k+1-u,1)$ is the $(k+1-u)$th largest element in $\hat{\mathbf{s}}(u,k+1-u,1) = (\mathbf{s}_{-1}(u,k+1-u), \Delta G(u+1-d), \Delta G(u+2-d), \ldots, \Delta G(k+1-d))$.

By Lemma A1, $\hat{s}_{[k+1-u]}(u,k+1-u,1) \geq \hat{s}_{[k-u]}(u,k-u,1)$, so $\max\left\{s_{u+1}, \hat{s}_{[k+1-u]}(u,k+1-u,1)\right\} \geq \max\left\{s_{u+1}, \hat{s}_{[k-u]}(u,k-u,1)\right\}$, which completes the proof. □

Proof of Theorem 2. (*i*) *Redefining Problem* $(\mathcal{P}_1)$. Replacing z with $k-u$ in Problem $(\mathcal{P}_1)$ yields

$$\begin{aligned} \max \quad & \sum_{i=1}^{u} \mu_i - c(k-u) + \mathbb{E}\left[F_{k-u}(u, \mathbf{S}(u, k-u))\right] \\ \text{s.t.} \quad & -u \in \{-n, -n+1, \ldots, 0\}, \\ & k \in \{u, u+1, \ldots, n\}. \end{aligned} \tag{A10}$$

However, the constraint set of $(-u, k)$, $\{-n, -n+1, \ldots, 0\} \times \{u, u+1, \ldots, n\}$ is not a lattice. To tackle this issue, we redefine (A10) as follows:

$$\begin{aligned} \max \quad & \sum_{i=1}^{u} \mu_i - c(k-u)^+ + \mathbb{E}\left[F_{(k-u)^+}(u, \mathbf{S}(u, (k-u)^+))\right] \\ \text{s.t.} \quad & -u \in \{-n, -n+1, \ldots, 0\}, \\ & k \in \{0, 1, \ldots, n\}. \end{aligned} \tag{A11}$$

One can verify that (A10) and (A11) are equivalent, and the new constraint set is a lattice.

We next show that the objective function of (A11) is supermodular in $(\sigma, -u, k)$. Let

$$\nu(\sigma, -u, k) = \mathbb{E}\left[F_{(k-u)^+}(u, \mathbf{S}(u, (k-u)^+))\right].$$

(*ii*) *Supermodularity in* (σ, k). Without loss of generality, let $u = 0$. It suffices to show that $\nu(\sigma, 0, k) - \nu(\sigma, 0, k-1)$ is increasing in σ for all $k \in \{1, 2, \ldots, n\}$. Part (*ii*) proceeds in three steps.

Step (*ii*)-1 (Decomposition). Applying (8) in Lemma 2 yields

$$\begin{aligned} \nu(\sigma, 0, k) - \nu(\sigma, 0, k-1) &= \mathbb{E}\left[(S_k - \hat{S}_{[k]}(0, k, k))^+ + F_{k-1}(0, \mathbf{S}(0, k-1)) - F_{k-1}(0, \mathbf{S}(0, k-1))\right] \\ &= \mathbb{E}\left[(S_k - \hat{S}_{[k]}(0, k, k))^+\right], \end{aligned} \tag{A12}$$

where $\hat{S}_{[k]}(0, k, k)$ is the kth order statistic of the random vector $\hat{\mathbf{S}}(0, k, k) = (\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_{2k-1}) = (S_1, S_2, \ldots, S_{k-1}, \Delta G(1-d), \Delta G(2-d), \ldots, \Delta G(k-d))$.

By (10), $S_i = \sigma\varepsilon_i + \mu_i$, so

$$(\text{A12}) = \mathbb{E}\left[(\sigma\varepsilon_k + \mu_k - \hat{S}_{[k]}(0, k, k))^+\right].$$

To highlight the dependence on σ, we slightly modify the notation and define $S_i^{-k}(\sigma)$ as the ith element and $S_{[k]}^{-k}(\sigma)$ as the kth order statistic of $\hat{\mathbf{S}}(0, k, k)$ when the standard deviation of S_i is σ. Therefore, it is equivalent to show that for any $\delta \geq 0$,

$$\mathbb{E}\left[((\sigma+\delta)\varepsilon_k + \mu_k - S_{[k]}^{-k}(\sigma+\delta))^+\right] \geq \mathbb{E}\left[(\sigma\varepsilon_k + \mu_k - S_{[k]}^{-k}(\sigma))^+\right].$$

In the sequel, we use $s_i^{-k}(\cdot)$ and $s_{[k]}^{-k}(\cdot)$ to represent the realizations of $S_i^{-k}(\cdot)$ and $S_{[k]}^{-k}(\cdot)$, respectively.

Step (*ii*)-2 (Asymmetric marginal values around zero). The key idea in this step is that, given any realization $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{k-1})$ and a *positive* realization ε_k, we can show that

$$\begin{aligned} & ((\sigma+\delta)\varepsilon_k + \mu_k - s_{[k]}^{-k}(\sigma+\delta))^+ - (\sigma\varepsilon_k + \mu_k - s_{[k]}^{-k}(\sigma))^+ \\ \geq\ & (-\sigma\varepsilon_k + \mu_k - s_{[k]}^{-k}(\sigma))^+ - (-(\sigma+\delta)\varepsilon_k + \mu_k - s_{[k]}^{-k}(\sigma+\delta))^+. \end{aligned} \tag{A13}$$

That is, the marginal value of σ for any positive realization ε_k is always larger than the marginal loss for its negative counterpart $-\varepsilon_k$. Let us for this moment suppose that (A13) is true. Then,

$$\begin{aligned}
&\mathbb{E}\left[((\sigma+\delta)\varepsilon_k+\mu_k-s_{[k]}^{-k}(\sigma+\delta))^+\right]\\
=&\int_{-\infty}^{0}((\sigma+\delta)\varepsilon_k+\mu_k-s_{[k]}^{-k}(\sigma+\delta))^+\phi_k(\varepsilon_k)\,\mathrm{d}\varepsilon_k+\int_{0}^{\infty}((\sigma+\delta)\varepsilon_k+\mu_k-s_{[k]}^{-k}(\sigma+\delta))^+\phi_k(\varepsilon_k)\,\mathrm{d}\varepsilon_k\\
=&\int_{0}^{\infty}((\sigma+\delta)(-\varepsilon_k)+\mu_k-s_{[k]}^{-k}(\sigma+\delta))^+\phi_k(\varepsilon_k)\,\mathrm{d}\varepsilon_k+\int_{0}^{\infty}((\sigma+\delta)\varepsilon_k+\mu_k-s_{[k]}^{-k}(\sigma+\delta))^+\phi_k(\varepsilon_k)\,\mathrm{d}\varepsilon_k\\
\geq&\int_{0}^{\infty}(-\sigma\varepsilon_k+\mu_k-s_{[k]}^{-k}(\sigma))^+\phi_k(\varepsilon_k)\,\mathrm{d}\varepsilon_k+\int_{0}^{\infty}(\sigma\varepsilon_k+\mu_k-s_{[k]}^{-k}(\sigma))^+\phi_k(\varepsilon_k)\,\mathrm{d}\varepsilon_k\\
=&\mathbb{E}\left[(\sigma\varepsilon_k+\mu_k-s_{[k]}^{-k}(\sigma))^+\right],
\end{aligned}$$

where the second equality follows from the symmetry of ε_k, and the inequality follows from (A13). Because this holds for any realization $(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_{k-1})$, we complete the proof for part (ii).

