Learning Optimal and Fair Policies for Online Allocation of Scarce Societal Resources from Data Collected in Deployment

Bill Tang
USC CAIS Center for Artificial Intelligence in Society, yongpeng@usc.edu

Çağıl Koçyiğit
Luxembourg Centre for Logistics and Supply Chain Management, University of Luxembourg, cagil.kocyigit@uni.lu

Eric Rice
USC CAIS Center for Artificial Intelligence in Society, ericr@usc.edu

Phebe Vayanos
USC CAIS Center for Artificial Intelligence in Society, phebe.vayanos@usc.edu

We study the problem of allocating scarce societal resources of different types (e.g., permanent housing, deceased donor kidneys for transplantation, ventilators) to heterogeneous allocatees on a waitlist (e.g., people experiencing homelessness, individuals suffering from end-stage renal disease, Covid-19 patients) based on their observed covariates. We leverage administrative data collected in deployment to design an online policy that maximizes expected outcomes while satisfying budget constraints, in the long run. Our proposed policy waitlists each individual for the resource maximizing the difference between their estimated mean treatment outcome and the estimated resource dual-price or, roughly, the opportunity cost of using the resource. Resources are then allocated as they arrive, in a first-come first-serve fashion. We demonstrate that our data-driven policy almost surely asymptotically achieves the expected outcome of the optimal out-of-sample policy under mild technical assumptions. We extend our framework to incorporate various fairness constraints. We evaluate the performance of our approach on the problem of designing policies for allocating scarce housing resources to people experiencing homelessness in Los Angeles based on data from the homeless management information system. In particular, we show that using our policies improves rates of exit from homelessness by 1.9% and that policies that are fair in either allocation or outcomes by race come at a very low price of fairness.

Key words: homelessness, fairness, efficiency, scarce resource allocation, data-driven optimization, causal inference

1. Introduction

We study the problem of allocating basic resources of different types to heterogeneous individuals within high-stakes social settings subject to budget and fairness constraints. We are particularly motivated by housing allocation for individuals experiencing homelessness in Los Angeles (LA)

County. According to the Los Angeles Homeless Services Authority (LAHSA), more than 69,000 people are currently experiencing homelessness in LA County as of February 2022 (LAHSA 2022a). While permanent housing interventions are a key aspect of addressing homelessness (U.S. HUD 2007), the number of individuals experiencing homelessness far exceeds the capacity of available resources. As of the latest Housing Inventory Count in March 2022 from LAHSA, there are only around 30,000 permanent housing units (LAHSA 2022b). Due to the scarcity of resources available relative to demand for them, there is a need to allocate housing in an effective way as individuals and housing resources arrive.

Additionally, since housing resources are a public good with significant impact for individuals experiencing homelessness, it is important that an allocation policy is, in some sense, fair. A lack of fairness in the system may erode the trust and effectiveness of any policy. Indeed, in our motivating example, LAHSA policymakers and community members are particularly concerned with issues of fairness in allocation and outcomes within homelessness services. A 2018 report from LAHSA's Ad Hoc Committee on Black People Experiencing Homelessness found that while Black individuals represented 9% of the overall population in LA County, they made up 40% of the population experiencing homelessness (LAHSA 2018b). Furthermore, while housing resources are allocated at equal rates across racial groups, Black individuals receiving Permanent Supportive Housing (PSH) experienced higher rates of returns to homelessness relative to other racial groups. These findings have led to further investigations into causes of racial inequities in outcomes within the current system (Milburn et al. 2021) and a policy focus on racial equity to address existing injustices. Therefore, an allocation system needs to be not just effective, but also fair to be adopted and trusted by policymakers and community members.

Currently in the U.S., many local communities pool resources into a centralized planning system known as Continuum of Care (CoC), which coordinates housing and services funding for people experiencing homelessness. Within CoCs, Coordinated Entry Systems (CES), a network of homeless service providers and funders, use a standardized process to manage housing and supportive services and connect individuals to appropriate interventions. Individuals seeking housing in the LA CoC have various entry points into the system, such as access centers, street outreach, or emergency shelters. Following entry, they may take a self-reported survey to be assessed for eligibility and vulnerability. The assessment survey, known as Vulnerability Index–Service Prioritization Decision Assistance Tool (VI-SPDAT), consists of a series of questions to measure vulnerability, such as prior housing history and disabilities, and weights responses to output a score between 0 and 17, where a higher score indicates greater vulnerability (OrgCode 2015). Generally, the score is used in determining and prioritizing the most vulnerable individuals to receive more supportive resources. Many communities, based on OrgCode recommendations, have used score cutoffs for prioritization or guidance in the past (LAHSA 2018a). Individuals scoring 8-17 are considered "high risk" and are prioritized for Permanent Supportive Housing (PSH), the most supportive and intensive type of housing resource. Individuals scoring 4-7 are recommended for Rapid-Rehousing (RRH), or short-term rental subsidies, while those scoring less than 4 are only eligible for services, which we refer to as Service Only (SO) (OrgCode and Community Solutions 2015). Based on these scores and the VI-SPDAT assessments, case managers, who are social service workers assisting individuals experiencing homelessness, will determine the appropriate resource to match an individual to.

