This page is intentionally blank. Proper e-companion title page, with INFORMS branding and exact metadata of the main paper, will be produced by the INFORMS office when the issue is being assembled.

EC.1. Further related work

Fairness, sequential decision making, & information acquisition. Fairness has been extensively studied in machine learning problems such as classification; see Barocas et al. (2017) for an overview. Fairness in offline allocation with limited resources has also been considered, starting with the seminal work of Baruah et al. (1993), Kumar and Kleinberg (2000), and Bertsimas et al. (2012). More recently, attention has turned to studying dynamic and operational considerations in fairness, such as sequential decisions and costly information acquisition; see surveys by Finocchiaro et al. (2021) and Nashed et al. (2023). To the best of our knowledge, our work is the first to study socially aware constraints in sequential search under costly inspection.

Several related strands are still worth noting. A number of papers study fairness in online selection problems for indivisible goods, such as secretary problems and prophet inequalities (Buchbinder et al., 2009; Correa et al., 2021; Arsenis and Kleinberg, 2022; Salem and Gupta, 2024). Others focus on dynamic allocation of divisible goods, designing policies to maximize egalitarian welfare (Lien et al., 2014; Sinclair et al., 2020; Manshadi et al., 2021). Fair division and market equilibrium problems (e.g., notions such as envy-freeness) have also been explored in dynamic settings (Walsh, 2011; Kash et al., 2014; Aleksandrov et al., 2015; Peysakhovich et al., 2023; Gao and Kroer, 2023). There are several other works that consider fairness in modern variants of online resource allocation, for example the work of Liao et al. (2022) analyzing fairness in allocating sequentially arriving items under non-stationarity, and Bateni et al. (2022) addressing fairness-efficiency trade-offs in online resource allocation with applications to online advertising. The interplay between information acquisition and fairness/discrimination has also been studied. The closest to us is the work of Cai et al. (2020), that revolves around achieving fairness through information acquisition. However, the setting differs from ours as it is concerned with targeted screening to improve information quality for a subset of individuals. Another conceptually related work—especially to our numerical study of long-term utilities and downstream outcomes–is Baek and Makhdoumi (2023) that highlights feedback loops exacerbating disparities in hiring, by showing that even a small initial difference in how well a firm can evaluate candidates from different groups can snowball into long-lasting hiring disparities due to a feedback loop.

Finally, another line of research that is conceptually related to us investigates fairness in bandit settings, introducing concepts such as fairness in exploration and hindsight (Cayci et al., 2020; Baek and Farias, 2021; Schumann et al., 2022; Li et al., 2025; Gupta and Kamble, 2021). Another example is the work of Komiyama and Noda (2024) that take a multi-armed bandit approach, showing how temporary affirmative-action subsidies when firms lack data about minorities can mitigate persistent statistical discrimination.

Fairness in operations and revenue management. There has been a growing literature studying various notions of fairness in different operational settings. For example, Cohen et al. (2022) explores the challenges and trade-offs involved in implementing price fairness constraints for different customer groups in the context of discriminatory pricing. Other examples study fairness-aware online price discrimination (Chen et al., 2021), fairness criteria in network revenue management with demand learning (Chen et al., 2021, 2022b), fairness in

assortment planning (Lu et al., 2023; Chen et al., 2022a), fair and dynamic rationing of scarce resources (Manshadi et al., 2021), group fairness in stochastic matching (Ma et al., 2022), group fairness in offline and online combinatorial optimization (Asadpour et al., 2023; Golrezaei et al., 2024; Tang and Yuan, 2023; Niazadeh et al., 2023), refugee resettlement (Freund et al., 2023), fair incentives in repeated engagements (Freund and Hssaine, 2025), individual fairness in revenue management (Arsenis and Kleinberg, 2022; Jaillet et al., 2024), and limitations of Rooney rules (Farajollahzadeh et al., 2025) in contexts such as employment or admission.

Extensions of sequential search: Pandora's box & beyond. Moving beyond fairness considerations, our work contributes to the rich literature on sequential search. Building on the seminal work of Weitzman (1979), numerous papers study sequential search in richer settings different from ours. Recent work in this direction includes studying settings with the option of selecting a box without inspection (Beyhaghi and Kleinberg, 2019; Alaei et al., 2021; Doval, 2018; Aouad et al., 2020), uncertainty in the availability of a box for inspection or selection (Brown and Uru, 2022), correlated or unknown reward distributions that should be learned or estimated via samples (Chawla et al., 2020; Gatmiry et al., 2022; Agarwal et al., 2024), and designing prior-independent search policies in the absence of prior information about candidate values (Brown and Uru, 2025). Modeling the stateful search as JMS enables us to capture several well-motivated variants of the Pandora's box problem while retaining the simple structure of the optimal policy, which is not necessarily the case in most of these other extensions. Another interesting extension of Pandora's box is when the ordering under which the boxes should be inspected is restricted by a given partial ordering, which is a model studied in Boodaghians et al. (2023). Despite this restriction, they show a polynomial-time computable index-based optimal policy for this extension. Beyond Pandora's box model, other models for sequential search and hiring have been studied that are conceptually related, e.g., Epstein and Ma (2024) consider a variant of the stochastic probing problem to model the ordering and selection in hiring pipelines.

Restless bandits and weakly-coupled MDPs. Also related to us, mostly in terms of the philosophy in designing algorithms, is the growing literature on Restless multi-armed bandits (RMAB) and weakly coupled MDPs. In the RMAB setting, even the arms that are not played may evolve according to potentially different transition probability kernels. In its most general form, finding an optimal policy is computationally hard. More specifically, Papadimitriou and Tsitsiklis (1999) prove that even under known transition kernels and infinite-horizon average reward, finding the optimal policy is PSPACE-hard. In response, several works aim to provide approximate optimality results. The technique of relaxing the ex-post constraints of the problem to hold in expectation over the horizon, and then Lagrangifying this relaxed constraint into the objective is predominantly employed to obtain approximate or heuristic policies in this literature (and also in the literature on weakly coupled MDPs, e.g., Hawkins (2003); Adelman and Mersereau (2008)). This technique, at a high-level, has similarities to how we address ex-ante constraints in both the Pandora's box setting and JMS. The most prevalent heuristic related to this technique is an index policy, called the *Whittle-index*, proposed by Whittle (1988), which is only well-defined under certain indexability conditions and may still be intractable to

compute in some cases (Niño-Mora, 2007). Nevertheless, Weber and Weiss (1990) show the asymptotic optimality of the Whittle-index policy in infinite-horizon average reward under certain assumptions. Other works include showing an unbounded gap for the Whittle-index under finite-horizon and discounted infinite-horizon settings—even if the other assumptions in Weber and Weiss (1990) hold—and instead providing alternative LP-based heuristics with sublinear regret bounds (Zhang and Frazier, 2021; Ghosh et al., 2023); online learning of the Whittle-index when transition kernels are unknown (Wang et al., 2023); and recent studies on incorporating fairness notions, such as a minimum fraction of times to pull each arm (Li et al., 2019; Wang et al., 2024)—for which they show sublinear regret—or bounding the maximum time since the last pull of any arm (Li and Varakantham, 2022). The literature on RMAB is massive and growing fast, and we refer the interested reader to this recent survey Niño-Mora (2023).

While at the surface level there seem to be some similarities between our framework and the RMAB/Weakly-coupled MDP setting, the two settings are both semantically and mathematically quite different. In fact, there are some fundamental distinctions between our Pandora's box and JMS models and the RMAB problem. For details, see the discussion in Section EC.1.1.

Primal-dual methods for learning in games and applications in fairness. The idea of using primal-dual method and online learning to solve games and linear programs goes to back to the seminal work of Plotkin et al. (1995) for solving fractional packing and covering LPs, with its roots in the classic work of Blackwell (1956) and Dantzig et al. (1956). In this framework, the problem, after Lagrangifying the constraints, can be reinterpreted as a min-max game played between the primal player (who proposes a feasible solution) and the dual player (who picks dual variables corresponding to the constraints). See Arora et al. (2012) for a survey. This technique has also manifested in various forms in the literature and has given rise to iterative primal-dual algorithms that rely on different first-order methods, e.g., Lyu et al. (2019); Jiang et al. (2019), or Blackwell approachability, e.g., Zhong et al. (2018) and Niazadeh et al. (2023). More recently, a similar approach has proven to be highly useful for near-optimally solving constrained online linear and convex programming problems (Agrawal et al., 2014; Agrawal and Devanur, 2014; Balseiro et al., 2023).

In terms of applications, such methods have been utilized in various applications in operations research and computer science, including bandit problems with knapsack constraints (Badanidiyuru et al., 2018), resource allocation problems (Devanur et al., 2011; Balseiro et al., 2023; Agrawal et al., 2014; Agrawal and Devanur, 2014), inventory pooling and capacity allocation problems with service level constraints (Lyu et al., 2019; Jiang et al., 2019), packing and covering problems (Plotkin et al., 1995), dynamic matching for refugee resettlement (Bansak et al., 2024), and classification problems or combinatorial optimization problems with subgroup fairness constraints (Kearns et al., 2018; Golrezaei et al., 2024). Lastly, the primal-dual method has also proved to be essential in designing online resource allocation and matching algorithms under adversarial arrival (Karp et al., 1990; Mehta et al., 2007; Buchbinder and Naor, 2009; Huang et al., 2019; Feng and Niazadeh, 2024; Ekbatani et al., 2025) and in different applications of these models (Golrezaei et al., 2014; Ma and Simchi-Levi,

2020; Gong et al., 2022; Feng et al., 2021; Delong et al., 2023; Udwani, 2024; Ekbatani et al., 2023; Feng et al., 2024). See Mehta et al. (2013) for a comprehensive survey on this topic.

We highlight that our primal-dual approach in Section 3, which leads to a near-optimal near-feasible solution, at a high-level, is built on this standard framework. However, there are important distinctions that make our algorithmic results novel. See the discussion in Section EC.1.2 for more details.

Exact and approximate algorithmic Carathéodory. The "Carathéodory problem" is a fundamental problem in geometry and polyhedral optimization, which dates back to the classic work of Carathéodory (1911). The basic proof of the Carathéodory theorem is constructive and based on an algorithm that has access to various types of oracles (separation oracle, optimization oracle, validity oracle, and membership oracle in each face of the polytope). See Grötschel et al. (1981, 2012); Vishnoi (2021) for details. More recently, this fundamental problem has been revisited through the lens of approximations, and several fast algorithms are designed by leveraging tools from online adversarial learning that provide approximate solutions (Mirrokni et al., 2017; Combettes and Pokutta, 2023). We remark that our algorithm (Algorithm 4) in Section EC.5 is essentially an algorithm for *exact* algorithmic Carathéodory problem that *only* uses (linear) optimization oracle. Compared to earlier methods mentioned, our method has the advantage of being simpler, using linear optimization oracle more efficiently, and not requiring any other oracle access to the polytope.

Computational aspects of Bayesian sequential decision making. To the best of our knowledge, our paper is the first to offer an FPTAS for JMS with on-average constraints. However, related computational settings have been explored in the literature. Segev and Singla (2021) present a framework for efficiently approximating fundamental stochastic combinatorial optimization problems, such as stochastic probing, improving approaches for various non-adaptive settings. Aouad et al. (2020) explore a generalization of the Pandora's box problem with sequential inspection, balancing information acquisition and cost efficiency, and offering near-optimal approximate schemes. Additionally, Anari et al. (2019) study Bayesian online allocations with laminar matroid constraints and provide an FPTAS when the matroid's depth is constant. Similar computational questions about the tractability of various Markov decision processes in Bayesian allocations have been studied in Papadimitriou and Tsitsiklis (1987); Papadimitriou et al. (2021). Conceptually, there are some similarities between our approach and the framework in Liu et al. (2018), which examines the delayed impact of fair machine learning, though the two settings are distinct and not directly comparable.

EC.1.1. Discussion on the Differences Between our Pandora/JMS Framework and Restless Bandits

Although our work employs index-based policies and uses Lagrangification technique to incorporate ex-ante constraints, our framework (in both the Pandora's box and JMS models) differs fundamentally from the Restless Multi-Armed Bandits (RMAB) framework. Below, we highlight the key distinctions:

- **Model primitives:** the two models are quite different in important ways, which drastically changes both the computational landscapes of these models, as well as the algorithm design principles:

(i) **Non-restless vs. restless arms:** In our Pandora's box and JMS settings, an arm (or Markov chain) *does not* evolve if we do not pull (inspect) it. The state changes occur only when we actively decide to inspect an arm and the global state is fully known at each step. By contrast, in RMAB problems, each arm's state continues to evolve (restlessly) even if it is not selected, which complicates the state space or leads to partial observability.

(ii) **Endogenous stochastic horizon vs. fixed/infinite exogenous horizon:** RMAB formulations consider an *exogenous* and *fixed* finite horizon T or an infinite horizon with discounting. Our JMS and Pandora's box settings, on the other hand, have a *stochastic* horizon by stopping once a Markov chain reaches a terminal state (or once the set of Markov chains at terminal states satisfies an ex-post matroid constraint, such as a capacity k). As a result, the horizon is *endogenous* to the algorithm.

(iii) **Different computational complexities:** General RMAB problems are known to be PSPACE-hard (Papadimitriou and Tsitsiklis, 1999, 1987); a well-known index-based policy in this setting based on *Whittle indices* is often just a (clever) heuristic, which can be optimal (or even near-optimal or approximately optimal) only in special cases or under strong assumptions and asymptotic scenarios. By contrast, under mild assumptions, our JMS and Pandora's box models admit a polynomial-time optimal index-based solution (a variant of the Gittins index-based policy) under matroid ex-post constraints on termination (see Section 3.2.1 and Section EC.8 for details; also see (Dumitriu et al., 2003)). Even after we add some ex-ante constraints on visit frequencies of states (which is an extra restriction on top of the ex-post termination constraint and the constraint to pull one arm at a time), we still maintain polynomial-time computability via dual-adjustments and specialized tie-breaking (Section 2), Carathéodory-type decompositions (Section 2.4.2), or FPTAS algorithms in case of convex constraints (Section 3).

We re-iterate that the key difference in computational tractability, together with quite different mathematical semantics of the two models as described above, underscore how the two frameworks, though superficially similar in their use of "index-like" ideas, are inherently different (in fact, JMS/Pandora settings are more like *Bayesian bandits* setting in terms of algorithmic landscapes than RMAB).

- **Nature of the Lagrangian approach:** while there are superficial similarities between the way that the Lagrangification techniques are used in our framework (for incorporating ex-ante constraints) and in the RMAB framework, in fact they are quite distinct and serve completely different purposes:

  (i) **Exact ex-post constraint vs. relaxed constraint**: In RMAB framework the ex-post constraint of "pull at most one arm each time" often gets relaxed into a single in-expectation constraint—e.g., "pull a total of T arms over T rounds in expectation." This relaxation allows one to apply Lagrangification, which *decouples* the problem across arms, leading to the well-known Whittle index approach. However, Whittle-index-based relaxation policies typically only guarantee feasibility in expectation (where all arms with a non-negative index are pulled at each time); additional rounding or scheduling heuristics (such as pulling highest index arm at each time, as in the classic Whittle index heuristic

policy) are required to restore ex-post feasibility, often yielding *approximate* guarantees or *asymptotic near-optimality* guarantees (Guha et al., 2010)) In contrast, we *always* pull one arm in each time (the arm with the highest Gittins index), and also never relax our ex-post constraint on termination (e.g., a capacity on how many terminal states can be accepted). Our capacity (or matroid) constraint is enforced on *every sample path*, and we only Lagrangify *extra ex-ante constraints* (like demographic parity or quota) that are meant to hold in expectation. As a result, our method yields a fully feasible optimal constrained policy for the original problem with the given ex-post capacity-like constraint and the extra ex-ante constraints, without requiring *any* further post-processing.

(ii) **Post-Lagrangification behavior (decoupling vs. adjusted instance):** In RMAB, once you Lagrangify the relaxed constraint, the problem "decouples" into single-armed subproblems—one for each arm. By solving each single-armed problem, one can define and calculated the Whittle index for each arm. By contrast, when we Lagrangify our ex-ante constraints, the resulting "adjusted" problem remains a single, integrated instance of JMS or Pandora's box with modified rewards or costs. It *does not* decouple into multiple subproblems, and we still need to solve that adjusted instance. As we show, the solution to this problem is a dual-adjusted index-based policy, along with a randomized tie-breaking to guarantee exact feasibility (or near-feasibility in case of convex constraints).

The above distinctions in the way that Lagrangification helps with designing algorithms capture essential differences between the two models, both in terms of the type of the constraints (i.e., relaxation of an ex-post constraint vs. an original ex-ante constraint) and also how they are handled via this technique. It also serves as another evidence why the RMAB's Whittle approach is not mathematically connected to the Pandora's box or JMS settings, with or without ex-ante constraints, and does not carry over there.

- **Multiple or more complex constraints:** RMAB literature usually focuses on a single ex-post constraint (e.g., at most one arm pulled per round) or a relaxed version of that constraint. In our setting, we can incorporate *multiple* ex-ante affine or convex constraints on the expected number of visits to any states (which can more complex than just a single affine constraint at the arm-level). We then solve the resulting constrained problem in polynomial time *exactly* (for affine constraints) or via a *near-optimal policy/FPTAS* (for convex constraints)—all while maintaining ex-post feasibility with respect to capacity or other structural constraints (like matroids). This level of generality and exact satisfaction of constraints distinguishes our framework even further from RMAB.

In conclusion, while there some similarities in the algorithmic philosophy used in our paper and this framework, the two models are mathematically and semantically different. As a result, to the best of our knowledge, there is no reduction or formal connection between our results and this literature.

EC.1.2. Discussion on Distinctions of G-RDIP (Algorithm 3) from the Standard Primal-Dual Method

Using primal-dual ideas from learning-in-games to solve convex-concave saddle-point problems—where a primal player best-responds iteratively and a dual player runs an adversarial online learning algorithm to find

the optimal dual that satisfies complementary slackness—is quite standard. In fact, as mentioned earlier, this approach dates back to seminal work on fractional packing and covering LPs (e.g., Plotkin et al. (1995); see also Arora et al. (2012) for a survey) and classic works of Blackwell (1956) and Dantzig et al. (1956). A similar algorithmic philosophy has also been extended to the online setting under i.i.d. stochastic arrivals (or variants such as random order or almost-i.i.d.), as in the work of Agrawal et al. (2014); Devanur et al. (2011); Agrawal and Devanur (2014). It is important to note that typically in this framework, given the optimal dual variables, the primal player's best response is simple, straightforward and computationally easy.

Although, at a high level, our G-RDIP approach (Algorithm 3) is built on this standard framework—and indeed, we employ Fenchel duality for reasons similar to those in the above papers—we believe that our work has the following distinguishing aspects, which highlight its novelty:

- **Computing the best response for the primal player:** Suppose that we only have ex-ante affine constraints. After Lagrangifying these constraints into the objective, the resulting best-response problem (i.e., maximizing the Lagrangian for a fixed set of dual variables) is equivalent to solving an unconstrained JMS problem with adjusted rewards. However, even this unconstrained JMS problem is nontrivial in the general setting, where some state rewards may be positive and others negative—situations that readily occur after dual adjustments. Previous work has analyzed this setting under the "No Free Lunch (NFL)" assumption on state rewards (see Section EC.8 and Definition EC.6), an assumption that can be violated after adjustments. To overcome this, in Section EC.8 we introduce a novel "collapsing reduction" that reduces an instance of JMS with arbitrary rewards to one that satisfies NFL in polynomial time. Consequently, we obtain a polynomial-time computable index-based algorithm for this more general version of the JMS problem, a result that we believe is of independent interest.
- **Two-Layer online learning and Fenchel duality for convex constraints:** Although we show how to compute a polynomial-time algorithm for the JMS problem with arbitrary rewards, this alone does not suffice when convex constraints are present. After Lagrangifying a convex constraint, the best-response problem becomes equivalent to solving an extension of the JMS problem where the objective is a concave function of the state visiting frequencies rather than a linear one. To our knowledge, no prior work addresses this specific problem, and it was not even known before our work that one could obtain a near-optimal solution for this problem. Inspired by the use of Fenchel duality in Agrawal and Devanur (2014) and related work such as Balseiro et al. (2023), we replace our convex constraints with their relaxations using Fenchel duals, introducing another set of dual variables. However, this leads to a technical challenge: the Lagrangian becomes linear in state frequencies but nonlinear (and indeed non-convex) in terms of the dual variables, so naively running online learning to minimize the Lagrangian would not work. We observe, however, that if we fix the Fenchel duals, the Lagrangian is linear in the remaining dual variables, and if we fix those, it is convex in the Fenchel duals. This observation suggests using "two layers of online learning" for the dual player and "best response" for the primal player to obtain a near-optimal solution. Accordingly, our final algorithm comprises an inner layer, where the Fenchel duals are learned, and an

outer layer, where the remaining dual variables are learned. For more details, please refer to Section 3.2.2 and Section 3.2, and see the analysis of G-RDIP in Section EC.9 (proof of Theorem 2).

EC.2. Missing Details of Section 2.2 and Section 2.3

EC.2.1. Missing Technical Details of Section 2.2

LEMMA EC.1 **(Checking for a Binding Constraint)**. *Suppose Problem* OPT-CONS *is feasible. If there exists an optimal solution π^* for Problem* OPT-UC *such that:*

$$\mathbf{E}\left[\sum_{i\in[n]}\theta_i^S\mathbb{A}_i^{\pi^*}+\sum_{i\in[n]}\theta_i^I\mathbb{I}_i^{\pi^*}\right]>b\,,$$

then there should exist an optimal solution $\hat{\pi}$ for Problem OPT-CONS *for which Constraint 2 is binding, that is,*

$$\mathbf{E}\left[\sum_{i\in[n]}\theta_i^S\mathbb{A}_i^{\hat{\pi}}+\sum_{i\in[n]}\theta_i^I\mathbb{I}_i^{\hat{\pi}}\right]=b\,,$$

Proof. We prove the claim by contradiction. Suppose that the claim does not hold. Then, for any optimal policy π' of Problem OPT-CONS—which exists due to the feasibility of Problem OPT-CONS—we have:

$$\mathbf{E}\left[\sum_{i\in[n]}\theta_i^S\mathbb{A}_i^{\pi'}+\sum_{i\in[n]}\theta_i^I\mathbb{I}_i^{\pi'}\right]<b\,.$$

Because π^* and π' have negative and positive slacks in Constraint 2, respectively, there exists a proper convex combination of these two policies for some $q\in(0,1)$, that is, a randomized policy $\tilde{\pi}$ that with probability q runs π' and with probability $1-q$ runs π^*, such that:

$$q\left(b-\mathbf{E}\left[\sum_{i\in[n]}\theta_i^S\mathbb{A}_i^{\pi'}+\sum_{i\in[n]}\theta_i^I\mathbb{I}_i^{\pi'}\right]\right)+(1-q)\left(b-\mathbf{E}\left[\sum_{i\in[n]}\theta_i^S\mathbb{A}_i^{\pi^*}+\sum_{i\in[n]}\theta_i^I\mathbb{I}_i^{\pi^*}\right]\right)=0\,.$$

Therefore, applying the linearity of the expectation, the resulting randomized policy $\tilde{\pi}$ satisfies Constraint 2 in its equality form. Moreover,

$$\textsc{Utility}(\tilde{\pi};\mathcal{I})=q\cdot\textsc{Utility}(\pi';\mathcal{I})+(1-q)\cdot\textsc{Utility}(\pi^*;\mathcal{I})\geq\textsc{Utility}(\pi';\mathcal{I})\,,$$

where $\mathcal{I}$ is the problem instance under consideration. The last inequality holds because $\textsc{Utility}(\pi^*;\mathcal{I})\geq\textsc{Utility}(\pi';\mathcal{I})$, since adding an ex-ante affine constraint to Problem OPT-UC can only lower the objective value. Thus, π' is also an optimal solution of Problem OPT-CONS for which Constraint 2 is binding, a contradiction. □

EC.2.2. Missing Technical Details of Section 2.3

PROPOSITION EC.1. *Algorithm 1 with $k=1$ implements the same ordering and stopping rule, up to tie-breaking, as in the optimal index-based policy in Weitzman (1979). Moreover, for $k>1$, this algorithm is exactly equivalent, again up to tie-breaking, to the optimal greedy (frugal) index-based policy in Kleinberg et al. (2016); Singla (2018).*

Proof. To see why, for a moment, suppose that there are no boxes with zero or negative cost or reward, there are no ties in the indices, and the model primitives are such that $\sigma_i \neq v_j$, for all $i, j \in [n]$. Then Algorithm 1 does the following: at each time if there is an opened box i^* in the set $\mathcal{C}$ (which only happens if v_{i^*} is larger than the indices of all the unopened boxes and the realized rewards of all the opened boxes so far), the algorithm stops and selects i^*; otherwise, it continues by opening an unopened box with the largest non-negative index. This algorithm is exactly equivalent to the optimal policy of Weitzman (1979) described earlier in Section 2.3.1.

Similarly, for $k > 1$, Algorithm 1 at each time greedily considers the unselected box i^* in $\mathcal{C}$ with the maximum option value o_i. If the box is already open, then it should be among the top k rewards in the set of opened boxes at the time of termination (as the algorithm terminates once the k^{th} largest reward among opened boxes is larger than all the remaining indices) and therefore will be selected. If the capacity k is reached after selection, then the algorithm terminates (as the k^{th} largest reward among opened boxes is now larger than all of the remaining indices); otherwise, it continues by selecting more opened boxes or opening an unopened box with the largest index. Again, this algorithm is exactly equivalent to the optimal policy of Kleinberg et al. (2016); Singla (2018), as described earlier in Section 2.3.1. □

LEMMA EC.2 **(Universality of the Refined Policy)**. *Any optimal policy for the Pandora's box problem (i.e., the special case of the unconstrained problem* OPT-UC *when $k = 1$) can be implemented by Algorithm 1 with a proper choice of tie-breaking rule τ.*

Proof. We start by recalling the definition of a "non-exposed" policy. A policy is said to be non-exposed if it is forced to eventually select any box i that the policy has opened and observed that $v_i > \sigma_i$. Note that any optimal policy π for the (the unconstrained version) of the Pandora's box problem must be non-exposed as shown in Kleinberg et al. (2016). Also, we remark that as stated in Kleinberg et al. (2016); Armstrong (2017), Singla (2018), the Pandora's box problem (with multiple selections) can be viewed as a static discrete choice problem (with multiple selections) in which this optimal policy *always* selects at most one (resp. k) of the boxes with the largest non-negative realized "random utility" defined as

$$\kappa_i \triangleq \min\{\sigma_i, v_i\} \tag{EC.1}$$

Therefore, for the special case of $k = 1$, any optimal policy should eventually select the box with the maximum non-negative κ_i (if any) in every sample path. See Kleinberg et al. (2016) for more details. Now compare any optimal policy π that satisfies the above properties with Algorithm 1. First of all, notice that any optimal policy has to inspect all of the negative-cost boxes. As a consequence, we can assume that policy π inspects them first without changing the rest of the search process. Now, under this convention, we fix a realization of all the random variables in our instance and run both policies π and Algorithm 1. Consider the first time step by which the policy π decides to choose a box i' that does not belong to the candidate set $\mathcal{C}$ (determined in line 5 of Algorithm 1), that is, it satisfies the following two conditions: (i) Box i' does not have the maximum option value, that is, $o_{i^*} > o_{i'}$ at that time, and also (ii) $i' \notin \{i \in [n] \setminus \mathcal{O} : \ c_i = 0\}$.

Let us consider two cases separately:

Case (a): Box i' has a negative or zero cost or it is already open. If box i' has negative cost, then, by our convention, it is open. If box i' has cost zero, then by condition (ii) – stated above – it is also open. Thus, choosing box i' implies that π selects i' with probability one. (Box i' can also be the outside option which is by default open and again choosing it means selecting it.)

Case (b): Box i' is not yet opened and has a positive cost. Again, we show that π will select i' with positive probability. Note that after i' is inspected by π, which occurs in the next step, the policy still would have enough capacity for selection (as it had before). Since $c_{i'} > 0$, there is a positive probability that the realization of $v_{i'}$ after inspecting box i' satisfies $v_{i'} > \sigma_{i'}$. Furthermore, by the non-exposedness property of π, this policy is forced to select i' in that case.

In summary, until now, we showed that there is a positive probability that π will be forced to eventually select box i'. In such sample paths (with nonzero measure), a few cases might arise:

Case (1): i^* has already been inspected when i' is chosen by π. First, this implies $v_{i^*} = o_{i^*} > o_{i'}$. In this case, if $v_{i^*} > \sigma_{i^*}$, then it implies that π is forced to also select i^* (by its non-exposed property). This results in a contradiction because π cannot select both i' and i^*. If $v_{i^*} \le \sigma_{i^*}$, then $\kappa_{i^*} = v_{i^*} > o_{i'} \ge \kappa_{i'}$, which is again a contradiction since it implies that the selected box i' by π cannot be the box with a maximum non-negative value of κ_i in certain sample paths.

Case (2): i^* is not inspected when i' is chosen by π. First, this implies $\sigma_{i^*} = o_{i^*} > o_{i'}$. Note also that $c_{i^*} \ge 0$, and hence there is a positive probability that $v_{i^*} \ge \sigma_{i^*}$. In that case, $\kappa_{i^*} = \sigma_{i^*} > o_{i'} \ge \kappa_{i'}$, which is again a contradiction since it implies that the selected box i' by π cannot be the box with a maximum non-negative value of κ_i in certain sample paths.

Putting everything together, all possible cases result in a contradiction, and hence we show that π cannot be an optimal policy, as desired. □

EC.2.3. Missing Proofs of Section 2.3

Proof of Proposition 1. We provide separate proofs for three parts of this lemma

(i) To prove this part, note that for any λ there exists always a deterministic policy that maximizes the Lagrangian $\mathcal{L}_{\text{CONS}}(\lambda, \pi)$. Therefore, w.l.o.g., we can restrict ourselves only to the set of all deterministic policies when computing $\mathcal{G}_{\text{CONS}}(\lambda)$. Now note that there are finitely many deterministic policies, simply because the number of boxes n and the number of possible value realizations are both finite (due to the discreteness assumption on the values). Fixing a deterministic policy π, the Lagrangian function is linear in λ. Therefore, the function $\mathcal{G}_{\text{CONS}}(\lambda)$ is the maximum over a finite number of linear functions, which implies that $\mathcal{G}_{\text{CONS}}(\lambda)$ is piecewise linear and convex.

(ii) Recall that $\mathcal{G}_{\text{CONS}}(\lambda)$ is always an upper-bound on the objective value of any feasible policy π in the primal problem, i.e., Problem OPT-CONS. Because this problem is assumed to be feasible (see the discussion

after Remark 1 and Remark 2 in Section 2.2), $\mathcal{G}_{\text{CONS}}$ is bounded from below. Given that $\mathcal{G}_{\text{CONS}}$ is piecewise linear and convex, it should always have a bounded minimizer λ^*.

(iii) For any given policy π, by taking partial derivative of $\mathcal{L}_{\text{CONS}}(\lambda;\pi)$ with respect to λ, we have:

$$\frac{\partial \mathcal{L}_{\text{CONS}}}{\partial \lambda} = b - \mathbf{E}\left[\sum_{i\in[n]} \theta_i^S \mathbb{A}_i^\pi + \sum_{i\in[n]} \theta_i^I \mathbb{I}_i^\pi\right] = \Delta_{\text{CONS}}^\pi \ .$$

Therefore, by a simple application of the envelope lemma, we have:

$$\Delta_{\text{CONS}}^{\pi^\lambda} \in \partial\left(\max_{\pi\in\Pi} \mathcal{L}_{\text{CONS}}(\cdot;\pi)\right)(\lambda) = \partial\left(\mathcal{L}_{\text{CONS}}(\cdot;\pi^\cdot)\right)(\lambda) = \partial\mathcal{G}_{\text{CONS}}(\lambda) \ ,$$

where π^λ is any policy maximizing the Lagrangian for a given λ, i.e., $\pi^\lambda \in \arg\max_{\pi\in\Pi} \mathcal{L}_{\text{CONS}}(\lambda;\pi)$, and $\partial f(x)$ is the set of all subgradients of f at point x.

□

Proof of Proposition 2. As mentioned in the proof sketch, this proof involves two steps.

Step 1: We show that there exists an optimal policy with a non-negative constraint slack. As we established in Proposition 1, the function $\mathcal{G}_{\text{CONS}}(\lambda)$ is the maximum of linear functions and, hence, is a convex piecewise-linear function. This piecewise-linear function has breakpoints at certain values, each corresponding to a λ at which the slope $\frac{\partial}{\partial\lambda}\mathcal{G}_{\text{CONS}}(\lambda)$ changes. Using part (iii) of Proposition 1, the slope of any differentiable line segment of $\mathcal{G}_{\text{CONS}}$ is equal to the slack $\Delta_{\text{CONS}}^{\pi^\lambda}$, where λ is an arbitrary point lying in that line segment and π^λ is an arbitrary optimal policy for the adjusted instance with respect to λ. Also, the space of possible policies (due to the finite and discrete support assumption on the rewards) is finite, and hence there are finitely many breaking points in this piecewise linear function.

Now, fixing any $\lambda^* \in \arg\min_{\lambda} \mathcal{G}_{\text{CONS}}(\lambda)$, there must exist a sufficiently small positive ε such that both λ^* and $\lambda^*+\varepsilon$ lie in the same line segment, or equivalently, the function $\mathcal{G}_{\text{CONS}}$ is a line in the interval $[\lambda, \lambda^*+\varepsilon]$. For this choice of ε, we show that any optimal policy $\pi^{\lambda^*+\varepsilon}$ for $\max_{\pi\in\Pi} \mathcal{L}_{\text{CONS}}(\lambda^*+\varepsilon,\pi)$ is also an optimal policy for $\max_{\pi\in\Pi} \mathcal{L}_{\text{CONS}}(\lambda^*,\pi)$. To see this, first note that $\pi^{\lambda^*+\varepsilon}$ is also an optimal policy for $\max_{\pi\in\Pi} \mathcal{L}_{\text{CONS}}(\lambda,\pi)$ for any choice of $\lambda \in (\lambda^*, \lambda^*+\varepsilon]$ (hence, $\mathcal{G}_{\text{CONS}}(\lambda) = \mathcal{L}_{\text{CONS}}(\lambda, \pi^{\lambda^*+\varepsilon})$ for any $\lambda \in (\lambda^*, \lambda^*+\varepsilon]$). This last statement holds because the optimal policy for the adjusted instance with any $\lambda \in (\lambda^*, \lambda^*+\varepsilon]$ has the same slack $\Delta_{\text{CONS}}^{\pi^\lambda} = \Delta_{\text{CONS}}^{\pi^{\lambda^*+\varepsilon}}$ regardless of the choice of λ. Therefore, if $\pi^{\lambda^*+\varepsilon}$ is not optimal for an adjusted instance with some $\lambda \in (\lambda^*, \lambda^*+\varepsilon)$, we should have:

$$\mathcal{L}_{\text{CONS}}(\lambda,\pi^\lambda) > \mathcal{L}_{\text{CONS}}(\lambda,\pi^{\lambda^*+\varepsilon}) \text{ and } \Delta_{\text{CONS}}^{\pi^\lambda} = \Delta_{\text{CONS}}^{\pi^{\lambda^*+\varepsilon}} \implies \textsc{Utility}(\pi^\lambda;\mathcal{I}) > \textsc{Utility}(\pi^{\lambda^*+\varepsilon};\mathcal{I}) \ .$$

However, this is in contradiction to the optimality of $\pi^{\lambda^*+\varepsilon}$ for the adjusted instance with $\lambda^*+\varepsilon$, simply because we can show:

$$\textsc{Utility}(\pi^\lambda;\mathcal{I}) > \textsc{Utility}(\pi^{\lambda^*+\varepsilon};\mathcal{I}) \text{ and } \Delta_{\text{CONS}}^{\pi^\lambda} = \Delta_{\text{CONS}}^{\pi^{\lambda^*+\varepsilon}} \implies \mathcal{L}_{\text{CONS}}(\lambda^*+\varepsilon,\pi^\lambda) > \mathcal{L}_{\text{CONS}}(\lambda^*+\varepsilon,\pi^{\lambda^*+\varepsilon}) \ .$$

Second, note that the function $\mathcal{G}_{\text{CONS}}$ is continuous and therefore we have:

$$\max_{\pi\in\Pi} \mathcal{L}_{\text{CONS}}(\lambda^*,\pi) = \mathcal{G}_{\text{CONS}}(\lambda^*) = \lim_{\delta\to 0^+} \mathcal{G}_{\text{CONS}}(\lambda^*+\delta)$$

Now, for $\delta\in(0,\varepsilon)$, we can replace $\mathcal{G}_{\text{CONS}}(\lambda^*+\delta)$ with $\mathcal{L}_{\text{CONS}}(\lambda^*+\delta,\pi^{\lambda^*+\varepsilon})$ as stated above. So we have:

$$\max_{\pi\in\Pi} \mathcal{L}_{\text{CONS}}(\lambda^*,\pi) = \lim_{\delta\to 0^+} \mathcal{L}_{\text{CONS}}(\lambda^*+\delta,\pi^{\lambda^*+\varepsilon}) = \mathcal{L}_{\text{CONS}}(\lambda^*,\pi^{\lambda^*+\varepsilon}) \ ,$$

where the last equality is retained due to the continuity of the Lagrangian function $\mathcal{L}_{\text{CONS}}(\lambda,\pi)$ with respect to λ for a fixed π (it is indeed a linear function). Therefore, $\pi^{\lambda^*+\varepsilon}$ is also an optimal policy for the adjusted instance with λ^*.

Furthermore, by convexity, all subgradients of $\mathcal{G}_{\text{CONS}}$ at $\lambda^*+\varepsilon$ must be nonnegative, implying that the constraint slack $\Delta_{\text{CONS}}^{\pi^{\lambda^*+\varepsilon}}$ of policy $\pi^{\lambda^*+\varepsilon}$ is nonnegative. Therefore, the policy $\pi^{\lambda^*+\varepsilon}$ is an adjusted optimal policy (with respect to λ^*) with nonnegative slack. Similarly, by repeating the same argument for $\lambda^*-\varepsilon$ for small enough ε, we conclude that there exists an adjusted optimal policy $\pi^{\lambda^*-\varepsilon}$ (with respect to λ^*) with nonpositive constraint slack, which completes the proof of this step.

Step 2: In this step, we show $\Delta_{\text{CONS}}^{\pi^+}\geq 0$ (resp. $0\geq\Delta_{\text{CONS}}^{\pi^-}$). To show this, we will look at the tie-breaking rule that arises from the perturbed adjusted problem with $\lambda=\lambda^*-\varepsilon$ for an *infinitesimal* $\varepsilon>0$ when we run Algorithm 2. More formally, we show that there exists a run of Algorithm 2 on the adjusted instance with $\lambda^*-\varepsilon$ that is exactly equivalent to running policy π^-, that is, a run of Algorithm 2 on the adjusted instance λ^* with the negative-extreme tie-breaking rule τ^- (which uses tie-breaking scores $\{s_i^-\}_{i\in\mathcal{C}}$ as in Definition 1).