Step (ii)-3 (Verification of (A13)). We first show that the left-hand side of (A13) is positive:

$$((\sigma+\delta)\varepsilon_k+\mu_k-s_{[k]}^{-k}(\sigma+\delta))^+-(\sigma\varepsilon_k+\mu_k-s_{[k]}^{-k}(\sigma))^+\geq 0. \tag{A14}$$

If $s_{[k]}^{-k}(\sigma)\geq\min\left\{\sigma\varepsilon_k+\mu_k,s_{[k]}^{-k}(\sigma+\delta)\right\}$, because $\varepsilon_k\geq 0$, (A14) holds. So we consider the case when $s_{[k]}^{-k}(\sigma)<\min\left\{\sigma\varepsilon_k+\mu_k,s_{[k]}^{-k}(\sigma+\delta)\right\}$. Recall that $\mu_1\geq\mu_2\geq\cdots\geq\mu_k$. For any $i\in\{1,2,\ldots,k-1\}$, if $\sigma\varepsilon_i+\mu_i\leq s_{[k]}^{-k}(\sigma)$, then $\sigma\varepsilon_i<\sigma\varepsilon_k+\mu_k-\mu_i\leq\sigma\varepsilon_k$. Because $\delta\geq 0$, $\delta\varepsilon_k\geq\delta\varepsilon_i$. Now, define set

$$\mathcal{A}=\left\{\delta\varepsilon_i\geq 0,\ i=1,2,\ldots,k-1:\sigma\varepsilon_i+\mu_i\leq s_{[k]}^{-k}(\sigma)\right\}.$$

We consider the following two cases.

(a) If $\mathcal{A}=\emptyset$, then $s_i^{-k}(\sigma+\delta)\leq s_{[k]}^{-k}(\sigma)$ for all $i\in\{1,2,\ldots,2k-1\}$ with $s_i^{-k}(\sigma)\leq s_{[k]}^{-k}(\sigma)$. Because the total number of these i's is more than k, one of the $s_i^{-k}(\sigma+\delta)$'s must be a candidate to be $s_{[k]}^{-k}(\sigma+\delta)$. Thus, $s_{[k]}^{-k}(\sigma+\delta)\leq s_{[k]}^{-k}(\sigma)$ (See the left arrow in Figure A1).

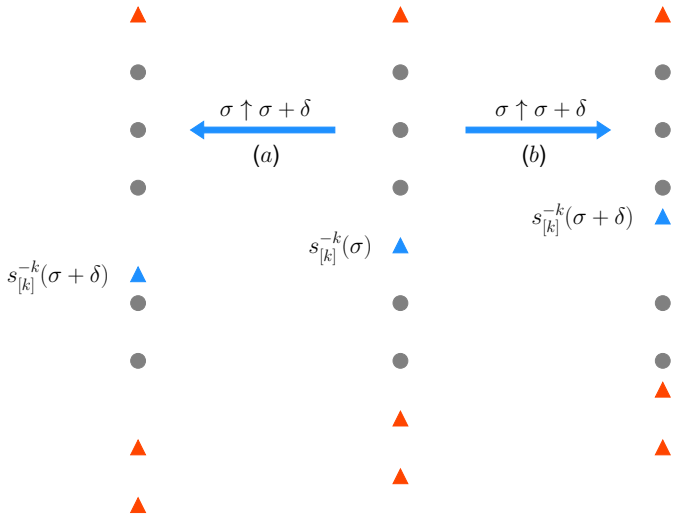
(b) If $\mathcal{A}\neq\emptyset$, then $s_{[k]}^{-k}(\sigma+\delta)-s_{[k]}^{-k}(\sigma)\leq\max\mathcal{A}\leq\delta\varepsilon_k$, where the first inequality follows from a contradiction argument: if $s_{[k]}^{-k}(\sigma+\delta)-s_{[k]}^{-k}(\sigma)>\max\mathcal{A}$, then $s_{[k]}^{-k}(\sigma+\delta)>s_i^{-k}(\sigma+\delta)$ for all $i\in\{1,2,\ldots,2k-1\}$ with $s_i^{-k}(\sigma)\leq s_{[k]}^{-k}(\sigma)$. Because the total number of these i's is more than k, no more than $(k-1)$ elements in $(s_1^{-k}(\sigma+\delta),s_2^{-k}(\sigma+\delta),\ldots,s_{2k-1}^{-k}(\sigma+\delta))$ are candidates to be $s_{[k]}^{-k}(\sigma+\delta)$, which is impossible (See the right arrow in Figure A1).

By cases (a) and (b), we have $s_{[k]}^{-k}(\sigma+\delta)-s_{[k]}^{-k}(\sigma)\leq\delta\varepsilon_k$. Therefore,

$$\begin{aligned}
&((\sigma+\delta)\varepsilon_k+\mu_k-s_{[k]}^{-k}(\sigma+\delta))^+-(\sigma\varepsilon_k+\mu_k-s_{[k]}^{-k}(\sigma))^+\\
=&((\sigma+\delta)\varepsilon_k+\mu_k-s_{[k]}^{-k}(\sigma+\delta))^+-(\sigma\varepsilon_k+\mu_k-s_{[k]}^{-k}(\sigma))\\
\geq&((\sigma+\delta)\varepsilon_k+\mu_k-s_{[k]}^{-k}(\sigma+\delta))^+-(\sigma\varepsilon_k+\mu_k-s_{[k]}^{-k}(\sigma+\delta)+\delta\varepsilon_k)\\
\geq&0.
\end{aligned}$$

Next, we verify (A13). If $s_{[k]}^{-k}(\sigma)\geq-\sigma\varepsilon_k+\mu_k$, then the right-hand side of (A13) is always negative. Because the left-hand side is positive (by (A14)), (A13) holds. Thus, in the sequel, we focus on the case where $s_{[k]}^{-k}(\sigma)<-\sigma\varepsilon_k+\mu_k$. In this case, we claim that $s_{[k]}^{-k}(\sigma+\delta)\leq s_{[k]}^{-k}(\sigma)$. Indeed, because $s_{[k]}^{-k}(\sigma)<-\sigma\varepsilon_k+\mu_k$,

Figure A1 Graphical Illustrations for Cases (a) and (b) in Step (ii)-3



Note. The circles denote $\Delta G(i-d)$'s and the triangles represent s_i's in the vector $\hat{\mathbf{s}}(0,k,k)$. The elements are arranged in descending order from top to bottom. The middle vector corresponds to the original vector $\hat{\mathbf{s}}(0,k,k)$ when the standard deviation of S_i is σ. The left arrow illustrates case (a) and the right arrow depicts case (b). In case (a), all s_i's that are less than the median $s_{[k]}^{-k}(\sigma)$ in the original vector decrease as σ increases to $\sigma+\delta$. In case (b), all s_i's that are less than $s_{[k]}^{-k}(\sigma)$ in the original vector cannot increase by more than $\delta\varepsilon_k$.

for any $i \in \{1,2,\ldots,k-1\}$ with $\sigma\varepsilon_i + \mu_i \le s_{[k]}^{-k}(\sigma)$, we have $\sigma\varepsilon_i < -\sigma\varepsilon_k + \mu_k - \mu_i \le -\sigma\varepsilon_k$, which implies that $\delta\varepsilon_i < -\delta\varepsilon_k \le 0$. Therefore, $s_{[k]}^{-k}(\sigma) \ge s_i^{-k}(\sigma+\delta)$ for all $i \in \{1,2,\ldots,2k-1\}$ with $s_i^{-k}(\sigma) \le s_{[k]}^{-k}(\sigma)$. Because the total number of these i's is more than k, one of the $s_i^{-k}(\sigma+\delta)$'s must be the candidate to be $s_{[k]}^{-k}(\sigma+\delta)$. Thus, $s_{[k]}^{-k}(\sigma+\delta) \le s_{[k]}^{-k}(\sigma) < -\sigma\varepsilon_k + \mu_k \le \sigma\varepsilon_k + \mu_k$. It follows that

$$
\begin{aligned}
&((\sigma+\delta)\varepsilon_k + \mu_k - s_{[k]}^{-k}(\sigma+\delta))^+ - (\sigma\varepsilon_k + \mu_k - s_{[k]}^{-k}(\sigma))^+ \\
&= ((\sigma+\delta)\varepsilon_k + \mu_k - s_{[k]}^{-k}(\sigma+\delta)) - (\sigma\varepsilon_k + \mu_k - s_{[k]}^{-k}(\sigma)) \\
&= \delta\varepsilon_k + s_{[k]}^{-k}(\sigma) - s_{[k]}^{-k}(\sigma+\delta) \\
&\ge \delta\varepsilon_k - s_{[k]}^{-k}(\sigma) + s_{[k]}^{-k}(\sigma+\delta) \\
&= (-\sigma\varepsilon_k + \mu_k - s_{[k]}^{-k}(\sigma)) - (-(\sigma+\delta)\varepsilon_k + \mu_k - s_{[k]}^{-k}(\sigma+\delta)) \\
&\ge (-\sigma\varepsilon_k + \mu_k - s_{[k]}^{-k}(\sigma))^+ - (-(\sigma+\delta)\varepsilon_k + \mu_k - s_{[k]}^{-k}(\sigma+\delta))^+,
\end{aligned}
$$

which completes the verification.