While the above prioritization policy, based on just score cuts, is interpretable, both the thresholds and the weights on responses to construct the score are not tied to any data on intervention outcomes nor to arrival rates of individuals and resources. This leads to three issues. First, it is not clear that assigning the most vulnerable individuals PSH, or even RRH, would lead to the best overall outcomes, such as reducing returns to homelessness after intervention. Second, there may not be enough housing available to serve all individuals prioritized for that resource because the thresholds do not account for arrival rates. Indeed, even OrgCode, the organization that developed the VI-SPDAT, has called for alternative approaches beyond the VI-SPDAT in 2020 (OrgCode 2020). Finally, a threshold prioritization tool cannot address stakeholder concerns around racial disparities in housing outcomes. An allocation system that is equitable in outcomes would need to account for the heterogeneity of treatment effects by protected groups such as race, gender, or age (Jo et al. 2023). In contrast to the existing tool, we seek to incorporate treatment outcomes and effects into the design of a policy for maximizing desirable societal outcomes and ensuring certain notions of fairness in allocation or outcomes.

Utilizing treatment outcomes and effects in policy design, however, is challenging since they are *counterfactual*, or "what if" quantities. This means we need to know what *would happen* to an individual if they are assigned to different resources. While running randomized control trials is the gold standard for inference of causal treatment effects, in high-stakes domains such as public health and social services, experimentation may be unethical or impractical. For example, it would be unethical to deny housing from individuals experiencing dangerous and vulnerable situations. A proper experiment may also take years before conclusive results, leading to negative life consequences for many individuals during that time span. Therefore, we aim to leverage observational data for learning treatment outcomes. The historical data that LAHSA and other communities have access to was collected in deployment from administered VI-SPDAT assessments, which contains an individual's covariates, and the LA County Homeless Management Information System (HMIS) database, which contains an individual's trajectory of interactions with the system. These interactions include events such as engagements with street outreach workers, emergency shelter access, and housing allocations, if any. From these trajectories, we construct a return to homelessness definition that is used as the outcome of interest.

Motivated by these issues with the existing housing allocation system, our goal is to design an efficient policy for allocating scarce housing resources to maximize positive expected outcomes, subject to budget and fairness constraints. Using prior observational data, we learn treatment outcomes based on an arriving individual's characteristics, such as prior housing history, to appropriately match individuals to housing resources. In particular, we want an online assignment policy rather than waiting for sufficient number of individuals and housing to aggregate in the system for offline matching. Community members are dissatisfied with scarce housing sitting empty for long periods of time given the severity of homelessness. Furthermore, individuals may experience adverse outcomes while waiting to be matched.

While we primarily focus on housing, our approach is applicable in general to allocation of scarce treatments in other public domains where system outcome efficiency and fairness are important. One example is the problem of kidney transplantations in which patients arrive to the system and join a waiting pool for a donor organ to be procured (Bertsimas et al. 2013, Dickerson and Sandholm 2015). Similar to our case, there is a significant organ shortage relative to the number of waiting patients and observational data exists from the United Network for Organ Sharing. Another example is the allocation of scarce healthcare equipment such as ventilators and critical care beds to patients arriving to the hospital with Covid-19 during spikes in hospitalizations. Observational data from electronic health records for hospitalized patients with Covid-19 can be used to learn policies for allocating equipment as patients arrive (Johnston et al. 2020, Radovanovic et al. 2021).

1.1. Related Works

Our work is closely related to literature on data-driven allocation of public resources, fair resource allocation, policy learning from observational data, and network revenue management.