First, suppose that ε is small enough so that the perturbation will not change any strict order among the possible realizations of the adjusted values $\widetilde{v}_i$ and the adjusted indices $\widetilde{\sigma}_i$ (and henceforth the adjusted option values $\widetilde{o}_i$), where the adjustment is with respect to λ^*. Moreover, we let ε be small enough so that the sign of no adjusted cost $\widetilde{c}_i$ changes. Therefore, if an adjusted cost $\widetilde{c}_i$ with respect to λ^* is strictly positive (resp. strictly negative), it remains strictly positive (resp. strictly negative) after adjustment with respect to $\lambda^*-\varepsilon$, regardless of the sign of θ_i^I. It is also important to note that if the adjusted cost $\widetilde{c}_i$ with respect to λ^* is exactly equal to zero, if we have $\theta_i^I<0$, then the adjusted cost with respect to $\lambda^*-\varepsilon$ turns out to be strictly positive (but infinitely close to zero), and if $\theta_i^I>0$, then the adjusted cost with respect to $\lambda^*-\varepsilon$ turns out to be strictly negative. Furthermore, if $\theta_i^I=0$ and the adjusted cost $\widetilde{c}_i$ with respect to λ^* is zero, it remains zero in the adjusted instance with respect to $\lambda^*-\varepsilon$. In the rest of the proof, we use the notation $\widetilde{v_i}^-$, $\widetilde{\sigma_i}^-$, $\widetilde{o_i}^-$, and $\widetilde{c_i}^-$ to denote adjusted values, adjusted reservation values (or indices), adjusted option values, and adjusted costs corresponding to the adjustment $\lambda^*-\varepsilon$.

Next, consider an optimal policy $\pi^{\lambda^*-\varepsilon}$ for the adjusted instance with $\lambda^*-\varepsilon$ for infinitesimal ε. Importantly, for any choice of tie-breaking rule for this policy, we have $\Delta_{\text{CONS}}^{\pi^{\lambda^*-\varepsilon}}\leq 0$. Now to compare this policy with π^-, we run both of these policies by using Algorithm 2 with the same instance with different adjustments as input, that is, we run π^- on the adjusted instance with λ^* and run $\pi^{\lambda^*-\varepsilon}$ on the adjusted instance with $\lambda^*-\varepsilon$. We also

couple the sample path realizations of the two runs. Now suppose inductively that the two policies have made exactly the same decisions up to some iteration of Algorithm 2. We then show that they can continue making the same decision while maintaining valid runs of both policies, which completes the proof. More precisely, suppose that policy π^- picks $i \in [n]$ from the set of candidates $\mathcal{C}$ to inspect (if i is not yet open) or add to the selection set (if i is already open) in the current iteration. We show that picking i in this iteration would be a valid choice for policy $\pi^{\lambda^*-\varepsilon}$.

To show the above claim, we consider the following cases:

- If box i is not yet open and $\widetilde{c}_i < 0$ (which implies $\widetilde{c}_i{}^- < 0$), then $\widetilde{\sigma}_i{}^- = \widetilde{o}_i{}^- = +\infty$ and hence this box will be among the boxes in $[n] \setminus \mathcal{S}$ (i.e., set of unselected boxes) with the maximum option value in the current iteration of $\pi^{\lambda^*-\varepsilon}$. Accordingly, the box i can be chosen by $\pi^{\lambda^*-\varepsilon}$ in this iteration, as desired.
- If box i is not yet open, $\widetilde{c}_i = 0$ and $\theta_i^I \geq 0$, then we either have $\widetilde{c}_i{}^- = 0$ (when $\theta_i^I = 0$) or $\widetilde{c}_i{}^- < 0$ (when $\theta_i^I > 0$). In the former case, box i will be among the candidate boxes in the current iteration of $\pi^{\lambda^*-\varepsilon}$, as it is an unopened zero-cost box. In the latter case, $\widetilde{\sigma}_i{}^- = \widetilde{o}_i{}^- = +\infty$ and this box will be among the boxes in $[n] \setminus \mathcal{S}$ with the maximum option value and, therefore, among the candidate boxes in the current iteration of $\pi^{\lambda^*-\varepsilon}$. When the two cases are combined, we conclude that box i can be chosen by policy $\pi^{\lambda^*-\varepsilon}$ in this iteration, as desired.
- If box i is not yet open and either $\widetilde{c}_i > 0$, or $\widetilde{c}_i = 0$ and $\theta_i^I < 0$, we first show that it should be among the boxes in $([n] \setminus \mathcal{S}) \cup \{0\}$ with the maximum option value in the run of π^-, and hence $\widetilde{\sigma}_i = \widetilde{o}_{\max}$, where

$$\widetilde{o}_{\max} = \max_{i' \in ([n] \setminus \mathcal{S}) \cup \{0\}} \widetilde{o}_{i'}$$

To prove this statement, note that if $\widetilde{c}_i > 0$ this statement is clearly true, as we know that i is among the candidate boxes of π^- in the current iteration. If $\widetilde{c}_i = 0$ and $\theta_i^I < 0$, note that $\widetilde{\sigma}_i$ is set to the upper-support $\bar{v}$ of the distribution of $\widetilde{v}_i$. If $\bar{v} = \widetilde{\sigma}_i < \widetilde{o}_{\max}$ then $\mathbf{Pr}[\widetilde{v}_i \geq \widetilde{o}_{\max}] = 0$. This is a contradiction, as according to Definition 1 the tie-breaking score s_i^- of box i should be set to $s_i^- = -\infty$ by π^-, so π^- is not allowed to choose i among the candidate boxes in $\mathcal{C}$ (note that there is always at least one box with a bounded tie-breaking score in $\mathcal{C}$).

Next, we show that this box i should also be among the boxes in $[n] \setminus \mathcal{S} \cup \{0\}$ with the maximum option value in $\pi^{\lambda^*-\varepsilon}$. This statement implies that box i is in the set of candidates in the current iteration of $\pi^{\lambda^*-\varepsilon}$ and therefore can be chosen by this policy in this iteration, as desired. First, observe that $\widetilde{c}_i{}^- > 0$, as either $\widetilde{c}_i > 0$, or $\widetilde{c}_i = 0$ and $\theta_i^I < 0$. Second, observe that $\widetilde{o}_i{}^- = \widetilde{\sigma}_i{}^-$, and we have:

$$\mathbf{E}\Big[\big(\widetilde{v}_i{}^- - \widetilde{\sigma}_i{}^-\big)^+\Big] = \mathbf{E}\Big[\big(\widetilde{v}_i + \varepsilon\theta_i^S - \widetilde{\sigma}_i{}^-\big)^+\Big] = \widetilde{c}_i{}^- = \widetilde{c}_i - \theta_i^I\varepsilon$$

Under our conditions in this case, if $\theta_i^I \geq 0$ (and hence $\widetilde{c}_i > 0$), we have $\widetilde{\sigma}_i{}^- = \widetilde{\sigma}_i + \varepsilon\theta_i^S + \delta$ for some infinitesimal $\delta \geq 0$. This simply holds because $\varepsilon > 0$ is infinitesimal. Furthermore, given that $\widetilde{c}_i > 0$, we have $\mathbf{Pr}[\widetilde{v}_i > \widetilde{\sigma}_i] > 0$, and therefore:

$$\mathbf{E}\Big[\big(\widetilde{v}_i{}^- - \widetilde{\sigma}_i{}^-\big)^+\Big] = \mathbf{E}\Big[\big(\widetilde{v}_i - \widetilde{\sigma}_i\big)^+\Big] - \delta \cdot \mathbf{Pr}[\widetilde{v}_i > \widetilde{\sigma}_i] = \widetilde{c}_i - \theta_i^I\varepsilon \;\Rightarrow\; \delta = \varepsilon\frac{\theta_i^I}{\mathbf{Pr}[\widetilde{v}_i > \widetilde{\sigma}_i]} = \varepsilon\frac{\theta_i^I}{\mathbf{Pr}[\widetilde{v}_i > \widetilde{o}_{\max}]}$$

Similarly, if $\theta_i^I < 0$ (and hence $\widetilde{c}_i \geq 0$), we have $\widetilde{\sigma}_i^{\,-} = \widetilde{\sigma}_i + \varepsilon\theta_i^S - \delta$ for some infinitesimal $\delta \geq 0$. Again, this holds because $\varepsilon > 0$ is infinitesimal. Moreover, $\mathbf{Pr}[\widetilde{v}_i \geq \widetilde{\sigma}_i] > 0$. This is true because either we have $\widetilde{c}_i > 0$, or $\widetilde{c}_i = 0$ and $\widetilde{\sigma}_i$ is set to the upper-support $\bar{v}$ of the distribution of $\widetilde{v}_i$ and hence $\mathbf{Pr}[\widetilde{v}_i \geq \widetilde{\sigma}_i] = \mathbf{Pr}[\widetilde{v}_i = \bar{v}] > 0$. Therefore, we have:

$$\mathbf{E}\Big[\big(\widetilde{v}_i^{\,-} - \widetilde{\sigma}_i^{\,-}\big)^+\Big] = \mathbf{E}\Big[\big(\widetilde{v}_i - \widetilde{\sigma}_i\big)^+\Big] + \delta \cdot \mathbf{Pr}[\widetilde{v}_i \geq \widetilde{\sigma}_i] = \widetilde{c}_i - \theta_i^I\varepsilon \;\Rightarrow\; \delta = -\varepsilon\frac{\theta_i^I}{\mathbf{Pr}[\widetilde{v}_i \geq \widetilde{\sigma}_i]} = \varepsilon\frac{\theta_i^I}{\mathbf{Pr}[\widetilde{v}_i \geq \widetilde{o}_{\max}]}$$

Putting the pieces together, the following holds for any unopened box with $\widetilde{c}_i > 0$, or $\widetilde{c}_i = 0$ and $\theta_i^I < 0$:

$$\widetilde{\sigma}_i^{\,-} = \widetilde{\sigma}_i + \varepsilon \cdot s_i^- \;, \tag{EC.2}$$

where s_i^- is the tie-breaking score of positive-extreme rule τ^- as in Definition 1. Now consider another box $i' \in [n] \setminus \mathcal{S} \cup \{0\},\ i' \neq i$. If $\widetilde{o}_i > \widetilde{o_{i'}}$, then $\widetilde{o}_i^{\,-} > \widetilde{o_{i'}}^{\,-}$ as ε is infinitesimal. Now suppose that $\widetilde{o}_i = \widetilde{o_{i'}}$ (and hence i' is also among the boxes with the maximum option value in $[n] \setminus \mathcal{S} \cup \{0\}$). First, note that if $\widetilde{c_{i'}} < 0$, or $\widetilde{c}_i = 0$ and $\theta_{i'}^I \geq 0$, then the tie-breaking score $s_{i'}^-$ of i' is set to $s_{i'}^- = +\infty$, so i' should be favored over i by π^- as $s_i^- < +\infty = s_{i'}^-$, a contradiction to the fact that π^- has picked i among the candidate boxes in $\mathcal{C}$. Therefore, $\widetilde{c_{i'}} > 0$, or $\widetilde{c_{i'}} = 0$ and $\theta_{i'}^I < 0$. Now consider two cases:

—If box i' is not open yet, we have:

$$\widetilde{o}_i^{\,-} = \widetilde{\sigma}_i^{\,-} = \widetilde{\sigma}_i + \varepsilon \cdot s_i^- \overset{(a)}{\geq} \widetilde{\sigma}_i + \varepsilon \cdot s_{i'}^- \overset{(b)}{=} \widetilde{\sigma_{i'}} + \varepsilon \cdot s_{i'}^- \overset{(c)}{=} \widetilde{\sigma_{i'}}^{\,-} = \widetilde{o_{i'}}^{\,-} \;,$$

where the inequality (a) holds as i has the maximum tie-breaking score s_i^- in the set of candidates $\mathcal{C}$, equality (b) holds as $\widetilde{\sigma}_i = \widetilde{o}_i = \widetilde{o_{i'}} = \widetilde{\sigma_{i'}}$ as i' is not open, and equality (c) holds due to Equation (EC.2) applied to box i' (as we proved earlier, this equation holds if $\widetilde{c_{i'}} > 0$, or $\widetilde{c_{i'}} = 0$ and $\theta_{i'}^I < 0$).

—If box i' is already open, or $i' = 0$ (outside option) we have:

$$\widetilde{o}_i^{\,-} = \widetilde{\sigma}_i^{\,-} = \widetilde{\sigma}_i + \varepsilon \cdot s_i^- \overset{(a)}{\geq} \widetilde{\sigma}_i + \varepsilon \cdot s_{i'}^- \overset{(b)}{=} \widetilde{v_{i'}} + \varepsilon \cdot s_{i'}^- \overset{(c)}{=} \widetilde{v_{i'}}^{\,-} = \widetilde{o_{i'}}^{\,-} \;,$$

where the inequality (a) holds as i has the maximum tie-breaking score s_i^- in the set of candidates $\mathcal{C}$, equality (b) holds as $\widetilde{\sigma}_i = \widetilde{o}_i = \widetilde{o_{i'}} = \widetilde{v_{i'}}$ as i' is open, and equality (c) holds because for an opened box i', $s_{i'}^- = \theta_i^S$ and $\widetilde{v_{i'}}^{\,-} = \widetilde{v_{i'}} + \varepsilon \cdot \theta_i^S$ (as a convention, set $\theta_0^S = 0$ for the outside option).

- If box i is already open or $i = 0$ (outside option), then it should be among the boxes in $[n] \setminus \mathcal{S} \cup \{0\}$ with the maximum option value in the run of π^-. We now show that this box will also be among the boxes in $[n] \setminus \mathcal{S} \cup \{0\}$ with the maximum option value in the run of $\pi^{\lambda^*-\varepsilon}$. Consider another box $i' \in [n] \setminus \mathcal{S} \cup \{0\},\ i' \neq i$. Similar to the previous case, if $\widetilde{o}_i > \widetilde{o_{i'}}$, then $\widetilde{o}_i^{\,-} > \widetilde{o_{i'}}^{\,-}$ as ε is infinitesimal. Now suppose that $\widetilde{o}_i = \widetilde{o_{i'}}$ (and hence i' is also among the boxes with the maximum option value in $[n] \setminus \mathcal{S} \cup \{0\}$). Similar to the previous case, the fact that i is picked over i' implies that $\widetilde{c_{i'}} > 0$, or $\widetilde{c_{i'}} = 0$ and $\theta_{i'}^I < 0$. Now we have two cases (as a convention, set $\theta_0^S = 0$ for the outside option):

— If box i' is not open yet, we have:

$$\widetilde{o_i}^- = \widetilde{v_i}^- \overset{(a)}{=} \widetilde{v_i} + \varepsilon \cdot s_i^- \overset{(b)}{\geq} \widetilde{v_i} + \varepsilon \cdot s_{i'}^- \overset{(c)}{=} \widetilde{\sigma_{i'}} + \varepsilon \cdot s_{i'}^- \overset{(d)}{=} \widetilde{\sigma_{i'}}^- = \widetilde{o_{i'}}^- \ ,$$

where the equality (a) holds as for the opened box i we have $\widetilde{v_i}^- = \widetilde{v_i} + \varepsilon \cdot \theta_i^S$ and $s_i^- = \theta_i^S$, inequality (b) holds as i has the maximum tie-breaking score s_i^- in the set of candidates $\mathcal{C}$, equality (c) holds as $\widetilde{v_i} = \widetilde{o_i} = \widetilde{o_{i'}} = \widetilde{\sigma_{i'}}$ as i' is not open, and equality (d) holds due to Equation (EC.2) applied to box i' (as we proved earlier, this equation holds if $\widetilde{c_{i'}} > 0$, or $\widetilde{c_{i'}} = 0$ and $\theta_{i'}^I < 0$).

— If box i' is already open, or $i' = 0$ (outside option) we have:

$$\widetilde{o_i}^- = \widetilde{v_i}^- \overset{(a)}{=} \widetilde{v_i} + \varepsilon \cdot s_i^- \overset{(b)}{\geq} \widetilde{v_i} + \varepsilon \cdot s_{i'}^- \overset{(c)}{=} \widetilde{v_{i'}} + \varepsilon \cdot s_{i'}^- \overset{(d)}{=} \widetilde{v_{i'}}^- = \widetilde{o_{i'}}^- \ ,$$

where the equality (a) holds as for the opened box i we have $\widetilde{v_i}^- = \widetilde{v_i} + \varepsilon \cdot \theta_i^S$ and $s_i^- = \theta_i^S$, inequality (b) holds as i has the maximum tie-breaking score s_i^- in the set of candidates $\mathcal{C}$, equality (c) holds as $\widetilde{v_i} = \widetilde{o_i} = \widetilde{o_{i'}} = \widetilde{v_{i'}}$ as i' is open, and equality (d) holds because for an opened box i', $s_{i'}^- = \theta_i^S$ and $\widetilde{v_{i'}}^- = \widetilde{v_{i'}} + \varepsilon \cdot \theta_i^S$.

Putting the above cases together, we have $\widetilde{o_i}^- \geq \widetilde{o_{i'}}^-$ for any box $i' \in [n] \setminus \mathcal{S} \cup \{0\},\ i' \neq i$, as desired.

The proof of the counterpart statement of $\Delta_{\text{CONS}}^{\pi^+} \geq 0$ follows a similar line of argument (and with exactly the same case analysis), which we omit for brevity. □

Proof of Remark 3. Let λ^* be a minimizer of $\mathcal{G}_{\text{CONS}}$ and π^* be any optimal policy for the adjusted instance with λ^* (where the adjustment is based on (8)). As stated in the proof of Proposition 2, any optimal policy $\pi^{\lambda^*+\varepsilon}$ for an adjusted instance with adjustment $\lambda^* + \varepsilon$ for infinitesimal $\varepsilon > 0$ is an optimal policy for the adjusted instance with adjustment λ^*. Note that $\mathcal{G}_{\text{CONS}}$ is differentiable at $\lambda^* + \varepsilon$ for infinitesimal ε. Then, applying the envelope theorem on $\mathcal{G}_{\text{CONS}}$ similar to part (iii) of Proposition 1, we conclude that the constraint slack of $\pi^{\lambda^*+\varepsilon}$ is equal to the slope of the convex piecewise linear function $\mathcal{G}_{\text{FS}}$ at point $\lambda^* + \varepsilon$. At the same time, as we showed in Proposition 2, this quantity is equal to $\Delta_{\text{CONS}}^{\pi^+}$, that is, the constraint slack of the policy π^+ with positive extreme positive tie-breakining rule τ^+ defined in Definition 1. Now, note that due to the convexity of $\mathcal{G}_{\text{CONS}}$, this slope is not lower than any subderivative / subgradient of $\mathcal{G}_{\text{CONS}}$ at λ^*. At the same time, the constraint slack $\Delta_{\text{CONS}}$of the optimal policy π^* for the adjusted instance with λ^* is equal to one of the subderivatives of $\mathcal{G}_{\text{CONS}}$ at λ^*. Therefore:

$$\Delta_{\text{CONS}}^{\pi^+} \geq \Delta_{\text{CONS}}^{\pi^*} \ ,$$

as desired. The counterpart argument $\Delta_{\text{CONS}}^{\pi^-} \leq \Delta_{\text{CONS}}^{\pi^*}$ can be proved in a similar fashion □

Proof of Theorem 1. First, note that the resulting randomized policy from Algorithm 2, which we denote by $\hat{\pi}$, constitutes an optimal solution of $\mathcal{G}_{\text{CONS}}(\lambda^*)$, as it randomizes over two such optimal solutions π^+ and π^-. The only part left to prove is that our randomized policy obtains an ex-ante constraint slack $\Delta_{\text{CONS}}^{\hat{\pi}}$ of exactly equal to zero. By construction,

$$\Delta_{\text{CONS}}^{\hat{\pi}} = b - \mathbf{E}\left[\sum_{i\in[n]} \theta_i^S \mathbb{A}_i^{\hat{\pi}} + \sum_{i\in[n]} \theta_i^I \mathbb{I}_i^{\hat{\pi}}\right] = \frac{\Delta_{\text{CONS}}^{\pi^+}}{\Delta_{\text{CONS}}^{\pi^+} - \Delta_{\text{CONS}}^{\pi^-}} \cdot \Delta_{\text{CONS}}^{\pi^-} - \frac{\Delta_{\text{CONS}}^{\pi^-}}{\Delta_{\text{CONS}}^{\pi^+} - \Delta_{\text{CONS}}^{\pi^-}} \cdot \Delta_{\text{CONS}}^{\pi^+} = 0 \ ,$$

as desired, hence finishing the proof. □

EC.3. More Intuitions and Managerial Insights from Section 2

In this section, we explore the implications of our results from Section 2, providing several managerial interpretations. We also present illustrative examples that demonstrate the necessity of our specific dual adjustments and randomized tie-breaking rules to achieve the optimal policy. These examples highlight that, although alternative policies or tie-breaking methods may be optimal in certain cases, they generally lead to suboptimal or in-feasible solutions.

EC.3.1. Implications for Demographic Group Fairness in Selection

Focusing on the special case of PARITY in selection, we have the following managerial observations:

(i) **More advantage to the under-represented group:** Ignoring tie-breaking, the adjustment in constructing the instance $\{(\widetilde{\mathcal{V}}_i, \widetilde{F}_i, c_i) \mid i \in [n]\}$ based on Equation (8) is both intuitive and economically interpretable. To illustrate, consider the optimal solution of the unconstrained problem. If the ex-ante number of selections from both groups is equal, the policy also satisfies parity in selection. Otherwise, suppose that group $\mathcal{Y}$ has a lower ex-ante number of selections, making it the *under-represented* group. In this case, $\lambda^* > 0$.[24] This implies that our adjustment (i) uniformly increases the rewards of those in $\mathcal{Y}$ by λ^*, (ii) uniformly decreases the rewards of those in $\mathcal{X}$ by λ^*, and (iii) does not adjust any costs.

(ii) **Preserving within-Group order and interleaving between groups:** In the adjusted instance, the indices are uniformly shifted: $\widetilde{\sigma}_i = \sigma_i + \lambda^*$ for all $i \in \mathcal{Y}$ and $\widetilde{\sigma}_i = \sigma_i - \lambda^*$ for all $i \in \mathcal{X}$. Ignoring tie-breaking, a key structural property of the optimal policy is that the within-group order of candidates is preserved after this adjustment. The only change in the search process pertains to the interleaved inspection order between the two groups.[25] Interleaving the relative inspection orderings of the two groups is a delicate aspect of our optimal policy. For example, consider a naive policy that achieves parity in selection by randomizing with probability $1/2$ between two search processes, each exclusively searching within one group and utilizing all available capacity. This policy does not interleave the inspection orderings of the groups and consequently suffers from an optimality gap, as illustrated in the example below.

EXAMPLE EC.1. Consider instance $\mathcal{I}$ for selecting one out of four candidates. Candidates 1 and 2 belong to $\mathcal{X}$, and 3 and 4 belong to $\mathcal{Y}$. All inspection costs are normalized to be 1, and,

$$v_1 = \begin{cases} 10 & \text{w.p. } 1/2 \\ 4 & \text{w.p. } 1/2 \end{cases}, \quad v_2 = \begin{cases} 9 & \text{w.p. } 1/2 \\ 3 & \text{w.p. } 1/2 \end{cases}, \quad v_3 = \begin{cases} 8 & \text{w.p. } 1/2 \\ 2 & \text{w.p. } 1/2 \end{cases}, \quad v_4 = \begin{cases} 7 & \text{w.p. } 1/2 \\ 1 & \text{w.p. } 1/2 \end{cases}.$$

Note that in such an example, $\sigma_1 = 8$, $\sigma_2 = 7$, $\sigma_3 = 6$ and $\sigma_4 = 5$. The optimal unfair policy π_{UNFAIR} inspects the candidates in the order $1 \to 2 \to 3 \to 4$, and $\text{UTILITY}(\pi_{\text{UNFAIR}}; \mathcal{I}) \approx 7.06$. However it is unfair: it selects from group $\mathcal{Y}$ with probability ≈ 0.1875. The naive fair policy π_{NAIVE} (defined above) flips a coin to decide

[24] For any $\lambda < 0$, we have $\mathcal{G}_{\text{CONS}}(\lambda) > \mathcal{G}_{\text{CONS}}(0)$ in this special case, thus λ cannot be a minimizer.

[25] As seen from our general adjustments in Equation (7) and the tie-breaking rules in Definition 1, the optimal policy for (OPT-CONS) employs a non-trivial adaptive ordering over the boxes in the general case, due to the non-linear relationships between costs and indices defined in Equation (3). For instance, in the case of PARITY in inspection, the optimal policy does not necessarily preserve the within-group order, unlike the optimal policy for PARITY in selection.

which group to consider and then inspects $\mathcal{X}$ in the order $1 \to 2$ and $\mathcal{Y}$ in the order $3 \to 4$. This policy selects exactly with probability 0.5 from each group, but only generates $\textsc{Utility}\left(\pi_{\text{NAIVE}}; \mathcal{I}\right) \approx 5.75$. Finally, our optimal fair policy π_{FAIR}, based on Algorithm 2, inspects the candidates in the order of $1 \to 3 \to 2 \to 4$ with probability 0.5, and in the order of $3 \to 1 \to 4 \to 2$ otherwise. It not only selects from each group with probability exactly 0.5, but also generates $\textsc{Utility}\left(\pi_{\text{FAIR}}; \mathcal{I}\right) \approx 6.5625$.

EC.3.2. The Necessity of Going Beyond Group-level Tie-breaking Rules

As discussed earlier in Section 2, for the special case of PARITY in selection, restricting to simple and intuitive group-level tie-breaking rules was sufficient to obtain two rules with opposite slack signs. This condition is both necessary and sufficient for constructing a randomized tie-breaking rule that exactly satisfies the ex-ante constraint. However, this simplifying property does not hold for all constraints, including PARITY in inspection. In this section, we provide a simple counterexample to illustrate this limitation. This example underscores the necessity of moving beyond group-level tie-breaking rules by specifying precise within-group order, to be able to satisfy the ex-ante constraint.

EXAMPLE EC.2. Consider instance $\mathcal{I}$ for selecting one out of four candidates. Candidates 1 and 2 belong to $\mathcal{X}$, and 3 and 4 belong to $\mathcal{Y}$. All inspection costs are normalized to be 1, and,

$$v_1 = \begin{cases} 7.5 & \text{w.p. } 2/3 \\ 4 & \text{w.p. } 1/3 \end{cases}, \quad v_2 = \begin{cases} 9 & \text{w.p. } 1/3 \\ 4 & \text{w.p. } 2/3, \end{cases}, \quad v_3 = \begin{cases} 10 & \text{w.p. } 1/4 \\ 4 & \text{w.p. } 3/4, \end{cases}, \quad v_4 = \begin{cases} 7 & \text{w.p. } 1/2 \\ 4 & \text{w.p. } 1/2, \end{cases}$$

Consider the problem of finding the optimal constrained policy subject to PARITY in inspection for this instance. It is easy to verity that $\lambda^* = 0$, indicating that there exists an optimal policy for the constrained problem that is also optimal for the unconstrained problem. By simple calculations, we find that $\sigma_1 = \sigma_2 = \sigma_3 = 6$ and $\sigma_4 = 5$. Moreover, no value realization of any candidate can equal these reservation values. Therefore, ties can only occur in the inspection order of boxes 1,2, and 3.

Suppose the optimal policy fixes the within-group inspection order in $\mathcal{X}$ such that candidate 2 is inspected before candidate 1. Recall the definition of the constraint slack:

$$\Delta_{\text{CONS}}^{i_1,i_2,i_3} \triangleq \mathbf{E}\left[\sum_{i\in\mathcal{Y}} \mathbb{I}_i^{\pi^{i_1,i_2,i_3}} - \sum_{i\in\mathcal{X}} \mathbb{I}_i^{\pi^{i_1,i_2,i_3}}\right], \tag{EC.3}$$

where i_1, i_2, i_3 is a permutation of candidates $\{1,2,3\}$ and π^{i_1,i_2,i_3} is the optimal policy that breaks the ties in the order $i_1 \succ i_2 \succ i_3$. Simple calculations show that:

$$\Delta_{\text{CONS}}^{3,2,1},\ \Delta_{\text{CONS}}^{2,3,1},\ \Delta_{\text{CONS}}^{2,1,3} < 0\,,$$

indicating that no optimal policy (deterministic or randomized) with this fixed within-group order in $\mathcal{X}$ can satisfy the constraint exactly. However, by occasionally changing the within-group order in $\mathcal{X}$ to have candidate 1 inspected before candidate 2, we observe that:

$$-\frac{1}{12} = \Delta_{\text{CONS}}^{3,2,1} < 0 < \Delta_{\text{CONS}}^{3,1,2} = \frac{1}{6}\,,$$

which implies that randomizing between the two orders $3 \succ 2 \succ 1$ and $3 \succ 1 \succ 2$—which have different within-group orders for boxes in $\mathcal{X}$—allows us to satisfy the constraint exactly, as expected.

EC.4. Technical Details of Section 2.4.1: Exact Optimal Policy for Pandora's Box with Value-Specific Ex-ante Affine Constraint

In this section, we provide all the technical details needed to extend our framework in Section 2.3 to incorporate value-specific constraints, as defined in Constraint 11. We start by providing some applications of this category of constraints. We then elaborate on how to generalize dual-based adjustments and extreme tie-breaking rules to this setting. We finish by providing the main result of this section, which is a characterization of the optimal constrained policy.

EC.4.1. Various Applications of value-specific constraints

Consider a threshold-based refinement of (PARITY), for selection or inspection, in which we set:

$$\theta_i^S = \begin{cases} -\mathbb{I}\{v_i \geq \underline{v}_{\mathcal{Y}}\} & i \in \mathcal{Y} \\ +\mathbb{I}\{v_i \geq \underline{v}_{\mathcal{X}}\} & i \in \mathcal{X} \end{cases}, \theta_i^I = 0 \quad \left(\text{or } \theta_i^I = \begin{cases} -\mathbb{I}\{v_i \geq \underline{v}_{\mathcal{Y}}\} & i \in \mathcal{Y} \\ +\mathbb{I}\{v_i \geq \underline{v}_{\mathcal{X}}\} & i \in \mathcal{X} \end{cases}, \theta_i^S = 0 \right),$$

where $\underline{v}_{\mathcal{Y}} \in \mathbb{R}_{\geq}$ (resp. $\underline{v}_{\mathcal{X}} \in \mathbb{R}_{\geq}$) is a threshold defining "acceptable" values for group $\mathcal{Y}$ (resp. $\mathcal{X}$). Typically, we would like to set the thresholds $\underline{v}_{\mathcal{Y}}, \underline{v}_{\mathcal{X}}$ high enough to exclude low-quality candidates and avoid issues such as token interviews as mentioned earlier. Alternatively, we can also consider a threshold-specific refinement of (QUOTA), again for both selection and inspection, in which we set:

$$\theta_i^S = \begin{cases} (\theta - 1) \cdot \mathbb{I}\{v_i \geq \underline{v}_{\mathcal{Y}}\} & i \in \mathcal{Y} \\ \theta \cdot \mathbb{I}\{v_i \geq \underline{v}_{\mathcal{X}}\} & i \in \mathcal{X} \end{cases}, \theta_i^I = 0 \quad \left(\text{or } \theta_i^I = \begin{cases} (\theta - 1) \cdot \mathbb{I}\{v_i \geq \underline{v}_{\mathcal{Y}}\} & i \in \mathcal{Y} \\ \theta \cdot \mathbb{I}\{v_i \geq \underline{v}_{\mathcal{X}}\} & i \in \mathcal{X} \end{cases}, \theta_i^S = 0 \right).$$

This focus on higher values achieves multiple objectives. First, it signals that opportunities (e.g., being interviewed or hired in the context of search and hiring) are accessible regardless of the demographic group, as long as the individual is considered as a top performer. Second, it promotes outcomes that are *truly* fair by eliminating the need for token interviews, as elaborated earlier.

Another significant application of this refined approach to fairness arises in scenarios involving high-cost minority candidates. For example, candidates residing in geographically challenging or inaccessible locations may incur higher inspection costs for the decision-maker. By incorporating a fairness constraint tailored to these high-cost individuals, a guaranteed level of opportunity—be it in the form of interviews or job offers—can be ensured for this group. For example, we can formulate a refinement of (QUOTA) for selection or inspection, in which we set:

$$\theta_i^S = \begin{cases} (\theta - 1) \cdot \mathbb{I}\{c_i \geq \underline{c}\} & i \in \mathcal{Y} \\ \theta & i \in \mathcal{X} \end{cases}, \theta_i^I = 0 \quad \left(\text{or } \theta_i^I = \begin{cases} (\theta - 1) \cdot \mathbb{I}\{c_i \geq \underline{c}\} & i \in \mathcal{Y} \\ \theta & i \in \mathcal{X} \end{cases}, \theta_i^S = 0 \right),$$

where $\underline{c}$ is the defining lower-limit of the cost for high-cost minority group. We note that this refinement mitigates the risk that these candidates are categorically overlooked due to cost considerations, thus adding another layer of nuance to fairness in hiring and search processes.

EC.4.2. Dual-based Adjustments & Extreme Tie-breaking Rules for Value-specific Constraints

To handle the refined Constraint (11), we first observe that for any adaptive feasible policy π, the indicator random variable $\mathbb{I}_i^\pi$ for inspecting box i is independent from the value v_i of the box. Therefore, by following exactly the same recipe as in Section 2.3 (i.e., Lagrangifying the constraint and re-arranging the terms in the Lagrangian function $\mathcal{L}_{\text{CONS}}$) and applying the law of iterated expectations, we get the following equivalent form for the Lagrangian function:

$$\mathcal{L}_{\text{CONS}}(\pi;\lambda) = \mathbf{E}\left[\sum_{i\in[n]} \mathbb{A}_i^\pi (v_i - \lambda\cdot\theta_i^S(v_i,c_i)) - \sum_{i\in[n]} \mathbb{I}_i^\pi \left(c_i + \lambda\cdot \mathbf{E}_{v_i\sim F_i}\left[\theta_i^I(v_i,c_i)\right]\right)\right] + \lambda\cdot b\,,$$

which in turn suggests the following refined dual-adjustment of the values and the costs given λ (cf. the earlier dual adjustment in (7)):

$$\widetilde{v_i} \triangleq v_i - \lambda\cdot\theta_i^S(v_i,c_i) \quad , \quad \widetilde{c_i} \triangleq c_i + \lambda\cdot \mathbf{E}_{v_i\sim F_i}\left[\theta_i^I(v_i,c_i)\right]\ . \tag{EC.4}$$

As before, the Lagrange dual function $\mathcal{G}_{\text{CONS}}$ can be defined as the minimizer of the Lagrangian function over all feasible policies. Moreover, by solving an adjusted instance based on the adjustment in (EC.4), we obtain query access to $\mathcal{G}_{\text{CONS}}$ (through the optimal objective value of the adjusted instance) and $\frac{\partial \mathcal{G}_{\text{CONS}}}{\partial \lambda}$ (through the corresponding constraint slack Δ_{CONS} of the optimal policy after adjustments). Finally, given the minimizer λ^* of $\mathcal{G}_{\text{CONS}}$, the two optimal policies π^+ and π^- corresponding to the perturbed adjusted instances with respect to $\lambda^*+\epsilon$ and $\lambda^*-\epsilon$, respectively, (i) will still be optimal for an instance with adjustment corresponding to λ^*, and (ii) will define the two extreme tie-breaking rules τ^+ and τ^- that guarantee positive and negative slacks, respectively (similar to Proposition 2).

Before explicitly characterizing these two extreme tie-breaking rules τ^+ and τ^-, let us first provide the required technical notation and setup. Consider the optimal adjusted instance, as defined in (EC.4) where $\lambda = \lambda^*$. Recall the definition of the maximum adjusted option value $\widetilde{o_{\max}} \triangleq \max_{i\in([n]\setminus\mathcal{S})\cup\{0\}} \widetilde{o_i}$, defined in any round in the execution of Algorithm 1 on the adjusted instance. Consider any candidate i and let $\{V_{i,1}, V_{i,2}, ..., V_{i,L}\}$, for some $L\in\mathbb{N}\cup\{0\}$, be all the values in $\mathcal{V}_i$ (i.e., the support of F_i) that after adjustment have all became equal to $\widetilde{o_{\max}}$, i.e., $\widetilde{V_{i,\ell}} \triangleq V_{i,\ell} - \lambda\cdot\theta_i^S(V_{i,\ell},c_i) = \widetilde{o_{\max}}, \forall \ell\in[L]$. Without loss, suppose that these values are sorted in decreasing order according to their $\theta_i^S(V_{i,\ell},c_i)$, that is, $\theta_i^S(V_{i,\ell},c_i) \geq \theta_i^S(V_{i,\ell'},c_i)$ if $\ell\leq\ell'$. With this in mind, consider two nested sequences $\mathcal{E}_0^- \subseteq \mathcal{E}_1^- \subseteq ... \subseteq \mathcal{E}_L^-$ and $\mathcal{E}_0^+ \subseteq \mathcal{E}_1^+ \subseteq ... \subseteq \mathcal{E}_L^+$ of subsets of support of F_i defined below:

$$\mathcal{E}_\ell^- \triangleq \left\{V\in\mathcal{V}_i : V - \lambda\cdot\theta_i^S(v,c_i) > \widetilde{o_{\max}}\right\} \cup \{V_{i,j}\}_{1\leq j\leq \ell}\,, \qquad \forall \ell: 0\leq\ell\leq L \tag{EC.5}$$

$$\mathcal{E}_\ell^+ \triangleq \left\{V\in\mathcal{V}_i : V - \lambda\cdot\theta_i^S(v,c_i) > \widetilde{o_{\max}}\right\} \cup \{V_{i,j}\}_{L+1-\ell\leq j\leq L}\,, \qquad \forall \ell: 0\leq\ell\leq L \tag{EC.6}$$

With these two sequences of subsets defined, we provide the exact characteristics of our two extreme tie-breaking rules in the following definition. To simplify the notation, we also slightly abuse the notation and just use θ_i^S (resp. θ_i^I) rather than $\theta_i^S(v_i,c_i)$ (resp. $\theta_i^I(v_i,c_i)$), while keeping in mind that these numbers are random variables for boxes that are yet to be opened.