(iii) *Supermodularity in* $(\sigma,-u)$. We shall show that $\nu(\sigma,-u,k) - \nu(\sigma,-u-1,k)$ is increasing in σ. When $k \le u$, the result trivially holds, so we assume that $k \ge u+1$. Without loss of generality, let $u=0$. This part also proceeds in three steps.

Step (iii)-1 (Decomposition). Using (A9) derived in the proof of Lemma 3, we obtain

$$
\begin{aligned}
\nu(\sigma,0,k) - \nu(\sigma,-1,k) &= \mathbb{E}\left[F_k(0,\mathbf{S}(0,k)) - F_{k-1}(1,\mathbf{S}(1,k-1))\right] \\
&= \mathbb{E}\left[\max\left\{S_1, \hat{S}_{[k]}(0,k,1)\right\}\right], \qquad \text{(A15)}
\end{aligned}
$$

where $\hat{S}_{[k]}(0,k,1)$ is the kth order statistic of the random vector $\hat{\mathbf{S}}(0,k,1) = (\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_{2k-1}) = (S_2, S_3, \ldots, S_k, \Delta G(1-d), \Delta G(2-d), \ldots, \Delta G(k-d))$.

Applying the same arguments as in Step (ii)-1, we further transform (A15) as follows:

$$(\text{A15}) = \mathbb{E}\left[\max\left\{\sigma\varepsilon_1 + \mu_1, \hat{S}_{[k]}(0,k,1)\right\}\right].$$

To emphasize the dependence on σ, we denote $S_i^{-1}(\sigma)$ as the ith element and $S_{[k]}^{-1}(\sigma)$ as the kth order statistic of $\hat{\mathbf{S}}(0,k,1)$ when the standard deviation of S_i is σ. Therefore, the goal is to show that for any $\delta \geq 0$,

$$\mathbb{E}\left[\max\left\{(\sigma+\delta)\varepsilon_1 + \mu_1, S_{[k]}^{-1}(\sigma+\delta)\right\}\right] \geq \mathbb{E}\left[\max\left\{\sigma\varepsilon_1 + \mu_1, S_{[k]}^{-1}(\sigma)\right\}\right].$$

In the sequel, we use $s_i^{-1}(\cdot)$ and $s_{[k]}^{-1}(\cdot)$ to represent the realizations of $S_i^{-1}(\cdot)$ and $S_{[k]}^{-1}(\cdot)$, respectively.

Step (iii)-2 (Asymmetric marginal values around zero). Following Step (ii)-2, given any realization $(\varepsilon_2, \varepsilon_3, \ldots, \varepsilon_k)$ and a positive realization ε_1, if we can show that

$$\begin{aligned} &\max\left\{(\sigma+\delta)\varepsilon_1 + \mu_1, s_{[k]}^{-1}(\sigma+\delta)\right\} - \max\left\{\sigma\varepsilon_1 + \mu_1, s_{[k]}^{-1}(\sigma)\right\} \\ \geq\ &\max\left\{-\sigma\varepsilon_1 + \mu_1, s_{[k]}^{-1}(\sigma)\right\} - \max\left\{-(\sigma+\delta)\varepsilon_1 + \mu_1, s_{[k]}^{-1}(\sigma+\delta)\right\}, \end{aligned} \tag{A16}$$

then, we have

$$\begin{aligned} &\mathbb{E}\left[\max\left\{(\sigma+\delta)\varepsilon_1 + \mu_1, s_{[k]}^{-1}(\sigma+\delta)\right\}\right] \\ =\ &\int_0^\infty \max\left\{-(\sigma+\delta)\varepsilon_1 + \mu_1, s_{[k]}^{-1}(\sigma+\delta)\right\}\phi_1(\varepsilon_1)\,\mathrm{d}\varepsilon_1 + \int_0^\infty \max\left\{(\sigma+\delta)\varepsilon_1 + \mu_1, s_{[k]}^{-1}(\sigma+\delta)\right\}\phi_1(\varepsilon_1)\,\mathrm{d}\varepsilon_1 \\ \geq\ &\int_0^\infty \max\left\{-\sigma\varepsilon_1 + \mu_1, s_{[k]}^{-1}(\sigma)\right\}\phi_1(\varepsilon_1)\,\mathrm{d}\varepsilon_1 + \int_0^\infty \max\left\{\sigma\varepsilon_1 + \mu_1, s_{[k]}^{-1}(\sigma)\right\}\phi_1(\varepsilon_1)\,\mathrm{d}\varepsilon_1 \\ =\ &\mathbb{E}\left[\max\left\{\sigma\varepsilon_1 + \mu_1, s_{[k]}^{-1}(\sigma)\right\}\right], \end{aligned}$$

where the first equality follows from the symmetry of ε_i, and the inequality follows from (A16). Because this holds for any realization $(\varepsilon_2, \varepsilon_3, \ldots, \varepsilon_k)$, we complete the proof for part (iii).

Step (iii)-3 (Verification of (A16)). We first show that the left-hand side of (A16) is positive:

$$\max\left\{(\sigma+\delta)\varepsilon_1 + \mu_1, s_{[k]}^{-1}(\sigma+\delta)\right\} - \max\left\{\sigma\varepsilon_1 + \mu_1, s_{[k]}^{-1}(\sigma)\right\} \geq 0. \tag{A17}$$

If $s_{[k]}^{-1}(\sigma) \leq \max\left\{\sigma\varepsilon_1 + \mu_1, s_{[k]}^{-1}(\sigma+\delta)\right\}$, then

$$\begin{aligned} &\max\left\{(\sigma+\delta)\varepsilon_1 + \mu_1, s_{[k]}^{-1}(\sigma+\delta)\right\} - \max\left\{\sigma\varepsilon_1 + \mu_1, s_{[k]}^{-1}(\sigma)\right\} \\ \geq\ &\max\left\{(\sigma+\delta)\varepsilon_1 + \mu_1, s_{[k]}^{-1}(\sigma+\delta)\right\} - \max\left\{\sigma\varepsilon_1 + \mu_1, s_{[k]}^{-1}(\sigma+\delta)\right\} \\ \geq\ &0, \end{aligned}$$

which implies that (A17) holds. We next show by contradiction that $s_{[k]}^{-1}(\sigma) > \max\left\{\sigma\varepsilon_1 + \mu_1, s_{[k]}^{-1}(\sigma+\delta)\right\}$ cannot hold. Suppose that it holds. For any $i \in \{2, 3, \ldots, k\}$, if $\sigma\varepsilon_i + \mu_i \geq s_{[k]}^{-1}(\sigma)$, then $\sigma\varepsilon_i > \sigma\varepsilon_1 + \mu_1 - \mu_i \geq \sigma\varepsilon_1$. Because $\delta \geq 0$ and $\varepsilon_1 \geq 0$, we have $\delta\varepsilon_i \geq \delta\varepsilon_1 \geq 0$. Now, define set

$$\mathcal{B} = \left\{\delta\varepsilon_i,\ i = 2, 3, \ldots, k : \sigma\varepsilon_i + \mu_i \geq s_{[k]}^{-1}(\sigma)\right\}.$$

We consider the following two cases.