Data-Driven Allocation of Public Resources. A closely related strand of literature focuses on data-driven allocation policies in scarce public resource settings such as kidney allocation (Bertsimas et al. 2013, Dickerson and Sandholm 2015), hospital ICUs (Grand-Clément et al. 2022), or homeless services (Azizi et al. 2018, Kube et al. 2019, Rahmattalabi et al. 2022, Kaya et al. 2022). Our methodology shares similarities with approaches using historical data to learn treatment outcome plug-ins for mixed-integer linear optimization formulations, which are then used to derive dual multipliers based assignment rules with various approximations. For example, Bertsimas et al. (2013) propose learning a scoring policy for online kidney allocation that approximately satisfies fairness constraints, but assumes knowledge of the treatment outcomes. Their scoring rule is a linear approximation of treatment outcomes minus the dual multipliers of various eligibility and fairness constraints. Azizi et al. (2018) propose an MIO formulation for learning fair and interpretable policies such as decision trees for prioritizing housing resources for youth and an approximation approach to solving the MIO formulation for tractability. In comparison to these works, we do not restrict ourselves to any particular functional approximations or scoring rules. Additionally, while MIO approaches generate interpretable policies at the cost of computational tractability, our approach still results in a relatively interpretable policy that is provably asymptotically optimal and only requires solving a linear program. In another related work, Bhattacharya and Dupas (2012) focus on maximizing welfare under a budget constraint for social programs with a binary treatment by learning asymptotically consistent treatment effects and propose an allocation policy based on data from a randomized experiment. In comparison, we allow for multiple scarce treatment types, are able to impose a variety of fairness constraints, and formulate an online implementation in which an individual is matched to an appropriate resource as they arrive to the system.

Fair Resource Allocation. Given the socially sensitive nature of the resource allocation systems that motivate our work, our paper also relates to research on fair decision-making in high-stakes domains. Generally, these works focus on imposing linear constraints limiting some notion of disparity between protected groups (Bertsimas et al. 2013, Corbett-Davies et al. 2017, Azizi et al. 2018, Nguyen et al. 2021, Rahmattalabi et al. 2022), on impossibility results of various fairness notions (Mashiat et al. 2022, Jo et al. 2023), or explicitly focus on algorithmic fair division (Rahmattalabi et al. 2021, Nguyen et al. 2021, Freeman et al. 2020, Manshadi et al. 2021). Within mechanism design literature, Athanassoglou and Sethuraman (2011) also consider designing an efficient (ordinal efficiency) and fair (no justified envy) algorithm for allocating collectively owned resources such as scarce housing. Our work follows in the vein of imposing linear fairness constraints such as parity in resource allocation rates between different racial groups since this is how policy-makers within LA homeless services evaluate fairness (LAHSA 2018b).

Policy Learning from Observational Data. Another strand of literature focuses on policy evaluation and learning from observational data by estimating counterfactuals or correcting for selection bias in historical treatment assignments. Various works focus on constructing unbiased estimators for policy evaluation from observational data under an unconfoundedness assumption, which can then be used for policy learning. These include the direct method of using plug-in estimates of counterfactual treatment outcomes (Qian and Murphy 2011), propensity score based reweighting strategies (Swaminathan and Joachims 2015), and doubly robust combinations of outcome estimates and propensity-based reweighting (Dudík et al. 2011). Kitagawa and Tetenov (2018) and Athey and Wager (2021) study regret bounds for policy learning under propensity-weighted or doubly robust objective functions relative to some well-specified policy class with structure while Zhou et al. (2022) extend the analysis to the multiple treatment case. Other works emphasize interpretable policy classes such as prescriptive trees (Kallus 2017, Bertsimas et al. 2019, Jo et al. 2021, Zhou et al. 2022). While these works derive important analytical results for policy classes with suitable structure, they either use a specified class such as trees in practice, leave the choice of policy class open ended, or use Π to denote domain constraints related to budget, fairness, functional form without considering specific implementations of those constraints. In contrast, we consider the general policy class of measurable functions and derive an explicit assignment policy parameterized by sample-based dual multipliers that is asymptotically optimal and can flexibly handle capacity and fairness constraints.