DEFINITION EC.1 (**REFINED EXTREME TIE-BREAKING RULE**). Given any set of candidates $\mathcal{C}$ for breaking ties at any point during the execution of Algorithm 1 (Line 7), the *negative-extreme rule*, denoted by τ^-, assigns a *tie-breaking score* $s_i^- \in \mathbb{R}$ to each $i \in \mathcal{C}$ as follows (here, $\mathcal{E}_0^- \subseteq \mathcal{E}_1^- \subseteq ... \subseteq \mathcal{E}_L^-$ and $\mathcal{E}_0^+ \subseteq \mathcal{E}_1^+ \subseteq ... \subseteq \mathcal{E}_L^+$ are two nested sequence of subsets of $\mathcal{V}_i$ at this point in the execution of the algorithm on adjusted instance, as defined in eq. (EC.5) and eq. (EC.6)):

- For $i \in \mathcal{C} \setminus \mathcal{O}$:
  - —If $c_i < 0$, set $s_i^- \leftarrow +\infty$.
  - —If $c_i \geq 0$, set $s_i^- \leftarrow \max_{0 \leq \ell \leq L} \left\{ \mathbf{E}_{v_i \sim F_i}\left[\theta_i^S | \mathcal{E}_\ell^-\right] + \frac{\mathbf{E}_{v_i \sim F_i}\left[\theta_i^I\right]}{\mathbf{Pr}\left[\mathcal{E}_\ell^-\right]} \right\}$
    $$\begin{pmatrix} s_i^- = +\infty & \text{if } \mathbf{Pr}\left[\mathcal{E}_0^-\right] = 0 \text{ and } \mathbf{E}_{v_i \sim F_i}\left[\theta_i^I\right] \geq 0 \\ s_i^- = -\infty & \text{if } \mathbf{Pr}\left[\mathcal{E}_L^-\right] = 0 \text{ and } \mathbf{E}_{v_i \sim F_i}\left[\theta_i^I\right] < 0. \end{pmatrix}$$
- For $i \in \mathcal{C} \cap \mathcal{O}$:
  - —If $i \neq 0$, set $s_i^- \leftarrow \theta_i^S$, and if $i = 0$ (that is, outside option), set $s_i^- \leftarrow 0$.

Similarly, the counterpart rule, calling it *positive-extreme rule* and denote it by τ^+, assigns a *tie-breaking score* $s_i^+ \in \mathbb{R}$ to each $i \in \mathcal{C}$ as follows:

- For $i \in \mathcal{C} \setminus \mathcal{O}$:
  - —If $c_i < 0$, set $s_i^+ \leftarrow +\infty$.
  - —If $c_i \geq 0$, set $s_i^+ \leftarrow \max_{0 \leq \ell \leq L} \left\{ -\mathbf{E}_{v_i \sim F_i}\left[\theta_i^S | \mathcal{E}_\ell^+\right] - \frac{\mathbf{E}_{v_i \sim F_i}\left[\theta_i^I\right]}{\mathbf{Pr}\left[\mathcal{E}_\ell^+\right]} \right\}$
    $$\begin{pmatrix} s_i^+ = +\infty & \text{if } \mathbf{Pr}\left[\mathcal{E}_0^+\right] = 0 \text{ and } \mathbf{E}_{v_i \sim F_i}\left[\theta_i^I\right] \leq 0 \\ s_i^+ = -\infty & \text{if } \mathbf{Pr}\left[\mathcal{E}_L^+\right] = 0 \text{ and } \mathbf{E}_{v_i \sim F_i}\left[\theta_i^I\right] > 0. \end{pmatrix}$$
- For $i \in \mathcal{C} \cap \mathcal{O}$:
  - —If $i \neq 0$, set $s_i^+ \leftarrow -\theta_i^S$, and if $i = 0$ (that is, outside option), set $s_i^+ \leftarrow 0$.

Then, the rule τ^- (resp. τ^+) breaks the ties in favor of scores $\{s_i^-\}_{i \in \mathcal{C}}$ (resp. $\{s_i^+\}_{i \in \mathcal{C}}$), that is, it returns any $i^* \in \arg\max_{i \in \mathcal{C}} s_i^-$ (resp. any $i^* \in \arg\max_{i \in \mathcal{C}} s_i^+$).

Given the above definitions of (i) dual-adjusted problem instance in Equation (EC.4) and (ii) extreme tie-breaking rules in Definition EC.1, we are ready to state and prove our main result for this section, which is Theorem EC.1.

THEOREM EC.1 (**Optimal Policy for Value-specific Constrained Problem**). *Consider a modified version of policy RDIP (described in Algorithm 2) in which:*

- *in line (2), the adjusted instances $\{(\widetilde{\mathcal{V}}_i, \widetilde{F}_i, \widetilde{c}_i) | i \in [n]\}$ is defined based on (EC.4), i.e., $\widetilde{v}_i \triangleq v_i - \lambda^* \cdot \theta_i^S(v_i, c_i)$ and $\widetilde{c}_i \triangleq c_i + \lambda^* \cdot \mathbf{E}_{v_i \sim F_i}[\theta_i^I(v_i, c_i)]$,*
- *in line (4), the extreme tie-breaking rules $\{\tau^+, \tau^-\}$ are defined based on the scoring rules introduced in Definition EC.1 in Section EC.4.*

Then this modified policy is optimal for the constrained Pandora's box problem with multiple selection, defined in (OPT-CONS), under a value-specific ex-ante affine constraint as in Equation (11).

Proof of Theorem EC.1. In general the proof of Theorem EC.1 is very similar to that of Proposition 2 and Theorem 1, as such we only provide the parts that have a non-trivial analog. In particular, the first step of Proposition 2 and the proof of Theorem 1 can also be used here. Therefore, the only part in which we need to provide details is the second step in the proof of Proposition 2. We show it here only for the negative-extreme rule, but the proof of the positive-extreme rule would be exactly the same (since the only difference is that instead of $-\varepsilon$ we have ε, which will just change the signs of all perturbations). Furthermore, all opened boxes, as well as all degenerate unopened boxes for which the score will be set to $+\infty$ or $-\infty$ will also be treated in the same way. As a result, the remaining part is to show that the score $\max_{0\le \ell\le L}\left\{\mathbf{E}_{v_i\sim F_i}\left[\theta_i^S|\mathcal{E}_\ell^-\right]+\frac{\mathbf{E}_{v_i\sim F_i}\left[\theta_i^I\right]}{\mathbf{Pr}\left[\mathcal{E}_\ell^-\right]}\right\}$ is in fact the correct amount for an unopened box whose perturbed adjusted cost $\widetilde{c_i}^-$, by perturbing λ with $-\varepsilon$, is positive and also is among $\underset{i\in([n]\setminus\mathcal{S})\cup\{0\}}{\arg\max}\ \{\widetilde{o_i}\}$, which means $\widetilde{\sigma_i}=\widetilde{o_i}=\widetilde{o_{\max}}$.

First, recall that the total number of possible deterministic policies is finite, indicating that there exists an $\bar{\varepsilon}>0$, such that the optimal policy for the perturbed instance $\lambda^*-\varepsilon$ remains the same for all $0\le\varepsilon\le\bar{\varepsilon}$. This shows that the ordering among all perturbed values and reservation values ($\widetilde{\sigma_i}$) will remain exactly the same during the entire perturbation interval $\varepsilon\in(0,\bar{\varepsilon})$.

Knowing that this ordering will remain unchanged over a sufficiently small interval, it is easy to verify that (i) the change in both $\widetilde{v_i}$ and $\widetilde{c_i}$ is linear in ε, and (ii) as a result, the change in $\widetilde{\sigma_i}$ (= $\widetilde{o_i}$) is linear in ε. Let $\widetilde{\sigma_i}^-(\varepsilon)$ be the adjusted reservation value of box i after perturbation $-\varepsilon$ (hence $\widetilde{\sigma_i}^-(0)=\widetilde{\sigma_i}$ and $\widetilde{\sigma_i}^-(\varepsilon)$ is linear in ε for $\varepsilon\in(0,\bar{\varepsilon})$. Given these linear functions $\widetilde{\sigma_i}^-$ for different boxes, among all boxes in $\mathcal{C}$ (those with a tie), the box i with the highest slope would be the one with the highest $\widetilde{o_i}$ after the perturbation, as all these boxes have the same adjusted reservation value $\widetilde{\sigma_i}=\widetilde{o_i}=\widetilde{o_{\max}}$ before the perturbation.

Thus, the only remaining part of the proof is to find an explicit formula for the slope of $\widetilde{\sigma_i}^-$ for such boxes in $\mathcal{C}$, for which we have $\widetilde{\sigma_i}=\widetilde{o_i}=\widetilde{o_{\max}}$. Let $\widetilde{\sigma_i}^-=\widetilde{\sigma_i}+\varepsilon\times\delta$, where δ is the slope of the linear function $\widetilde{\sigma_i}^-$. Denoting the adjusted cost and adjusted values of box i after perturbation $-\varepsilon$ (according to Equation (EC.4)) by $\widetilde{c_i}^-$ and $\widetilde{v_i}^-$, respectively, the following equation should hold:

$$\widetilde{c_i}-\varepsilon\times\mathbf{E}_{v_i\sim F_i}\left[\theta_i^I(v_i,c_i)\right]\overset{(1)}{=}\widetilde{c_i}^-\overset{(2)}{=}\mathbf{E}_{v_i\sim F_i}\left[(\widetilde{v_i}^--\widetilde{\sigma_i}^-)^+\right]=\sum_{V_i\in\text{supp}(F_i):\widetilde{V_i}^-\ge\widetilde{\sigma_i}^-}(\widetilde{V_i}^--\widetilde{\sigma_i}^-)\mathbf{Pr}[V_i]\ ,\quad\text{(EC.7)}$$

where $\widetilde{V_i}^-\equiv\widetilde{V_i}+\varepsilon\times\theta_i^S(V_i,c_i)$, Equation (1) holds due to the definition of adjusted cost after perturbation in Equation (EC.4), and Equation (2) holds due to the definition of the reservation value. Note that for a value V_i in the support of F_i, if $\widetilde{V_i}>\widetilde{\sigma_i}$ then $\widetilde{V_i}^-\ge\widetilde{\sigma_i}^-$ for sufficiently small ε. First, suppose that there is no value V_i in the support of box i such that $\widetilde{V_i}\equiv V_i-\lambda\cdot\theta_i^S(V_i,c_i)=\widetilde{\sigma_i}$. In this case, taking the derivative with respect to ε of both sides of Equation (EC.7) and rearranging the terms, it is easy to show that $\delta=\mathbf{E}_{v_i\sim F_i}\left[\theta_i^S|\mathcal{E}_0^-\right]+\frac{\mathbf{E}_{v_i\sim F_i}\left[\theta_i^I\right]}{\mathbf{Pr}\left[\mathcal{E}_0^-\right]}$ (similar to the way we calculated δ in the proof of step 2 in Proposition 2).

Now, assume that there are $L\ge1$ values $\{V_{i,1},V_{i,2},...,V_{i,L}\}$ in the support of F_i that satisfy $\widetilde{V_{i,\ell}}\equiv V_{i,\ell}-\lambda\cdot\theta_i^S(V_{i,\ell},c_i)=\widetilde{\sigma_i}$. To find a similar characterization for δ using Equation (EC.7), we have to find values in $\{V_{i,1},V_{i,2},...,V_{i,L}\}$ for which $\widetilde{V_{i,\ell}}^-\equiv\widetilde{V_{i,\ell}}+\varepsilon\times\theta_i^S(V_{i,\ell},c_i)$ is no smaller than $\widetilde{\sigma_i}^-\equiv\widetilde{\sigma_i}+\varepsilon\times\delta$. Note that

$\widetilde{V_{i,\ell}} = \widetilde{\sigma}_i$ for all $\ell \in [1:L]$, and therefore $\widetilde{V_{i,\ell}}^- \geq \widetilde{\sigma}_i^-$ if and only if $\theta_i^S(V_{i,\ell}, c_i) \geq \delta$. Also, recall that the values $\{V_{i,1}, V_{i,2}, ..., V_{i,L}\}$ are sorted in the decreasing order of the slopes $\theta_i^S(V_{i,\ell}, c_i)$. As a result, there should exist a unique $0 \leq \ell \leq L$ such that $\theta_i^S(V_{i,\ell'}, c_i) \geq \delta$ if and only if $0 \leq \ell' \leq \ell$, or equivalently $\theta_i^S(V_{i,\ell}, c_i) \geq \delta > \theta_i^S(V_{i,\ell+1}, c_i)$. Putting everything together, the set of values V_i in the support of F_i whose adjustment after perturbation would be higher than $\widetilde{\sigma}_i^-$ (adjusted reservation value $\widetilde{\sigma}_i$ after perturbation) is *exactly* the subset $\mathcal{E}_\ell^-$. We can now find the slope δ using Equation (EC.7). More precisely, the slope δ should satisfy the following chain of equations:

$$\begin{aligned}
\widetilde{c}_i - \varepsilon \times \mathbf{E}_{v_i \sim F_i}\left[\theta_i^I(v_i, c_i)\right] &= \widetilde{c}_i^- \\
&= \mathbf{E}_{v_i \sim F_i}\left[(\widetilde{v}_i^- - \widetilde{\sigma}_i^-)^+\right] \\
&= \sum_{V_i \in \mathcal{E}_\ell^-} (\widetilde{V}_i^- - \widetilde{\sigma}_i^-)\mathbf{Pr}[V_i] \\
&= \sum_{V_i \in \mathcal{E}_\ell^-} \left(\left(\widetilde{V}_i - \widetilde{\sigma}_i\right) \times \mathbf{Pr}[V_i] + \varepsilon \times \left(\theta_i^S(V_i, c_i) - \delta\right) \times \mathbf{Pr}[V_i]\right) \\
&= \mathbf{E}_{v_i \sim F_i}\left[(\widetilde{v}_i - \widetilde{\sigma}_i)^+\right] + \varepsilon \times \mathbf{Pr}\left[\mathcal{E}_\ell^-\right]\left(\mathbf{E}_{v_i \sim F_i}\left[\theta_i^S(v_i, c_i) | \mathcal{E}_\ell^-\right] - \delta\right) \\
&= \widetilde{c}_i + \varepsilon \times \mathbf{Pr}\left[\mathcal{E}_\ell^-\right]\left[\mathbf{E}_{v_i \sim F_i}\left(\theta_i^S(v_i, c_i) | \mathcal{E}_\ell^-\right) - \delta\right].
\end{aligned}$$

If we cancel $\widetilde{c}_i$ from both RHS and LHS and then divide by $\varepsilon \times \mathbf{Pr}\left[\mathcal{E}_\ell^-\right]$, we get

$$\delta = \mathbf{E}_{v_i \sim F_i}\left[\theta_i^S(v_i, c_i) | \mathcal{E}_\ell^-\right] + \frac{\mathbf{E}_{v_i \sim F_i}\left[\theta_i^I(v_i, c_i)\right]}{\mathbf{Pr}\left[\mathcal{E}_\ell^-\right]}.$$

For simplicity, let us define $d_\ell \triangleq \mathbf{E}_{v_i \sim F_i}\left[\theta_i^S(v_i, c_i) | \mathcal{E}_\ell^-\right] + \frac{\mathbf{E}_{v_i \sim F_i}\left[\theta_i^I(v_i, c_i)\right]}{\mathbf{Pr}\left[\mathcal{E}_\ell^-\right]}$ for all $\ell, 0 \leq \ell \leq L$. With this, the problem reduces to a verification problem, wherein we should verify that for which ℓ the following inequalities hold:

$$\theta_i^S(V_{i,\ell}, c_i) \geq d_\ell > \theta_i^S(V_{i,\ell+1}, c_i). \tag{EC.8}$$

Importantly, it turns out that ℓ satisfies Equation (EC.8) if and only if $d_\ell = \max_{0 \leq s \leq L} d_s$. This can be easily derived from the combination of the following four properties; and thus, we skip the rest of the details for the sake of brevity.

1) By definition, the sequence $\theta_i^S(V_{i,\ell}, c_i)$ is a (weakly) decreasing sequence w.r.t. ℓ.

2) $d_{\ell+1}$ is a convex combination of d_ℓ and $\theta_i^S(V_{i,\ell+1}, c_i)$.

3) Combining 1 and 2, we get that the sequence d_ℓ is a (weakly) increasing sequence up until some $\hat{\ell}$, and then it will become a (weakly) decreasing sequence. This also tells us that $d_{\hat{\ell}} = \max d_\ell$.

4) $\hat{\ell}$, and also any ℓ for which d_ℓ still remains equal to $d_{\hat{\ell}}$, are the only ℓ's that satisfy (EC.8). More specifically, any smaller ℓ does not satisfy the second inequality, any larger ℓ does not satisfy the first inequality, and all $\ell \in \arg\max_s d_s$ do satisfy both of the inequalities in Equation (EC.8).

With this we immediately conclude that the correct slope δ would be:

$$\delta = \max_{\ell} d_\ell = \max_{0\le \ell \le L} \left\{ \mathbf{E}_{v_i\sim F_i}\left[\theta_i^S | \mathcal{E}_\ell^-\right] + \frac{\mathbf{E}_{v_i\sim F_i}\left[\theta_i^I\right]}{\mathbf{Pr}\left[\mathcal{E}_\ell^-\right]} \right\},$$

which is exactly the amount that we set to our score s_i^- in such scenarios. The rest would again be quite similar to what we did in Proposition 2, and we show that this scoring rule enables us to run exactly the optimal policy π^- corresponding to the perturbed adjusted instance by $\lambda - \varepsilon$. Hence, we conclude the proof. □

EC.5. Technical Details of Section 2.4.2: Exact Optimal Policy with Multiple Ex-ante Affine Constraints

In this section, we provide all the technical details for the results promised in Section 2.4.2. In particular, we show how to "properly" generalize our approach from Section 2 and Section 3 to handle multiple ex-ante affine constraints *exactly*, that is, without any slack—resulting in a polynomial-time algorithm that computes an optimal policy satisfying all the ex-ante affine constraints with no additive error.

Our generalized approach involves reducing the problem to a variant of the classical (algorithmic) *exact Carathéodory problem* (Carathéodory, 1911). We first explain this reduction in Section EC.5.1. We then introduce specific oracle algorithms in Section EC.5.2 that are polynomial-time computable within both our Pandora's box setting and its generalization to joint Markovian scheduling. Next, in Section EC.5.3, we demonstrate how to solve the reduced exact Carathéodory problem in polynomial time, given access to these oracles in a blackbox manner. Lastly, in Section EC.5.5, we present a simple example showing that the natural extension of "extreme tie-breaking rules" from Section 2 fails, even when applied to settings with two ex-ante affine constraints, indicating that our reduction to exact algorithmic Carathéodory is crucial for solving the problem with multiple affine ex-ante constraints.

In the remainder of this section, we focus on the Pandora's box problem with multiple selections (described in Section 2.1) and its generalization to the joint Markovian scheduling problem with multiple selections or a matroid constraint (described in Section 3.1). We assume we are given $m \in \mathbb{N}$ ex-ante affine constraints, analogous to Constraint 2 for the Pandora's box problem or the ex-ante affine constraints defined in Section 3.1 for the joint Markovian scheduling problem. We first focus on the special case where all constraints are equalities. Later, in Section EC.5.4, we demonstrate how to reduce the problem with m general ex-ante affine constraints, where some are equalities and others are inequalities, to a problem with m ex-ante affine equality constraints.

EC.5.1. Reduction to the Exact Carathéodory Problem

For some notation throughout this section, given an admissible policy π, we denote the constraint slack vector by $\mathbf{\Delta}^\pi_{\text{CONS}} = \left(\Delta^\pi_{\text{CONS},j}\right)_{j\in[m]}$. Here, $\Delta^\pi_{\text{CONS},j}$ represents the slack of the j^{th} ex-ante affine constraint—for example, for the case of the Pandora's box problem, it is defined in Equation (10). For the joint Markovian scheduling problem with m ex-ante affine constraints, this slack vector can be defined similarly. Note that for now we have assumed all constraints are in the equality form. The main objective of this section is to compute a randomized admissible policy π^* that maximizes the expected utility of the search while ensuring that $\mathbf{\Delta}^{\pi^*}_{\text{CONS}} = \mathbf{0} \in \mathbb{R}^m$.

Notably, if the randomized optimal policy π^* is a convex combination (or equivalently, a randomization) of finitely many deterministic admissible policies $\{\pi^{(i)}\}_{i\in S}$ for a finite set S, then $\boldsymbol{\Delta}_{\text{CONS}}^{\pi^*}$ will be the same convex combination of constraint slack vectors $\{\boldsymbol{\Delta}_{\text{CONS}}^{\pi^{(i)}}\}_{i\in S}$, and therefore:

$$\boldsymbol{\Delta}_{\text{CONS}}^{\pi^*} = \mathbf{0} \in \text{Conv}\left(\{\boldsymbol{\Delta}_{\text{CONS}}^{\pi^{(i)}}\}_{i\in S}\right) \ ,$$

where $\text{Conv}(\cdot)$ denotes the convex hull of its input argument.

Now, let us focus on the Pandora's box setting first. The argument for the case of the joint Markovian scheduling is exactly identical and omitted for brevity. Following the same approach as in the case of the single affine constraint, given the vector of dual variables $\boldsymbol{\lambda} \in \mathbb{R}^m$, we define the Lagrangian relaxation function $\mathcal{L}_{\text{CONS}}$ and the Lagrangian dual function $\mathcal{G}_{\text{CONS}}$ as follows:

$$\mathcal{G}_{\text{CONS}}(\boldsymbol{\lambda}) \triangleq \max_{\pi\in\Pi} \ \mathcal{L}_{\text{CONS}}(\pi;\boldsymbol{\lambda}) \ . \tag{EC.9}$$

The function $\mathcal{G}_{\text{CONS}}$ will have same properties as before (such as being a piece-wise affine convex function). Moreover, it continues to hold that an optimal index-based policy (similar to Algorithm 1) in the Lagrangian adjusted version of the problem would be the maximizer solution in Equation (EC.9), providing us with polynomial-time access to both the value and sub-gradients of $\mathcal{G}_{\text{CONS}}(\boldsymbol{\lambda})$, as before. By applying standard methods in convex optimization, we can efficiently find the vector of optimal dual variables $\boldsymbol{\lambda}^*$ minimizing the Lagrangian dual function $\mathcal{G}_{\text{CONS}}$. However, to find the optimal constrained policy we essentially need to find the "saddle point"— a randomized policy π^* that maximizes the Lagrangian relaxation (in expectation) against the worst-case choice of $\boldsymbol{\lambda}$, that is,

$$\pi^* \in \arg\max_{\pi\in\Delta(\Pi)} \min_{\boldsymbol{\lambda}} \mathbf{E}\left[\mathcal{L}_{\text{CONS}}(\pi;\boldsymbol{\lambda})\right] \ .$$

By applying strong-duality (i.e., a weaker version of Sion's minimax theorem (Sion, 1958)), the resulting randomized policy π^* would be a convex combination of (deterministic) maximizer policies in Equation (EC.9) when $\boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda}^*$, and satisfies $\boldsymbol{\Delta}_{\text{CONS}}^{\pi^*} = \mathbf{0}$. However, it remains a challenge to compute this convex combination, as there may be exponentially many such maximizer policies.

Let N be the number of these deterministic maximizer policies denoted by $\{\pi^{(i)}\}_{i\in[N]}$.[26] Each policy $\pi^{(i)}$ is an optimal dual-adjusted index-based policy with respect to $\boldsymbol{\lambda}^*$, corresponding to a certain deterministic tie-breaking rule $\tau^{(i)}$ and associated with a particular slack vector $\boldsymbol{\Delta}_{\text{CONS}}^{\pi^{(i)}} \in \mathbb{R}^m$. The goal here is to select a handful of these policies in a computationally efficient way, so that by randomizing over them, we can achieve slack of $\mathbf{0}$. In other words, we would like to find a small subset $S \subseteq [N]$, such that:

$$\mathbf{0} \in \text{Conv}\left(\{\boldsymbol{\Delta}_{\text{CONS}}^{\pi^{(i)}}\}_{i\in S}\right) \ .$$

[26] As mentioned earlier in Section 2.1, there are finitely many index-based policies in the Pandora's box with multiple selections when value distributions have finite discrete support. Moreover, as we mentioned in Section 3.1, there are also finitely many index-based policies in the JMS problem, simply because we assume the underlying MCs have finite state spaces. All of our results in this section extend to the setting with continuous distributions through proper adjustments and formalizations, which we omit for the sake of simplicity.

Note that because we assume the problem is feasible, there should exist a saddle point solution, or equivalently, a randomized optimal constrained policy π^*. Therefore, we already know that we can obtain a slack of $\mathbf{0}$ by randomizing over all of these maximizer policies, i.e.,

$$\mathbf{0} \in \text{Conv}\left(\{\boldsymbol{\Delta}_{\text{CONS}}^{\pi^{(i)}}\}_{i\in N}\right) .$$

With this formulation of our problem, one can think of the m-dimensional polytope $\mathcal{P} = \text{Conv}(V)$, where $V \triangleq \{\boldsymbol{\Delta}_{\text{CONS}}^{\pi^{(i)}}\}_{i\in N} \subset \mathbb{R}^m$. Now our problem of finding S as described above is, in fact, an instance of the *exact algorithmic Carathéodory problem*: given the polytope $\mathcal{P}$ that contains $\mathbf{0}$, find a "small" subset of points $V' \subseteq V$ in polynomial-time such that $\mathbf{0} \in \text{Conv}(V')$.[27]

We recall that the polytope $\mathcal{P}$ described above can have exponentially many vertices in the parameters of the problem. Even though the classical Carathéodory theorem (Carathéodory, 1911) implies that there should exist $m+1$ vertices of this m-dimensional polytope $\mathcal{P}$ that cover $\mathbf{0}$ (their convex hull includes $\mathbf{0}$), it is not even clear whether we can find a polynomial number of points in V that can cover $\mathbf{0}$. If we can find such a set of points (and therefore their corresponding policies and constraint slack vectors), then by using linear programming we can find the desired convex combination to satisfy the slack of $\mathbf{0}$, and therefore we will have a randomized optimal policy for our problem (i.e., a randomization over policies uncovered, with the resulting convex combination obtained through solving a feasibility LP) that satisfies all the constraints exactly.

In what follows, we provide an affirmative answer by showing a polynomial-time algorithm for our specific instance of the exact Carathéodory problem. In particular, in Section EC.5.2 we show that linear optimization over polytope $\mathcal{P}$ is equivalent to finding the dual-adjusted index-based policy corresponding to a certain perturbation of $\boldsymbol{\lambda}^*$, which can be done in polynomial-time (as we showed in Section 2.3 for the Pandora's box problem, and in Section 3.2.1 and Section EC.8 for the joint Markovian scheduling problem). Having blackbox access to this polynomial-time oracle, in Section EC.5.3 we show how to solve the exact algorithmic Carathéodory problem.

EC.5.2. Linear Optimization Oracle: Basic and Extended

Consider the polytope $\mathcal{P} \subset \mathbb{R}^m$ defined earlier in Section EC.5.1, and an arbitrary direction $\boldsymbol{\omega} \in \mathbb{R}^m$. The goal of this section is to implement a "linear optimization oracle" over $\mathcal{P}$, denoted by LIN-ORACLE, in polynomial time. This simple oracle is formally defined as follows.

DEFINITION EC.2 (LINEAR OPTIMIZATION ORACLE). Given the polytope $\mathcal{P} = \text{Conv}(V) \subset \mathbb{R}^m$, the oracle LIN-ORACLE$(\cdot;\mathcal{P})$ is defined by the following input-output relationship:

- **input:** a direction $\boldsymbol{\omega}$ in $\mathbb{R}^m$.
- **output:** a point $\mathbf{v} \in V$ such that $\mathbf{v} \in \arg\max_{\mathbf{u}\in\mathcal{P}} \boldsymbol{\omega}\cdot\mathbf{u}$.

[27] An alternative way of defining the goal in the algorithmic Carathéodory problem is identifying a subset of *extreme points (i.e., vertices)* of $\mathcal{P}$ that their convex hull includes the target point. Note that not all the points in V are the vertices of $\mathcal{P}$. However, the two versions of the problem are mathematically equivalent, as long as the oracles the algorithm uses always return a vertex, which is without loss of generality by applying standard arguments (see Section EC.5.3 for more details).

By convention, if the input vector is empty, the oracle returns an arbitrary point $\mathbf{v} \in V$.

In order to implement the above linear optimization oracle for our polytope $\mathcal{P}$, we use the structure of this polytope. More specifically, given the optimal dual variables $\boldsymbol{\lambda}^*$, we show that we can find a policy $\hat{\pi}$ in polynomial time such that: (i) the policy $\hat{\pi}$ is an optimal dual-adjusted index-based policy corresponding to $\boldsymbol{\lambda}^*$, and (ii) among such policies, it maximizes $\boldsymbol{\omega} \cdot \boldsymbol{\Delta}^{\hat{\pi}}_{\text{CONS}}$. Formally speaking, we have the following proposition.

PROPOSITION EC.2. *Let the polytope $\mathcal{P}$ be as defined in Section EC.5.1. For any given direction $\boldsymbol{\omega} \in \mathbb{R}^m$, the output of the oracle* LIN-ORACLE$(\boldsymbol{\omega}; \mathcal{P})$ *can be computed in polynomial time.*

Proof. Consider perturbing the vector of optimal dual variables $\boldsymbol{\lambda}^*$ by a perturbation vector $\varepsilon\boldsymbol{\omega}$, where $\varepsilon > 0$ is an infinitesimal scalar. For any admissible policy π for the Pandora's box problem with multiple selections (or similarly, for any admissible policy for the JMS problem), we have:

$$\mathcal{L}_{\text{CONS}}(\pi; \boldsymbol{\lambda}^* + \varepsilon\boldsymbol{\omega}) = \mathcal{L}_{\text{CONS}}(\pi; \boldsymbol{\lambda}^*) + \varepsilon\boldsymbol{\omega} \cdot \boldsymbol{\Delta}^{\pi}_{\text{CONS}}. \tag{EC.10}$$

Let $\pi^{(\varepsilon)} \in \arg\max_{\pi \in \Pi} \mathcal{L}_{\text{CONS}}(\pi; \boldsymbol{\lambda}^* + \varepsilon\boldsymbol{\omega})$. First, $\pi^{(\varepsilon)}$ will be a dual-adjusted index-based optimal policy corresponding to $\boldsymbol{\lambda}^* + \varepsilon\boldsymbol{\omega}$, and thus it is polynomial-time computable. Second, since $\pi^{(\varepsilon)}$ maximizes the right-hand side of (EC.10) for an infinitesimal ε, it must maximize $\mathcal{L}_{\text{CONS}}(\pi; \boldsymbol{\lambda}^*)$. Moreover, it should be the policy π that maximizes $\boldsymbol{\omega} \cdot \boldsymbol{\Delta}^{\pi}_{\text{CONS}}$ among all policies π in $\arg\max_{\pi' \in \Pi} \mathcal{L}_{\text{CONS}}(\pi'; \boldsymbol{\lambda}^*)$.

Combining these observations, for sufficiently small $\varepsilon > 0$, $\pi^{(\varepsilon)}$ is a dual-adjusted index-based optimal policy corresponding to $\boldsymbol{\lambda}^*$, and among such policies, which differ in their tie-breaking rules, it uses a (deterministic) tie-breaking rule that maximizes $\boldsymbol{\omega} \cdot \boldsymbol{\Delta}^{\pi}_{\text{CONS}}$.[28] Hence:

$$\boldsymbol{\Delta}^{\pi^{(\varepsilon)}}_{\text{CONS}} \in \arg\max_{\mathbf{u} \in \mathcal{P}} \boldsymbol{\omega} \cdot \mathbf{u} \quad \text{and} \quad \boldsymbol{\Delta}^{\pi^{(\varepsilon)}}_{\text{CONS}} \in V,$$

allowing us to implement LIN-ORACLE$(\boldsymbol{\omega}; \mathcal{P})$ by returining $\boldsymbol{\Delta}^{\pi^{(\varepsilon)}}_{\text{CONS}}$ (and its corresponding policy $\pi^{(\varepsilon)}$) in polynomial time, as required. □

Before proceeding to the next part, we also introduce the notion of an "extended linear optimization oracle," denoted by EXT-LIN-ORACLE, which slightly generalizes the standard oracle LIN-ORACLE that solves linear optimization over the polytope $\mathcal{P}$. Later, we show that this oracle is not a strict generalization and is indeed *equivalent* to LIN-ORACLE through a simple polynomial-time reduction. Consequently, if linear optimization over $\mathcal{P}$ can be solved in polynomial time, then EXT-LIN-ORACLE can also be implemented as a polynomial-time oracle algorithm.

DEFINITION EC.3 (EXTENDED LINEAR OPTIMIZATION ORACLE). Given the polytope $\mathcal{P} = \text{Conv}(V) \subset \mathbb{R}^m$, the oracle EXT-LIN-ORACLE$(\cdot; \mathcal{P})$ is defined by this input-output relationship:

- **Input:** A tuple of k directions $(\boldsymbol{\omega}_i)_{i \in [k]} = (\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_k)$ for some $k \in \mathbb{N}$, where each $\boldsymbol{\omega}_i \in \mathbb{R}^m$.

[28] More specifically, we can find a closed-form tie-breaking rule for any perturbation of the form $\varepsilon\boldsymbol{\omega}$ using our extreme tie-breaking rules defined in Definition 1. This can be done by simply considering a single ex-ante affine constraint corresponding to a linear combination of our m constraints with coefficients $\{\omega_i\}_{i \in [m]}$.

- **Output:** A point $\mathbf{v} \in \mathcal{A}_k \cap V$, where $\mathcal{A}_k \subseteq \mathcal{A}_{k-1} \subseteq \cdots \subseteq \mathcal{A}_0 \triangleq \mathcal{P}$, and for each $i \in [k]$:

$$\mathcal{A}_i \triangleq \arg\max_{\mathbf{u} \in \mathcal{A}_{i-1}} \boldsymbol{\omega}_i \cdot \mathbf{u}.$$

By convention, if the input tuple is empty, the oracle returns an arbitrary point $\mathbf{v} \in V$.

We note that if we give a single direction $\boldsymbol{\omega} \in \mathbb{R}^m$ as input to the oracle EXT-LIN-ORACLE$(\boldsymbol{\omega}; \mathcal{P})$, then it returns a point $\mathbf{v} \in \mathbb{R}^m$ such that:

$$\mathbf{v} \in \arg\max_{\mathbf{u} \in V} \boldsymbol{\omega} \cdot \mathbf{u} \equiv \left(\arg\max_{\mathbf{u} \in \mathcal{P}} \boldsymbol{\omega} \cdot \mathbf{u} \right) \cap V.$$

Therefore, it can implement the linear optimization oracle LIN-ORACLE over the polytope $\mathcal{P}$ as a special case. The following lemma shows that the oracle EXT-LIN-ORACLE is in fact (computationally) equivalent to the linear optimization oracle LIN-ORACLE.

LEMMA EC.3. *Given the polytope $\mathcal{P} = Conv(V)$, for any tuple of directions $(\boldsymbol{\omega}_i)_{i \in [k]} = (\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_k)$, the output of the oracle* EXT-LIN-ORACLE$((\boldsymbol{\omega}_i)_{i \in [k]}; \mathcal{P})$ *can be computed by a single query to* LIN-ORACLE$(\cdot; \mathcal{P})$.

Proof. Given the k directions $(\boldsymbol{\omega}_i)_{i \in [k]}$, consider a single direction $\boldsymbol{\omega}^{(\varepsilon)} \triangleq \sum_{i \in [k]} \varepsilon^{(i-1)} \boldsymbol{\omega}_i$, where $\varepsilon > 0$ is an infinitesimal scalar. Let $\mathbf{v} \in V$ be the output of LIN-ORACLE$(\boldsymbol{\omega}^{(\varepsilon)}; \mathcal{P})$, i.e.,

$$\mathbf{v} \in \arg\max_{\mathbf{u} \in \mathcal{P}} \boldsymbol{\omega}^{(\varepsilon)} \cdot \mathbf{u} \equiv \arg\max_{\mathbf{u} \in \mathcal{P}} \boldsymbol{\omega}_1 \cdot \mathbf{u} + \varepsilon(\boldsymbol{\omega}_2 \cdot \mathbf{u}) + \varepsilon^2(\boldsymbol{\omega}_3 \cdot \mathbf{u}) + \cdots + \varepsilon^{k-1}(\boldsymbol{\omega}_k \cdot \mathbf{u}).$$

For sufficiently small ε, if $\mathbf{v}$ maximizes $\boldsymbol{\omega}^{(\varepsilon)} \cdot \mathbf{u}$ over $\mathcal{P}$, it must also maximize $\boldsymbol{\omega}_1 \cdot \mathbf{u}$ over $\mathcal{P}$. Let $\mathcal{A}_1$ be the set of all such maximizers. Then:

$$\mathbf{v} \in \arg\max_{\mathbf{u} \in \mathcal{A}_1} \boldsymbol{\omega}_2 \cdot \mathbf{u} + \varepsilon(\boldsymbol{\omega}_3 \cdot \mathbf{u}) + \varepsilon^2(\boldsymbol{\omega}_4 \cdot \mathbf{u}) + \cdots + \varepsilon^{k-2}(\boldsymbol{\omega}_k \cdot \mathbf{u}).$$

Applying a similar argument recursively, for small enough ε, $\mathbf{v}$ also maximizes $\boldsymbol{\omega}_2 \cdot \mathbf{u}$ within $\mathcal{A}_1$. Recalling $\mathcal{A}_0 = \mathcal{P}$ and $\mathcal{A}_i = \arg\max_{\mathbf{u} \in \mathcal{A}_{i-1}} \boldsymbol{\omega}_i \cdot \mathbf{u}$ in Definition EC.3, we conclude that for all $i \in [k]$

$$\mathbf{v} \in \mathcal{A}_i \ .$$

Thus, a single call to LIN-ORACLE suffices to implement EXT-LIN-ORACLE for any tuple of directions, as desired. □

In the remainder of this section, we assume blackbox access to the oracle EXT-LIN-ORACLE. Based on our earlier discussion, if an algorithm uses EXT-LIN-ORACLE in a computationally efficient manner, it can be implemented in polynomial time due to Proposition EC.2 and Lemma EC.3.

EC.5.3. The Exact Algorithmic Carathéodory Problem with Oracle Access: Formal Statement & Solution

We are now ready to formally state the problem we aim to solve:

Problem Statement (Exact Algorithmic Carathéodory): *Given blackbox access to the extended linear optimization oracle* EXT-LIN-ORACLE *(as in Definition EC.3) for a polytope $\mathcal{P} = Conv(V) \subset \mathbb{R}^m$, and knowing that $\mathcal{P}$ includes the origin $\mathbf{0} \in \mathbb{R}^m$, find a polynomial-size subset $V' \subseteq V$ of points (returned by the oracle), in polynomial time, such that $\mathbf{0} \in Conv(V')$.*

Review of basic concepts: We start by reviewing some basic concepts and definitions in polyhedral geometry and linear algebra that we will use throughout the remainder of this section.

DEFINITION EC.4 (CONE, DUAL CONE, POLAR CONE). Let $V' \neq \emptyset$ be a bounded subset of $\mathbb{R}^m$. Then, we have the following definitions:

- *Conic hull of V'*, denoted by $\text{Cone}(V')$:

$$\text{Cone}(V') \triangleq \left\{ \sum_{i\in[k]} \alpha_i \mathbf{v}_i : \forall i,\ \mathbf{v}_i \in V',\ \alpha_i \in \mathbb{R}_{\geq 0},\ k \in \mathbb{N} \right\}.$$

- *Dual cone of V'*, denoted by $\text{Dual-Cone}(V')$:

$$\text{Dual-Cone}(V') \triangleq \{\mathbf{u} \in \mathbb{R}^m : \mathbf{u}\cdot\mathbf{v} \geq 0 \text{ for all } \mathbf{v} \in V'\}.$$

- *Polar cone of V'*, denoted by $\text{Polar-Cone}(V')$:

$$\text{Polar-Cone}(V') \triangleq \{\mathbf{u} \in \mathbb{R}^m : \mathbf{u}\cdot\mathbf{v} \leq 0 \text{ for all } \mathbf{v} \in V'\} = -\text{Dual-Cone}(V').$$

By convention, we also set $\text{Cone}(\emptyset) = \{\mathbf{0}\}$ and $\text{Dual-Cone}(\emptyset) = \text{Polar-Cone}(\emptyset) = \mathbb{R}^m$.