(*a*) If $\mathcal{B}=\emptyset$, then $s_i^{-1}(\sigma+\delta)\geq s_{[k]}^{-1}(\sigma)$ for all $i\in\{1,2,\ldots,2k-1\}$ with $s_i^{-1}(\sigma)\geq s_{[k]}^{-1}(\sigma)$. Because the total number of these i's is more than k, one of the $s_i^{-1}(\sigma+\delta)$'s must be a candidate to be $s_{[k]}^{-1}(\sigma+\delta)$. Thus, $s_{[k]}^{-1}(\sigma+\delta)\geq s_{[k]}^{-1}(\sigma)$, which is a contradiction.

(*b*) If $\mathcal{B}\neq\emptyset$, then $s_{[k]}^{-1}(\sigma+\delta)-s_{[k]}^{-1}(\sigma)\geq\min\mathcal{B}\geq 0$, which contradicts $s_{[k]}^{-1}(\sigma)>s_{[k]}^{-1}(\sigma+\delta)$. Here, the first inequality follows from a contradiction argument: if $s_{[k]}^{-1}(\sigma+\delta)-s_{[k]}^{-1}(\sigma)<\min\mathcal{B}$, then $s_{[k]}^{-1}(\sigma+\delta)<s_i^{-1}(\sigma+\delta)$ for all $i\in\{1,2,\ldots,2k-1\}$ with $s_i^{-1}(\sigma)\geq s_{[k]}^{-1}(\sigma)$. Because the total number of these i's is more than k, and $s_{[k]}^{-1}(\sigma+\delta)$ is strictly less than these $s_i^{-1}(\sigma+\delta)$'s, $s_{[k]}^{-1}(\sigma+\delta)$ is impossible to take.

By cases (*a*) and (*b*), we conclude that $s_{[k]}^{-1}(\sigma)>\max\left\{\sigma\varepsilon_1+\mu_1,s_{[k]}^{-1}(\sigma+\delta)\right\}$ cannot hold.

Next, we verify (A16). If $s_{[k]}^{-1}(\sigma)\leq-\sigma\varepsilon_1+\mu_1$, then

$$\begin{aligned}
&\max\left\{(\sigma+\delta)\varepsilon_1+\mu_1,s_{[k]}^{-1}(\sigma+\delta)\right\}-\max\left\{\sigma\varepsilon_1+\mu_1,s_{[k]}^{-1}(\sigma)\right\}\\
\geq\;&(\sigma+\delta)\varepsilon_1+\mu_1-(\sigma\varepsilon_1+\mu_1)\\
\geq\;&-\sigma\varepsilon_1+\mu_1-\max\left\{-(\sigma+\delta)\varepsilon_1+\mu_1,s_{[k]}^{-1}(\sigma+\delta)\right\}\\
=\;&\max\left\{-\sigma\varepsilon_1+\mu_1,s_{[k]}^{-1}(\sigma)\right\}-\max\left\{-(\sigma+\delta)\varepsilon_1+\mu_1,s_{[k]}^{-1}(\sigma+\delta)\right\},
\end{aligned}$$

which implies that (A16) holds.

If $s_{[k]}^{-1}(\sigma)>-\sigma\varepsilon_1+\mu_1$, because $s_{[k]}^{-1}(\sigma)\leq\max\left\{\sigma\varepsilon_1+\mu_1,s_{[k]}^{-1}(\sigma+\delta)\right\}$, we only need to consider two cases. (1) If $s_{[k]}^{-1}(\sigma)\leq s_{[k]}^{-1}(\sigma+\delta)$, then the right-hand side of (A16) is always negative. Because the left-hand side is positive (by (A17)), (A16) holds. (2) If $s_{[k]}^{-1}(\sigma+\delta)<s_{[k]}^{-1}(\sigma)\leq\sigma\varepsilon_1+\mu_1$, we claim that $s_{[k]}^{-1}(\sigma)-s_{[k]}^{-1}(\sigma+\delta)\leq\delta\varepsilon_1$. Indeed, because $s_{[k]}^{-1}(\sigma)>-\sigma\varepsilon_1+\mu_1$, for any $i\in\{2,3,\ldots,k\}$ with $\sigma\varepsilon_i+\mu_i\geq s_{[k]}^{-1}(\sigma)$, we have $\sigma\varepsilon_i>-\sigma\varepsilon_1+\mu_1-\mu_i\geq-\sigma\varepsilon_1$, which implies that $\delta\varepsilon_i\geq-\delta\varepsilon_1$. Therefore, $s_i^{-1}(\sigma+\delta)\geq s_i^{-1}(\sigma)-\delta\varepsilon_1\geq s_{[k]}^{-1}(\sigma)-\delta\varepsilon_1$ for all $i\in\{1,2,\ldots,2k-1\}$ with $s_i^{-1}(\sigma)\geq s_{[k]}^{-1}(\sigma)$. Because the total number of these i's is more than k, one of the $s_i^{-1}(\sigma+\delta)$'s must be a candidate to be $s_{[k]}^{-1}(\sigma+\delta)$. Thus, $s_{[k]}^{-1}(\sigma+\delta)\geq s_{[k]}^{-1}(\sigma)-\delta\varepsilon_1$. It follows that

$$\begin{aligned}
&\max\left\{(\sigma+\delta)\varepsilon_1+\mu_1,s_{[k]}^{-1}(\sigma+\delta)\right\}-\max\left\{\sigma\varepsilon_1+\mu_1,s_{[k]}^{-1}(\sigma)\right\}\\
=\;&\delta\varepsilon_1\\
\geq\;&s_{[k]}^{-1}(\sigma)-s_{[k]}^{-1}(\sigma+\delta)\\
=\;&\max\left\{-\sigma\varepsilon_1+\mu_1,s_{[k]}^{-1}(\sigma)\right\}-s_{[k]}^{-1}(\sigma+\delta)\\
\geq\;&\max\left\{-\sigma\varepsilon_1+\mu_1,s_{[k]}^{-1}(\sigma)\right\}-\max\left\{-(\sigma+\delta)\varepsilon_1+\mu_1,s_{[k]}^{-1}(\sigma+\delta)\right\},
\end{aligned}$$

which completes the verification.

(*iv*) Because $v(\sigma,-u,k)$ is supermodular in (σ,k), $(\sigma,-u)$, and $(-u,k)$ (by Lemma 3), we conclude that $v(\sigma,-u,k)$ is supermodular in $(\sigma,-u,k)$. Therefore, the optimal policy (u^*,z^*,k^*) has the desired monotonicity properties with respect to σ^2 (Topkis 1998). □

Proof of Proposition 1. The result for Problem $(\mathcal{P}_0^u)$ is straightforward. The result for Problem $(\mathcal{P}_0^z)$ can be shown by using the argument for the second part of the proof of Theorem 1. Therefore, we omit the proof for brevity. □

Proof of Lemma 4. We first show the concavity. Without loss of generality, let $u=0$. By Lemma 2, for any $z\in\{1,2,\ldots,n\}$,

$$\mathbb{E}\left[F_z(0,\mathbf{S}(0,z))\right]-\mathbb{E}\left[F_{z-1}(0,\mathbf{S}(0,z-1))\right]=\mathbb{E}\left[(S_z-\hat{S}_{[z]}(0,z,z))^+\right], \tag{A18}$$

where $\hat{S}_{[z]}(0,z,z)$ is the zth order statistic in the random vector $\hat{\mathbf{S}}(0,z,z)=(\hat{S}_1,\hat{S}_2,\ldots,\hat{S}_{2z-1})=(S_1,S_2,\ldots,S_{z-1},\Delta G(1-d),\Delta G(2-d),\ldots,\Delta G(z-d))$. It suffices to show that (A18) is decreasing in z.