Network Revenue Management. A fourth related stream of literature is network revenue management (NRM), which focuses on the sale of products comprised of scarce resources to customers who request a specific product with an offer price (Talluri and Van Ryzin 2004). In domains such as airlines and hotels, the system must decide to accept or reject an arriving customer's offer and the decision depends on the stochastic demand and revenue for products relative to resource capacities. In particular, our approach is inspired by work on bid-price controls, a class of policies where an offer is accepted only if the revenue generated is greater than the sum of optimal dual variables (bid-prices) associated with the resource capacity constraints (Talluri and Van Ryzin 1998). Early work by Talluri and Van Ryzin (1998) showed that such a control policy is asymptotically optimal as the number of requests and resources grow as long as the relative relationship of supply and demand stays constant. Since NRM problems are often formulated as dynamic programs, recent works have focused on settings where large state spaces cannot be solved tractably and computing optimal bid-prices is challenging. Tractable solutions are generally heuristics, approximate dynamic programming, or Lagrangian relaxation methods (Adelman 2007, Kunnumkal and Talluri 2016, Zhang and Weatherford 2017, Li et al. 2023). While our approach is inspired by bid-price controls within NRM, our setting differs in that the system must choose amongst a set of resources to match an arrival to rather than an accept or reject decision for a specific requested resource.

1.2. Contributions and Organization of the Paper

Our contributions in this paper are:

- We introduce an online policy called the dual-price queuing policy for matching individuals with limited resources as they arrive, assuming full distributional information. Under this policy, each individual is waitlisted for the treatment that maximizes the difference between their expected outcome given their covariates and the optimal resource dual-price. The proposed policy ensures that capacity constraints are met. With mild technical assumptions, we demonstrate that this policy achieves maximum expected average outcome over an infinite time horizon.

- We develop a sample-based analogue of the dual-price queuing policy that relies solely on historical (observational) data and does not require distributional knowledge. We prove that, under mild technical assumptions, this policy almost surely asymptotically achieves the optimal expected average outcome as the number of historical data samples tends to infinity.
- We extend our framework to incorporate various fairness constraints, including statistical parity in allocations and outcomes between groups of a protected characteristic (e.g., race, gender, or age). We introduce fairness-constrained extensions of the sample-based dual-price queuing policy. While these extensions are designed to meet the respective fairness constraints in-sample, we demonstrate through numerical experiments that they can also improve fairness out-of-sample.
- We demonstrate through computational experiments on synthetic and real data that our method's guarantees of asymptotic optimality lead to improvements in out-of-sample performance, even in finite samples. Specifically, we evaluate the policy's performance in allocating scarce housing resources to people experiencing homelessness in Los Angeles, based on data from the HMIS database. We show that using our policies improves rates of exit from homelessness by 1.9%, and policies that are fair in either allocation or outcomes by race come at a very low price of fairness.

The remainder of this paper is structured as follows. Section 2 introduces the policy design problem and outlines the technical assumptions required for our optimality results. Sections 3 and 4 study the dual-price queuing policy under full distributional information and its sample-based variant, respectively, along with their optimality guarantees. Section 5 studies the fairness-constrained extensions of the dual-price queuing policy. Finally, Section 6 presents numerical results based on synthetic and real data. All proofs are relegated to the online appendix.

Notation. All random variables in this paper are defined as measurable functions on an abstract probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and are capitalized (*e.g.*, X) while their realizations are denoted by the same symbols in lowercase letters (*e.g.*, x) unless stated otherwise. Throughout the paper, (in)equalities containing random variables should be interpreted to hold $\mathbb{P}$ almost surely. We let $\mathcal{L}_\infty(\mathcal{X}, \mathcal{C})$ denote the set of all bounded Borel-measurable functions from a Borel set $\mathcal{X}$ to a Borel set $\mathcal{C}$. For any set $\mathcal{A}$, we use $|\mathcal{A}|$ to denote the cardinality of the set. We use $\mathbb{1}[\cdot]$ to denote the indicator function of a logical expression E, where $\mathbb{1}[E] = 1$ if E is true and $\mathbb{1}[E] = 0$ otherwise.

2. Problem Formulation and Preliminaries

We consider an online allocation system where heterogeneous individuals and treatments of different types arrive sequentially over time and focus on an infinite horizon setting in order to model long-run performance. We can assign each individual a treatment from the finite set $\mathcal{T} = \{0, 1, \ldots, m\}$