We also use the abbreviated notation $C_{V'}$, $C^*_{V'}$, and $C^\circ_{V'}$ to denote the conic hull, the dual cone, and the polar cone of V', respectively. When it is clear from the context, we may also drop the subscript V' from this notation. See Figure EC.1 for a geometric visualization of these cones.

Also, recall definitions of the *linear span* of a set $Y \subseteq \mathbb{R}^m$:

$$\text{Span}(Y) \triangleq \left\{ \sum_{i\in[k]} \alpha_i \mathbf{v}_i : \forall i,\ \mathbf{v}_i \in Y,\ \alpha_i \in \mathbb{R},\ k \in \mathbb{N} \right\},$$

the *orthogonal complement* of a linear subspace X:

$$X^\perp \triangleq \{\mathbf{u} \in \mathbb{R}^m : \mathbf{u}\cdot\mathbf{v} = 0 \text{ for all } \mathbf{v} \in X\},$$

and the *orthogonal projection* of a set $Y \subseteq \mathbb{R}^m$ onto a linear subspace $X \subseteq \mathbb{R}^m$:

$$\text{Proj}_X(Y) \triangleq \left\{\mathbf{u} \in X : \mathbf{u} = \mathbf{y} + \mathbf{v},\ \mathbf{v} \in X^\perp,\ \mathbf{y}\ \in Y\right\}.$$

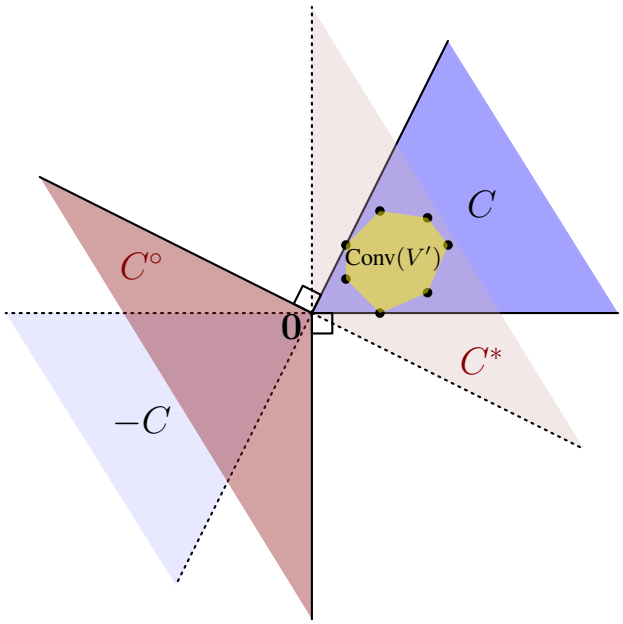


Figure EC.1 **Cone C (dark blue), polar cone C° (dark red), negative cone $-C$ (light blue), and dual cone C^* (light red) of subset of points V', with the convex hull $\text{Conv}(V')$ (yellow).**

EC.5.3.1. The Algorithm Before describing our main algorithm in this section, we prove the following key technical lemma, which is crucial in both the design and analysis of our algorithm.

LEMMA EC.4. *Suppose $\mathbf{0} \in \mathcal{P}$ for a given polytope $\mathcal{P} \subset \mathbb{R}^m$. For any given subset of points $\hat{V} \subseteq \mathcal{P}$, let $U \subseteq \mathcal{P}$ be the set of all points $\left\{\mathbf{u} \in \mathcal{P} : \exists\, \boldsymbol{\omega} \in \text{Proj}_{\text{Span}(\mathcal{P})}(C^\circ_{\hat{V}}) \setminus \{\mathbf{0}\},\ \mathbf{u} \in \arg\max_{\mathbf{u}' \in \mathcal{P}}\ \boldsymbol{\omega} \cdot \mathbf{u}'\right\}$, that is, the convex hull of all points in the polytope $\mathcal{P}$ that are maximizers along some non-zero direction in the projection of the polar cone $C^\circ_{\hat{V}}$ onto the linear subspace spanning $\mathcal{P}$. Then we have:*

$$\mathbf{0} \in \text{Conv}(U \cup \hat{V}).$$

Proof. We assume $\hat{V} \neq \emptyset$, $U \neq \emptyset$, $\mathbf{0} \notin \text{Conv}(\hat{V})$, and $\mathbf{0} \notin \text{Conv}(U)$, as otherwise we have:

(i) if $\hat{V} = \emptyset$, then $C_{\hat{V}} = \{\mathbf{0}\}$ and $C^*_{\hat{V}} = C^\circ_{\hat{V}} = \mathbb{R}^m$. Hence, $\text{Conv}(U) = U = \mathcal{P} \ni \mathbf{0}$, and we are done.

(ii) if $U = \emptyset$, then $\text{Proj}_{\text{Span}(\mathcal{P})}(C^\circ_{\hat{V}}) = \mathbf{0}$. Denoting $\text{Span}(\mathcal{P})^\perp$ by $\mathcal{W}$ (which implies $\text{Span}(\mathcal{P}) = \mathcal{W}^\perp$), we conclude that $C^\circ_{\hat{V}} \subseteq \mathcal{W}$. We now claim that $C_{\hat{V}} = \mathcal{W}^\perp$. First, note that $C_{\hat{V}} \subseteq \mathcal{W}^\perp$, as $\hat{V} \subseteq \mathcal{P} \subset \mathcal{W}^\perp$. Moreover, if $\mathbf{v} \in \mathcal{W}^\perp$, then $\mathbf{v} \cdot \mathbf{u} = 0$ for all $\mathbf{u} \in C^\circ_{\hat{V}}$, and therefore, $\mathbf{v} \in \text{Polar-Cone}(C^\circ_{\hat{V}}) = C_{\hat{V}}$. This implies that $\mathcal{W}^\perp \subseteq C_{\hat{V}}$, which proves our claim. Now, if $\hat{V} = \{\mathbf{0}\}$ we are done. Otherwise, pick an arbitrary $\mathbf{v} \in \hat{V}, \mathbf{v} \neq \mathbf{0}$. Note that $\mathbf{v} \in C_{\hat{V}}$, $C_{\hat{V}} = \mathcal{W}^\perp$, and $\mathcal{W}^\perp$ is a linear subspace. Therefore we should have $-\mathbf{v} \in C_{\hat{V}}$, and hence we should be able to write $-\mathbf{v}$ as a conic combination of vectors in $\hat{V}$. Note that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$. Therefore, there exists a non-zero conic combination of vectors in $\hat{V}$ that is equal to $\mathbf{0}$. By normalizing the corresponding non-negative coefficients to sum up to 1, we have $\mathbf{0} \in \text{Conv}(\hat{V})$ and hence we are done.

(iii) if $\mathbf{0} \in \text{Conv}(\hat{V})$ or $\mathbf{0} \in \text{Conv}(U)$, then clearly $\mathbf{0} \in \text{Conv}(U \cup \hat{V})$ and we are done.

Given these assumptions, we obtain the following equivalent condition for the statement of the lemma that we want to prove:

$$\mathbf{0} \in \text{Conv}(U \cup \hat{V}) \iff (-C_{\hat{V}}) \cap \text{Conv}(U) \neq \emptyset. \tag{EC.11}$$

To see the $\Rightarrow$ direction of this equivalence, note that if $\mathbf{0} \in \text{Conv}(\hat{V} \cup U)$, then there exists a non-zero conic combination of points in U that can be written as the negative of a non-zero conic combination of points in $\hat{V}$, simply because $\hat{V}, U \neq \emptyset$ and $\mathbf{0} \notin \text{Conv}(\hat{V}), \text{Conv}(U)$. Therefore, after normalization, there exists a convex combination of points in U that can be written as a conic combination of points in $-\hat{V}$, hence $(-C_{\hat{V}}) \cap \text{Conv}(U) \neq \emptyset$. To see the $\Leftarrow$ direction of the equivalence, note that if $(-C_{\hat{V}}) \cap \text{Conv}(U) \neq \emptyset$, then there exists a convex combination of points in U that is equal to a non-zero conic combination of points in $-\hat{V}$, as $\mathbf{0} \notin \text{Conv}(U)$. Therefore, there exists a non-zero conic combinations of points in $U \cup \hat{V}$ that is equal to $\mathbf{0}$, and hence after normalization, there exists a convex combination of points in $U \cup \hat{V}$ that is equal to $\mathbf{0}$. Therefore, $\mathbf{0} \in \text{Conv}(U \cup \hat{V})$.

Having the above equivalence, to finish the proof of the lemma, we prove that the RHS of Equation (EC.11) holds by contradiction. Suppose $(-C_{\hat{V}}) \cap \text{Conv}(U) = \emptyset$. Then there exists a strict separating hyperplane H that separates $-C_{\hat{V}}$ and $\text{Conv}(U)$, as both of them are non-empty closed convex sets and $\text{Conv}(U)$ is compact. Note that $\mathbf{0} \notin \text{Conv}(U)$ and $\mathbf{0}$ is the only vertex of the cone $-C_{\hat{V}}$. Therefore, without loss of generality we can assume that H passes through $\mathbf{0}$. Let $\boldsymbol{\omega} \neq \mathbf{0}$ be the normal vector of H, pointing to the side of H that includes $-C_{\hat{V}}$ (i.e., the upper-half). Therefore, for any $\mathbf{v} \in -C_{\hat{V}}$, we have that $\boldsymbol{\omega} \cdot \mathbf{v} \geq 0$, and for any $\mathbf{u} \in \text{Conv}(U)$ we have that $\boldsymbol{\omega} \cdot \mathbf{u} < 0$.

Now decompose $\boldsymbol{\omega}$ into $\boldsymbol{\omega} = \boldsymbol{\omega}_{\mathcal{W}} + \boldsymbol{\omega}_{\mathcal{W}^\perp}$, where $\boldsymbol{\omega}_{\mathcal{W}} \in \mathcal{W}$ and $\boldsymbol{\omega}_{\mathcal{W}^\perp} \in \mathcal{W}^\perp$. Observe that $U \subseteq \mathcal{P} \subset \mathcal{W}^\perp$ and $\hat{V} \subseteq \mathcal{P} \subset \mathcal{W}^\perp$, hence $-C_{\hat{V}} \subseteq \mathcal{W}^\perp$ and $\text{Conv}(U) \subset \mathcal{W}^\perp$, implying that:

$$\forall \mathbf{v} \in -C_{\hat{V}} : \ \boldsymbol{\omega} \cdot \mathbf{v} = \boldsymbol{\omega}_{\mathcal{W}^\perp} \cdot \mathbf{v} \quad , \quad \forall \mathbf{u} \in \text{Conv}(U) : \ \boldsymbol{\omega} \cdot \mathbf{u} = \boldsymbol{\omega}_{\mathcal{W}^\perp} \cdot \mathbf{u}$$

As a result, the hyperplane H' passing through $\mathbf{0}$ with normal vector $\boldsymbol{\omega}_{\mathcal{W}^\perp} \neq \mathbf{0}$ should also be a strict separating hyperplane that separates $-C_{\hat{V}}$ and $\text{Conv}(U)$, with $-C_{\hat{V}}$ being in the upper-half.

We now consider a point $\mathbf{u}^* \in \arg\max_{\mathbf{u}' \in \mathcal{P}} \boldsymbol{\omega}_{\mathcal{W}^\perp} \cdot \mathbf{u}'$. As $\mathbf{u}^*, \mathbf{0} \in \mathcal{P}$, we should have:

$$\boldsymbol{\omega}_{\mathcal{W}^\perp} \cdot \mathbf{u}^* \geq \boldsymbol{\omega}_{\mathcal{W}^\perp} \cdot \mathbf{0} = 0 \ .$$

At the same time, because $-C_{\hat{V}}$ is in the upper-half of hyperplane H', $\boldsymbol{\omega}_{\mathcal{W}^\perp} \cdot \mathbf{v} \geq 0$ for all the points $\mathbf{v} \in -C_{\hat{V}}$, and therefore we have $\boldsymbol{\omega}_{\mathcal{W}^\perp} \in C^\circ_{\hat{V}}$. We conclude that $\boldsymbol{\omega}_{\mathcal{W}^\perp} \in \text{Proj}_{\mathcal{W}^\perp}(C^\circ_{\hat{V}}) = \text{Proj}_{\text{Span}(\mathcal{P})}(C^\circ_{\hat{V}})$ and $\boldsymbol{\omega}_{\mathcal{W}^\perp} \neq \mathbf{0}$, and therefore $\mathbf{u}^* \in U$ (by the definition of the set U). Because U is on the opposite side of hyperplane H compared to $-C_{\hat{V}}$, i.e., in the lower-half of hyperplane H', and H' is a strict separating hyperplane, we have

$$\boldsymbol{\omega}_{\mathcal{W}^\perp} \cdot \mathbf{u}^* < 0 \ ,$$

a contradiction, which finishes the proof of the lemma. □

Now that we have proved Lemma EC.4, we will formally present our algorithm, named *Ellipsoid-based Exact Carathéodory (EEC)*, in Algorithm 4. Given blackbox oracle access to EXT-LIN-ORACLE (which can be implemented in polynomial-time due to Lemma EC.3 and Proposition EC.2), this algorithm recovers a subset of points $\hat{V} \subseteq V$ returned by the oracle such that $\mathbf{0} \in \text{Conv}(\hat{V})$, by only sending polynomial-number of queries to the oracle EXT-LIN-ORACLE and some additional polynomial-time computation. Intuitively speaking, inspired by our key technical lemma in Lemma EC.4, the algorithm is designed to identify a face of the polytope $\mathcal{A} \ni \mathbf{0}$ and a subset of points $\hat{V} \subseteq V \cap \mathcal{A}$, such that if we invoke Lemma EC.4 on $\mathcal{A}$ and $\hat{V}$, we have $U = \emptyset$—and hence we can conclude that $\mathbf{0} \in \text{Conv}(\hat{V})$. We formalize this statement in Section EC.5.3.2.

Before proceeding further, we remark on the connection between our method and the celebrated *"ellipsoid method"* (Khachiyan, 1979), which we use in our analysis.

REMARK EC.1. At a high level, our iterative EEC algorithm is based on the classical ellipsoid method. We use certain mathematical properties of this method in both the algorithm and its analysis, in particular, maintaining an ellipsoid E_t as search space in each iteration t and updating E_t by (i) identifying its center, (ii) slicing E_t using a hyperplane H passing through the center with a normal vector $\boldsymbol{\omega}$, and (iii) explicitly computing the next ellipsoid E_{t+1} so that it is the *minimal ellipsoid* containing one of the two slices produced by cutting E_t with H. Moreover, the ellipsoid method guarantees that the volume of the resulting minimal ellipsoid shrinks exponentially fast. Specifically, if the dimension of E_t (and E_{t+1}) is d, then

$$\frac{\text{Vol}(E_{t+1})}{\text{Vol}(E_t)} \leq e^{-\frac{1}{2(d+1)}}.$$

For additional details, we refer the reader to Vishnoi (2021).

EC.5.3.2. Analysis of the EEC Algorithm The EEC algorithm always maintains a tuple of orthogonal directions Ω. This tuple remains unchanged during the iterations of the inner loop and is only updated at the end of the inner loop (equivalently, when the algorithm goes to the next outer loop iteration). Let $\mathcal{W}$ denote the linear subspace generated by spanning the directions in Ω and let $\mathcal{W}^\perp$ denote its orthogonal complement. Each time the inner loop ends, the algorithm adds the last direction $\boldsymbol{\omega}$ to the current set Ω and updates $\mathcal{W}$ and $\mathcal{W}^\perp$ accordingly, unless $E_t = \{\mathbf{0}\}$, in which case the algorithm terminates. As an important invariant, the EEC algorithm should maintain a subset $V_t \subseteq V$ of points at each iteration t of the inner loop, such that $V_t \subset \mathcal{W}^\perp$ (as we show next).

Consider the tuple $\Omega = (\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_i)$ at some point during the execution of the algorithm, where i is the size of Ω, together with the corresponding linear subspace $\mathcal{W}$ and its orthogonal complement $\mathcal{W}^\perp$. These components help the algorithm identify a suitable (lower-dimensional) face of the polytope $\mathcal{P}$, namely, the face $\mathcal{A}_i$ derived in Definition EC.3 given the directions in Ω. As another important invariant of the algorithm, this face should contain the origin $\mathbf{0}$ and satisfy $\mathcal{A}_i \subset \mathcal{W}^\perp$. The algorithm then "zooms in" on $\mathcal{A}_i$ and the search is restricted to finding a subset of points $\hat{V} \subseteq V \cap \mathcal{A}_i$ such that $\mathbf{0} \in \text{Conv}(\hat{V})$—in other words, the algorithm re-starts the search starting from $\mathcal{A}_i$ as if $\mathcal{A}_i$ was the initial polytope. We note that these face polytopes are nested, that is, $\mathcal{A}_i \subset \mathcal{A}_{i-1} \subset \ldots \mathcal{A}_0 \equiv \mathcal{P}$ at any point during the execution of the algorithm. The following lemma (Lemma EC.5) formalizes this connection and shows they satisfy such desired properties mentioned.

Algorithm 4: Ellipsoid-based Exact Carathéodory (EEC)

input : m-dimensional polytope $\mathcal{P} = \text{Conv}(V) \ni \mathbf{0}$, oracle access to EXT-LIN-ORACLE.

output: set of points $\hat{V} \subseteq V$ such that $\mathbf{0} \in \text{Conv}(\hat{V})$.

Initialize $t \leftarrow 0$, direction tuple $\Omega \leftarrow \emptyset$, ellipsoid $E_0 \leftarrow \mathcal{B}_2^m(1) \triangleq$ unit ℓ_2-ball in $\mathbb{R}^m$, $E_{-1} \leftarrow \{\mathbf{0}\}$, $V_0 \leftarrow \emptyset$. /* Let $C_{V_0} \leftarrow \text{Cone}(V_0) = \{\mathbf{0}\}$, $C^\circ_{V_0} \leftarrow \text{Polar-Cone}(V_0) = \mathbb{R}^m$ */

while $E_t \neq \{\mathbf{0}\}$ **do**

  /* **Outer loop** */

  Set $\mathcal{W} \leftarrow \text{Span}(\Omega)^\perp$ and $E_0 \leftarrow \text{Proj}_{\mathcal{W}}(\mathcal{B}_2^m(1))$ /* i.e., factoring out directions in Ω, hence $E_0 \perp \Omega$; the search zooms in on a face $\mathcal{A}_i \subset \mathcal{W}^\perp$ (Lemma EC.5) */

  Set $t \leftarrow 0$

  while $E_t \neq \{\mathbf{0}\}$ *and* $E_t \neq E_{t-1}$ **do**

    /* **Inner loop** */

    if *center of ellipsoid* $E_t \neq \mathbf{0}$ **then**

      Set $\boldsymbol{\omega} \leftarrow$center of ellipsoid E_t

    else

      Set $\boldsymbol{\omega} \leftarrow$arbitrary non-zero direction in E_t.

      /* the direction $\boldsymbol{\omega}$ is always in E_t and hence orthogonal to Ω */

    Find $\mathbf{v}_t = \text{EXT-LIN-ORACLE}((\Omega, \boldsymbol{\omega}); \mathcal{P})$ and set $V_{t+1} \leftarrow V_t \cup \{\mathbf{v}_t\}$.

    /* the point $\mathbf{v}_t$ is always in the face $\mathcal{A}_i \subset \mathcal{W}^\perp$ (Lemma EC.5). */

    if $\mathbf{v}_t = \mathbf{0}$ **then**

      return $\hat{V} = \{\mathbf{0}\}$. /* the algorithm terminates as $\mathbf{0} \in \hat{V}$. */

    else

      Set $H_t \leftarrow$ hyperplane with normal vector $\mathbf{v_t}$ passing through the center of E_t.

      Set $H_t^- \leftarrow$negative half-space of H_t (i.e., in the opposite direction of $\mathbf{v}_t$).

      /* Let $C_{V_{t+1}} \leftarrow \text{Cone}(V_{t+1})$, $C^\circ_{V_{t+1}} \leftarrow \text{Polar-Cone}(V_{t+1}) = C^\circ_{V_t} \cap \{\mathbf{u} : \mathbf{u} \cdot \mathbf{v}_t \leq 0\}$ */

      Set $E_{t+1} \leftarrow$minimal ellipsoid containing $E_t \cap H_t^-$. /* By following the exact construction as in the ''ellipsoid method'' (Khachiyan, 1979). */

      $t \leftarrow t + 1$.

  if $E_t = E_{t-1}$ **then**

    Set $\Omega \leftarrow (\Omega, \boldsymbol{\omega})$. /* reducing the dimension of the search space. */

return $\hat{V} = V_t$.

LEMMA EC.5. *Given a tuple of directions $\Omega = (\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_i)$ for any $i \in \mathbb{Z}_{\geq 0}$ at any point during the execution of Algorithm 4, and its corresponding orthogonal complement space $\mathcal{W}_i^{\perp}$, we have*

$$\mathbf{0} \in \mathcal{A}_i \;\; \textit{and} \;\; \mathcal{A}_i \subset \mathcal{W}_i^{\perp} \, ,$$

where $\mathcal{A}_i$ is defined as in Definition EC.3 for EXT-LIN-ORACLE $(\Omega; \mathcal{P})$.

Proof. First, we prove $\mathbf{0} \in \mathcal{A}_i$ by induction on i, the size of Ω. If $i = 0$ (that is, the tuple is empty, which happens at the beginning of the execution of the algorithm), we know $\mathbf{0} \in \mathcal{P} = \mathcal{A}_0$. Now suppose that $\mathbf{0} \in \mathcal{A}_i$ with $\Omega = (\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_i)$ at the beginning of some outer iteration of the algorithm. At the end of this outer iteration, we will either terminate or update the set of directions Ω to $(\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_i, \boldsymbol{\omega})$, where $\boldsymbol{\omega}$ corresponds to the direction identified in lines 6-9 of Algorithm 4 in the last iteration of the inner loop. If this update occurs, it has to be the case that $E_{t+1} = E_t$, which only happens if the hyperplane H_t contains the entire E_t rather than cutting it through. Consequently, $\mathbf{v}_t$ should be orthogonal to E_t, and in particular to $\boldsymbol{\omega} \in E_t$, implying $\boldsymbol{\omega} \cdot \mathbf{v}_t = 0$. Furthermore, by applying the induction hypothesis, we already know $\mathbf{0} \in \mathcal{A}_i$. Therefore, by construction of $\mathbf{v}_t$, that is, $\mathbf{v}_t =$ EXT-LIN-ORACLE $((\Omega, \boldsymbol{\omega}); \mathcal{P})$, we can write

$$\mathbf{v}_t \in \mathcal{A}_{i+1} = \arg\max_{\mathbf{u}' \in \mathcal{A}_i} \boldsymbol{\omega} \cdot \mathbf{u}' \;\rightarrow\; \max_{\mathbf{u}' \in \mathcal{A}_i} \boldsymbol{\omega} \cdot \mathbf{u}' = \boldsymbol{\omega} \cdot \mathbf{v}_t = 0 = \boldsymbol{\omega} \cdot \mathbf{0} \, . \tag{EC.12}$$

This shows $\mathbf{0} \in \mathcal{A}_{i+1}$, which completes the induction proof.

Second, we also prove $\mathcal{A}_i \subset \mathcal{W}_i^{\perp}$ by induction. As for the base of induction, $\mathcal{A}_0 = \mathcal{P} \subset \mathbb{R}^m = \mathcal{W}_0^{\perp}$. Now assume $\mathcal{A}_i \subset \mathcal{W}_i^{\perp}$, and we show that $\mathcal{A}_{i+1} \subset \mathcal{W}_{i+1}^{\perp}$. Note that $\mathcal{A}_{i+1} \subset \mathcal{A}_i \subset \mathcal{W}_i^{\perp}$. Therefore, it is enough to show that $\forall \mathbf{v} \in \mathcal{A}_{i+1}$, $\boldsymbol{\omega} \cdot \mathbf{v} = 0$, and hence $\mathcal{A}_{i+1} \subset \mathcal{W}_{i+1}^{\perp}$. Now, if $\mathbf{v} \in \mathcal{A}_{i+1}$, then $\boldsymbol{\omega} \cdot \mathbf{v} = \max_{\mathbf{u}' \in \mathcal{A}_i} \boldsymbol{\omega} \cdot \mathbf{u}' = 0$ (as we showed earlier in eq. (EC.12)), finishing the induction proof. □

Suppose that upon termination of the algorithm, we have $\Omega = (\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_k)$ for some $k \in \mathbb{Z}_{\geq 0}$. Let $\hat{V} = V_\tau \subseteq V \cap \mathcal{A}_k$ be the final set of points returned by the algorithm, where $t = \tau$ is the last index of the last inner loop before termination. By applying the key Lemma EC.4 to the face polytope $\mathcal{A}_k$, which includes $\mathbf{0}$ as stated in Lemma EC.5 we have

$$\mathbf{0} \in \text{Conv}(V_\tau \cup U_\tau),$$

where $U_\tau \subseteq \mathcal{A}_k$ is defined as in Lemma EC.4, i.e., the convex hull of all points in the face polytope $\mathcal{A}_k$ that are maximizers along some non-zero direction in the polar cone $C_{V_\tau}^{\circ}$ of V_τ projected onto the linear span of $\mathcal{A}_k$. If we manage to show that (i) the EEC algorithm (Algorithm 4) terminates after a polynomial number of iterations, and (ii) at termination $\text{Proj}_{\text{Span}(\mathcal{A}_k)}(C_{V_\tau}^{\circ}) = \{\mathbf{0}\}$ and therefore $U_\tau = \emptyset$, then we have a polynomial-time algorithm recovering a subset $V_\tau \subseteq V$ such that $\mathbf{0} \in \text{Conv}(V_\tau)$.

In order to show property (ii) above, it is enough to show that $C_{V_\tau}^{\circ} \subseteq \text{Span}(\mathcal{A}_k)^{\perp}$ and therefore $\text{Proj}_{\text{Span}(\mathcal{A}_k)}(C_{V_\tau}^{\circ}) = \{\mathbf{0}\}$. To establish this claim, we present and prove two simple lemmas.

The first lemma (Lemma EC.6), intuitevly speaking, controls the "projected volume" of the polar cone $C_{V_\tau}^{\circ}$, which turns out to be crucial for establishing the claim.

LEMMA EC.6. *At any iteration t of any of the inner loops of Algorithm 4, we have $C^\circ_{V_t} \cap E_0 \subseteq E_t$.*

Proof. Fix an inner loop of the algorithm (corresponding to a particular outer iteration). We prove the claim by induction on t. As for the base of the induction, for $t=0$ we clearly have $C^\circ_{V_0} \cap E_0 \subseteq E_0$. Now suppose $C^\circ_{V_t} \cap E_0 \subseteq E_t$ at iteration t, and we show that $C^\circ_{V_{t+1}} \cap E_0 \subseteq E_{t+1}$.

To see this, let $\mathbf{c}_t$ denote the center of E_t. First of all, if $\mathbf{c}_t = \mathbf{0}$, then

$$H_t^- = \{\mathbf{u} : \mathbf{u} \cdot \mathbf{v}_t \le 0\}. \tag{EC.13}$$

If $\mathbf{c}_t \neq \mathbf{0}$, then the algorithm sets the direction $\boldsymbol{\omega}$ in that inner iteration to $\mathbf{c}_t$. At the same time, if $\Omega = (\boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_i)$ at the beginning of this inner loop, then based on Lemma EC.5 we know $\mathbf{0} \in \mathcal{A}_i$ (recall the definition of $\mathcal{A}_i$ in Definition EC.3). As a result, according to the definition of $\mathbf{v}_t$, we have $\mathbf{c}_t \cdot \mathbf{v}_t \ge \mathbf{c}_t \cdot \mathbf{0} = 0$, and therefore

$$H_t^- = \{\mathbf{u} : \mathbf{u} \cdot \mathbf{v}_t \le \mathbf{c}_t \cdot \mathbf{v}_t\} \supseteq \{\mathbf{u} : \mathbf{u} \cdot \mathbf{v}_t \le 0\} \tag{EC.14}$$

By combining the induction hypothesis $C^\circ_{V_t} \cap E_0 \subseteq E_t$ with eq. (EC.13) (or eq. (EC.14)), noting that $C^\circ_{V_{t+1}} = C^\circ_{V_t} \cap \{\mathbf{u} : \mathbf{u} \cdot \mathbf{v}_t \le 0\}$ and $E_{t+1} \supseteq E_t \cap H_t^-$, we have

$$C^\circ_{V_{t+1}} \cap E_0 = C^\circ_{V_t} \cap E_0 \cap \{\mathbf{u} : \mathbf{u} \cdot \mathbf{v}_t \le 0\} \subseteq E_t \cap H_t^- \subseteq E_{t+1}\ ,$$

and therefore $C^\circ_{V_{t+1}} \cap E_0 \subseteq E_{t+1}$, as desired. □

Using Lemma EC.6 and the fact that algorithm EEC only terminates when the ellipsoid of the last iteration satisfies $E_\tau = \{\mathbf{0}\}$, we conclude that:

$$C^\circ_{V_\tau} \cap E_0 = E_\tau = \{\mathbf{0}\} \tag{EC.15}$$

We now have our second lemma (Lemma EC.7) that builds on this conclusion to show that $C^\circ_{V_\tau} \subseteq \mathcal{W}$. In fact, we prove a slightly stronger claim.

LEMMA EC.7. *If $C^\circ_{V_\tau} \cap E_0 = \{\mathbf{0}\}$, then $C^\circ_{V_\tau} = Span(\mathcal{A}_k)^\perp = \mathcal{W}$.*

Proof. First, note that $V_\tau \subseteq \mathcal{A}_k \subset \mathcal{W}^\perp \equiv \mathrm{Span}(E_0)$ by construction of V_τ and Lemma EC.5; therefore, for any $\boldsymbol{\omega}' \in \mathcal{W}$, we have:

$$\forall\, \mathbf{v} \in V_\tau :\ \boldsymbol{\omega}' \cdot \mathbf{v} = 0\ ,$$

and hence $\boldsymbol{\omega}' \in C^\circ_{V_\tau}$. This implies $\mathcal{W} \subseteq \mathrm{Span}(\mathcal{A}_k)^\perp \subseteq C^\circ_{V_\tau}$.

Second, we prove $\mathcal{W} \supseteq C^\circ_{V_\tau}$ by contradiction. Suppose that there exists a (non-zero) direction $\boldsymbol{\omega}' \in C^\circ_{V_\tau} \setminus \mathcal{W}$. Then $\boldsymbol{\omega}'$ can be decomposed into $\boldsymbol{\omega}' = \boldsymbol{\omega}'_{\mathcal{W}} + \boldsymbol{\omega}'_{\mathcal{W}^\perp}$, where $\boldsymbol{\omega}'_{\mathcal{W}} \in \mathcal{W}$, $\boldsymbol{\omega}'_{\mathcal{W}^\perp} \in \mathcal{W}^\perp$, and $\boldsymbol{\omega}'_{\mathcal{W}^\perp} \neq \mathbf{0}$. Now, for any $\mathbf{v} \in V_\tau$, noting that $V_\tau \subset \mathcal{W}^\perp$ and thus $\boldsymbol{\omega}'_{\mathcal{W}} \cdot \mathbf{v} = 0$, we have:

$$\boldsymbol{\omega}'_{\mathcal{W}^\perp} \cdot \mathbf{v} = \boldsymbol{\omega}' \cdot \mathbf{v} \le 0\ ,$$

where the last inequality holds due to the definition of the polar cone (Definition EC.4). Consequently, $\boldsymbol{\omega}'_{\mathcal{W}^\perp} \in C^\circ_{V_\tau}$, and hence $\boldsymbol{\omega}'_{\mathcal{W}^\perp}/\|\boldsymbol{\omega}'_{\mathcal{W}^\perp}\| \in C^\circ_{V_\tau}$. Moreover, the vector $\boldsymbol{\omega}'_{\mathcal{W}^\perp}/\|\boldsymbol{\omega}'_{\mathcal{W}^\perp}\|$ clearly belongs to the projection of the unit ℓ_2-ball onto the linear subspace $\mathcal{W}^\perp$ (which is E_0). Therefore, we have $C^\circ_{V_\tau} \cap E_0 \ni \boldsymbol{\omega}'_{\mathcal{W}^\perp}/\|\boldsymbol{\omega}'_{\mathcal{W}^\perp}\| \neq \mathbf{0}$, a contradiction with $C^\circ_{V_\tau} \cap E_0 = \{\mathbf{0}\}$. □

Putting everything together—in particular, having Lemmas EC.5, EC.6, and EC.7—we are now ready to show the following main theorem of this section.

THEOREM EC.2. *Algorithm EEC (4) terminates in polynomial time w.r.t. the size of the problem instance. Furthermore, if $\hat{V} \subseteq V$ is the final set of points returned by the algorithm, we have:*

$$\mathbf{0} \in \mathit{Conv}(\hat{V}). \tag{EC.16}$$

Proof. To show that the algorithm terminates in polynomial time, notice that we are essentially following the update rule of the "ellipsoid method" to obtain E_{t+1} from E_t at each iteration t of each inner loop of the algorithm (which corresponds to a fixed outer iteration). As such, we know that the volume of the ellipsoid will be geometrically shrinking (except only in the last inner iteration). Therefore, for every inner-loop, the number of iterations of the algorithm in that inner loop is upper bounded by $\mathcal{O}\left(\log \frac{1}{\delta}\right)$, where $\delta = 2^{-b}$ is the numerical precision of the input instance and b is the bit complexity of the input instance. Hence, the number of iterations in each inner-loop is polynomial in b. Also, when each inner loop terminates, the hyperplane H_t contains the entire E_t rather than cutting it through (and therefore, $E_t = E_{t+1}$)—unless $E_t = \{\mathbf{0}\}$, in which case the algorithm would terminate. This former case only happens when E_t is in lower dimension and $\mathbf{v}_t$ is orthogonal to it. As the dimension of E_0 goes down by exactly 1 at each outer iteration, and the starting dimension $\dim(\mathcal{P}) = m$, we can only have at most m number of outer iterations. Putting these pieces together, the algorithm terminates after sending polynomial number of queries to the oracle EXT-LIN-ORACLE and polynomial-time extra computation, as desired.

To show that $\mathbf{0} \in \mathrm{Conv}(\hat{V})$, as mentioned earlier, we first invoke the key technical lemma (Lemma EC.4) for the face polytope $\mathcal{A}_k \ni \mathbf{0}$ (Lemma EC.5) and the final set of points $\hat{V} = V_\tau \subseteq V \cap \mathcal{A}_k$. We note that $C^\circ_{V_\tau} = \mathrm{Span}(\mathcal{A}_k)^\perp = \mathcal{W}$ (Lemma EC.6, eq. (EC.15), and Lemma EC.7), and therefore $\mathrm{Proj}_{\mathrm{Span}(\mathcal{A}_k)}(C^\circ_{V_\tau}) = \{\mathbf{0}\}$. As a result $U_\tau = \emptyset$, and hence because of the covering guarantee of Lemma EC.4 we should have:

$$\mathbf{0} \in \mathrm{Conv}(V_\tau \cup U_\tau) = \mathrm{Conv}(\hat{V}) \ ,$$

which finishes the proof of the theorem. □

We conclude this section by remarking that although one could potentially solve this variant of the exact algorithmic Carathéodory problem in polynomial time using standard (but more involved) reductions from optimization to separation (and vice versa)—since algorithmic Carathéodory via a combination of separation and membership oracles is well-known—the resulting algorithm would be quite complicated and would not match the running time of our Algorithm 4. Moreover, our algorithm is able to recover an almost-linear $\tilde{\mathcal{O}}(m)$ number of policies in the resulting randomized tie-breaking rule (thinking of the encoding bit complexity of the problem instance as a constant ϵ, to reduce the volume of the initial ellipsoid in each inner-iteration of Algorithm 4 to ϵ^m, we need $\mathcal{O}\left(m \log(1/\epsilon)\right)$ number of iterations), or in other words, the size of the uncovered convex combination by our algorithm as a function of the dimension m is almost linear. Note that based on

the Carathéodory theorem, this is almost the best possible, as every point can be represented by a convex combination of at most $m+1$ vertices in an m-dimensional polytope (and this is tight). Also, our algorithm is simple, structured, and interpretable—which is completely in contrast to any other known method for solving these types of Carathéodory problems in the literature, e.g., Grötschel et al. (1981, 2012).

EC.5.4. Extension to Multiple General Affine Constraints

So far we have assumed that all affine constraints are equalities. We now show how to reduce the problem with m general affine constraints, some of which are inequalities, to a problem with only equality affine constraints.

Suppose that our original problem has m_n inequality and m_e equality constraints ($m = m_n + m_e$). For each constraint slack vector $\mathbf{\Delta}^{\pi}_{\text{CONS}} = (\Delta^{\pi}_{\text{CONS},j})_{j\in[m]} \in \mathbb{R}^m$, let the first m_n coordinates correspond to the inequality constraints, and the rest correspond to the equality constraints. In the presence of inequality constraints, the goal of our problem is to find a small subset of policies $\{\pi^{(i)}\}_{i\in S}$, with $S \subseteq [N]$, such that:

$$\text{Conv}\left(\{\mathbf{\Delta}^{\pi^{(i)}}_{\text{CONS}}\}_{i\in S}\right) \cap Q \neq \emptyset,$$

where $Q \triangleq \{\mathbf{v} \in \mathbb{R}^m : \forall i \in [m_n], \mathbf{v}_i \geq 0,\ \forall j \in [m_e], \mathbf{v}_{m_n+j} = 0\}$. We also have the guarantee that $\mathcal{P} \cap Q \neq \emptyset$ before finding the set S, where $\mathcal{P} = \text{Conv}(\{\mathbf{\Delta}^{\pi^{(i)}}_{\text{CONS}}\}_{i\in[N]})$, as before. We remark that in the special case with $m_n = 0$, we have $Q = \{\mathbf{0}\}$ and this problem becomes the exact Carathéodory problem in Section EC.5.1. In what follows, we basically show that this new problem is not a strict generalization, and there is a bi-directional polynomial-time reduction from this problem to the exact Carathéodory problem.