Because $S_{z+1}\leq_{\text{st}} S_z$, and the three random variables $\hat{S}_{[z]}(0,z,z)$, S_z, and S_{z+1} are mutually independent, we have $(S_{z+1}-\hat{S}_{[z]}(0,z,z))^+\leq_{\text{st}}(S_z-\hat{S}_{[z]}(0,z,z))^+$ (Shaked and Shanthikumar 2007, Theorem 1.A.3 (a) and (b), p. 6). In addition, by the second inequality in Lemma A1, $\hat{S}_{[z]}(0,z,z)\leq\hat{S}_{[z+1]}(0,z+1,z+1)$, almost surely. Thus, $(S_{z+1}-\hat{S}_{[z+1]}(0,z+1,z+1))^+\leq(S_{z+1}-\hat{S}_{[z]}(0,z,z))^+$, almost surely. It follows that

$$\mathbb{E}\left[(S_{z+1}-\hat{S}_{[z+1]}(0,z+1,z+1))^+\right]\leq\mathbb{E}\left[(S_{z+1}-\hat{S}_{[z]}(0,z,z))^+\right]\leq\mathbb{E}\left[(S_z-\hat{S}_{[z]}(0,z,z))^+\right].$$

Next, we show the decreasing differences; that is,

$$\mathbb{E}\left[F_{z+1}(u+1,\mathbf{S}(u+1,z+1))\right]-\mathbb{E}\left[F_z(u+1,\mathbf{S}(u+1,z))\right]\leq\mathbb{E}\left[F_{z+1}(u,\mathbf{S}(u,z+1))\right]-\mathbb{E}\left[F_z(u,\mathbf{S}(u,z))\right].$$

Letting $z=k-u$, we have

$$\begin{aligned}
&\mathbb{E}\left[F_{z+1}(u+1,\mathbf{S}(u+1,z+1))\right]-\mathbb{E}\left[F_z(u+1,\mathbf{S}(u+1,z))\right]\\
&=\mathbb{E}\left[F_{k+1-u}(u+1,\mathbf{S}(u+1,k+1-u))\right]-\mathbb{E}\left[F_{k-u}(u+1,\mathbf{S}(u+1,k-u))\right]\\
&\leq\mathbb{E}\left[F_{k+2-u}(u,\mathbf{S}(u,k+2-u))\right]-\mathbb{E}\left[F_{k+1-u}(u,\mathbf{S}(u,k+1-u))\right]\\
&=\mathbb{E}\left[F_{z+2}(u,\mathbf{S}(u,z+2))\right]-\mathbb{E}\left[F_{z+1}(u,\mathbf{S}(u,z+1))\right]\\
&\leq\mathbb{E}\left[F_{z+1}(u,\mathbf{S}(u,z+1))\right]-\mathbb{E}\left[F_z(u,\mathbf{S}(u,z))\right],
\end{aligned}$$

where the first inequality follows because $\mathbb{E}\left[F_{k-u}(u,\mathbf{S}(u,k-u))\right]$ is submodular in (u,k) by Lemma 3, and the second inequality follows because $\mathbb{E}\left[F_z(u,\mathbf{S}(u,z))\right]$ is discrete concave in z. □

Proof of Theorem 3. The proof is established via two steps. In the first step, we show that $u^*\leq u'\leq u^*+z^*$, and in the second step, we show that $u^*+z^*\leq z'$ if $z^*>0$.

Step 1 ($u^*\leq u'\leq u^*+z^*$). We first prove the first inequality, and then the second.

(i) The proof for $u^*\leq u'$ is by contradiction. Suppose that $u^*>u'$. The idea of the proof is to construct a feasible solution $(\tilde{u},\tilde{z})$ for Problem $(\mathcal{P}_1)$ that achieves a higher reward than that under (u^*,z^*). Let $(\tilde{u},\tilde{z})=(u',z^*)$. Because $\tilde{z}=z^*\leq n-u^*\leq n-u'$, policy $(\tilde{u},\tilde{z})$ is feasible. Computing the difference of the rewards for the two policies, we obtain

$$\begin{aligned}
&\sum_{i=1}^{\tilde{u}}\mu_i-c\tilde{z}+\mathbb{E}\left[F_{\tilde{z}}(\tilde{u},\mathbf{S}(\tilde{u},\tilde{z}))\right]-\sum_{i=1}^{u^*}\mu_i+cz^*-\mathbb{E}\left[F_{z^*}(u^*,\mathbf{S}(u^*,z^*))\right]\\
&=\mathbb{E}\left[F_{z^*}(u',\mathbf{S}(u',z^*))\right]-\mathbb{E}\left[F_{z^*}(u^*,\mathbf{S}(u^*,z^*))\right]+\sum_{i=1}^{u'}\mu_i-\sum_{i=1}^{u^*}\mu_i\\
&=\mathbb{E}\left[\max_{0\leq h\leq z^*}\left\{\sum_{i=1}^{h}S_{[i]}(u',z^*)-G(u^*+h-d)-G(u'+h-d)+G(u^*+h-d)\right\}\right]
\end{aligned}$$

$$
\begin{aligned}
&- \mathbb{E}\left[F_{z^*}(u^*, \mathbf{S}(u^*, z^*))\right] + \sum_{i=1}^{u'} \mu_i - \sum_{i=1}^{u^*} \mu_i \\
&\geq \mathbb{E}\left[F_{z^*}(u^*, \mathbf{S}(u', z^*))\right] - \mathbb{E}\left[F_{z^*}(u^*, \mathbf{S}(u^*, z^*))\right] + \sum_{i=1}^{u'} \mu_i - G(u' - d) - \sum_{i=1}^{u^*} \mu_i + G(u^* - d) \\
&> 0,
\end{aligned}
$$

which contradicts the optimality of (u^*, z^*) for Problem $(\mathcal{P}_1)$. Here, the first inequality follows from the convexity of $G(\cdot)$ and $u^* > u'$. For the second inequality, because $u^* > u'$, $\mu_{u'+i} \geq \mu_{u^*+i}$ for all $i \in \{1, 2, \ldots, z^*\}$. Then, by Lemma 1,

$$
\mathbb{E}\left[F_{z^*}(u^*, \mathbf{S}(u', z^*))\right] - \mathbb{E}\left[F_{z^*}(u^*, \mathbf{S}(u^*, z^*))\right] \geq 0.
$$

In addition, by the optimality of u' for Problem $(\mathcal{P}_1^u)$,

$$
\sum_{i=1}^{u'} \mu_i - G(u' - d) - \left(\sum_{i=1}^{u^*} \mu_i - G(u^* - d) \right) > 0.
$$

Therefore, $u^* \leq u'$.

(ii) The proof for $u' \leq u^* + z^*$ is by contradiction. Suppose that $u' > u^* + z^*$. Similarly, the idea of the proof is to construct a feasible solution $(\tilde{u}, \tilde{z})$ for Problem $(\mathcal{P}_1)$ that achieves a higher reward than that under (u^*, z^*). Let $(\tilde{u}, \tilde{z}) = (u' - z^*, z^*)$. Because $\tilde{u} = u' - z^* > u^* \geq 0$ and $\tilde{u} + \tilde{z} = u' \leq n$, policy $(\tilde{u}, \tilde{z})$ is feasible. Computing the difference of the rewards for the two policies, we obtain

$$
\begin{aligned}
&\sum_{i=1}^{\tilde{u}} \mu_i - c\tilde{z} + \mathbb{E}\left[F_{\tilde{z}}(\tilde{u}, \mathbf{S}(\tilde{u}, \tilde{z}))\right] - \sum_{i=1}^{u^*} \mu_i + cz^* - \mathbb{E}\left[F_{z^*}(u^*, \mathbf{S}(u^*, z^*))\right] \\
&= \sum_{i=1}^{u'-z^*} \mu_i + \mathbb{E}\left[F_{z^*}(u' - z^*, \mathbf{S}(u' - z^*, z^*))\right] - \sum_{i=1}^{u^*} \mu_i - \mathbb{E}\left[F_{z^*}(u^*, \mathbf{S}(u^*, z^*))\right] \\
&= \mathbb{E}\left[\max_{0 \leq h \leq z^*} \left\{ \sum_{i=1}^{h} S_{[i]}(u' - z^*, z^*) - G(u^* + h - d) - G(u' - z^* + h - d) + G(u^* + h - d) \right\} \right] \\
&\qquad - \mathbb{E}\left[F_{z^*}(u^*, \mathbf{S}(u^*, z^*))\right] + \sum_{i=1}^{u'-z^*} \mu_i - \sum_{i=1}^{u^*} \mu_i \\
&\geq \mathbb{E}\left[F_{z^*}(u^*, \mathbf{S}(u' - z^*, z^*))\right] - \mathbb{E}\left[F_{z^*}(u^*, \mathbf{S}(u^*, z^*))\right] - G(u' - d) + G(u^* + z^* - d) + \sum_{i=1}^{u'-z^*} \mu_i - \sum_{i=1}^{u^*} \mu_i \\
&\geq 0,
\end{aligned}
$$