To see this reduction, consider adding m_n dummy vectors $\{-\mathbf{e}^{(i)}\}_{i\in[m_n]}$ to the original set of slack vectors $\{\mathbf{\Delta}^{\pi^{(i)}}_{\text{CONS}}\}_{i\in[N]}$, where $\mathbf{e}^{(i)} \in \mathbb{R}^m$ is the standard unit vector for coordinate i. Define

$$\bar{\mathcal{P}} = \text{Conv}\left(\{\mathbf{\Delta}^{\pi^{(i)}}_{\text{CONS}}\}_{i\in[N]} \cup \{-\mathbf{e}^{(i)}\}_{i\in[m_n]}\right).$$

Note that $Q = \text{Cone}(\{\mathbf{e}^{(i)}\}_{i\in[m_n]})$. Therefore, we have the following equivalence:

$$\mathcal{P} \cap Q \neq \emptyset \quad \Longleftrightarrow \quad \mathbf{0} \in \bar{\mathcal{P}}.$$

Moreover, given oracle access to the linear optimization oracle LIN-ORACLE$(\cdot;\mathcal{P})$ for the polytope $\mathcal{P} = \text{Conv}(V)$, we can easily solve linear optimization over polytope $\bar{\mathcal{P}}$ along some direction $\boldsymbol{\omega}$ by first calling LIN-ORACLE$(\boldsymbol{\omega};\mathcal{P})$ to return $\mathbf{v}^* \in V$ and then comparing $\boldsymbol{\omega} \cdot \mathbf{v}^*$ with $\boldsymbol{\omega} \cdot -\mathbf{e}^{(i)}$ for all $i \in [m_n]$ and then returning the one with the maximum dot product. Using the equivalence of linear optimization oracle and the extended linear optimization oracle as in Lemma EC.3, we can construct the extended linear optimization oracle EXT-LIN-ORACLE$(\cdot;\bar{\mathcal{P}})$ for $\bar{\mathcal{P}}$.

Putting all the pieces together, our reduction is as follows: we know $\mathcal{P} \cap Q \neq \phi$, so we know $\mathbf{0} \in \bar{\mathcal{P}}$. Now, by using EXT-LIN-ORACLE$(\cdot;\bar{\mathcal{P}})$, we can efficiently find a polynomial-sized subset of points $V' \subseteq V \cup \{-\mathbf{e}^{(i)}\}_{i\in[m_n]}$, such that $\mathbf{0} \in \text{Conv}(V')$. Nonetheless, we can partition $V' = \hat{V} \cup \{-e^{(i_\ell)}\}_\ell$, where $\hat{V} \subseteq V$. Note that $\mathbf{0} \in \text{Conv}(V')$. Therefore, there exists a convex combination of points in $\hat{V}$ that is equal to a conic combination of points $\{e^{(i_\ell)}\}_\ell \subseteq \text{Cone}(\{\mathbf{e}^{(i)}\}_{i\in[m_n]}) = Q$, implying $\text{Conv}(\hat{V}) \cap Q \neq \emptyset$ and completing our reduction.

EC.5.5. Failure of the Extreme Tie-breaking Rules for Multiple Constraints

In this section, we provide an illustrative example demonstrating why a simple extension of our extreme tie-breaking rules for a single constraint in the Pandora's box setting may not work even in the case with $m = 2$ ex-ante affine constraints—highlighting the importance of our earlier approach by solving the problem via a reduction to the exact algorithmic Carathéodory problem with oracle access.

EC.5.5.1. Overview of the Suggested Approach Recall that when we had a single affine constraint ($m = 1$), we observed in Section 2.3.3 that the two extreme tie-breaking rules induced by the perturbed optimal dual variables $\lambda^* - \varepsilon$ and $\lambda^* + \varepsilon$ indeed achieve the two extreme slack values (highest and lowest) in the constraint among all possible tie-breaking rules. Consequently, one slack should be non-negative while the other should be non-positive (as the problem is feasible), thus enabling zero slack by randomizing over them.

Given the success of extreme tie-breaking rules in the special case of $m = 1$, it might seem reasonable to generalize this idea to settings with $m > 1$ constraints. To this end, we consider 2^m dual-adjusted optimal policies corresponding to specific perturbed versions of vector $\boldsymbol{\lambda}^* \in \mathbb{R}^m$, that is, vectors of the form $\boldsymbol{\lambda}^* + \boldsymbol{\varepsilon}$, where the perturbation vector has the form $\boldsymbol{\varepsilon} \in \{-\varepsilon, +\varepsilon\}^m$ for an infinitesimal scalar $\varepsilon > 0$. One might hope that with a proper randomization over this set, we can make the slack of all constraints zero. However, this approach fails due to the following two reasons.

First, even if this approach can yield zero slacks for all the constraints simultaneously, the computational complexity of this method is significant. The running time is exponential with respect to m, making it impractical for real-world implementations when m is large.

Second, ignoring its computational complexity, there is a deeper issue with this approach and can fail accordingly. In the remainder, we explain the issue first, and then show a simple example in which it arises.

EC.5.5.2. A Geometric Interpretation of the Issue Recall the definition of polytope $\mathcal{P} = \text{Conv}(V)$ from Section EC.5.1. Each point $\mathbf{v} \in V$ corresponds to the slack vector of a dual-adjusted optimal policy $\pi^{(i)}$ with some tie-breaking rule. When we focus only on the extreme tie-breaking rules discussed above, then we essentially have a subset of size 2^m of the points in V, corresponding to the slacks of tie-breaking rules derived from perturbations $\boldsymbol{\lambda}^* + \boldsymbol{\varepsilon}$ for $\boldsymbol{\varepsilon} \in \{\pm\varepsilon\}^m$.

For $m > 1$ affine equality constraints, we can show that the above subset may *not* contain $\mathbf{0}$ in its convex hull, indicating that there is no randomization over the 2^m corresponding dual-adjusted optimal policies that can satisfy all constraints exactly. The following illustrates this phenomenon with a simple and concrete example.

A simple counterexample: In the following, we present a simple parametric example that rigorously demonstrates the phenomenon described earlier. This example has only 3 candidates, where candidates 1 and 2 belong to $\mathcal{X}$ and candidate 3 belongs to $\mathcal{Y}$. Each candidate/box has a binary random value as follows:

$$v_1 = \begin{cases} H_1 = 6 & \text{w.p. } 1 \\ L_1 = 1 & \text{w.p. } 0 \end{cases}, \; v_2 = \begin{cases} H_2 = 14 & \text{w.p. } p_2, \\ L_2 = 2 & \text{w.p. } 1 - p_2. \end{cases}, \; v_3 = \begin{cases} H_3 = 9 & \text{w.p. } p_3 \\ L_3 = 3 & \text{w.p. } 1 - p_3 \end{cases}. \tag{EC.17}$$

Also let the inspection costs be $c_1 = 2$, $c_2 = 10p_2$, and $c_3 = 5p_3$, respectively. This implies $\sigma_1 = \sigma_2 = \sigma_3 = 4$, where σ_i is the reservation value of candidate i, as defined in Equation (3). Hence, it holds that:

$$\min(H_1, H_2, H_3) > \sigma_1 = \sigma_2 = \sigma_3 > L_3 > \max(L_1, L_2). \tag{EC.18}$$

We note that our counterexample remains valid if (EC.18) holds, which can be satisfied by several other choices of numerical values as well. Moreover, (EC.18) ensures that an optimal policy would stop if and only if (i) it sees a high value and selects it, or (ii) it has already inspected all three candidates (all with low values) and then selects the third candidate.

As for the constraints, suppose the decision maker wants to simultaneously satisfy (normalized versions of) demographic parity in both selection and inspection, that is,

$$\mathbf{E}[\mathbb{A}_1^\pi + \mathbb{A}_2^\pi] = 2\,\mathbf{E}[\mathbb{A}_3^\pi] \quad \text{and} \quad \mathbf{E}[\mathbb{I}_1^\pi + \mathbb{I}_2^\pi] = 2\,\mathbf{E}[\mathbb{I}_3^\pi]\,. \tag{EC.19}$$

With this setting in mind, we highlight that in this example the loss due to fairness is designed to be 0, which means that $\lambda^* = \mathbf{0}$ (this can easily be verified). In other words, one of the (randomized) optimal unconstrained policies is also feasible. Thus, the remaining question is to find exactly which tie-breaking rules to choose.

Because all σ_i's are equal, the policy can inspect the candidates in any of the $3! = 6$ possible permutations over $\{1, 2, 3\}$ (and that is the only source of tie in this example). After simple calculations for each of these permutations, the resulting slack in EC.19 for parity in inspection and selection, denoted by Δ_I and Δ_S, respectively, are as follows (here, we choose $p_2 = 0.1$ and $p_3 = 0.8$):

- $1 \to 2 \to 3$: $\Delta_I = 1, \Delta_S = 1$
- $2 \to 1 \to 3$: $\Delta_I = 2 - p_2 = 1.9, \Delta_S = 1$
- $3 \to 1 \to 2$: $\Delta_I = -(1 + p_3) = -1.8, \Delta_S = 1 - 3p_3 = -1.4$
- $3 \to 2 \to 1$: $\Delta_I = p_2 p_3 - p_2 - 2p_3 = -1.62, \Delta_S = 1 - 3p_3 = -1.4$
- $1 \to 3 \to 2$: $\Delta_I = 1, \Delta_S = 1$
- $2 \to 3 \to 1$: $\Delta_I = p_2 p_3 + p_2 - p_3 = -0.62, \Delta_S = 1 - 3p_3 + 3p_2 p_3 = -1.16$

Nevertheless, it turns out that if we have $0 < p_2, p_3 < 1$, then permutations $2 \to 1 \to 3$ and $3 \to 1 \to 2$ are the only ones that can be obtained by the four perturbations $(\pm\varepsilon, \pm\varepsilon)$ corresponding to the extreme tie-breaking rules. Moreover, it is easy to verify that for a general setup of parameters $p_2, p_3 \in (0, 1)$, having only the second and third permutations are not enough to cover $\mathbf{0}$, but once we also include the slack of other permutations, then we will be able to cover $\mathbf{0}$. Figure EC.2 illustrates this point.

EC.6. Constraint Formulations, Examples

In this section we exemplify some of the applications we mentioned in Section 3.1 for the types of constraints that we consider. In order to provide better intuition, we focus on the Markov chains corresponding to our Example 1 as illustrated in Figure 2b.

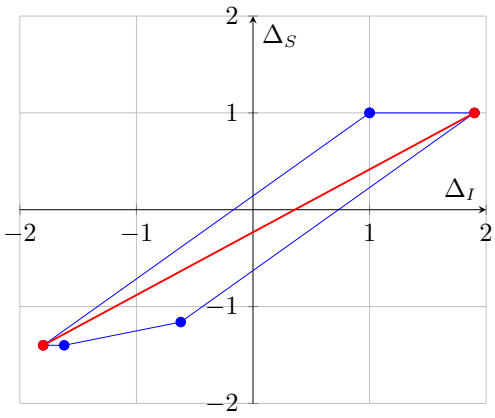


Figure EC.2 Configuration of the slacks for all 6 orders, under $p_2 = 0.1, p_3 = 0.8$. The top right and bottom left belong to the two distinct tie-breaks achievable by $\{-\varepsilon, +\varepsilon\}^2$. The fact that red line does not pass through origin shows the insufficiency of this approach.

EC.6.1. Affine Constraints

For simplicity of exposition, let us first define $S_i^{\text{phone}}, S_i^{\text{onsite}}, S_i^{\text{offer}}$ as the set of all vertices (i.e., states) for $\mathcal{G}_i$ (candidate i's Markov chain) corresponding to phone, onsite, and offer stages, respectively. Note that for every candidate i both S_i^{phone} and S_i^{onsite} consist of only a single state in this particular example (see Figure 2b), while S_i^{offer} includes multiple states. Additionally, we denote by $\boldsymbol{\theta}_{i,v}$, the coefficient in an affine constraint corresponding to the node $v \in \mathcal{G}_i$.

In the following, we formally elaborate some of the mentioned affine constraints in the main body.

Minority group quota on offer. In this case, we add a single constraint to set a quota on the minority group $\mathcal{Y}$ candidates' expected number of offers:

$$\boldsymbol{\theta}_{i,v} = \begin{cases} (\theta - 1) & \text{if } i \in \mathcal{Y} \ \& \ v \in S_i^{\text{offer}} \\ \theta & \text{if } i \in \mathcal{X} \ \& \ v \in S_i^{\text{offer}} \\ 0 & \text{o.w.} \end{cases}, \tag{EC.20}$$

with the constraint being $\boldsymbol{\theta} \cdot \mathbf{p} \leq 0$.

Onsite interview budget. Here, again we add a single constraint on all candidates' onsite interview stage:

$$\boldsymbol{\theta}_{i,v} = \begin{cases} 1 & \text{if } v \in S_i^{\text{onsite}} \\ 0 & \text{o.w.} \end{cases}, \tag{EC.21}$$

with the constraint being: $\boldsymbol{\theta} \cdot \mathbf{p} \leq b$ where b is our desired average budget for onsite interview (in terms of number of people).

Individual fairness in phone interview (lower bounding the minimum opportunity). For this application, we should add a constraint for every candidate $j \in [n]$, denoted by $\boldsymbol{\theta}^j$, such that only the coefficients corresponding to the j^{th} candidate's phone interview stage would appear in it:

$$\forall j \in [n], \quad \boldsymbol{\theta}^j_{i,v} = \begin{cases} 1 & \text{if } i = j \ \& \ v \in S_i^{\text{phone}} \\ 0 & \text{o.w.} \end{cases}, \tag{EC.22}$$

with the constraints being: $\boldsymbol{\theta}^j \cdot \mathbf{p} \geq b, \forall j \in [n]$, where b is some desired lower bound for each individual's phone interview chance.

EC.6.2. Convex Constraints

Here we formulate some examples of convex constraints that our model is able to capture, including but not limited to Nash social welfare, negative entropy, and the generalized mean (a.k.a Hölder mean), which are standard notions for capturing various forms of egalitarian welfare, e.g., see Roughgarden (2010).

Individual-level Nash social welfare on phone interview. We need a single constraint on all candidates' phone interview stage:

$$b \leq F(\mathbf{p}) \triangleq \sqrt[n]{\prod_{i=1}^{n}\Big(\sum_{v\in S_i^{\text{phone}}} \mathbf{p}_{i,v}\Big)}.$$

Group-level negative entropy on onsite interview. We add a single constraint on all groups' onsite interview stage:

$$b \geq F(\mathbf{p}) \triangleq \mathbf{p}_{\mathcal{X}}^{\text{onsite}} \log\left(\mathbf{p}_{\mathcal{X}}^{\text{onsite}}\right) + \mathbf{p}_{\mathcal{Y}}^{\text{onsite}} \log\left(\mathbf{p}_{\mathcal{Y}}^{\text{onsite}}\right),$$

where $\mathbf{p}_{\mathcal{X}}^{\text{onsite}} \triangleq \sum_{i\in\mathcal{X}, v\in S_i^{\text{onsite}}} \mathbf{p}_{i,v}$ and $\mathbf{p}_{\mathcal{Y}}^{\text{onsite}} \triangleq \sum_{i\in\mathcal{Y}, v\in S_i^{\text{onsite}}} \mathbf{p}_{i,v}$.

Group-level generalized mean on offer. We consider a single constraint on all groups' offer stage:

$$F(\mathbf{p}) \triangleq \left(\frac{(\mathbf{p}_{\mathcal{X}}^{\text{onsite}})^q + (\mathbf{p}_{\mathcal{Y}}^{\text{onsite}})^q}{2}\right)^{1/q},$$

where $\mathbf{p}_{\mathcal{X}}^{\text{onsite}} \triangleq \sum_{i\in\mathcal{X}, v\in S_i^{\text{onsite}}} \mathbf{p}_{i,v}$ and $\mathbf{p}_{\mathcal{Y}}^{\text{onsite}} \triangleq \sum_{i\in\mathcal{Y}, v\in S_i^{\text{onsite}}} \mathbf{p}_{i,v}$. Note that for $q < 1$, $F(\mathbf{p})$ is a concave function—which includes geometric mean ($q = 0$) and maximally egalitarian welfare ($q = -\infty$). Therefore, we would set a lower bound for it in the constraint to make sure enough fairness in the allocation of utilities.

Note that for $q > 1$, $F(\mathbf{p})$ is a convex function—which includes max functions ($q = \infty$)—and it is not playing the role of an equalitarian welfare anymore. However, such a function now somewhat captures the amount of "unfairness" in the utilities, that is, how far the utilities are from all being equal. Therefore, we would want to set an upper bound for it in order to encourage fairness. Our framework is general enough that can capture such constraints as well.

Finally, note that one can always add a small quadratic function to the above convex/concave functions with the correct sign, so as to make the resulting function strongly convex/concave as well—which is a requirement we need in our near-optimal near-feasible approximation scheme.

EC.7. A Premier on Fenchel Duality and its Implications

In this supplemental section, we provide more details regarding Fenchel duality and provide a lemma that is crucial in our analysis in proof of Theorem 2 in Section EC.9.

DEFINITION EC.5 (**FENCHEL CONJUGATE (BUBECK ET AL., 2015)**). Given a convex function $F : \mathbb{R}^d \to \mathbb{R}$, the Fenchel conjugate function $F^* : \mathbb{R}^d \to \mathbb{R}$ is defined as:

$$\forall \boldsymbol{\mu} \in \mathbb{R}^d : F^*(\boldsymbol{\mu}) \triangleq \sup_{\mathbf{p}\in\mathbb{R}^d} \left(\boldsymbol{\mu}\cdot\mathbf{p} - F(\mathbf{p})\right)$$

LEMMA EC.8 **(an adaptation of a similar lemma in Bubeck et al. (2015)).** *Suppose (i)* $F:\mathbb{R}^d\to\mathbb{R}$ *is strictly convex, (ii) admits continuous first partial derivatives, (iii)* $\lim_{\|\mathbf{p}\|\to\infty}\|\nabla F_j(\mathbf{p})\|=+\infty$*, and (iv) there exists constants* $H_\mu, L_p>0$ *such that* $\|\nabla F(\mathbf{p})\|_\infty\le H_\mu$ *for every* $\mathbf{p}\in[0,H_p]^d$ *and* $\forall\mathbf{p}:\|\mathbf{p}\|_\infty>L_p$ *we have* $\|\nabla F(\mathbf{p})\|_\infty>H_\mu$*. Then we have:*

(I) The Fenchel conjugate function F^* *is strictly convex with continuous first partial derivatives.*

(II) The conjugate of F^* *is the function* F *itself, i.e.* $(F^*)^*(\mathbf{p})=F(\mathbf{p})$ *for all* $\mathbf{p}\in\mathbb{R}^d$*.*

(III) (envelop theorem) $\nabla F^*(\boldsymbol{\mu})=\mathbf{p}^*(\boldsymbol{\mu})$*, where* $\mathbf{p}^*(\boldsymbol{\mu})\triangleq\underset{\mathbf{p}\in\mathbb{R}^d}{argmax}\ (\boldsymbol{\mu}\cdot\mathbf{p}-F(\mathbf{p}))$*, and* $\nabla F(\mathbf{p})=\boldsymbol{\mu}^*(\mathbf{p})$*, where* $\boldsymbol{\mu}^*(\mathbf{p})\triangleq\underset{\boldsymbol{\mu}\in\mathbb{R}^d}{argmax}\ (\boldsymbol{\mu}\cdot\mathbf{p}-F^*(\boldsymbol{\mu}))$*.*

(IV) The gradient map $\nabla F:\mathbb{R}^d\to\mathbb{R}^d$ *is a bijection (i.e., an invertible and surjective map) and* $(\nabla F)^{-1}=\nabla F^*$*. Moreover, when the map is restricted to the domain* $[0,H_p]^d$*, its image is a subset of* $[-H_\mu,H_\mu]^d$*, and for any point* $\boldsymbol{\mu}\in[-H_\mu,H_\mu]^d$*,* $\nabla F^*(\boldsymbol{\mu})\in[-L_p,L_p]^d$*.*

Proof. Our assumptions (i), (ii) and (iii) guarantee that F is a Legendre map/mirror map, and hence satisfies (I) and (II), and the first part of (IV). See Definition 1 and Lemma 1 in Audibert et al. (2014). (III) is a simple consequence of applying envelop theorem for high-dimensional differentiable functions, applied to F and F^*. Finally, the second part of (IV) holds as $\|\nabla F(\mathbf{p})\|_\infty\le H_\mu$ for every $\mathbf{p}\in[0,H_p]^d$ due to (iv), and last part of (IV) holds as if $\|\nabla F(\mathbf{p})\|_\infty\le H_\mu$ we have $\forall\mathbf{p}:\|\mathbf{p}\|_\infty\le L_p$ due to (iv). □

EC.8. Optimal policy for a General JMS

The first step in solving (OPT-JMS-CONS) is solving the same problem with no ex-ante constraints, that is, finding a policy π that maximizes $\mathbf{E}[R_\pi]$. Importantly, we allow the rewards $\mathbf{R}=[R_i(s)]_{i\in[n],s\in\mathcal{S}_i}$ to take negative or positive rewards in the JMS instance, which proves to be crucial for incorporating ex-ante constraints, as we have already seen in Section 2 and we will also see later when we define dual-adjusted rewards (Section 3.2) for JMS. We sketch how to devise a polynomial-time algorithm for this problem, even in such an instance.

Past work studying the JMS problem with linear rewards characterize the optimal policy by either assuming negative rewards (i.e., costs) for intermediate states and only allowing positive rewards for the terminal states (see, e.g., Dumitriu et al. (2003); Gupta et al. (2019)), or considering the more general so called *No Free Lunch (NFL)* assumption on the state-reward structure of the Markov chains (see, e.g., Gittins (1979); Kleinberg and Slivkins (2017)) and showing a similar analysis extends.

DEFINITION EC.6 (**NFL (KLEINBERG AND SLIVKINS, 2017)**). An alternative $\mathcal{G}$ satisfies NFL if for any state $s\in\mathcal{S}$ with $R(s)>0$, there exists a terminal state $t\in\mathcal{T}$ such that $A(s,t)>0$.

Intuitively speaking, the NFL assumption implies that there shall be no opportunity to receive a positive reward from an intermediary state without risking a transition to a terminal state, thereby terminating the search.

Under NFL assumption, the earlier work established the optimality of the *Gittins index policy* (Gittins, 1979; Dumitriu et al., 2003), which is a generalization of the optimal index-based policy of Weitzman for the Pandora's box problem: Given an instance $\{\mathcal{G}_i\}_{i\in[n]}$, there exists an index mapping $\sigma:\cup_{i\in[n]}\mathcal{S}_i\to\mathbb{R}$ such that choosing the Markov chain $\mathcal{G}_i$ with maximum $\sigma_i(s_i)$ to inspect given states $\{s_i\}_{i\in[n]}$ at each time, until either

k number of the Markov chains enter a terminal state or all remaining indices become non-positive (hence termination), is an optimal policy.

For completeness, in the following, we revisit how the Gittins indices are defined in this more general model under NFL assumption. For each state s in the MC i (satisfying NFL), we define the $\sigma_i(s)$ as the smallest real number such that the following property holds: Consider a new JMS problem that only subsumes MC i, starting from state s, as well as another Markov chain that consists of only 2 states, both with zero rewards, an initial state $s^{'}$ and the terminal state $T^{'}$ with a transition probability of 1 from $s^{'}$ to $T^{'}$. Now if we subtract the amount $\sigma_i(s)$ from the rewards of all of the terminal states in MC i, then there exists no policy that can achieve positive expected reward for this new instance of JMS.

These amounts are, in fact, the Gittins indices of the corresponding states in the JMS instance. We highlight that these indices can be computed in polynomial time using backward induction, as shown in Gittins (1979); Kleinberg and Slivkins (2017). Consequently, they proved this theorem:

THEOREM EC.3 **(Gittins Index Policy for JMS under NFL).** *The index-based policy, which at each time t inspects the Markov chain whose $\sigma_i(s_i^t)$ is the highest across all Markov chains, until either k number of the Markov chains enter a terminal state or all remaining indices become non-positive (hence termination), is an optimal policy for the JMS instance satisfying NFL assumption in Definition EC.6.*

As we show in the remainder of this section, still a refinement of the Gittins index policy above (after proper pre-processing on the Markov chains) can solve the linear optimization over the space of randomized policies Π in polynomial-time for arbitrary reward vectors $\mathbf{R}$, where this time $\{R_i(s)\}$ can be arbitrarily positive or negative. In fact, we show how to reduce the problem in polynomial-time to the special case satisfying NFL by introducing the idea of a *collapsed instance*.

THEOREM EC.4 **(Optimal Policy for JMS with Arbitrary Rewards).** *Given any instance $\{\mathcal{G}_i\}_{i\in[n]}$ of the JMS problem, there exists a polynomial-time reduction that: (i) generates a new instance of the JMS problem satisfying NFL, called "collapsed instance", and (ii) by computing the Gittins indices of the collapsed instance, it returns a new set of indices σ such that the index-based policy corresponding to σ is optimal for the original instance of the JMS with arbitrary rewards.*

In the following subsection, we elaborate on the above discussion and the statement of Theorem EC.4. We then provide proof of Theorem EC.4.

EC.8.1. Collapsing Reduction and Analysis of Theorem EC.4

In this subsection, we characterize the optimal policy for a JMS problem with general rewards, and thus we prove Theorem EC.4. To that end, we start by formally defining "free-lunch", or "FL", states as follows: For a given MC, any state $s \in \mathcal{S}$ is FL iff it violates the NFL condition given in Definition EC.6, i.e., $R(s) > 0$ and there does not exist a terminal state $t \in \mathcal{T}$ such that $A(s,t) > 0$.

Given any problem instance $\{\mathcal{G}_i\}_{i\in[n]}$ that may also include some FL states, we now present a reduction, called "Collapsing", which results in a new JMS instance denoted by $\{\mathcal{G}_i^{\mathcal{C}}\}_{i\in[n]}$ where each $\mathcal{G}_i^{\mathcal{C}}$ contains no FL state.

DEFINITION EC.7 (**Collapsed MC and JMS**). For any $\mathcal{G}_i$ with general rewards, we construct its collapsed version, $\mathcal{G}_i^{\mathcal{C}}$, by iteratively collapsing FL states until there remains no FL state in the resulting MC, $\mathcal{G}_i^{\mathcal{C}}$. In particular, at each iteration, select a FL state, say s. Let $\mathcal{S}^p$ (resp. $\mathcal{S}^c$) be the set of all parents (resp. children) states of s in the current MC, excluding s itself.[29]

1. **State elimination:** Remove state s from the current MC.
2. **Updating reward:** For any $s^p \in \mathcal{S}^p$, add $\frac{R_i(s)}{1-A_i(s,s)}$ to $R_i(s^p)$.
3. **Updating transition probabilities:** For any $s^p \in \mathcal{S}^p$ and $s^c \in \mathcal{S}^c$, add $\frac{A_i(s^p,s)A_i(s,s^c)}{1-A_i(s,s)}$ to $A_i(s^p, s^c)$.

After completing this iterative process for each MC, we arrive at the JMS $\{\mathcal{G}_i^{\mathcal{C}}\}_{i\in[n]}$, the collapsed version of $\{\mathcal{G}_i\}_{i\in[n]}$. This process will end in at most d iterations, as we are removing one state at each iteration.

Based on this process, for any $\mathcal{G}_i$, $i \in [n]$, we define the set of "non-collapsed" states, denoted by $\mathcal{NC}_i \subseteq \mathcal{S}_i$, as the set of all states of $\mathcal{G}_i^{\mathcal{C}}$. We call states in $\mathcal{S}_i \setminus \mathcal{NC}_i$, "collapsed" states. With these definitions, we next establish an equivalence between stationary policies for $\{\mathcal{G}_i^{\mathcal{C}}\}_{i\in[n]}$ and the class of stationary "efficient" policies for the original JMS instance $\{\mathcal{G}_i\}_{i\in[n]}$, as defined below:

DEFINITION EC.8 (**Efficient Policy**). We call a policy "efficient", if it never terminates when (i) it has remaining capacity for selection and (ii) there exists at least one $\mathcal{G}_i$ whose current state is in $\mathcal{S}_i \setminus \mathcal{NC}_i$. In other words, as long as there exists Markov chains whose current states are collapsed states, an efficient policy will always inspect one of such MCs as long as it has not run out of capacity k for slection.

Note that there exists an optimal policy of $\{\mathcal{G}_i\}_{i\in[n]}$ that is efficient. To see why, first note that if there exists an MC in a collapsed state, it is always strictly better to inspect such an MC to accrue its positive expected reward before terminating. Next, notice that the order of inspecting MCs at collapsed states do not impact the expected reward, because (i) all of them have to eventually be inspected, and (ii) inspecting a Markov chain at a collapsed state will not result in terminating the search process, as it does not cause any of the MCs to go to a terminal state. We state the aforementioned equivalence in the following claim.

CLAIM EC.1. *Consider any general instance of JMS, $\{\mathcal{G}_i\}_{i\in[n]}$, with starting states $(s_1^{(0)}, \ldots, s_n^{(0)})$ where $\forall i \in [n] : s_i^{(0)} \in \mathcal{NC}_i$. Then for any stationary efficient policy π of $\{\mathcal{G}_i\}_{i\in[n]}$ with $(s_1^{(0)}, \ldots, s_n^{(0)})$, there exists a stationary policy for $\{\mathcal{G}_i^{\mathcal{C}}\}_{i\in[n]}$ with $(s_1^{(0)}, \ldots, s_n^{(0)})$, that achieves the same expected reward, and vice versa.*

Proof. Define $\pi^{\mathcal{C}}$ as the restriction of π on only the non-collapsed states $\prod_{i\in[n]} \mathcal{NC}_i$. In other words, policy $\pi^{\mathcal{C}}$ for $\{\mathcal{G}_i^{\mathcal{C}}\}_{i\in[n]}$ at any state will make the exact same decision as π does in that state of the original JMS $\{\mathcal{G}_i\}_{i\in[n]}$. To see why the expected rewards under policy π (for $\{\mathcal{G}_i\}_{i\in[n]}$) and $\pi^{\mathcal{C}}$ (for $\{\mathcal{G}_i^{\mathcal{C}}\}_{i\in[n]}$) are the same, note that if there is a MC in $\{\mathcal{G}_i\}_{i\in[n]}$ at a collapsed state, π will inspect that (by definition of being efficient). Further, as noted above, the order of inspecting MCs at collapsed states does not impact the expected reward. As such, the expected reward accrued during inspection of MCs at collapsed states will be the same as the increase in the rewards of non-collapsed states determined in the reduction of $\{\mathcal{G}_i\}_{i\in[n]}$ to $\{\mathcal{G}_i^{\mathcal{C}}\}_{i\in[n]}$ (as in Definition EC.7). For the reverse direction, we define π for $\{\mathcal{G}_i\}_{i\in[n]}$ as the policy which makes the same decision as $\pi^{\mathcal{C}}$ does,

[29] Note that s can have a self-loop and thus be its own parent and child

if every $\mathcal{G}_i$ is at a state in $\mathcal{NC}_i$. Otherwise, it will inspect a $\mathcal{G}_i$ whose state is not in $\mathcal{NC}_i$. By a similar line of reasoning, the expected reward under the newly-constructed π (for $\{\mathcal{G}_i\}_{i\in[n]}$) will be the same as that under $\pi^{\mathcal{C}}$ (for $\{\mathcal{G}_i^{\mathcal{C}}\}_{i\in[n]}$). □

Building on Claim EC.1, in the next claim we complete the proof of Theorem EC.4 by giving an optimal index-based policy for the original JMS, $\{\mathcal{G}_i\}_{i\in[n]}$.

CLAIM EC.2. *Let $\sigma_i^{\mathcal{C}}(s)$, $\forall i\in[n], \forall s\in\mathcal{NC}_i$, be the Gittins indices defined in Kleinberg and Slivkins (2017); Gupta et al. (2019) for the JMS, $\{\mathcal{G}_i^{\mathcal{C}}\}_{i\in[n]}$, which satisfies the NFL condition. Then, the index-based (greedy) policy for selecting at most k number of MCs based on the following indices is an optimal policy for the original JMS, $\{\mathcal{G}_i\}_{i\in[n]}$.*

$$\sigma_i(s)\triangleq\begin{cases}\sigma_i^{\mathcal{C}}(s) & s\in\mathcal{NC}_i\ ,\\ +\infty & o.w.\end{cases} \tag{EC.23}$$

Proof. To prove this claim, consider any stationary efficient optimal policy π^* for $\{\mathcal{G}_i\}_{i\in[n]}$.[30] Since it is efficient, by Claim EC.1, its equivalent "collapsed" policy $\pi^{\mathcal{C}}$ for $\{\mathcal{G}_i^{\mathcal{C}}\}_{i\in[n]}$ achieves the same expected reward. Now consider the index-based policy based on the $\sigma_i(s)$ introduced above; we call it $\widetilde{\pi}$. First, notice that this policy is also an efficient policy: by definition of indices, if there is a MC at a collapsed state (thus with index $+\infty$), then this policy will inspect such a MC. Since $\widetilde{\pi}$ is stationary and efficient, again by Claim EC.1 its equivalent "collapsed" policy $\widetilde{\pi}^{\mathcal{C}}$ for $\{\mathcal{G}_i^{\mathcal{C}}\}_{i\in[n]}$ achieves the same expected reward. Second, notice that $\widetilde{\pi}^{\mathcal{C}}$ was nothing but the optimal Gittins index policy for $\{\mathcal{G}_i^{\mathcal{C}}\}_{i\in[n]}$, implying that its expected reward cannot be less than $\pi^{\mathcal{C}}$. Hence, we can conclude that for any starting states $(s_1^{(0)},\ldots,s_n^{(0)})$, where $\forall i\in[n]: s_i^{(0)}\in\mathcal{NC}_i$, the expected reward of $\widetilde{\pi}$ is at least that of π^*. Finally, suppose there are some Markov chains whose starting states are collapsed states. Then, since both $\widetilde{\pi}$ and π^* are efficient, both will inspect those Markov chains until they reach a state $(s_1,\ldots,s_n)$, where $\forall i\in[n]: s_i\in\mathcal{NC}_i$. As a result, from any starting state the expected reward of $\widetilde{\pi}$ would be at least that of π^*, implying that $\widetilde{\pi}$ is also an optimal policy. This will conclude the proof of this claim. □

In the last part of this section, for the sake of completeness, we restate the definition of the Gittins indices, $\sigma_i^{\mathcal{C}}(s)$, for the collapsed JMS $\{\mathcal{G}_i^{\mathcal{C}}\}_{i\in[n]}$, which satisfied the NFL assumption (Dumitriu et al., 2003; Kleinberg and Slivkins, 2017).

DEFINITION EC.9 (**Gittins indices of $\{\mathcal{G}_i^{\mathcal{C}}\}_{i\in[n]}$**). For any Markov chain $\mathcal{G}_i^{\mathcal{C}}$ and state $s\in\mathcal{NC}_i$, we define the $\sigma_i^{\mathcal{C}}(s)$ as the smallest real number such that this property holds: *Consider a new JMS problem with only two Markov chains $\left\{\widetilde{\mathcal{G}}_j\right\}_{j\in[2]}$, where $\widetilde{\mathcal{G}}_1\triangleq\mathcal{G}_i^{\mathcal{C}}$, and $\widetilde{\mathcal{G}}_2\triangleq(\mathcal{S}=\{s',t'\},\mathcal{T}=\{t'\},A(s_1,s_2)=\mathbb{1}\{s_2=t'\},R=(0,0))$. Now if we subtract $\sigma_i^{\mathcal{C}}(s)$ from the rewards of all of the terminal states in $\widetilde{\mathcal{G}}_1$, then there exists no policy that can achieve positive expected reward for this new instance of JMS.*

We conclude by noting that we can establish the polynomial-time computability of the Gittins indices defined above by a simple adjustment to the polynomial-time algorithm given in Dumitriu et al. (2003). We omit the details for the sake of brevity.

[30] See Dumitriu et al. (2003) for existence of an optimal stationary policy.

EC.9. Missing Proofs of Section 3

Proof of Theorem 2. Let the parameters be chosen as (i) $H_\lambda = H_\beta = \frac{dH_p+\varepsilon}{\delta}$, (ii) $K_{\rm I} = \left(\frac{6dH_\mu(H_p+L_p)H_\beta m_c}{\varepsilon}\right)^2 = \mathcal{O}\left(\frac{1}{\delta^2\epsilon^2}\right)$ and $K_{\rm O} = \left(\frac{3\max\left(2H_\lambda m_a, H_\beta m_c\right)}{\varepsilon}\right)^2 = \mathcal{O}\left(\frac{1}{\delta^2\epsilon^2}\right)$, and (iii) $\gamma_{\rm I} = \frac{2H_\mu}{(H_p+L_p)H_\beta}\frac{1}{\sqrt{K_{\rm I}}}$, $\gamma_{{\rm O},\lambda} = \frac{H_\lambda}{2\sqrt{K_{\rm O}}}$, and $\gamma_{{\rm O},\beta} = \frac{H_\beta}{\sqrt{K_{\rm O}}}$. We start by considering the inner-loop of the G-RDIP policy (Algorithm 3). Fix an iteration (m,ℓ) of the inner loop. For any $i\in[m_c]$ and $\boldsymbol{\mu}_i\in[-H_\mu,H_\mu]^d$ we have:

$$\begin{aligned}
\beta_i^{(m)}(\boldsymbol{\mu}_i^{(m,\ell)}-\boldsymbol{\mu}_i)\cdot\left(\nabla F_i^*(\boldsymbol{\mu}_i^{(m,\ell)})-\mathbf{p}_{\pi^{(m,\ell)}}\right) &= \frac{1}{\gamma_{\rm I}}(\boldsymbol{\mu}_i^{(m,\ell)}-\boldsymbol{\mu}_i)\cdot(\boldsymbol{\mu}_i^{(m,\ell)}-\boldsymbol{\omega}_i^{(m,\ell)})\\
&\overset{(1)}{=}\frac{1}{2\gamma_{\rm I}}\left(\|\boldsymbol{\mu}_i^{(m,\ell)}-\boldsymbol{\mu}_i\|_2^2+\|\boldsymbol{\mu}_i^{(m,\ell)}-\boldsymbol{\omega}_i^{(m,\ell)}\|_2^2-\|\boldsymbol{\omega}_i^{(m,\ell)}-\boldsymbol{\mu}_i\|_2^2\right)\\
&\overset{(2)}{\le}\frac{1}{2\gamma_{\rm I}}\left(\|\boldsymbol{\mu}_i^{(m,\ell)}-\boldsymbol{\mu}_i\|_2^2+\|\boldsymbol{\mu}_i^{(m,\ell)}-\boldsymbol{\omega}_i^{(m,\ell)}\|_2^2-\|\boldsymbol{\mu}_i^{(m,\ell+1)}-\boldsymbol{\mu}_i\|_2^2\right)\\
&\overset{(3)}{\le}\frac{1}{2\gamma_{\rm I}}\left(\|\boldsymbol{\mu}_i^{(m,\ell)}-\boldsymbol{\mu}_i\|_2^2-\|\boldsymbol{\mu}_i^{(m,\ell+1)}-\boldsymbol{\mu}_i\|_2^2+d\gamma_{\rm I}^2H_\beta^2(H_p+L_p)^2\right),
\end{aligned}\tag{EC.24}$$

where equality (1) holds due to the Pythagorean's lemma, inequality (2) holds by the fact that $\boldsymbol{\mu}_i^{(m,\ell+1)}$ is the projection of $\boldsymbol{\omega}_i^{(m,\ell)}$ onto $[-H_\mu,H_\mu]^d$, and inequality (3) holds as $\|\mathbf{p}_{\pi^{(m,\ell)}}-\nabla F_i^*(\boldsymbol{\mu}_i^{(m,\ell)})\|_\infty\le(H_p+L_p)$, sbecause $\|\nabla F_i^*(\boldsymbol{\mu}_i^{(m,\ell)})\|_\infty\le L_p$ for all $\boldsymbol{\mu}_i^{(m,\ell)}\in[-H_\mu,H_\mu]^d$ by applying Lemma EC.8. By averaging both hand sides of (EC.24) over $\ell\in[K_{\rm I}]$, rearranging the terms, and finally setting $\gamma_{\rm I}=\frac{2H_\mu}{(H_p+L_p)H_\beta}\frac{1}{\sqrt{K_{\rm I}}}$ we have:

$$\begin{aligned}
\frac{1}{K_{\rm I}}\sum_{\ell\in[K_{\rm I}]}\beta_i^{(m)}(\boldsymbol{\mu}_i^{(m,\ell)}-\boldsymbol{\mu}_i)\cdot\left(\nabla F_i^*(\boldsymbol{\mu}_i^{(m,\ell)})-\mathbf{p}_{\pi^{(m,\ell)}}\right) &\le \frac{1}{2K_{\rm I}\gamma_{\rm I}}\|\boldsymbol{\mu}_i^{(m,1)}-\boldsymbol{\mu}_i\|_2^2-\frac{1}{2K_{\rm I}\gamma_{\rm I}}\|\boldsymbol{\mu}_i^{(m,K_{\rm I}+1)}-\boldsymbol{\mu}_i\|_2^2\\
&\quad+\frac{\gamma_{\rm I}}{2}d(H_p+L_p)^2H_\beta^2\\
&\le\frac{2dH_\mu^2}{K_{\rm I}\gamma_{\rm I}}+\frac{\gamma_{\rm I}}{2}d(H_p+L_p)^2H_\beta^2=\frac{2dH_\mu(H_p+L_p)H_\beta}{\sqrt{K_{\rm I}}}
\end{aligned}\tag{EC.25}$$

Now, denote by $\overline{\boldsymbol{\mu}}_i^{(m)}$ the average of $\boldsymbol{\mu}_i^{(m,\ell)}$ during the outer iteration k, i.e., $\overline{\boldsymbol{\mu}}_i^{(m)}=\frac{1}{K_{\rm I}}\sum_{\ell\in[K_{\rm I}]}\boldsymbol{\mu}_i^{(m,\ell)}$ and denote by $\overline{\mathbf{p}}^{(m)}$ the average expected visit numbers vector of $\pi^{(m,\ell)}$ during the outer iteration k, i.e., $\overline{\mathbf{p}}^{(m)}=\frac{1}{K_{\rm I}}\sum_{\ell\in[K_{\rm I}]}\mathbf{p}_{\pi^{(m,\ell)}}$. Using these notations, we can further inequality (EC.25). To do so, by incorporating the convexity of the conjugate function F_i^* (Lemma EC.8) we have:

$$\begin{aligned}
\frac{1}{K_{\rm I}}\sum_{\ell\in[K_{\rm I}]}\beta_i^{(m)}(\boldsymbol{\mu}_i^{(m,\ell)}-\boldsymbol{\mu}_i)\cdot\nabla F_i^*(\boldsymbol{\mu}_i^{(m,\ell)}) &\ge\frac{1}{K_{\rm I}}\sum_{\ell\in[K_{\rm I}]}\beta_i^{(m)}\left(F_i^*(\boldsymbol{\mu}_i^{(m,\ell)})-F_i^*(\boldsymbol{\mu}_i)\right)\\
&\ge\beta_i^{(m)}\left(F_i^*(\overline{\boldsymbol{\mu}}_i^{(m)})-F_i^*(\boldsymbol{\mu}_i)\right)
\end{aligned}\tag{EC.26}$$

Also, we have:

$$\frac{1}{K_{\rm I}}\sum_{\ell\in[K_{\rm I}]}\beta_i^{(m)}(\boldsymbol{\mu}_i^{(m,\ell)}-\boldsymbol{\mu}_i)\cdot\mathbf{p}_{\pi^{(m,\ell)}}=\frac{1}{K_{\rm I}}\sum_{\ell\in[K_{\rm I}]}\beta_i^{(m)}\boldsymbol{\mu}_i^{(m,\ell)}\cdot\mathbf{p}_{\pi^{(m,\ell)}}-\beta_i^{(m)}\boldsymbol{\mu}_i\cdot\overline{\mathbf{p}}^{(m)}\tag{EC.27}$$

By combining (EC.25), (EC.26), and (EC.27), and summing up the terms for all $i\in[m_c]$, we have:

$$\begin{aligned}
&\sum_{i\in[m_c]}\beta_i^{(m)}\left(F_i^*(\overline{\boldsymbol{\mu}}_i^{(m)})-F_i^*(\boldsymbol{\mu}_i)\right)+\sum_{i\in[m_c]}\beta_i^{(m)}\boldsymbol{\mu}_i\cdot\overline{\mathbf{p}}^{(m)}-\sum_{i\in[m_c]}\frac{1}{K_{\rm I}}\sum_{\ell\in[K_{\rm I}]}\beta_i^{(m)}\boldsymbol{\mu}_i^{(m,\ell)}\cdot\mathbf{p}_{\pi^{(m,\ell)}}\\
\le&\frac{2dH_\mu(H_p+L_p)H_\beta m_c}{\sqrt{K_{\rm I}}},
\end{aligned}\tag{DUAL-BEST-RESPONSE-INNER}$$

for any $\boldsymbol{\mu}_i \in [-H_\mu, H_\mu]^d$. On the other hand, our algorithm also selects the optimal policy $\pi^{(m,\ell)} \in \Pi$ for the adjusted rewards $\widetilde{\mathbf{R}}^{(m,\ell)} = \mathbf{R} - \sum_{j\in[m_a]} \lambda_j^{(m)}\boldsymbol{\theta}_j - \sum_{i\in[m_c]} \beta_i^{(m)}\boldsymbol{\mu}_i^{(m,\ell)}$ at any iteration (m,ℓ). Hence:

$$\frac{1}{K_{\mathrm{I}}}\sum_{\ell\in[K_{\mathrm{I}}]} \widetilde{\mathbf{R}}^{(m,\ell)}\cdot\mathbf{p} \le \frac{1}{K_{\mathrm{I}}}\sum_{\ell\in[K_{\mathrm{I}}]} \widetilde{\mathbf{R}}^{(m,\ell)}\cdot\mathbf{p}_{\pi^{(m,\ell)}}\ , \tag{EC.28}$$

or equivalently:

$$\mathbf{R}\cdot\mathbf{p} - \sum_{j\in[m_a]}\lambda_j^{(m)}\boldsymbol{\theta}_j\cdot\mathbf{p} - \sum_{i\in[m_c]}\beta_i^{(m)}\overline{\boldsymbol{\mu}}_i^{(m)}\cdot\mathbf{p} \le \mathbf{R}\cdot\overline{\mathbf{p}}^{(m)} - \sum_{j\in[m_a]}\lambda_j^{(m)}\boldsymbol{\theta}_j\cdot\overline{\mathbf{p}}^{(m)} - \sum_{i\in[m_c]}\frac{1}{K_{\mathrm{I}}}\sum_{\ell\in[K_{\mathrm{I}}]}\beta_i^{(m)}\boldsymbol{\mu}_i^{(m,\ell)}\cdot\mathbf{p}_{\pi^{(m,\ell)}}\ ,$$
(PRIMAL-BEST-RESPONSE-INNER)

where the above inequalities hold for any $\mathbf{p}\in\mathcal{P}$. Finally, by combining the two inequalities in (DUAL-BEST-RESPONSE-INNER) and (PRIMAL-BEST-RESPONSE-INNER), and rearranging the terms, for any set of vectors $\mathbf{p}\in\mathcal{P}$ and $\boldsymbol{\mu}_i\in[-H_\mu,H_\mu]^d$, and for $i\in[m_c], m\in[K_{\mathrm{O}}]$ we have:

$$\overline{\mathcal{L}}_{\text{JMS-CONS}}\left(\mathbf{p};\{\lambda_j^{(m)}\},\{\beta_i^{(m)}\},\{\overline{\boldsymbol{\mu}}_i^{(m)}\}\right) - \overline{\mathcal{L}}_{\text{JMS-CONS}}\left(\overline{\mathbf{p}}^{(m)};\{\lambda_j^{(m)}\},\{\beta_i^{(m)}\},\{\boldsymbol{\mu}_i\}\right) \le \frac{2dH_\mu(H_p+L_p)H_\beta m_c}{\sqrt{K_{\mathrm{I}}}} \equiv \varepsilon_1.$$
(INNER-APPROXIMATE-EQUILIBRIUM)

Next, we look at the outer loop of G-RDIP policy. Fix an iteration k of the outer loop, and consider the way $\beta_i^{(m)}$ and $\lambda_j^{(m)}$ are updated in this iteration for each $i\in[m_c]$ and $j\in[m_a]$. First, define $\forall i\in[m_c],\ \hat{\beta}_i^k := \beta_i^{(m)} + \gamma_{\mathrm{O},\beta}F_i(\overline{\mathbf{p}}^{(m)})$. Then for any $i\in[m_c]$ and $\beta_i\in[0,H_\beta]$ we have:

$$\begin{aligned}
-\left(\beta_i^{(m)}-\beta_i\right)F_i\left(\overline{\mathbf{p}}^{(m)}\right) &= \frac{1}{\gamma_{\mathrm{O},\beta}}\left(\beta_i^{(m)}-\beta_i\right)\left(\beta_i^{(m)}-\hat{\beta}_i^{(m)}\right)\\
&= \frac{1}{2\gamma_{\mathrm{O},\beta}}\left(\left(\beta_i^{(m)}-\beta_i\right)^2 + \left(\beta_i^{(m)}-\hat{\beta}_i^{(m)}\right)^2 - \left(\hat{\beta}_i^{(m)}-\beta_i\right)^2\right)\\
&\le \frac{1}{2\gamma_{\mathrm{O},\beta}}\left(\left(\beta_i^{(m)}-\beta_i\right)^2 - \left(\beta_i^{(m+1)}-\beta_i\right)^2 + \gamma_{\mathrm{O},\beta}^2\right)\ .
\end{aligned} \tag{EC.29}$$

By averaging both hand sides of (EC.29) for $m\in[K_{\mathrm{O}}]$, rearranging the terms, and finally setting $\gamma_{\mathrm{O},\beta} = \frac{H_\beta}{\sqrt{K_{\mathrm{O}}}}$ we obtain the following inequality:

$$-\frac{1}{K_{\mathrm{O}}}\sum_{m\in[K_{\mathrm{O}}]}\left(\beta_i^{(m)}-\beta_i\right)F_i\left(\overline{\mathbf{p}}^{(m)}\right) \le \frac{1}{2K_{\mathrm{O}}\gamma_{\mathrm{O},\beta}}\left(\beta_i^{(1)}-\beta_i\right)^2 + \frac{\gamma_{\mathrm{O},\beta}}{2} \le \frac{H_\beta^2}{2K_{\mathrm{O}}\gamma_{\mathrm{O},\beta}} + \frac{\gamma_{\mathrm{O},\beta}}{2} = \frac{H_\beta}{\sqrt{K_{\mathrm{O}}}}\ . \tag{EC.30}$$

Similarly, by considering the update equation of $\lambda_j^{(m)}$ for any $j\in[m_a]$, following exactly the same lines as in the above argument, and finally setting $\gamma_{\mathrm{O},\lambda} = \frac{H_\lambda}{2\sqrt{K_{\mathrm{O}}}}$, for any $\lambda_j\in[0,H_\lambda]$ we have:

$$\frac{1}{K_{\mathrm{O}}}\sum_{m\in[K_{\mathrm{O}}]}\left(\lambda_j^{(m)}-\lambda_j\right)\left(b_j-\boldsymbol{\theta}_j\cdot\overline{\mathbf{p}}^{(m)}\right) \le \frac{1}{2K_{\mathrm{O}}\gamma_{\mathrm{O},\lambda}}\left(\lambda_j^{(1)}-\lambda_j\right)^2 + 2\gamma_{\mathrm{O},\lambda} \le \frac{H_\lambda^2}{2K_{\mathrm{O}}\gamma_{\mathrm{O},\lambda}} + 2\gamma_{\mathrm{O},\lambda} = \frac{2H_\lambda}{\sqrt{K_{\mathrm{O}}}}$$
(EC.31)

where in the first inequality we have used the assumption that $|b_j|\le 1$, and that for every $\mathbf{p}\in[0,H_p]^d$ we have $|\boldsymbol{\theta}_j\cdot\mathbf{p}|\le 1$. Now, $\forall i\in[m_c]$ denote by $\overline{\beta}_i$ the average of $\beta_i^{(m)}$ over all outer iterations, i.e., $\overline{\beta}_i = \frac{1}{K_{\mathrm{O}}}\sum_{m\in[K_{\mathrm{O}}]}\beta_i^{(m)}$, and $\forall j\in[m_a]$ denote by $\overline{\lambda}_j$ the average of $\lambda_j^{(m)}$ over all outer iterations, i.e., $\overline{\lambda}_j =$

$\frac{1}{K_O}\sum_{m\in[K_O]}\lambda_j^{(m)}$. Also, denote by $\overline{\mathbf{p}}$ the average of all vectors of expected visit numbers across all iterations of our algorithm, i.e., $\overline{\mathbf{p}}=\frac{1}{K_O}\sum_{m\in[K_O]}\overline{\mathbf{p}}^{(m)}\equiv\frac{1}{K_O K_I}\sum_{m\in[K_O]}\sum_{\ell\in[K_I]}\mathbf{p}_{\pi^{(m,\ell)}}$. Note that due to the convexity of F_i, $F_i(\overline{\mathbf{p}})\le\frac{1}{K_O}\sum_{m\in[K_O]}F_i(\overline{\mathbf{p}}^{(m)})$. Using this fact, and by summing up both hand sides of (EC.30) for $i\in[m_c]$, we obtain this inequality for any choice of $\beta_i\in[0,H_\beta]$ and for $i\in[m_c]$:

$$\sum_{i\in[m_c]}\beta_i F_i(\overline{\mathbf{p}})-\sum_{i\in[m_c]}\frac{1}{K_O}\sum_{m\in[K_O]}\beta_i^{(m)}F_i(\overline{\mathbf{p}}^{(m)})\le\frac{H_\beta m_c}{\sqrt{K_O}}\equiv\varepsilon_2\ .\quad\text{(DUAL-BEST-RESPONSE-OUTER-I)}$$

Similarly, by summing up both hand sides of (EC.31) for $j\in[m_a]$, we obtain the following inequality for any choice of $\lambda_j\in[0,H_\lambda]$ and for $j\in[m_a]$:

$$\sum_{j\in[m_a]}\frac{1}{K_O}\sum_{m\in[K_O]}\lambda_j^{(m)}\left(b_j-\boldsymbol{\theta}_j\cdot\overline{\mathbf{p}}^{(m)}\right)-\sum_{j\in[m_a]}\lambda_j\left(b_j-\boldsymbol{\theta}_j\cdot\overline{\mathbf{p}}\right)\le\frac{2H_\lambda m_a}{\sqrt{K_O}}\equiv\varepsilon_3\ .$$
(DUAL-BEST-RESPONSE-OUTER-II)

To put all the pieces together and obtain the final result, first note that at any iteration k, one can consider the assignment $\forall i\in[m_c],\ \boldsymbol{\mu}_i\leftarrow\nabla F_i(\overline{\mathbf{p}}^{(m)})\in[-H_\mu,H_\mu]^d$, in (INNER-APPROXIMATE-EQUILIBRIUM). Due to Lemma EC.8, $F_i(\overline{\mathbf{p}}^k)=\boldsymbol{\mu}_i\cdot\overline{\mathbf{p}}^{(m)}-F_i^*(\boldsymbol{\mu}_i)$, and therefore:

$$\overline{\mathcal{L}}_{\text{JMS-CONS}}\left(\overline{\mathbf{p}}^{(m)};\{\lambda_j^{(m)}\},\{\beta_i^{(m)}\},\{\nabla F_i(\overline{\mathbf{p}}^{(m)})\}\right)=\mathcal{L}_{\text{JMS-CONS}}\left(\overline{\mathbf{p}}^{(m)};\{\lambda_j^{(m)}\},\{\beta_i^{(m)}\}\right)\ .$$

Therefore, we obtain the following inequality

$$\forall m\in[K_O],\ \mathbf{p}\in\mathcal{P}:\ \ \overline{\mathcal{L}}_{\text{JMS-CONS}}\left(\mathbf{p};\{\lambda_j^{(m)}\},\{\beta_i^{(m)}\},\{\overline{\boldsymbol{\mu}}_i^{(m)}\}\right)-\mathcal{L}_{\text{JMS-CONS}}\left(\overline{\mathbf{p}}^{(m)};\{\lambda_j^{(m)}\},\{\beta_i^{(m)}\}\right)\le\varepsilon_1 \tag{EC.32}$$

Recall the definition of the optimal fair policy π^*_{CONS} in (OPT-JMS-CONS). Such a policy exists as the Markovian game instance $\{\mathcal{G}_i\}_{i\in[n]}$ is feasible (Assumption 2). Let $\text{OPT}_{\text{CONS}}\triangleq\mathbf{E}\left[R_{\pi^*_{\text{CONS}}}\right]=\mathbf{R}\cdot\mathbf{p}_{\pi^*_{\text{CONS}}}$. By setting $\mathbf{p}=\mathbf{p}_{\pi^*_{\text{CONS}}}$ in inequality (EC.32), and using the fact that $\overline{\mathcal{L}}_{\text{JMS-CONS}}$ is a relaxation of the optimal policy for any *feasible* choice of dual variables, i.e., $\boldsymbol{\lambda},\boldsymbol{\beta}\ge0$, and $\forall i\in[m_c]:\ \boldsymbol{\mu}_i\in\mathbb{R}^d$, we have:

$$\forall m\in[K_O]:\ \ \text{OPT}_{\text{CONS}}-\mathcal{L}_{\text{JMS-CONS}}\left(\overline{\mathbf{p}}^{(m)};\{\lambda_j^{(m)}\},\{\beta_i^{(m)}\}\right)\le\varepsilon_1 \tag{EC.33}$$

Now, by averaging over all iterations $m\in[K_O]$ in (EC.33), we obtain the following inequality:

$$\text{OPT}_{\text{CONS}}-\mathbf{R}\cdot\overline{\mathbf{p}}-\sum_{j\in[m_a]}\frac{1}{K_O}\sum_{m\in[K_O]}\lambda_j^{(m)}\left(b_j-\boldsymbol{\theta}_j\cdot\overline{\mathbf{p}}^{(m)}\right)+\sum_{i\in[m_c]}\frac{1}{K_O}\sum_{m\in[K_O]}\beta_i^{(m)}F_i(\overline{\mathbf{p}}^{(m)})\le\varepsilon_1 \tag{EC.34}$$

To conclude, we add up both hand sides of inequalities (EC.34), (DUAL-BEST-RESPONSE-OUTER-I), and (DUAL-BEST-RESPONSE-OUTER-II), so that we obtain the following final inequality (which holds for *any assignment* of $\lambda_j\in[0,H_\lambda],j\in[m_a]$ and $\beta_i\in[0,H_\beta],i\in[m_c]$):

$$\text{OPT}_{\text{CONS}}-\mathbf{R}\cdot\overline{\mathbf{p}}-\sum_{j\in[m_a]}\lambda_j\left(b_j-\boldsymbol{\theta}_j\cdot\overline{\mathbf{p}}\right)+\sum_{i\in[m_c]}\beta_i F_i(\overline{\mathbf{p}})\le\varepsilon_1+\varepsilon_2+\varepsilon_3\le\varepsilon\ , \tag{EC.35}$$

Now, by setting $\lambda_j = 0$ for all $j \in [m_a]$ and $\beta_i = 0$ for all $i \in [m_c]$, the expected reward objective of the G-RDIP policy $\hat{\pi}$ (returned by Algorithm 3) is bounded below by:

$$\mathbf{E}[R_{\hat{\pi}}] = \frac{1}{K_{\mathrm{O}} K_{\mathrm{I}}} \sum_{m \in [K_{\mathrm{O}}]} \sum_{\ell \in [K_{\mathrm{I}}]} \mathbf{R} \cdot \mathbf{p}_{\pi^{(m,\ell)}} = \mathbf{R} \cdot \overline{\mathbf{p}} \geq \mathrm{OPT}_{\mathrm{CONS}} - \varepsilon$$

At the same time, notice that $|\mathrm{OPT}_{\mathrm{CONS}} - \mathbf{R} \cdot \overline{\mathbf{p}}| \leq dH_p$. Hence, for all $j \in [m_a]$ we should have that $b_j - \boldsymbol{\theta}_j \cdot \overline{\mathbf{p}} \geq -\delta$, because if the converse holds for some j then we can set $\lambda_j = \frac{dH_p + \varepsilon}{\delta} \leq H_\lambda$ (and all other $\lambda_{j'}$'s and β_i's are set to zero), which violates (EC.35). Similarly, for all $i \in [m_c]$ we should have $F_i(\overline{\mathbf{p}}) \leq \delta$, because if the converse holds for some i then we can set $\beta_i = \frac{dH_p + \varepsilon}{\delta} \leq H_\beta$ (and all other $\beta_{i'}$'s and λ_j's are set to zero), which violates (EC.35). This completes the proof of the first part of the theorem.

Regarding running time, $K_I = \mathcal{O}(d^4)$ and $K_O = \mathcal{O}(d^2)$, and therefore the total number of iterations of our algorithm is $\mathcal{O}(K_I K_O) = \mathcal{O}(d^6)$. We note that our algorithm needs to solve a JMS instance for the primal player and compute the gradient for the dual player in each iteration. In general, solving an instance of JMS requires an extra polynomial-time computation. This extra computation depends on two factors: (i) the amount of time it takes to compute the indices for each individual arm—which is polynomial-time; if the Markov chain i has ℓ_i number of nodes/states, the running time is at most $\mathcal{O}(\ell_i^5)$ to solve for the indices using dynamic programming (Dumitriu et al., 2003) (note that $\sum_i \ell_i = d$). (ii) the amount of time it takes to compute the gradient of $\bar{\mathcal{L}}_{\text{JMS-CONST}}$, which requires computing the expected number of visits of different states under the optimal index-based policy (computed earlier). The latter quantity depends on the absorption time of the underlying kernels of the Markov chains, which is $\mathcal{O}(d)$, as it is assumed that the Markov chain is finite and absorbing, as there is a constant H_p such that the expected number of visits of each state before absorption is bounded above by H_p. We note that the eventual running time will be polynomial in d, n, $\frac{1}{\varepsilon}$ and $\frac{1}{\delta}$, as desired. □

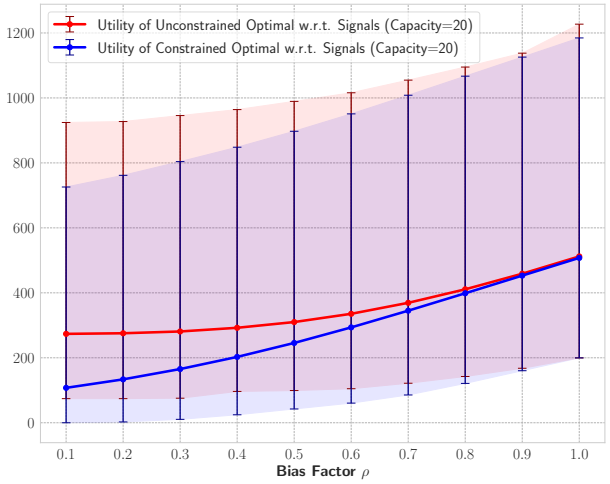
EC.10. Supplemental Numerical Simulations

In this section, we provide supplementary materials for the numerical study in Section 4. In particular, we include the following discussions and simulations:

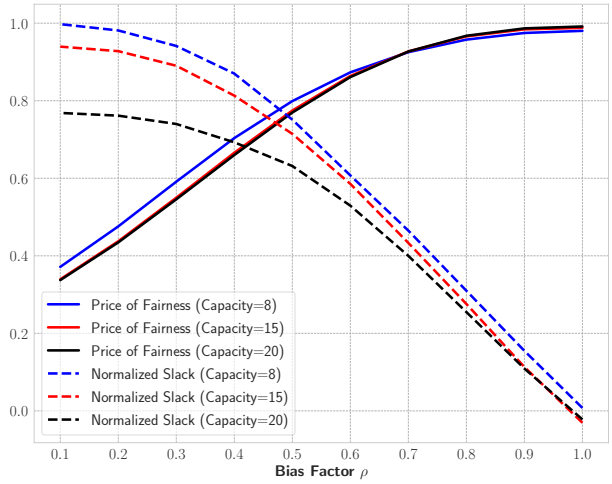
- Further discussions and simulations on unintended consequences of our socially-aware constraints (Section EC.10.1); additional experiments related to demographic parity in selection (Section EC.10.2); additional numerical results for the average quota constraint QUOTA (Section EC.10.3); the average budget for subsidization constraint BUDGET (Section EC.10.4); and several more detailed performance metrics (Section EC.10.5);
- Robustness checks of our numerical findings under uniform distributions for candidates' values (instead of normal), examining both short-term effects (Section EC.10.6.1) and long-term impacts (Section EC.10.6.2) under constraint PARITY, as well as under constraint QUOTA (Section EC.10.6.3);
- Experiments on the JMS model for Example 1 under multiple affine constraints, specifically imposing one constraint of type PARITY at each stage of the search process (Section EC.10.7).

EC.10.1. Unintended Consequences of Demographic Parity: a Dichotomy

Regardless of considering biased observable signals or unbiased unobservable true values to evaluate performances, in scenarios that the cost of search is high, we may observe a certain type of *unintended inefficiency* in the performance of optimal constrained policy for excessively small values of ρ. To see this, we consider the same setup as before but with the only difference being that we increase the inspection costs to be drawn independently from a uniform distribution over $[c_l = 25, c_h = 35]$ rather than $[c_l = 3, c_h = 6]$: in both Figure EC.3 and Figure EC.4 (analogs of Figure 4 and Figure 5, respectively), for sufficiently small values of ρ, say $\rho \in [0, 0.3]$, the expected utility of optimal constrained policy drops drastically as ρ becomes smaller, while the expected utility of optimal unconstrained policy remains almost unchanged.

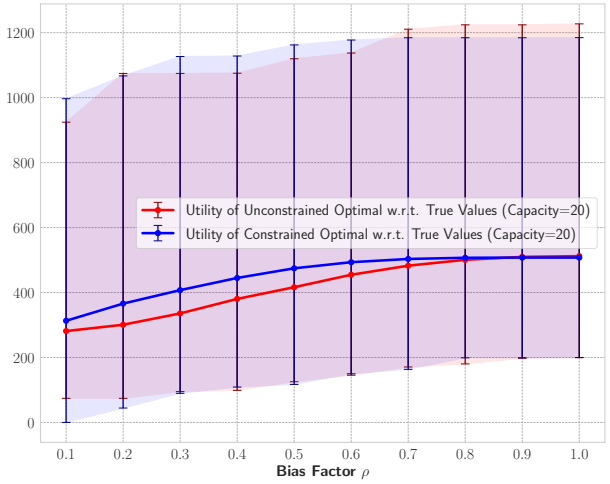


(a) Expected utilities calculated based on signals $\{v_i\}_{i\in[n]}$ for the unconstrained optimal policy (red) and the constrained optimal policy (blue).

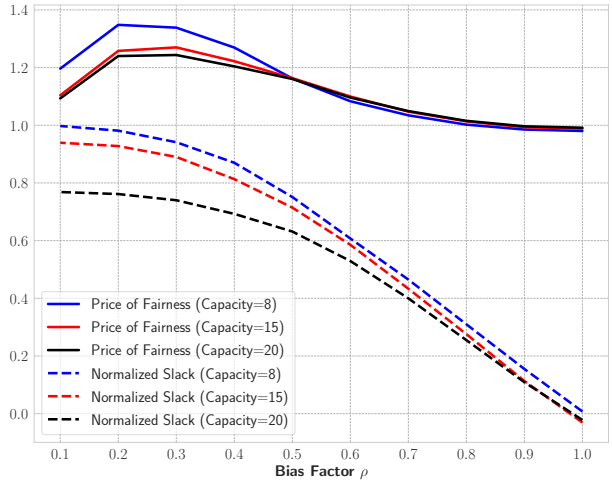


(b) Price of fairness ratio calculated based on signals $\{v_i\}_{i\in[n]}$ (solid lines) and the normalized constraint slack of unconstrained optimal policy (dashed lines).

Figure EC.3 Comparing the short-term outcomes of unconstrained and constrained optimal policies, when inspection costs are high.



(a) Expected utilities calculated based on true values $\{v_i^\dagger\}_{i\in[n]}$ for the unconstrained optimal policy (red) and the constrained optimal policy (blue).



(b) Price of fairness ratio calculated based on true values $\{v_i^\dagger\}_{i\in[n]}$ (solid lines) and the normalized constraint slack of unconstrained optimal policy (dashed lines).

Figure EC.4 Comparing the long-term outcomes of unconstrained and constrained optimal policies, when inspection costs are high.

Why would imposing the parity constraint have a different "calibrating effect" for bias factors 0.7 and 0.3? It turns out that the optimal constrained policy ends up not filling the entire capacity when inspection costs are high and ρ is small. This under-allocation is because the (observable and biased) signal distributions suggest that if demographic parity is enforced, it is less costly to leave the capacity unused than inspecting and then hiring "seemingly" low-quality high-cost candidates. In Figure EC.5, we plot the fraction of unallocated capacity by the optimal constrained policy as parameters ρ and k vary, which clearly shows the existence of this unintended effect for small values of ρ (and that it intensifies for larger values of k). Lastly, we note that this is in contrast with the behavior of the optimal unconstrained policy, for it continues to fill most of its capacity even if $\rho = 0.1$, as can be seen from its normalized slack in both Figure EC.3 and Figure EC.4.

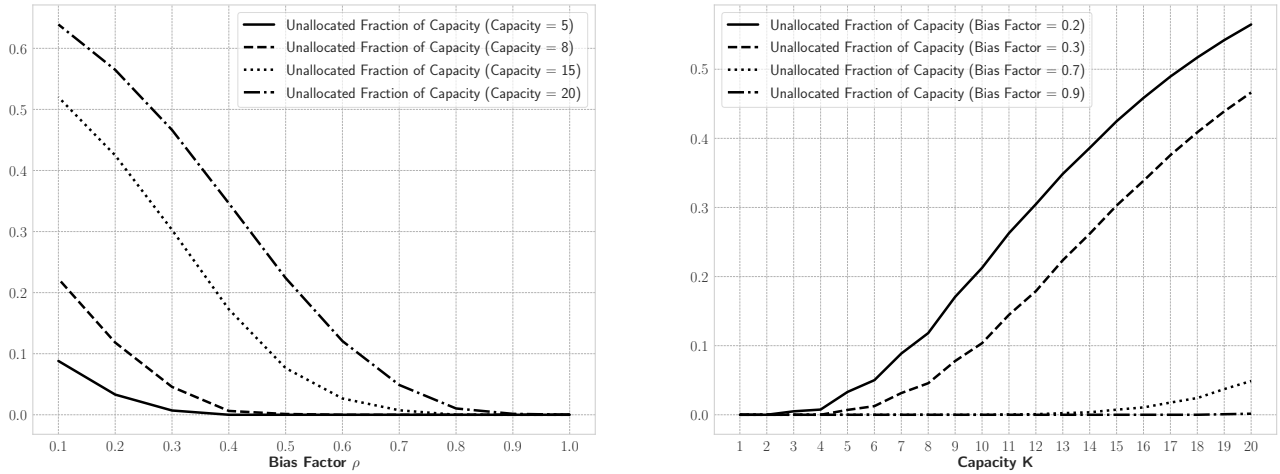


Figure EC.5 The unallocated fraction of the capacity by the optimal constrained policy for PARITY in selection.

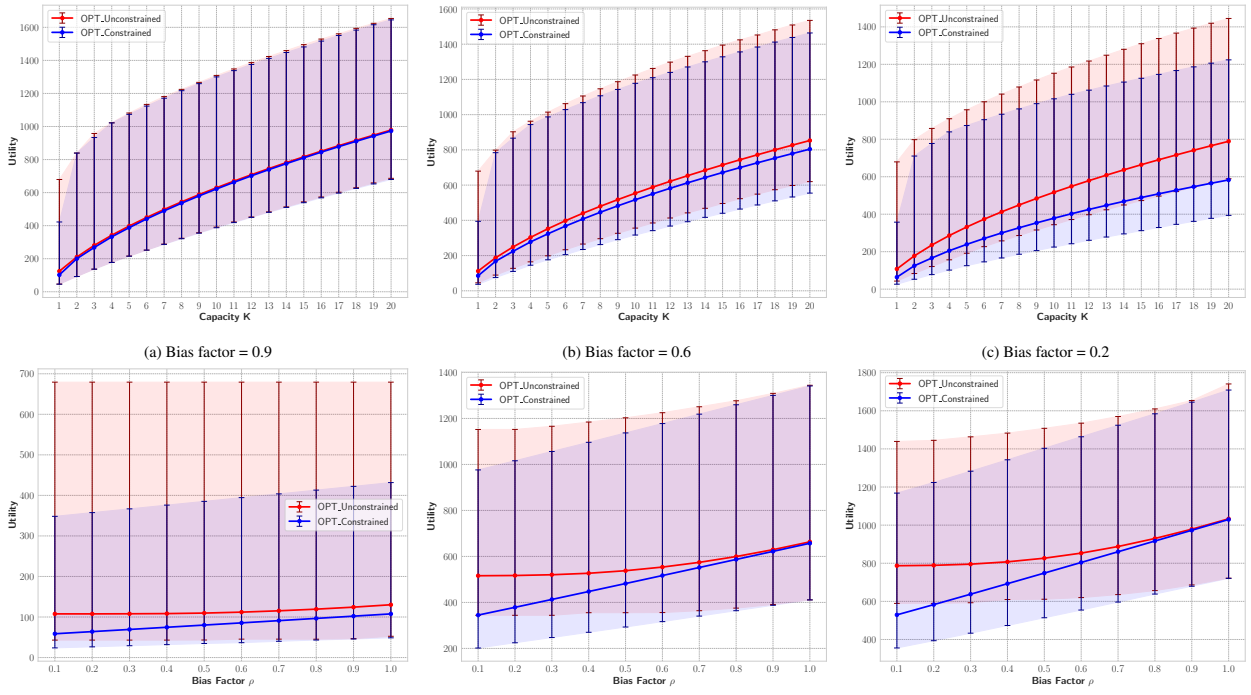
EC.10.2. Demographic parity in selection

We first study the effect of demographic parity in selection, that is, the constraint (PARITY) for selection. In Section EC.10.2.1, we consider the short-term effect of imposing demographic parity. To do so, we use observable signals (which are biased for the minority group $\mathcal{Y}$) as the only surrogate for true values and measure the utilities using these signals. We then consider the long-term effects of imposing demographic parity in Section EC.10.2.2, by measuring the utilities with respect to true values and not biased signals. Note that in our setting, the true values of $\mathcal{Y}$ are not statistically different from the true values of $\mathcal{X}$. In both settings, we compare the optimal unconstrained policy, that is, the solution to (OPT-UC), with the optimal constrained policy, that is, the solution to (OPT-CONS), where both policies have only access to observable biased signals upon inspection. Finally, we study the effect of increasing capacity on the "price of fairness" in Section EC.10.2.3. In Section EC.10.5, we also report the running time of these optimal policies; see Figure EC.20.

EC.10.2.1. Short-term performance – the effect of changing capacity and bias factor.

In our first scenario, we compare the two optimal policies as capacity $k \in [1:20]$ and bias factor $\rho \in \{0.1, 0.2, \cdots, 1\}$ vary. In Figure EC.6, we plot the net short-term utilities (calculated based on biased observable signals $\{v_i\}_{i\in[n]}$) as a function of k for different bias factors and as a function of ρ for different capacities. Furthermore, in Figure EC.7, we plot the ratio of the net utility of the optimal constrained policy over that of the

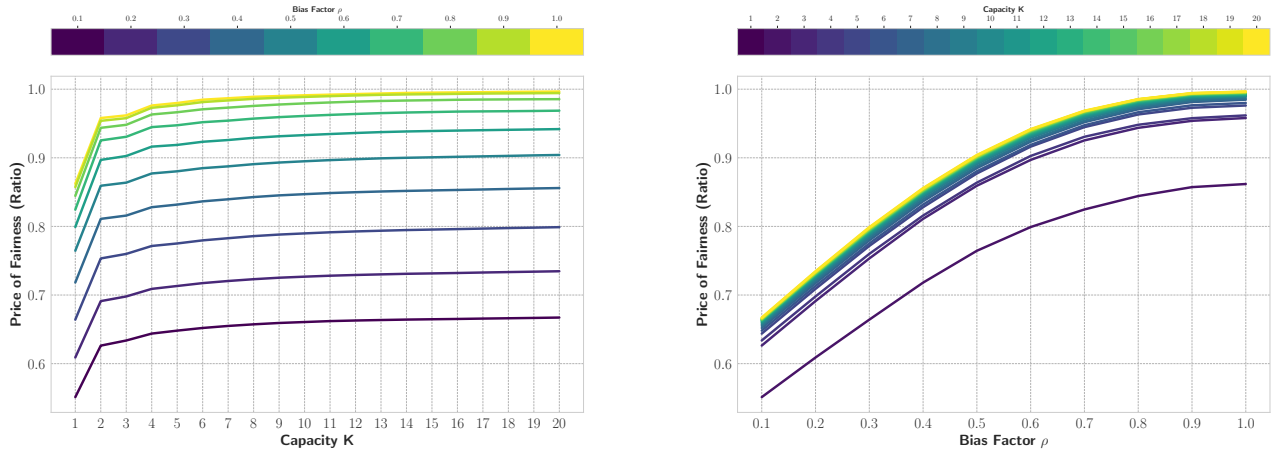
optimal unconstrained policy, again as a function of both k and ρ. To see a similar plot for utility differences, refer to Figure EC.21 in Section EC.10.5. Lastly, in Figure EC.8, we plot the constraint slack of the optimal unconstrained policy, as a function of k and also as a function of ρ. See Figure EC.22 in Section EC.10.5 for the graph of the dual adjustment λ^* required to fix the disparity (see Equation (8)), as a function of capacity k and bias factor ρ. Next, we discuss some managerial insights that are derived from these simulations.



(a) Bias factor = 0.9 (b) Bias factor = 0.6 (c) Bias factor = 0.2

(d) Capacity = 1 (e) Capacity = 10 (f) Capacity = 20

Figure EC.6 Comparing the short-term utilities of optimal unconstrained (red) and constrained (blue) policies for demographic parity in selection, for different capacities k and bias factors ρ.