which contradicts the optimality of (u^*, z^*) for Problem $(\mathcal{P}_1)$. Here, the first inequality follows from the convexity of $G(\cdot)$ and $u' - z^* > u^*$. For the second inequality, because $u' - z^* > u^*$, $\mu_{u^*+i} \geq \mu_{u'-z^*+i}$ for all $i \in \{1, 2, \ldots, z^*\}$. Then, by Lemma 1,

$$
\mathbb{E}\left[F_{z^*}(u^*, \mathbf{S}(u' - z^*, z^*))\right] - \mathbb{E}\left[F_{z^*}(u^*, \mathbf{S}(u^*, z^*))\right] \geq \sum_{i=u'-z^*+1}^{u'} \mu_i - \sum_{i=u^*+1}^{u^*+z^*} \mu_i,
$$

and by the optimality of u' for Problem $(\mathcal{P}_1^u)$,

$$
\sum_{i=1}^{u'-z^*} \mu_i - \sum_{i=1}^{u^*} \mu_i + \sum_{i=u'-z^*+1}^{u'} \mu_i - \sum_{i=u^*+1}^{u^*+z^*} \mu_i - G(u' - d) + G(u^* + z^* - d)
$$

$$
\begin{aligned}
&= \sum_{i=1}^{u'} \mu_i - G(u'-d) - \left(\sum_{i=1}^{u^*+z^*} \mu_i - G(u^*+z^*-d) \right) \\
&\geq 0.
\end{aligned}
$$

Therefore, $u' \leq u^* + z^*$.

Step 2 ($u^* + z^* \leq z'$ if $z^* > 0$). The proof is by contradiction. Suppose that $z' < u^* + z^*$. The idea of the proof is to construct a feasible solution $(\tilde{u}, \tilde{z})$ for Problem $(\mathcal{P}_1)$ that achieves a higher reward than that under (u^*, z^*). We consider the following two cases.

Case 1. $z' \geq u^*$. Let $(\tilde{u}, \tilde{z}) = (u^*, z' - u^*)$. Because $0 \leq \tilde{z} = z' - u^* < z^*$, policy $(\tilde{u}, \tilde{z})$ is feasible. Computing the difference of the rewards for the two policies, we obtain

$$
\begin{aligned}
& -c\tilde{z} + \mathbb{E}\left[F_{\tilde{z}}(\tilde{u}, \mathbf{S}(\tilde{u}, \tilde{z}))\right] + cz^* - \mathbb{E}\left[F_{z^*}(u^*, \mathbf{S}(u^*, z^*))\right] \\
= & -c(z'-u^*) + \mathbb{E}\left[F_{z'-u^*}(u^*, \mathbf{S}(u^*, z'-u^*))\right] + cz^* - \mathbb{E}\left[F_{z^*}(u^*, \mathbf{S}(u^*, z^*))\right] \\
\geq & -cz' + \mathbb{E}\left[F_{z'}(0, \mathbf{S}(0, z'))\right] + c(u^*+z^*) - \mathbb{E}\left[F_{u^*+z^*}(0, \mathbf{S}(0, u^*+z^*))\right] \\
> & \ 0,
\end{aligned}
$$

where the first inequality follows by Lemma 3 (submodularity), and the second inequality follows by the optimality of z' for Problem $(\mathcal{P}_1^z)$.

Case 2. $z' < u^*$. Let $(\tilde{u}, \tilde{z}) = (u^*, 0)$. Clearly, policy $(\tilde{u}, \tilde{z})$ is feasible. In addition, because $z^* > 0$, policy $(\tilde{u}, \tilde{z})$ is not identical to (u^*, z^*). Computing the difference of the rewards for the two policies, we obtain

$$
\begin{aligned}
& \mathbb{E}\left[F_{\tilde{z}}(\tilde{u}, \mathbf{S}(\tilde{u}, \tilde{z}))\right] + cz^* - \mathbb{E}\left[F_{z^*}(u^*, \mathbf{S}(u^*, z^*))\right] \\
= & \ \mathbb{E}\left[F_0(u^*, \mathbf{S}(u^*, 0))\right] + cz^* - \mathbb{E}\left[F_{z^*}(u^*, \mathbf{S}(u^*, z^*))\right] \\
\geq & \ \mathbb{E}\left[F_{u^*}(0, \mathbf{S}(0, u^*))\right] + cz^* - \mathbb{E}\left[F_{u^*+z^*}(0, \mathbf{S}(0, u^*+z^*))\right] \\
\geq & -cz' + \mathbb{E}\left[F_{z'}(0, \mathbf{S}(0, z'))\right] + c(z'+z^*) - \mathbb{E}\left[F_{z'+z^*}(0, \mathbf{S}(0, z'+z^*))\right] \\
> & \ 0,
\end{aligned}
$$

where the first inequality follows by Lemma 3 (submodularity), the second inequality follows by Lemma 4 (concavity), and the third inequality follows by the optimality of z' for Problem $(\mathcal{P}_1^z)$ and $z^* > 0$. □

Proof of Theorem 4. The proof consists of two steps. In the first step, we show an unconditional version of (17) that

$$
\mathbb{P}\left(\left| \frac{\mathcal{G}_z(u, \tilde{\mathbf{S}}(u,z))}{\mathbb{E}\left[\mathcal{G}_z(u, \tilde{\mathbf{S}}(u,z))\right]} - 1 \right| \geq \frac{\log(z)}{\sqrt{z}} \right) \leq \frac{C_2}{\log(z)}, \tag{A19}
$$

where C_2 is a positive constant that does not depend on z, n, and d. In the second step, we derive a similar bound for $\mathbb{E}[\mathcal{G}_z(u, \tilde{\mathbf{S}}(u,z)) \mid \mathbf{X}_1] / \mathbb{E}[\mathcal{G}_z(u, \tilde{\mathbf{S}}(u,z))]$. By the law of total variance, $\mathrm{Var}(\mathbb{E}[\mathcal{G}_z(u, \tilde{\mathbf{S}}(u,z)) \mid \mathbf{X}_1]) \leq \mathrm{Var}(\mathcal{G}_z(u, \tilde{\mathbf{S}}(u,z)))$. Thus, the second step is a direct consequence of the first step.

To verify (A19), we first show that $\mathrm{Var}(\mathcal{G}_z(u, \tilde{\mathbf{S}}(u,z))) = O(z\log(z))$. By (16), we have

$$
\mathcal{G}_z(u, \tilde{\mathbf{S}}(u,z)) - \mathbb{E}\left[\mathcal{G}_z(u, \tilde{\mathbf{S}}(u,z))\right] = \sum_{i=1}^{z} A_i + \sum_{i=1}^{z} B_i,
$$

where for $i=1,2,\ldots,z$,

$$\begin{cases} A_i = (\tilde{S}_{u+i} - \hat{S}_{[i]}(u,i,i))^+ - \mathbb{E}\left[(\tilde{S}_{u+i} - \hat{S}_{[i]}(u,i,i))^+ \middle| \mathcal{F}_{i-1}\right], \\ B_i = \mathbb{E}\left[(\tilde{S}_{u+i} - \hat{S}_{[i]}(u,i,i))^+ \middle| \mathcal{F}_{i-1}\right] - \mathbb{E}\left[(\tilde{S}_{u+i} - \hat{S}_{[i]}(u,i,i))^+\right]. \end{cases}$$

Here, $\mathcal{F}_i$, $i=1,2,\ldots$, is the σ-field generated by the data $(\tilde{S}_{u+1}, \tilde{S}_{u+2}, \ldots, \tilde{S}_{u+i})$, and $\mathcal{F}_0$ is the trivial σ-field. Clearly, $(\mathcal{F}_i)_{i\geq 0}$ is a filtration. Note that $\sum_{l=1}^{i} A_l$ is a martingale with respect to $\mathcal{F}_i$. Then, the martingale differences A_i and A_j for any $i \neq j$ are uncorrelated (Durrett 2019). It follows that

$$\begin{aligned} &\mathrm{Var}(\mathcal{G}_z(u, \tilde{\mathbf{S}}(u,z))) \\ &= \mathrm{Var}\left(\sum_{i=1}^{z} A_i + \sum_{i=1}^{z} B_i\right) \\ &\leq 2\mathrm{Var}\left(\sum_{i=1}^{z} A_i\right) + 2\mathrm{Var}\left(\sum_{i=1}^{z} B_i\right) \\ &= 2\sum_{i=1}^{z} \mathrm{Var}(A_i) + 2\mathbb{E}\left[\left(\sum_{i=1}^{z} B_i\right)^2\right] \\ &\leq 2\sum_{i=1}^{z} \mathrm{Var}(A_i) + 2z\sum_{i=1}^{z} \mathrm{Var}\left(B_i\right), \end{aligned} \tag{A20}$$

where the two inequalities follow from Cauchy–Schwarz inequality, and the second equality follows from the uncorrelation among A_i's.

We now show that $\sup_i \mathrm{Var}(A_i) < \infty$. We have

$$\begin{aligned} \mathrm{Var}(A_i) &= \mathbb{E}\left[\left((\tilde{S}_{u+i} - \hat{S}_{[i]}(u,i,i))^+\right)^2\right] - 2\mathbb{E}\left[\mathbb{E}\left[(\tilde{S}_{u+i} - \hat{S}_{[i]}(u,i,i))^+ \mathbb{E}\left[(\tilde{S}_{u+i} - \hat{S}_{[i]}(u,i,i))^+ \middle| \mathcal{F}_{i-1}\right] \middle| \mathcal{F}_{i-1}\right]\right] \\ &\quad + \mathbb{E}\left[\left(\mathbb{E}\left[(\tilde{S}_{u+i} - \hat{S}_{[i]}(u,i,i))^+ \middle| \mathcal{F}_{i-1}\right]\right)^2\right] \\ &= \mathbb{E}\left[\left((\tilde{S}_{u+i} - \hat{S}_{[i]}(u,i,i))^+\right)^2\right] - \mathbb{E}\left[\left(\mathbb{E}\left[(\tilde{S}_{u+i} - \hat{S}_{[i]}(u,i,i))^+ \middle| \mathcal{F}_{i-1}\right]\right)^2\right] \\ &\leq \mathbb{E}\left[(\tilde{S}_{u+i} - \hat{S}_{[i]}(u,i,i))^2 \mathbf{1}\left(\tilde{S}_{u+i} \geq \hat{S}_{[i]}(u,i,i)\right)\right], \end{aligned}$$

where the first equality follows from the law of total expectation, and $\mathbf{1}(\cdot)$ is an indicator function. In addition, by definition, it is easy to see that the median $\hat{S}_{[i]}(u,i,i)$ satisfies

$$\Delta G(u+1-d) \leq \hat{S}_{[i]}(u,i,i) \leq \Delta G(u+i-d), \text{ almost surely.} \tag{A21}$$

Then, it suffices to show that $\sup_{\xi\in[\Delta G(u+1-d), \Delta G(u+i-d)]} \mathbb{E}[(\tilde{S}_{u+i} - \xi)^2 \mathbf{1}(\tilde{S}_{u+1} \geq \xi)] < \infty$. By taking derivative with respect to ξ, it can be shown that the expectation is decreasing in ξ. Therefore, the supremum is achieved at $\xi = \Delta G(u+1-d)$; that is,

$$\mathrm{Var}(A_i) \leq \mathbb{E}\left[(\tilde{S}_{u+i} - \Delta G(u+1-d))^2 \mathbf{1}\left(\tilde{S}_{u+i} \geq \Delta G(u+1-d)\right)\right],$$

which is uniformly bounded because $\tilde{S}_{u+i}$'s are i.i.d. with a finite variance and $\Delta G(u+1-d)$ is bounded.

To analyze the second term in (A20), we denote the expression form of the conditional expectation term in B_i as

$$\psi(\hat{S}_{[i]}(u,i,i)) := \int_{\hat{S}_{[i]}(u,i,i)}^{\infty} \left(\tilde{s}_{u+i} - \hat{S}_{[i]}(u,i,i)\right) f_S(\tilde{s}_{u+i})\, \mathrm{d}\tilde{s}_{u+i},$$

where $f_S(\cdot)$ is the conditional density of $\tilde{S}_{u+i}$, given that its corresponding initial score, X_{j1} for some j, satisfies the condition $X_{[u]1} \geq X_{j1} \geq X_{[u+z]1}$. Then, we have

$$\psi'(\hat{S}_{[i]}(u,i,i)) = -\int_{\hat{S}_{[i]}(u,i,i)}^{\infty} f_S(\tilde{s}_{u+i})\,\mathrm{d}\tilde{s}_{u+i},$$

which is bounded by 1 in absolute value. By the mean value theorem,

$$\psi(\hat{S}_{[i]}(u,i,i)) = \psi(0) + \hat{S}_{[i]}(u,i,i)\int_0^1 \psi'(r\hat{S}_{[i]}(u,i,i))\,\mathrm{d}r,$$

which implies that $\mathrm{Var}(B_i) \leq \mathrm{Var}(\hat{S}_{[i]}(u,i,i))$. Because $d \asymp n$ and $\hat{S}_{[i]}(u,i,i)$ is the median of a $(2i-1)$-dimensional random vector composed of $(i-1)$ i.i.d. random variables alongside i constants, $\mathrm{Var}(\hat{S}_{[i]}(u,i,i))$ scales as $O(1/i)$ (Shao 2003, Theorem 5.10, p. 353). Wrapping up, we have $\sum_{i=1}^{z}\mathrm{Var}(B_i) = O(\log(z))$. Thus, $\mathrm{Var}(\mathcal{G}_z(u,\tilde{\mathbf{S}}(u,z))) = O(z\log(z))$.

Next, we show that $|\mathbb{E}[\mathcal{G}_z(u,\tilde{\mathbf{S}}(u,z))]| = \Omega(z)$. Because $\mathbb{E}[\mathcal{G}_z(u,\tilde{\mathbf{S}}(u,z))]$ is discrete concave and increasing in z, we have

$$\left|\mathbb{E}\left[\mathcal{G}_z(u,\tilde{\mathbf{S}}(u,z))\right]\right| \geq \frac{z}{n-u}\mathbb{E}\left[\mathcal{G}_{n-u}(u,\tilde{\mathbf{S}}(u,n-u))\right] = \frac{z}{n-u}\sum_{i=1}^{n-u}\mathbb{E}\left[(\tilde{S}_{u+i} - \hat{S}_{[i]}(u,i,i))^+\right].$$

By the second inequality in (A21), $\mathbb{E}[(\tilde{S}_{u+i} - \hat{S}_{[i]}(u,i,i))^+] \geq \mathbb{E}[(\tilde{S}_{u+i} - \Delta G(u+i-d))^+]$, which is uniformly bounded. This implies that $|\mathbb{E}[\mathcal{G}_z(u,\tilde{\mathbf{S}}(u,z))]| = \Omega(z)$.