(a) (b)

Figure EC.7 Short-term price of fairness of demographic parity in selection in terms of utility ratio; (a) as a function of capacity k, and (b) as a function of bias factor ρ.

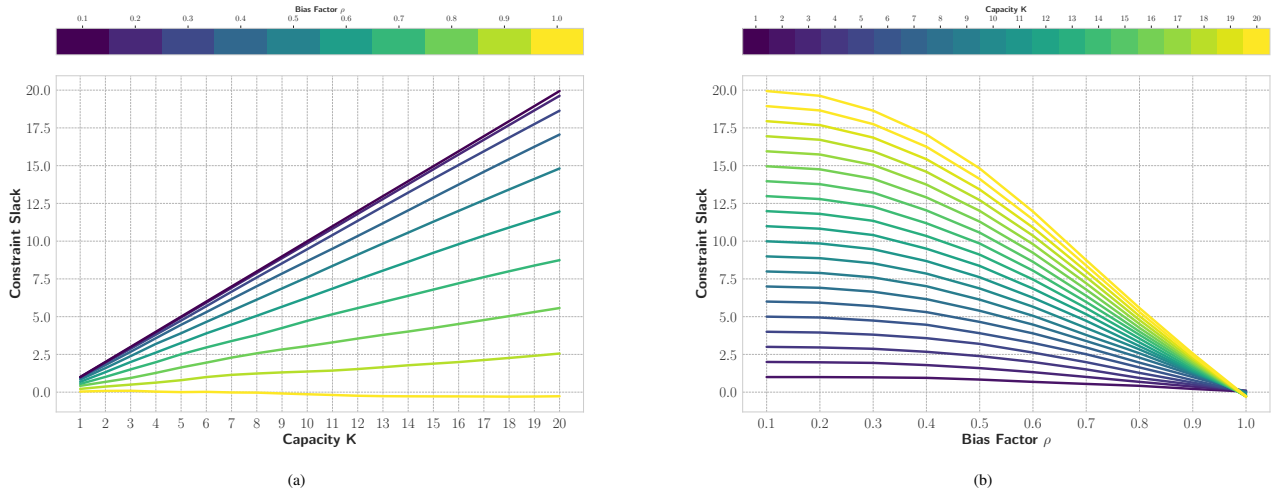


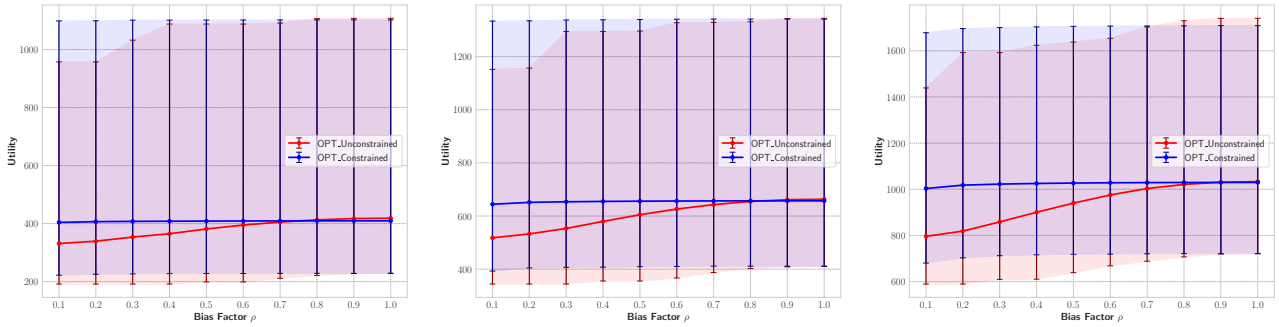
Figure EC.8 The constraint slack of the unconstrained optimal policy in demographic parity in selection; (a) as a function of capacity k, and (b) as a function of bias factor ρ.

First, clearly the short-term utility gap between the two policies increases as ρ decreases (which means more bias) and the optimal adjustment λ^* increases; nevertheless, for a moderate value of ρ, say $\rho \in [0.7, 1]$, the performance gap is quite small, while the constraint slack of the optimal unconstrained policy is still quite considerable. We have investigated this managerial insight in detail in Section 4.1.

Second, for large enough values of ρ, say $\rho \in [0.3, 1]$, the utility of both optimal policies increases as k increases; nevertheless, for small ρ, the performance of optimal constrained policy becomes constant after some $k \approx 6$ as it begins to suffer from a new form of inefficiency: due to the significant difference between the two groups, this policy decides *not to fill* its capacity to satisfy (PARITY). We have already investigated this source of inefficiency in detail in Section EC.10.1.

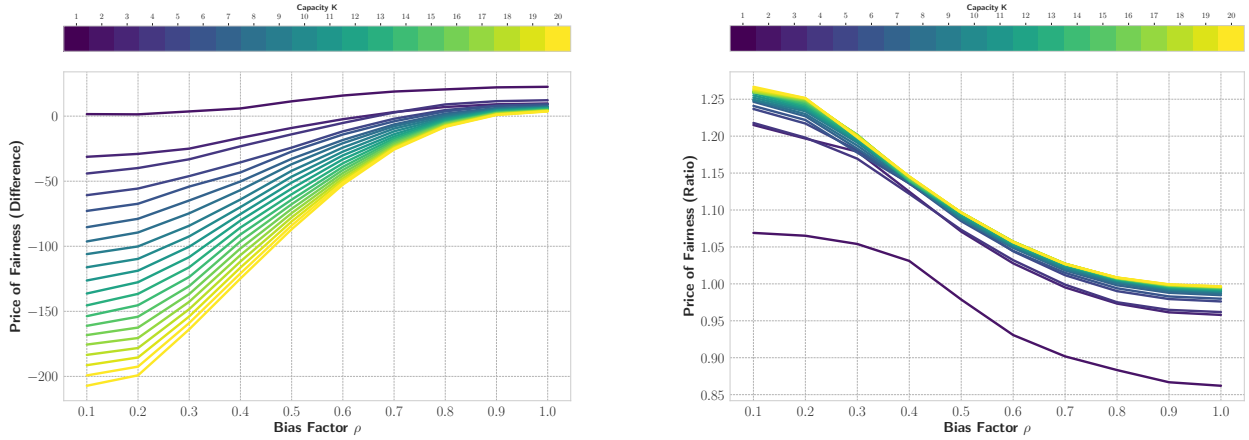
EC.10.2.2. Long-term performance – the effect of changing capacity and bias factor Now, we compare the expected long-term utilities of the optimal constrained and the optimal unconstrained policies in Figure EC.9, where the long-term utilities are calculated based on the true values. The price of fairness with respect to the true values, in terms of the ratio of optimal constrained to optimal unconstrained and also their difference, is reported in Figure EC.10. Interestingly, we observe that the true utility of the optimal constrained policy *dominates* that of the optimal unconstrained policy, as long as the bias factor is not very small (e.g., $\rho \geq 0.07$ for $k = 5$, $\rho \geq 0.18$ for $k = 10$ and $\rho \geq 0.28$ for $k = 20$). We have investigated this managerial insight in Section 4.2.

For small bias factors, even when the true values are unbiased, the constrained optimal policy might decide *not to fill* its capacity to satisfy (PARITY) – hence it suffers from a similar form of inefficiency as mentioned earlier. See more details in Section EC.10.1.



(a) Capacity = 5 (b) Capacity = 10 (c) Capacity = 20

Figure EC.9 Comparing the long-term utilities of optimal unconstrained (red) and constrained (blue) policies for demographic parity in selection, for different capacities k and bias factors ρ.



(a) Difference in utilities (b) Ratio of utilities

Figure EC.10 Long-term price of fairness of demographic parity in selection (with biased signals and unbiased true values) as a function of bias factor ρ for different capacities k in terms of (a) utility differences (b) utility ratios.

EC.10.2.3. A few positions more. Given the previous investigation, an intriguing question can be asked: how many additional units of capacity should be used to impose demographic parity in selection without any loss in short-term or long-term utility? To answer this question, in Figure EC.11, we plot the number of extra units k' used for each capacity k, so that the optimal constrained policy with capacity k has at least the same net utility as the optimal unconstrained policy with capacity $k-k'$. We study both settings: short-term utilities calculated using biased signals (part (a)) and long-term utilities calculated using unbiased true values (part (b)).

Interestingly, we observe in Figure EC.11(a) that for a moderate bias factor, say $\rho \approx 0.75$, around 23% extra capacity can ensure that demographic parity in selection would not harm short-term utility at all. This percentage decreases to less than 5% for $\rho = 0.9$ and increases to approximately 50% (with a sharp increase) when $\rho = 0.6$. This sharp increase, combined with the inefficiency caused by the unused capacity mentioned earlier in this section, suggests that when there is a significant bias in the signals of one of the groups, the decision maker might be better off focusing on more relaxed notions of fairness than (PARITY), for example

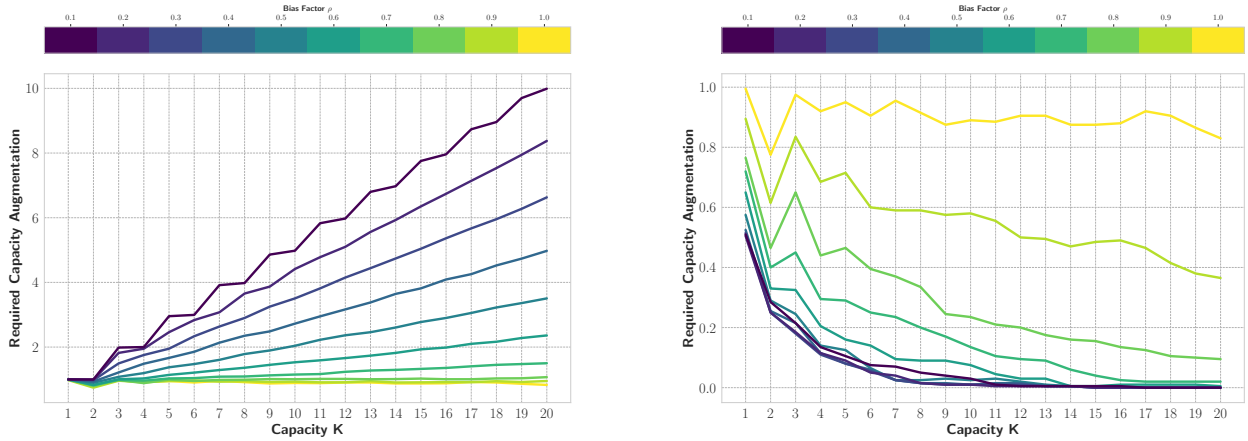


Figure EC.11 The required extra capacity as a function of capacity k for different bias factors ρ, to compensate for the (a) short-term utility reduction, and (b) long-term utility reduction.

(QUOTA) with $\theta \ll 0.5$. We further investigate this phenomenon in Section EC.10.3. See also Section EC.10.1 for a more in-depth discussion on how/why to adjust the quota parameter θ as a function of bias factor ρ.

Switching to the case of long-term utilities, which are calculated based on unbiased true values, the earlier observation that a few more positions can drastically help with the price of fairness becomes amplified: for a wide range of bias factors (e.g., $\rho \in [0.18, 1]$ for $k = 10$), the optimal constrained policy dominates the optimal unconstrained policy in terms of long-term utility. Furthermore, under significantly biased signals where this domination does not occur (e.g., $\rho = 0.1$ or $\rho = 0.2$), increasing k by a small amount goes a long way: Figure EC.11(b) suggests that increasing the capacity by 11% for $\rho = 0.2$ and by 38% for $\rho = 0.1$ increases the true utility of the optimal constrained policy to more than that of the optimal unconstrained policy. We investigate how these percentages change as we switch to more relaxed notions of fairness, for example, (QUOTA) with $\theta \ll 0.5$., in Section EC.10.3.

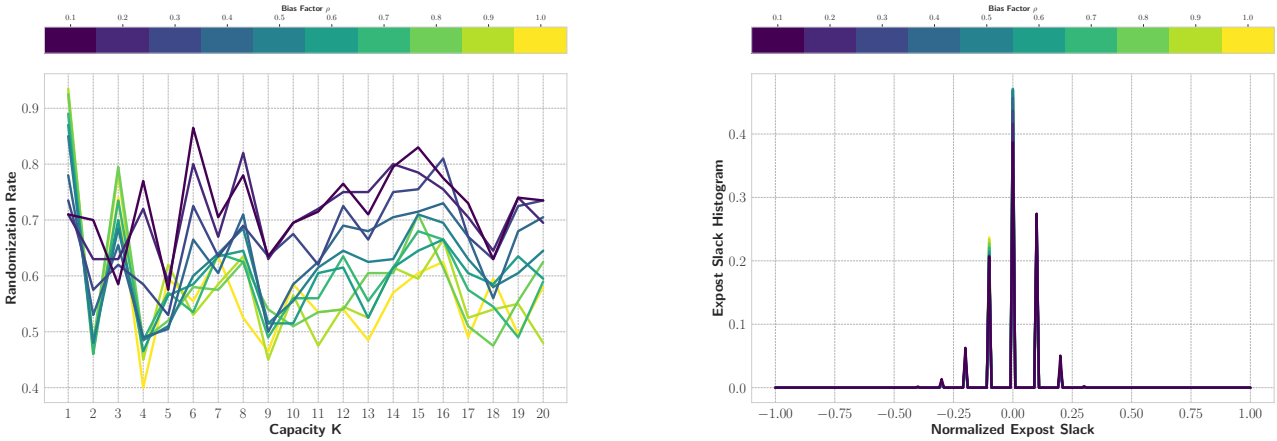
EC.10.2.4. Additional Notes. Figure EC.12a emphasizes on the significance of randomization, by showing that the optimal policy does, indeed, randomize over 2 extreme tie-breaking rules in majority of the instances. Figure EC.12b demonstrates the histogram of the normalized ex-post slack in the constraint, measured by the formula:

$$\frac{\sum_{i\in\mathcal{X}} \mathbb{A}_i^\pi - \sum_{i\in\mathcal{Y}} \mathbb{A}_i^\pi}{\sum_{i\in\mathcal{X}} \mathbb{A}_i^\pi + \sum_{i\in\mathcal{Y}} \mathbb{A}_i^\pi}.$$

As can be seen in the plot, there is a fast decay in the tail of the distribution. This implies, even though our optimal policy is designed to only satisfy the ex-ante constraint, its ex-post slack is also very close to 0 in most of the practical instances.

EC.10.3. Average quota in selection

Next, we study the average quota constraint in selection, that is, (QUOTA) for selection with parameter $\theta \in [0, 1]$. Importantly, $\theta = 0.5$ corresponds to demographic parity, while $\theta \in [0, 0.5)$ (resp., $\theta \in (0.5, 1]$) is more



(a) Fraction of instances that randomization was needed in the optimal policy.

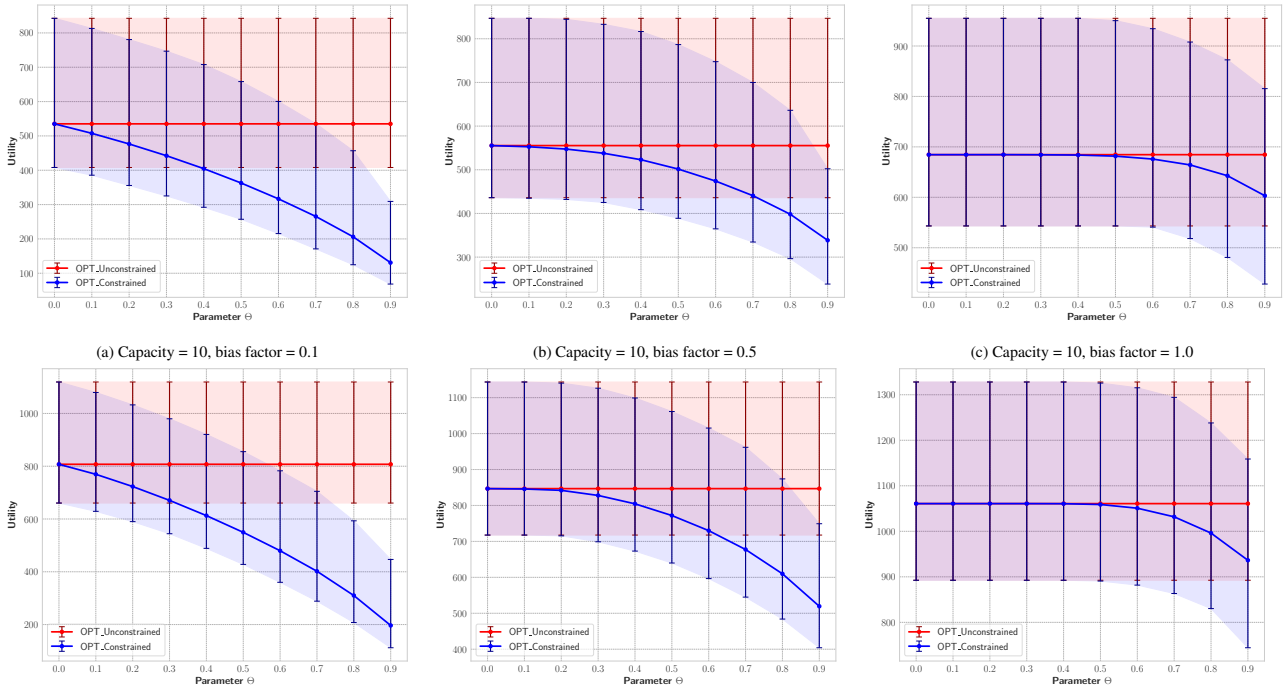
(b) Histogram of normalized ex-post slack, Capacity = 20

Figure EC.12 Measuring the necessity for randomness and the ex-post statistics of the constraint slack. (a) fraction of instances where randomization was employed, and (b) empirical distribution of ex-post slack.

relaxed (resp. more restricting) than demographic parity. We repeat the same simulation scenarios as before in Section EC.10.3.1, Section EC.10.3.2, and Section EC.10.3.3.

EC.10.3.1. Short-term performance – the effect of changing capacity and bias factor. In Figure EC.13, we plot short-term utilities (calculated based on biased observable signals) as a function of the quota parameter θ for different values of capacity k and bias factor ρ. In Section EC.10.5, we also plot the short-term price of fairness ratio (Figure EC.23) and the optimal dual adjustment λ^* (Figure EC.24) as a function of θ. First, we observe that the short-term utility gap between optimal constrained and unconstrained policies is increasing in θ, as expected. However, we also observe that for smaller values of bias factor, for example $\rho \in [0, 0.3]$, the utility decreases dramatically as θ increases. This observation suggests that when there is a significant asymmetry between the two groups, a smaller choice of $\theta \ll 0.5$ is a better choice from the perspective of short-term utility. On the other hand, for higher values of ρ, higher values of θ are admissible to obtain the same short-term price of fairness. See Section EC.10.1 for more details on the choice of θ as a function of ρ to mitigate the unintended under-allocations mentioned earlier.

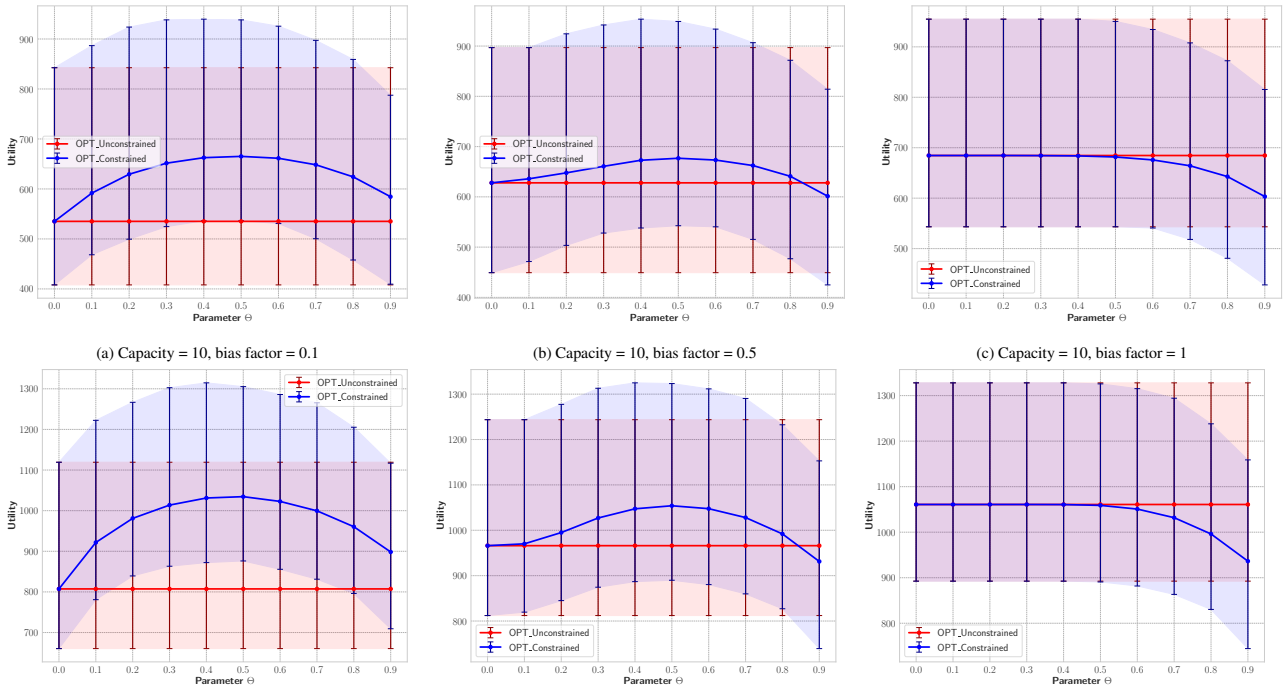
EC.10.3.2. Long-term performance – the effect of changing capacity and bias factor. We now consider a setting similar to Section EC.10.2.2 with biased signals and unbiased values and study the long-term performance of our policies. See Figure EC.14 for a comparison of long-term utilities (calculated based on the true values) under optimal constrained and unconstrained policies. In Section EC.10.5, we further plot the price of fairness ratio with respect to the true values as a function of θ (Figure EC.25). As before, the optimal constrained policy dominates the optimal unconstrained policy with respect to the true values in a wide range of parameters. Moreover, as can be seen in all these graphs, if the bias in the signals decreases (that is, the bias factor ρ increases), the range of parameter θ in which the domination occurs expands. Our results suggest that (i) adding an average quota with parameter $\theta \in [0, 1]$, similar to demographic parity, can help increase long-term utilities (with respect to true values), and (ii) tuning parameter θ based on the bias in the signals can drastically amplify this effect. We further discuss this managerial insight in Section EC.10.1.



(a) Capacity = 10, bias factor = 0.1 (b) Capacity = 10, bias factor = 0.5 (c) Capacity = 10, bias factor = 1.0

(d) Capacity = 20, bias factor = 0.1 (e) Capacity = 20, bias factor = 0.5 (f) Capacity = 20, bias factor = 1.0

Figure EC.13 Comparing the short-term utilities of optimal unconstrained (red) and constrained (blue) policies for the average quota in selection, for different capacities $k \in [1:20]$ and bias factors $\rho \in \{0.1, \cdots, 1\}$; both curves are functions of the quota parameter $\theta \in [0,1]$ in (QUOTA) ($\theta = 0.5$ corresponds to demographic parity).

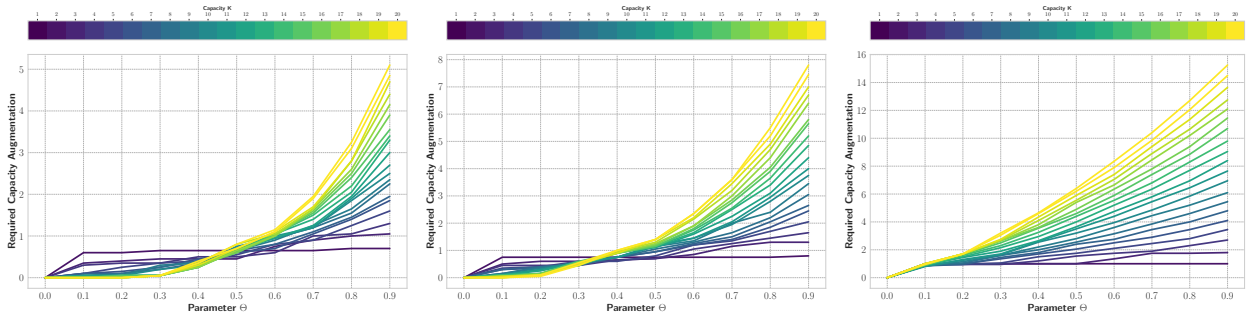


(a) Capacity = 10, bias factor = 0.1 (b) Capacity = 10, bias factor = 0.5 (c) Capacity = 10, bias factor = 1

(d) Capacity = 20, bias factor = 0.1 (e) Capacity = 20, bias factor = 0.5 (f) Capacity = 20, bias factor = 1

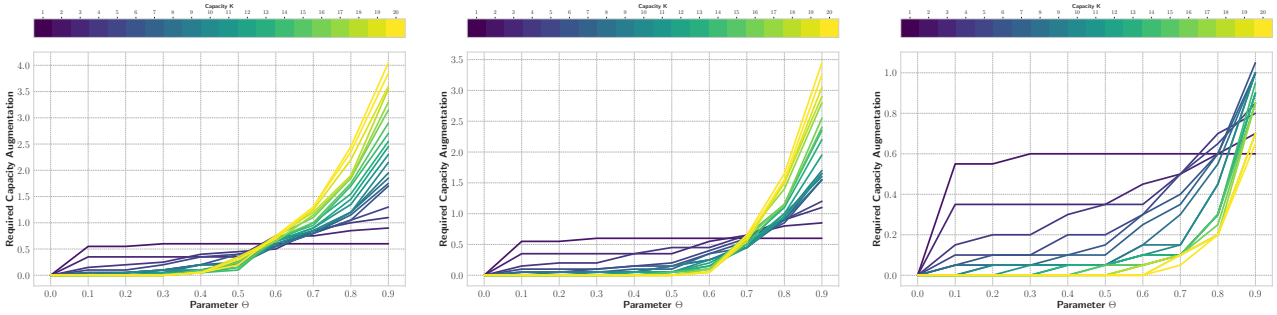
Figure EC.14 Comparing the long-term utilities of optimal unconstrained (red) and constrained (blue) policies for the average quota in selection, for different capacities $k \in [1:20]$ and bias factors $\rho \in \{0.1, \cdots, 1\}$; both curves are functions of the quota parameter $\theta \in [0,1]$ in (QUOTA) ($\theta = 0.5$ corresponds to demographic parity).

EC.10.3.3. A few positions more. We now consider the setting in Section EC.10.2.3, but this time considering the quota selection constraint as in (QUOTA), and study the effect of enhancing the capacity. See Figure EC.15 for the effect on short-term utilities and Figure EC.16 for the effect on long-term utilities. Comparing these two graphs with Figure EC.11, we observe that (i) for small values of θ (e.g., $\theta \leq 0.5$ for $\rho = 0.9$ and $\theta \leq 0.3$ for $\rho = 0.7$), increasing the capacity by less than 5% is enough to ensure that the optimal constrained policy dominates the optimal unconstrained policy in terms of short-term utilities; (ii) the impact of capacity enhancement increases drastically when measuring the performance of policies based on their long-term utilities. For example, with $\theta \leq 0.7$ for $\rho = 0.7$ and $\theta \leq 0.5$ for $\rho = 0.9$, increasing the capacity by less than 2.5% is enough to ensure that the long-term utility of the optimal constrained policy dominates the long-term utlitiy of the optimal unconstrained policy.



(a) Bias factor = 0.9 (b) Bias factor = 0.7 (c) Bias factor = 0.3

Figure EC.15 The required extra capacity to compensate for the short-term utility reduction due to imposing the average quota in selection, as a function of the quota parameter $\theta \in [0, 1]$ in (QUOTA) ($\theta = 0.5$ corresponds to demographic parity) for different values of bias factor ρ and capacity k.



(a) Bias factor = 0.9 (b) Bias factor = 0.7 (c) Bias factor = 0.3

Figure EC.16 The required extra capacity to compensate for the long-term utility reduction due to imposing the average quota in selection, as a function of the quota parameter $\theta \in [0, 1]$ in (QUOTA) ($\theta = 0.5$ corresponds to demographic parity) for different values of bias factor ρ and capacity k.

EC.10.3.4. Additional Notes. Similar to Section EC.10.2.4 we plot the normalized ex-post slack in Figure EC.17, measured by the formula:

$$\frac{\theta(\sum_{i\in\mathcal{X}}\mathbb{A}_i^\pi)-(1-\theta)(\sum_{i\in\mathcal{Y}}\mathbb{A}_i^\pi)}{\theta(\sum_{i\in\mathcal{X}}\mathbb{A}_i^\pi)+(1-\theta)(\sum_{i\in\mathcal{Y}}\mathbb{A}_i^\pi)}.$$

As can be seen, for each θ, the distribution has a low variance. Additionally, the reason the histogram of some of the smaller θ values are skewed toward left is due to the fact that the quota constraint becomes less binding as we decrease θ, and the optimal unconstrained policy itself may be a feasible policy.

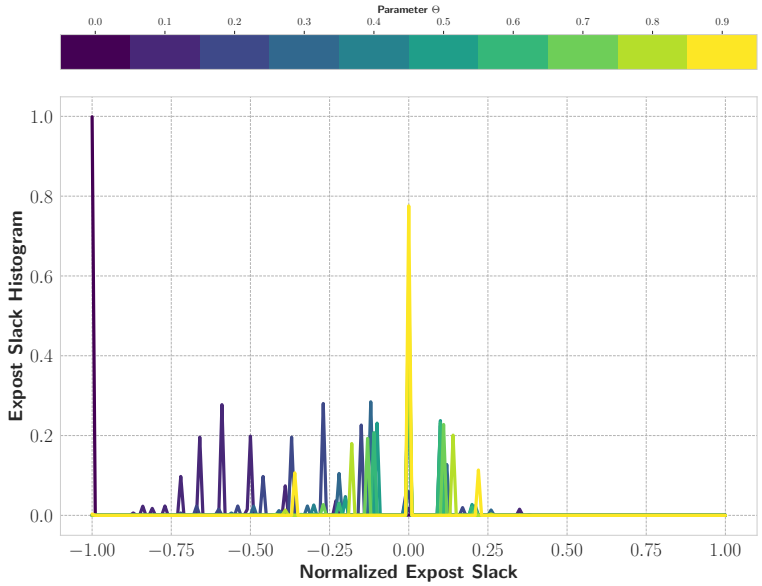


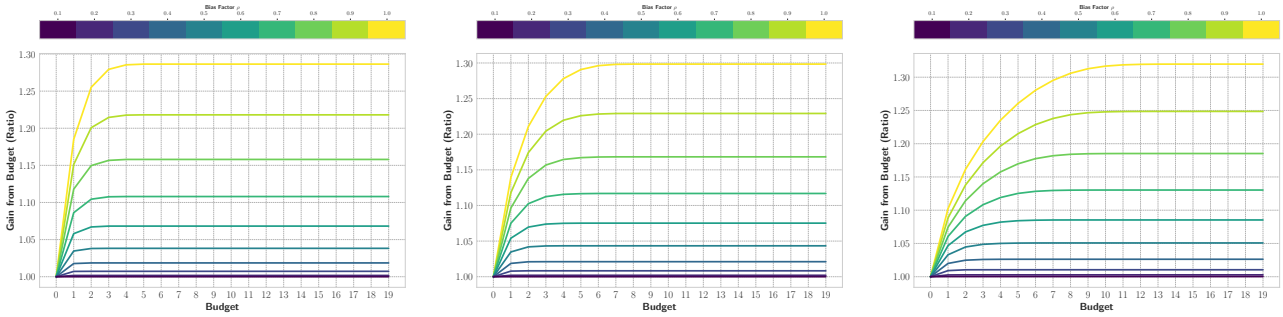
Figure EC.17 Histogram of normalized ex-post slack, Capacity = 20, Bias Factor = 0.7

EC.10.4. Average budget for subsidization

We finally study the effect of the average budget to subsidize the hiring expenses of underprivileged applicants. To model this, we consider two groups of candidates $\mathcal{Y}$ and $\mathcal{X}$. Given an average budget b, we consider a variant of (BUDGET) in selection where the decision maker has to select no more than b candidates from $\mathcal{Y}$ in expectation, while it has to satisfy an overall ex-post capacity constraint k among all individuals.

Clearly, any non-zero budget increases the search utility. Therefore, we define "gain from budget" as the ratio of the utility of the optimal constrained policy with the budget b to that of the optimal policy without any budget — which means that it cannot select anyone from group $\mathcal{Y}$. See Figure EC.18 to learn how the gain from the budget increases as a function of b, for various parameter choices for bias factor ρ and capacity k. In Section EC.10.5, we further compare the utility of the optimal constrained policy with budget b versus that of the optimal policy with unlimited budget (Figure EC.26) as well as the optimal dual adjustment λ^* (Figure EC.27), both as a function of the budget b.

Given the above setup, we can also ask an intriguing question. How valuable is the average budget? In particular, starting from the initial total capacity k and the average budget b for hiring from the group $\mathcal{Y}$, how



(a) Capacity = 5 (b) Capacity = 10 (c) Capacity = 20

Figure EC.18 The the gain from budget, i.e., the ratio of utilities of optimal constrained policy with a given average budget to the optimal policy with no budget, for the average budget in selection subsidization (with biased values) problem, for different capacities k and bias factors ρ, a function of average budget b on excepted selections from group $\mathcal{Y}$.

do we compare the utility gain from employing one additional unit of capacity with the utility gain derived from the increase of the average budget by one? In Table EC.1 we try to answer this question. Each entry corresponds to a pair (k, b), where k is the capacity and b is the average budget under the current system. The number written in each entry is the gain from a unit increase in the budget minus the gain from a unit increase the capacity, where gain is defined in terms of the ratio of utilities of optimal budget-constrained to optimal without budget. Our results suggest that (i) for small values of the current budget, the value of one extra unit of the average budget is considerably more than one additional unit of capacity, e.g., see the column corresponding to $b = 0$ or $b = 1$; (ii) the extra gain for each unit of budget is decreasing in b. Combining these two observations, we conclude that a little bit of average budget can go a long way — not only does it help with more representation from the underprivileged group $\mathcal{Y}$, but also it allows selections from (potentially top) members of group $\mathcal{Y}$ and leading to increasing the overall efficiency.

k \ b	0	1	2	3	4
1	0.291	0.040	0.003	0.003	0.003
2	0.251	0.067	0.009	0.001	0.001
3	0.221	0.079	0.021	0.004	0.001
4	0.201	0.081	0.028	0.007	0.001
5	0.185	0.082	0.034	0.010	0.002
6	0.173	0.082	0.039	0.015	0.003
7	0.163	0.082	0.043	0.019	0.005
8	0.154	0.081	0.047	0.024	0.009
9	0.146	0.079	0.049	0.028	0.013
10	0.139	0.077	0.049	0.031	0.016

Table EC.1 Gain of extra budget - Gain of extra capacity in terms of price of fairness.

EC.10.4.1. Additional Notes. Similar to Section EC.10.2.4 we plot the normalized ex-post slack in Figure EC.19, measured by the formula:

$$\frac{\sum_{i\in\mathcal{Y}} \mathbb{A}_i^\pi - b}{\sum_{i\in\mathcal{Y}} \mathbb{A}_i^\pi + b}.$$

As can be seen, for each budget b, the distribution has a low variance. Additionally, the reason the histogram of some of the larger budgets b are skewed toward left is due to the fact that the budget constraint becomes less binding as we increase the budget b, and the optimal unconstrained policy itself may be a feasible policy.

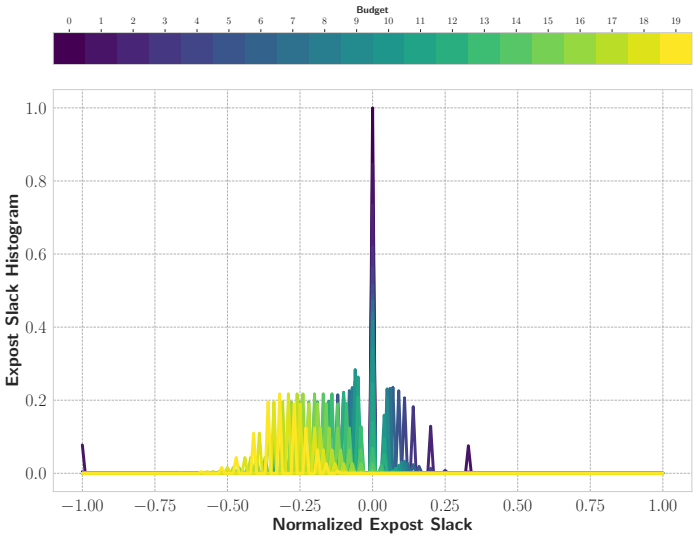
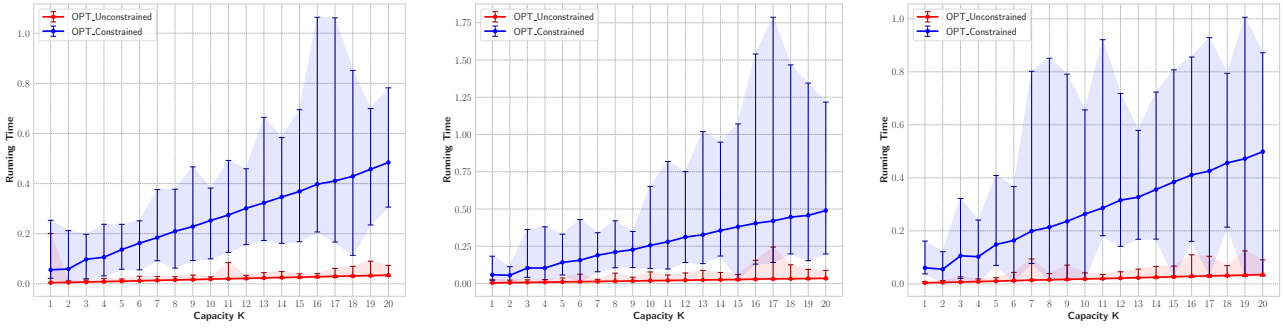


Figure EC.19 Histogram of normalized ex-post slack, Capacity = 20, Bias Factor = 1.0

EC.10.5. Missing Figures and Discussions of Section EC.10

We provide all the missing figures in our numerical simulations for Pandora's box under normal values here.



(a) Bias Factor = 0.3 (b) Bias Factor = 0.7 (c) Bias Factor = 1.0

Figure EC.20 Comparison of running times of optimal constrained and unconstrained policies (in seconds).

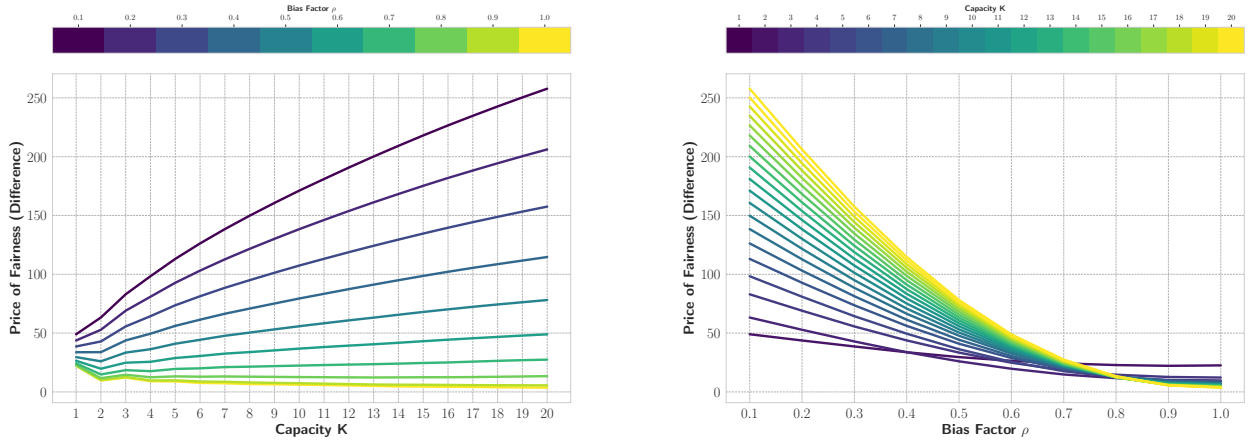


Figure EC.21 **Short-term price of fairness of demographic parity in selection in terms of utility differences; (a) as a function of capacity k, and (b) as a function of bias factor ρ.**

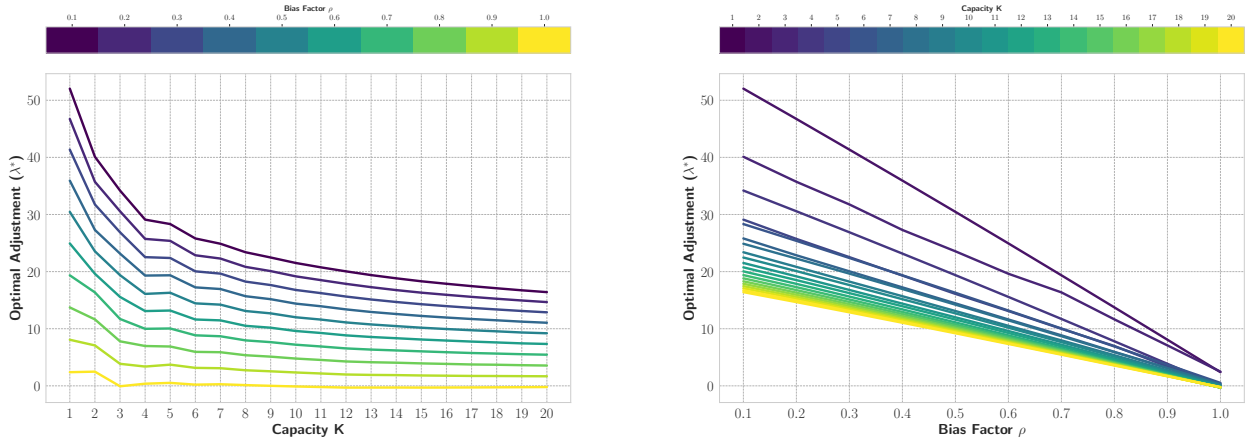


Figure EC.22 **The optimal dual adjustment λ^* in demographic parity in selection; (a) as a function of capacity k, and (b) as a function of bias factors ρ.**

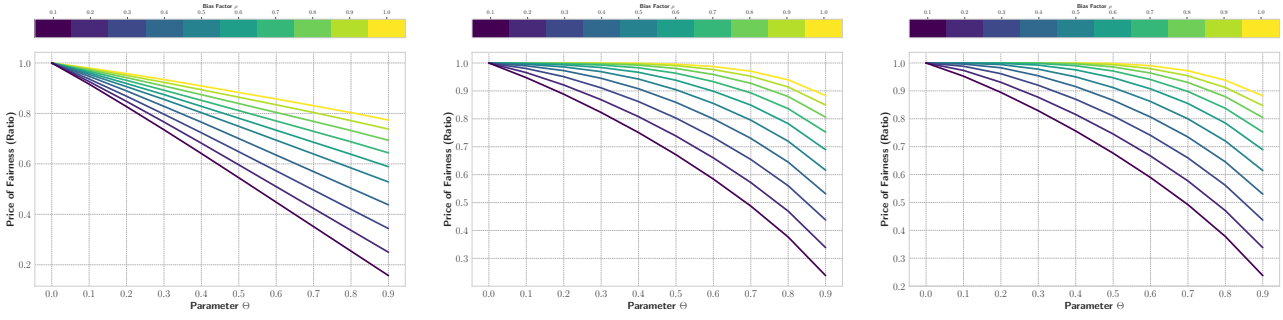
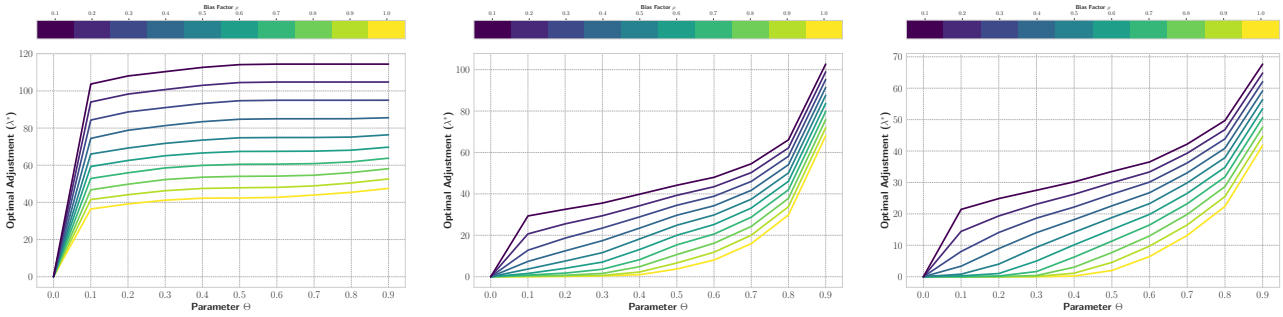
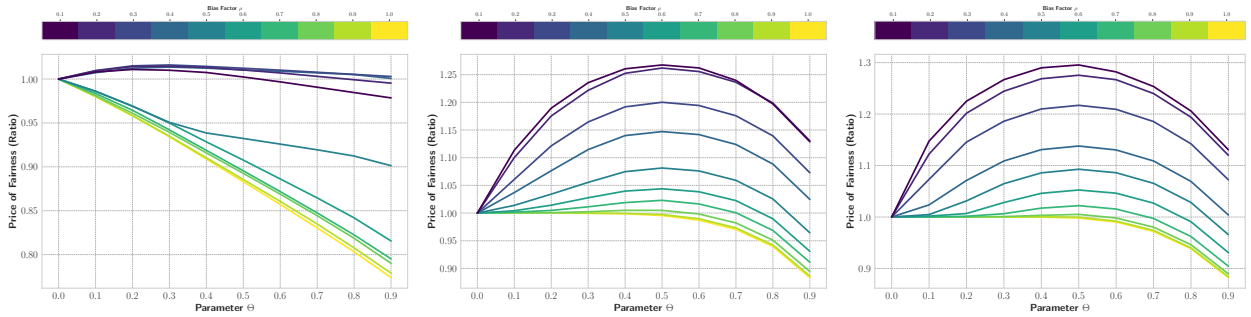


Figure EC.23 **Short-term price of fairness of average quota in selection in terms of utility ratio, as a function of quota parameter $\theta \in [0, 1]$ in (QUOTA) ($\theta = 0.5$ corresponds to demographic parity).**



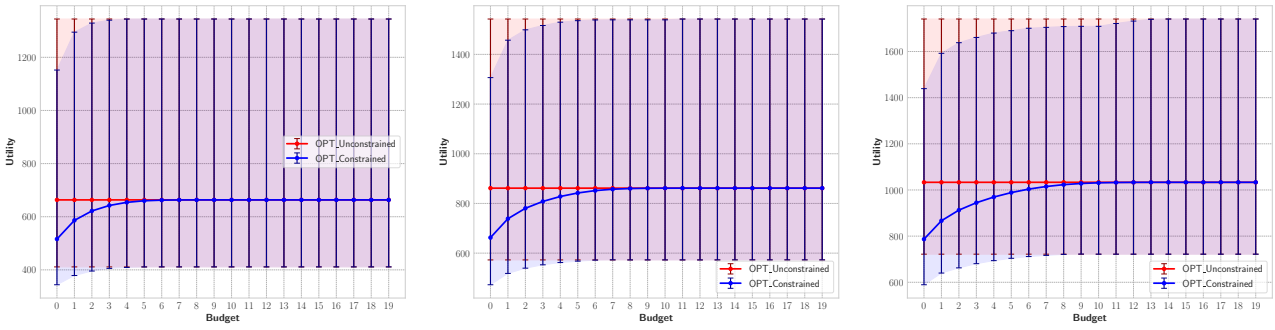
(a) Capacity = 1 (b) Capacity = 10 (c) Capacity = 20

Figure EC.24 The optimal dual adjustment λ^* for quota in selection constraint, as a function of quota parameter $\theta \in [0,1]$ in (QUOTA) ($\theta = 0.5$ corresponds to demographic parity).



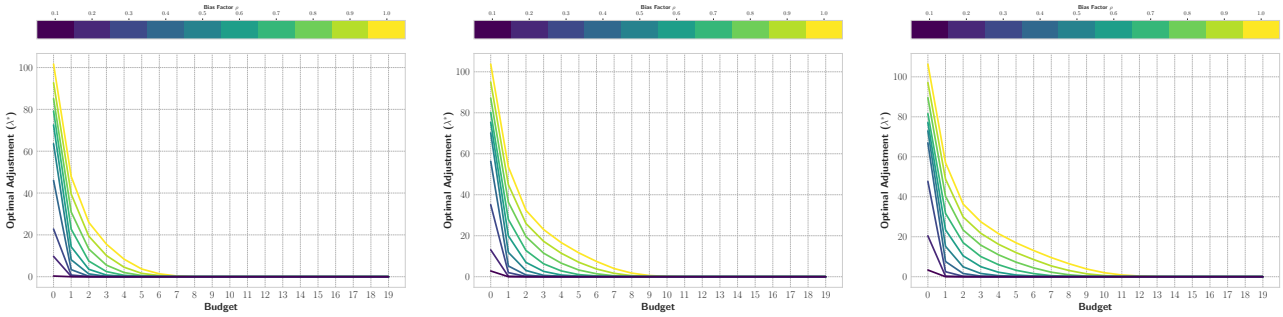
(a) Capacity = 1 (b) Capacity = 10 (c) Capacity = 20

Figure EC.25 Long-term price of fairness of average quota in selection in terms of utility ratio, as a function of quota parameter $\theta \in [0,1]$ in (QUOTA) ($\theta = 0.5$ corresponds to demographic parity).



(a) Capacity = 10, bias factor = 1.0 (b) Capacity = 15, bias factor = 1.0 (c) Capacity = 20, bias factor = 1.0

Figure EC.26 Comparing the utilities of optimal unconstrained (red) and optimal constrained (red) policies for the average budget in selection subsidization, for different capacities k, as a function of average budget b on excepted selections from group $\mathcal{Y}$ (red curve corresponds to unlimited budget).



(a) Capacity = 10 (b) Capacity = 15 (c) Capacity = 20

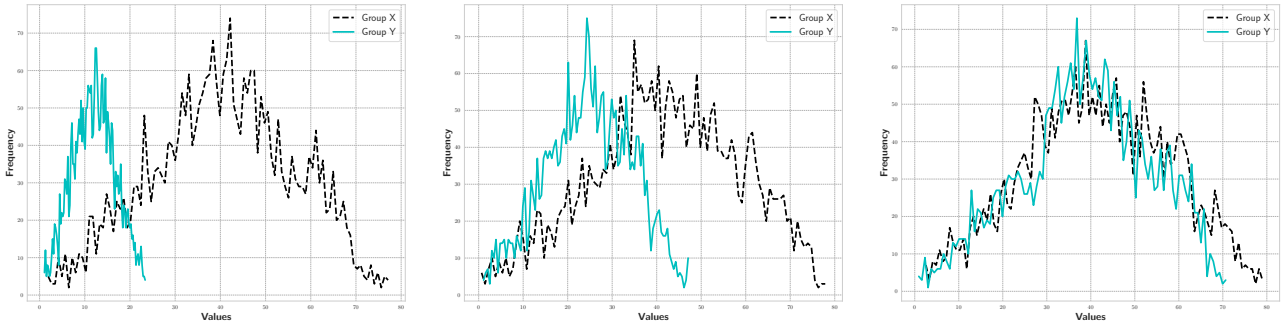
Figure EC.27 The optimal dual adjustment λ^* for the average budget in selection subsidization, for different capacities k , as a function of average budget b on excepted number of selections from group $\mathcal{Y}$.

EC.10.6. Numerical Simulations - Uniform Values

In this section we will demonstrate the robustness of the insights derived from our numerical simulations before, by conducting the same set of experiments but now under a different set of instances. As can be seen from the following figures, all of our previous managerial insights continue to hold, with some small changes to the exact numbers.

Basic simulation setup: The setup and structure of the problem is mainly as same as that of Section 4, except for the value distributions for the alternatives. In particular, this time we generate the mean values of the alternatives independently from a $\mathrm{Uniform}[20, 60]$ distribution (as opposed to the $\mathrm{LogNormal}$ distribution before), and then add an independent $\mathrm{Uniform}[-20, 20]$ noise on top of its mean (as opposed to the Gaussian noise before) to construct the value distribution for the corresponding alternatives. However, the rest of the setup, including the cost parameters, remain the same as before.

In Figure EC.28, you can observe the histogram of the generated values across all individuals, under three different bias levels.

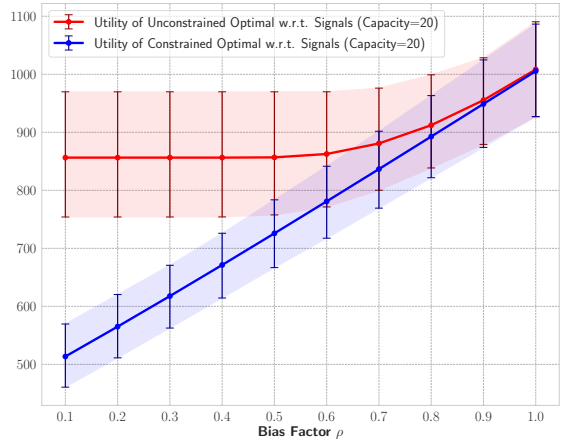


(a) Bias factor = 0.3 (b) Bias factor = 0.6 (c) Bias factor = 0.9

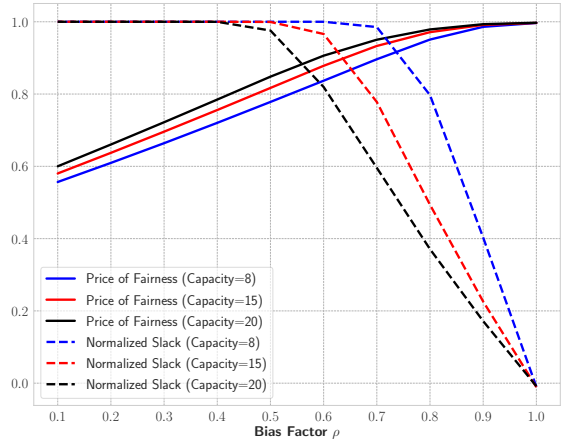
Figure EC.28 Sample histograms of the generated values $\{v_i\}_{i\in[1:60]}$ for the groups $\mathcal{Y}$ (cyan) and $\mathcal{X}$ (black).

In the remainder of this part, we list the results of all the new simulations for this new instance. We encourage the reader to compare these results with Section 4 and Section EC.10. As can be seen, while the resulting curves are (obviously) slightly different, there is no qualitative difference between these simulations and our previous set of simulations, indicating the robustness of our numerical results.

EC.10.6.1. Short-term outcomes: (Surprisingly) small utilitarian loss Figure EC.29a illustrates the short-term utilities of both optimal unconstrained policy and our proposed optimal constrained policy. Moreover, Figure EC.29b shows the price of fairness together with the normalized slack of the optimal unconstrained policy.



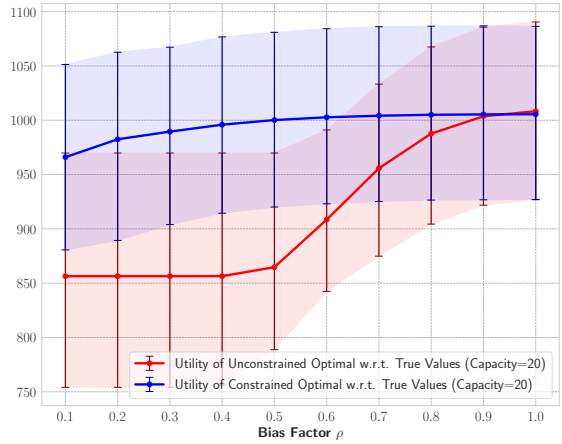
(a) Expected utilities calculated based on signals $\{v_i\}_{i\in[n]}$ for the unconstrained optimal policy (red) and the constrained optimal policy (blue).



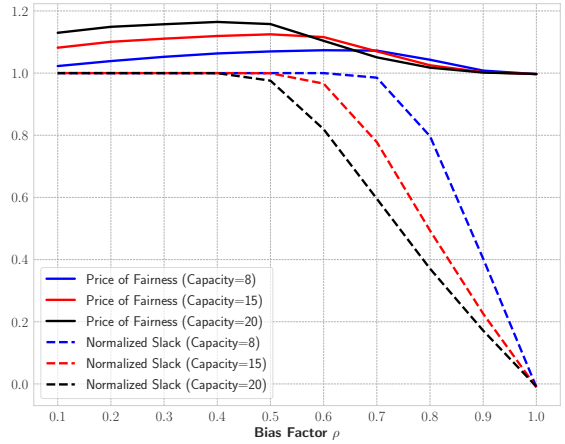
(b) Price of fairness ratio calculated based on signals $\{v_i\}_{i\in[n]}$ (solid lines) and the normalized constraint slack of unconstrained optimal policy (dashed lines).

Figure EC.29 Comparing the short-term outcomes of unconstrained and constrained optimal policies.

EC.10.6.2. Long-term outcome: potential utilitarian gain Figure EC.30a illustrates the long-term utilities of both optimal unconstrained policy and our proposed optimal constrained policy. Moreover, Figure EC.30b shows the price of fairness together with the normalized slack of the optimal unconstrained policy.



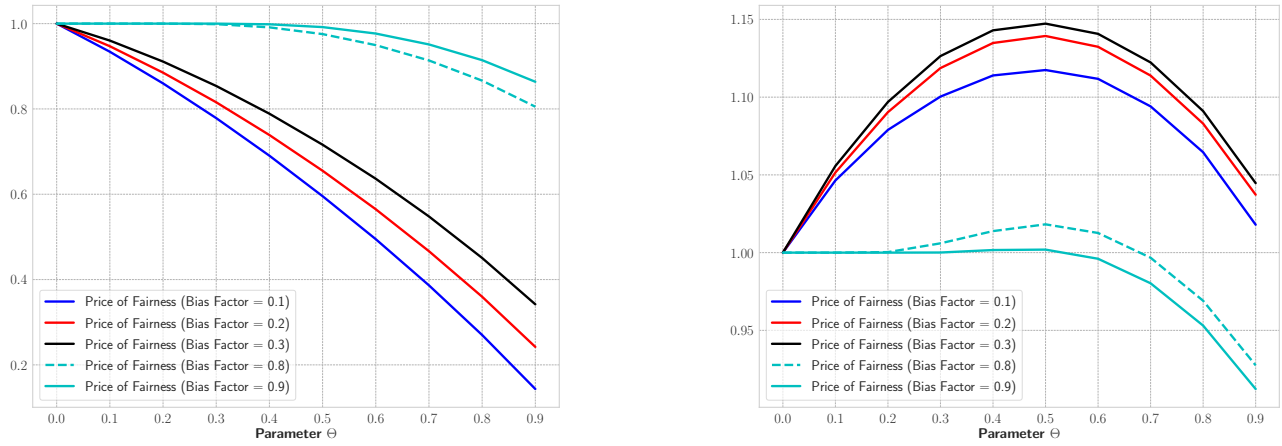
(a) Expected utilities calculated based on true values $\{v_i^\dagger\}_{i\in[n]}$ for the unconstrained optimal policy (red) and the constrained optimal policy (blue).



(b) Price of fairness ratio calculated based on true values $\{v_i^\dagger\}_{i\in[n]}$ (solid lines) and the normalized constraint slack of unconstrained optimal policy (dashed lines).

Figure EC.30 Comparing the long-term outcomes of unconstrained and constrained optimal policies

EC.10.6.3. Minimum Quota Constraint Figure EC.31 illustrates short-term and long-term price of fairness across various quota parameters.



(a) Short-term price of fairness

(b) Long-term price of fairness

Figure EC.31 Performance of optimal constrained policy for QUOTA in selection with parameter θ (k=20).

EC.10.7. Numerical Simulations - JMS

In this section, we study and analyze the performance of Algorithm 3 on a set of instances for the JMS problem. The main goal of our simulations in this section is to study the running time and convergence of Algorithm 3 as an iterative algorithm and a FPTAS to the optimal policy, but we also study the utility of the search obtained by this policy (with respect to the observable signals).

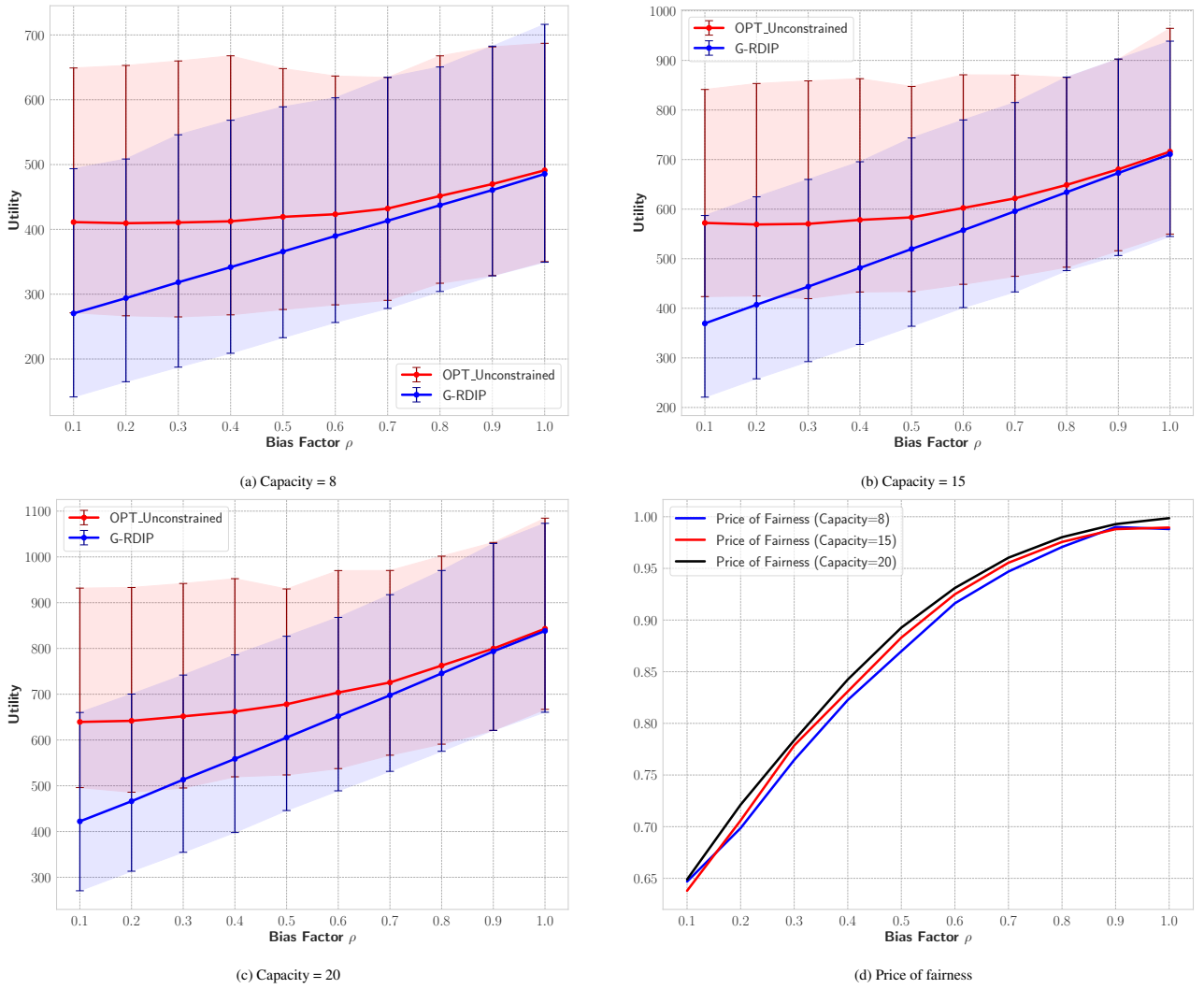
More specifically, we consider the JMS instance provided in Example 1 (illustrated in Figure 2b), which was a two-stage search with the possibility of rejection. We then consider three constraints that we would like to satisfy all at the same time. More specifically, we want to satisfy the PARITY in selection in all three stages of the search process, namely "phone interviews", "on-site interviews" and "offers".

Basic simulation setup: The value distribution for each of the alternatives is, in fact, generated the same way as in Section 4. As for the costs and transition probabilities for the extra stages that are apparent in this problem, we use the following setup:

- Cost of phone interview stage: $\text{Uniform}[1,2]$
- Cost of onsite interview stage: $\text{Uniform}[2,4]$
- Cost of offer to each individual = 3
- Probability of passing the phone interview = 80%
- Probability offer getting accepted = 90%

In order to evaluate the performance of the algorithm in expectation, we run a Monete-Carlo simulation with 20 instances.

Short-term price of fairness: Figure EC.32 shows the short-term utilities of both our near-optimal constrained policy and unconstrained optimal policy, as well as their ratio (price of fairness), under different capacities. As can be seen from the plots, the price of fairness is quite small, especially for $\rho \in [0.6, 1]$. This shows that the negative externalities due to fairness considerations are small and negligible in practical instances similar to those we consider here.



(a) Capacity = 8

(b) Capacity = 15

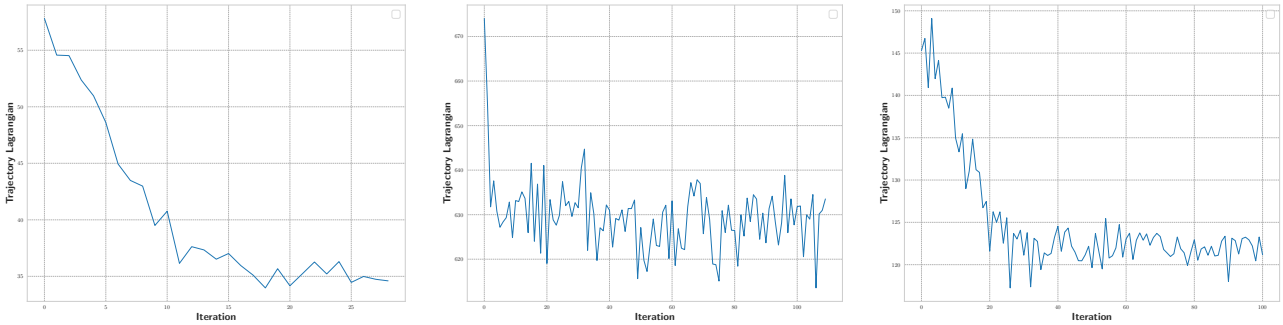
(c) Capacity = 20

(d) Price of fairness

Figure EC.32 (JMS simulation) (a,b,c) show the expected utilities calculated based on signals $\{v_i\}_{i\in[n]}$ for the unconstrained optimal policy (red) and the constrained optimal policy (blue); and (d): Price of fairness ratio calculated based on signals $\{v_i\}_{i\in[n]}$ (solid lines) and the normalized constraint slack of unconstrained optimal policy (dashed lines).

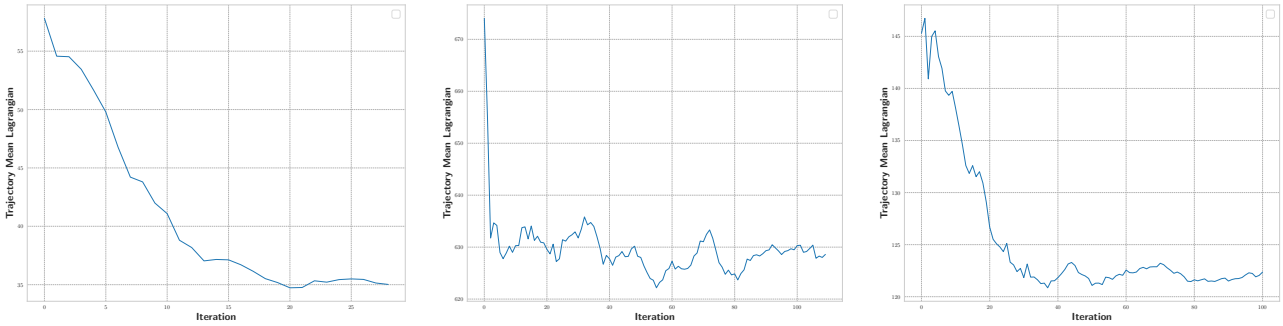
Convergence trajectory: The following Figures EC.33 to EC.38 illustrates the trajectory of Lagrangian, mean (over the past iterations) of Lagrangian, mean (over the past iterations) of slacks for each of the three constraints (each corresponding to the parity at one of the stages), as well as the dual adjustments $\boldsymbol{\lambda}$. As shown by all the figures, we can see that all these metrics will converge to their goal in around $K_{\mathrm{O}} = 20$ to $K_{\mathrm{O}} = 40$ number of iterations. Note that these are only outer iterations of Algorithm 3, as we do not have any convex constraints in this set of simulations and there is no need for the inner-loop. This demonstrates that, even though the theoretical number of iterations derived in Theorem 2 can be quite large, the actual number of iterations need for convergence is quite small under practical instances.

Running times: Figure EC.39 shows the running times of Algorithm 3, at different bias levels. Although the running times are considerably longer—compared to the results for Algorithm 2 when we had a single affine constraint and a simple single-stage search problem, as demonstrated in Figure EC.20—they all take less than



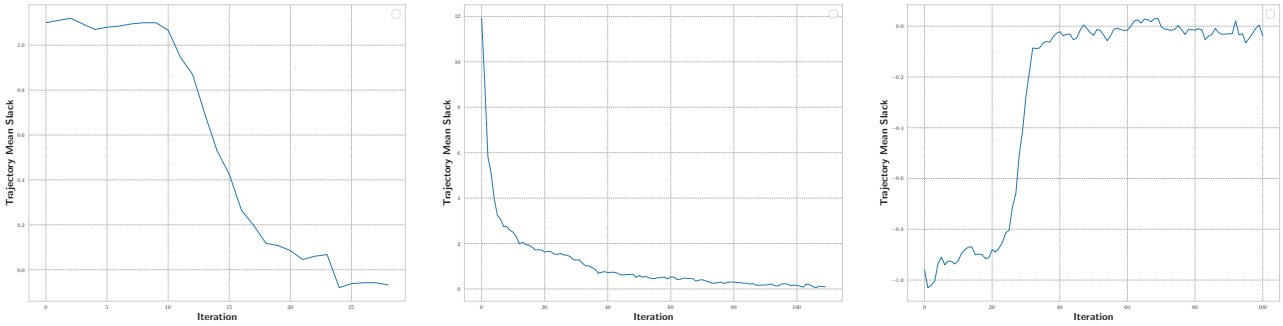
(a) Bias factor = 0.3 (b) Bias factor = 0.7 (c) Bias factor = 0.9

Figure EC.33 (JMS simulation) Trajectory of Lagrangian under capacity $k = 20$



(a) Bias factor = 0.3 (b) Bias factor = 0.7 (c) Bias factor = 0.9

Figure EC.34 (JMS simulation) Trajectory of mean Lagrangian under capacity $k = 20$

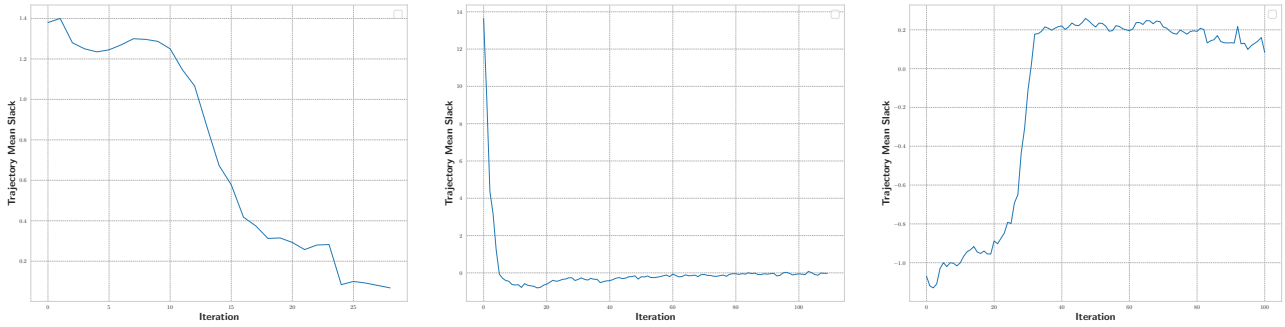


(a) Bias factor = 0.3 (b) Bias factor = 0.7 (c) Bias factor = 0.9

Figure EC.35 (JMS simulation) Trajectory of mean slack for the parity at the offer stage under capacity $k = 20$

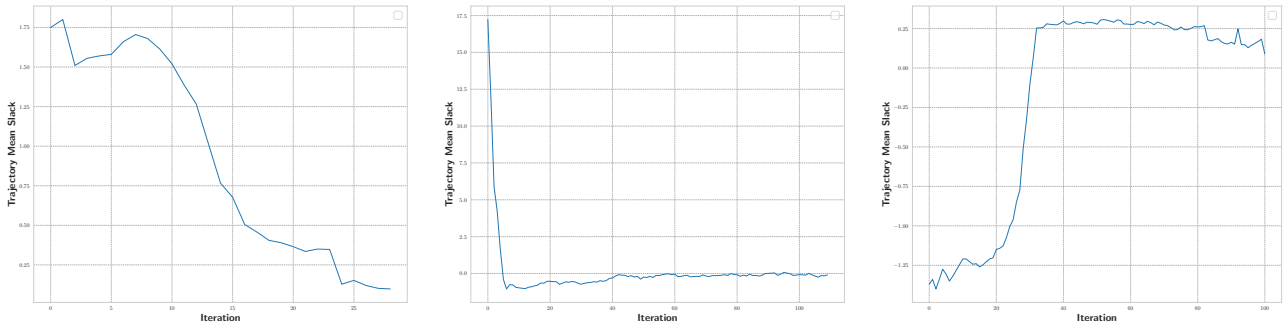
a minute for any given instance of the problem on the computer we used for our simulations (the same as the one we used for our earlier simulations).[31] This is especially important because we only need to run our algorithms once to find the (near-optimal) policy in any application, and after that the policy can be executed on each instantiation of the problem instance.

[31] We used a MacbookPro with 2.3 GHz Quad-core Intel Core i7 CPU, with 16GB of 3733 MHZ LDDR4X Memory for all of the simulations throughout the paper.



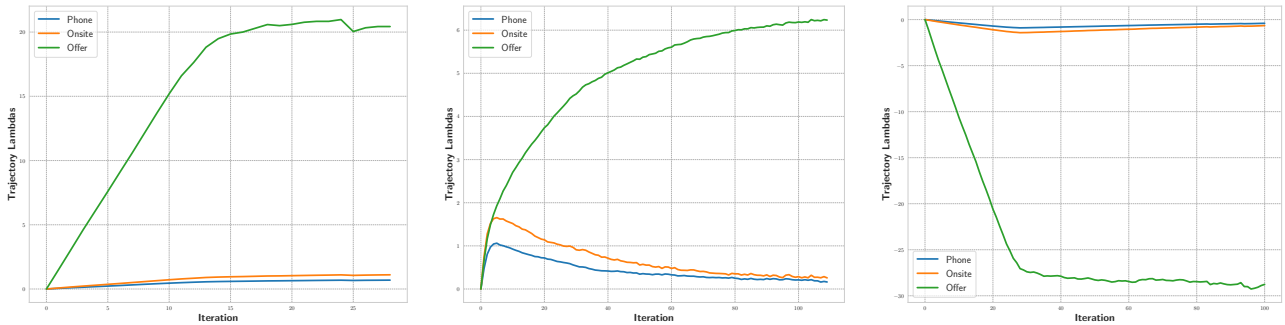
(a) Bias factor = 0.3 (b) Bias factor = 0.7 (c) Bias factor = 0.9

Figure EC.36 (JMS simulation) Trajectory of mean slack for the parity at the onsite stage under capacity $k = 20$



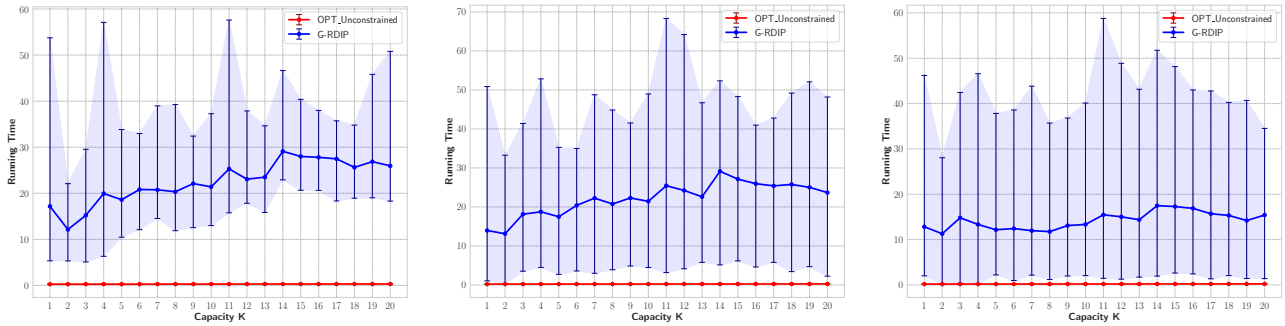
(a) Bias factor = 0.3 (b) Bias factor = 0.7 (c) Bias factor = 0.9

Figure EC.37 (JMS simulation) Trajectory of mean slack for parity at phone interview stage, capacity $k = 20$



(a) Bias factor = 0.3 (b) Bias factor = 0.7 (c) Bias factor = 0.9

Figure EC.38 Trajectory of dual adjustment λ under capacity $k = 20$



(a) Bias factor = 0.3 (b) Bias factor = 0.7 (c) Bias factor = 1.0

Figure EC.39 The comparison of running times of near-optimal constrained (G-RDIP) and optimal unconstrained policies (in seconds).