Finally, the concentration bound (A19) follows from Chebyshev's inequality. □

Appendix C: Proofs for the Auxiliary Results

Proof of (9) ***in Section 5.*** By the conditional variance formula,

$$\begin{aligned}
&\mathrm{Var}(g(X_{i1},X_{i2}) \mid X_{i1}) \\
&= \mathbb{E}\left[g^2(X_{i1},X_{i2}) \mid X_{i1}\right] - f^2(X_{i1}) \\
&= \mathbb{E}\left[Y_i^2 - 2Y_i\epsilon_i' + \epsilon_i'^2 \mid X_{i1}\right] - f^2(X_{i1}) \\
&= \mathbb{E}\left[2Y_i(\epsilon_i - \epsilon_i') - \epsilon_i^2 + \epsilon_i'^2 \mid X_{i1}\right] \\
&= \mathbb{E}\left[2(f(X_{i1}) + \epsilon_i)\epsilon_i - 2(g(X_{i1},X_{i2}) + \epsilon_i')\epsilon_i' - \epsilon_i^2 + \epsilon_i'^2 \mid X_{i1}\right] \\
&= -2\mathbb{E}\left[g(X_{i1},X_{i2})\epsilon_i' \mid X_{i1}\right] + \mathbb{E}\left[\epsilon_i^2 \mid X_{i1}\right] - \mathbb{E}\left[\epsilon_i'^2 \mid X_{i1}\right] \\
&= -2\mathbb{E}\left[\mathbb{E}\left[g(X_{i1},X_{i2})\epsilon_i' \mid X_{i1},X_{i2}\right] \mid X_{i1}\right] + \mathbb{E}\left[\epsilon_i^2 \mid X_{i1}\right] - \mathbb{E}\left[\epsilon_i'^2 \mid X_{i1}\right] \\
&= -2\mathbb{E}\left[g(X_{i1},X_{i2})\mathbb{E}\left[\epsilon_i' \mid X_{i1},X_{i2}\right] \mid X_{i1}\right] + \mathbb{E}\left[\epsilon_i^2 \mid X_{i1}\right] - \mathbb{E}\left[\epsilon_i'^2 \mid X_{i1}\right] \\
&= \mathbb{E}\left[\epsilon_i^2 \mid X_{i1}\right] - \mathbb{E}\left[\epsilon_i'^2 \mid X_{i1}\right],
\end{aligned}$$

where the fifth equality follows from $\mathbb{E}\left[f(X_{i1})\epsilon_i \mid X_{i1}\right] = f(X_{i1})\mathbb{E}\left[\epsilon_i \mid X_{i1}\right] = 0$, the sixth follows from the law of iterated expectations, and the last follows from $\mathbb{E}\left[\epsilon_i' \mid X_{i1},X_{i2}\right] = 0$. □

Proof of Proposition A1. For Theorem 1 to hold, it suffices to examine whether the second-stage problem (6) still satisfies the marginal properties presented in Lemma 1. By Lemma 2, the second-stage problem becomes

$$\begin{aligned}
&\mathbb{E}\left[\max_{0\leq h\leq|\mathcal{Z}|}\left\{\sum_{j=1}^{h} S'_{[j]}(\mathbf{x}_1,\mathcal{Z}) - G(|\mathcal{U}|+h-d)\right\}\right] \\
&= \mathbb{E}\left[(S'_i(x_{i1}) - \hat{S}'_{[|\mathcal{Z}|]})^+\right] + \mathbb{E}\left[\max_{0\leq h\leq|\mathcal{Z}|-1}\left\{\sum_{j=1}^{h} S'_{[j]}(\mathbf{x}_1,\mathcal{Z}\setminus\{i\}) - G(|\mathcal{U}|+h-d)\right\}\right], \qquad \text{(A22)}
\end{aligned}$$

where $\hat{S}'_{[|\mathcal{Z}|]}$ is the $|\mathcal{Z}|$th order statistic of the random vector $((S'_j(x_{j1}))_{j\in\mathcal{Z}\setminus\{i\}}, \Delta G(|\mathcal{U}|+1-d), \Delta G(|\mathcal{U}|+2-d), \ldots, \Delta G(|\mathcal{U}|+|\mathcal{Z}|-d))$, which is independent of $S'_i(x_{i1})$.

Take any realization $\hat{s}'_{[|\mathcal{Z}|]}$ of $\hat{S}'_{[|\mathcal{Z}|]}$. Letting $S'_i(x_{i1}) = f(x_{i1}) + \sigma(x_{i1})\epsilon''_i$, we take the first-order derivative of (A22) with respect to x_{i1} as follows:

$$\frac{\partial}{\partial x_{i1}}\mathbb{E}\left[(S'_i(x_{i1}) - \hat{s}'_{[|\mathcal{Z}|]})^+\right] = f'(x_{i1})\mathbb{P}\left(S'_i(x_{i1}) \geq \hat{s}'_{[|\mathcal{Z}|]}\right) + \sigma'(x_{i1})\mathbb{E}\left[\epsilon''_i \mathbf{1}\left(S'_i(x_{i1}) \geq \hat{s}'_{[|\mathcal{Z}|]}\right)\right].$$

Then, the marginal properties in Lemma 1 are equivalent to

$$0 \leq f'(x_{i1})\mathbb{P}\left(S'_i(x_{i1}) \geq \hat{s}'_{[|\mathcal{Z}|]}\right) + \sigma'(x_{i1})\mathbb{E}\left[\epsilon''_i \mathbf{1}\left(S'_i(x_{i1}) \geq \hat{s}'_{[|\mathcal{Z}|]}\right)\right] \leq f'(x_{i1}).$$

Rearranging it yields

$$\begin{aligned}
&\frac{-f'(x_{i1})\mathbb{P}\left(S'_i(x_{i1}) \geq \hat{s}'_{[|\mathcal{Z}|]}\right)}{\mathbb{E}\left[\epsilon''_i \mathbf{1}\left(S'_i(x_{i1}) \geq \hat{s}'_{[|\mathcal{Z}|]}\right)\right]} \leq \sigma'(x_{i1}) \leq \frac{f'(x_{i1})\mathbb{P}\left(S'_i(x_{i1}) < \hat{s}'_{[|\mathcal{Z}|]}\right)}{\mathbb{E}\left[\epsilon''_i \mathbf{1}\left(S'_i(x_{i1}) \geq \hat{s}'_{[|\mathcal{Z}|]}\right)\right]} \\
\iff \quad &\frac{-f'(x_{i1})}{\mathbb{E}\left[\epsilon''_i \mid S'_i(x_{i1}) \geq \hat{s}'_{[|\mathcal{Z}|]}\right]} \leq \sigma'(x_{i1}) \leq \frac{f'(x_{i1})\mathbb{P}\left(S'_i(x_{i1}) < \hat{s}'_{[|\mathcal{Z}|]}\right)}{\mathbb{E}\left[\epsilon''_i\right] - \mathbb{E}\left[\epsilon''_i \mathbf{1}\left(S'_i(x_{i1}) < \hat{s}'_{[|\mathcal{Z}|]}\right)\right]} \\
\iff \quad &\frac{-f'(x_{i1})}{\mathbb{E}\left[\epsilon''_i \mid S'_i(x_{i1}) \geq \hat{s}'_{[|\mathcal{Z}|]}\right]} \leq \sigma'(x_{i1}) \leq \frac{-f'(x_{i1})}{\mathbb{E}\left[\epsilon''_i \mid S'_i(x_{i1}) < \hat{s}'_{[|\mathcal{Z}|]}\right]} \\
\iff \quad &\frac{1}{\mathbb{E}\left[\epsilon''_i \mid f(x_{i1}) + \sigma(x_{i1})\epsilon''_i < \hat{s}'_{[|\mathcal{Z}|]}\right]} \leq -\frac{\sigma'(x_{i1})}{f'(x_{i1})} \leq \frac{1}{\mathbb{E}\left[\epsilon''_i \mid f(x_{i1}) + \sigma(x_{i1})\epsilon''_i \geq \hat{s}'_{[|\mathcal{Z}|]}\right]}.
\end{aligned}$$

Because it suffices to require the above inequalities to hold for any realization $\hat{s}'_{[|\mathcal{Z}|]}$ of $\hat{S}'_{[|\mathcal{Z}|]}$, we finalize the proof by taking the supremum for the lower bound and the infimum for the upper bound. □