of candidate treatments, and we use treatment 0 to represent the no-treatment option (or control). Each individual is characterized by a (random) vector $X \in \mathcal{X} \subseteq \mathbb{R}^d$ of covariates and their potential outcomes $\{Y^t\}_{t\in\mathcal{T}}$, where $Y^t \in \mathbb{R}$ represents the outcome of receiving treatment $t \in \mathcal{T}$ under the potential outcomes framework (Hernan and Robins 2023). The covariate vector X, which includes characteristics needed to learn treatment outcomes and assignments, may include protected features $G \in \mathcal{G}$ such as race, gender, or sex. We will formulate fairness constraints using the protected features G in Section 5. We define $m^t(x) = \mathbb{E}[Y^t|X = x]$ as the conditional average treatment outcome for treatment $t \in \mathcal{T}$ and covariate vector $x \in \mathcal{X}$, and we assume that $m^t \in \mathcal{L}_\infty(\mathcal{X}, \mathbb{R})$ for all $t \in \mathcal{T}$. We denote by $\mathcal{K}$ the set of arriving individuals and assume throughout that $\{(X_k, \{Y_k^t\}_{t\in\mathcal{T}})\}_{k\in\mathcal{K}}$ are independent and identically distributed (i.i.d.) across individuals. We will assess the effectiveness of a policy based on its performance in the long run, specifically as $|\mathcal{K}| \to \infty$. We emphasize that we do not know the joint distribution of $(X, \{Y^t\}_{t\in\mathcal{T}})$.

We are interested in the scarce resource setting where each treatment type $t \in \mathcal{T} \setminus \{0\}$ has limited availability, i.e., the arrival rate of treatments of type $t \in \mathcal{T} \setminus \{0\}$ is less than the arrival rate of individuals. More specifically, we consider the setting where for treatment $t \in \mathcal{T}$, $b^t \in [0, 1]$ denotes the asymptotic capacity of treatment t per individual. For example, if $b^t = 0.3$, then we can allocate treatment t to at most 30% of all individuals. Since treatment 0 is the no-treatment case, we set $b^0 = 1$.

Since the effectiveness of an assignment depends on an individual's characteristics and their matched treatment, our goal is to design a policy $\pi \in \mathcal{L}_\infty(\mathcal{X}, [0, 1]^{m+1})$ for assigning treatments to different individuals conditional on their covariates. Specifically, $\pi^t(x)$ denotes the probability of assigning treatment t to an individual with covariates x and $\sum_{t\in\mathcal{T}} \pi^t(x) = 1$ for all $x \in \mathcal{X}$. The objective of the policy design problem is to maximize the asymptotic expected average outcome of assigning treatments to arriving individuals. We can formulate the expected average outcome under a policy π as

$$\lim_{|\mathcal{K}|\to\infty} \mathbb{E}\left[\frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}} \pi^t(X_k)Y_k^t\right] = \lim_{|\mathcal{K}|\to\infty} \mathbb{E}\left[\frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}} \pi^t(X_k)\mathbb{E}[Y_k^t|X_k]\right] = \mathbb{E}\left[\sum_{t\in\mathcal{T}} \pi^t(X)m^t(X)\right],$$

where the last equality follows from the definition of m^t and the assumption that $\{(X_k, \{Y_k^t\}_{t\in\mathcal{T}})\}_{k\in\mathcal{K}}$ are i.i.d. across different individuals.

Due to the limited availability of treatments, policy π is required to satisfy capacity constraints of the form.

$$\lim_{|\mathcal{K}|\to\infty} \frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}} \pi^t(X_k) \le b^t \quad \forall t \in \mathcal{T}.$$

As $\{(X_k, \{Y_k^t\}_{t\in\mathcal{T}})\}_{k\in\mathcal{K}}$ are i.i.d. across different individuals and π is bounded, by the strong law of large numbers we have

$$\lim_{|\mathcal{K}|\to\infty} \frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}} \pi^t(X_k) = \mathbb{E}[\pi^t(X)] \quad \forall t\in\mathcal{T}.$$

The policy design problem is thus given by

$$\begin{aligned} z^\star = \max_{\pi\in\Pi} \quad & \mathbb{E}\left[\sum_{t\in\mathcal{T}} \pi^t(X) m^t(X)\right] \\ \text{s.t.} \quad & \mathbb{E}\left[\pi^t(X)\right] \le b^t \quad \forall t\in\mathcal{T}, \end{aligned} \tag{$\mathcal{P}$}$$

where

$$\Pi = \left\{\pi \in L_\infty(\mathcal{X}, [0,1]^{m+1}) : \sum_{t\in\mathcal{T}} \pi^t(x) = 1 \quad \forall x\in\mathcal{X}\right\}.$$

Without loss of generality, we take m^t to be non-negative for all $t\in\mathcal{T}$. In fact, if m^t did take on negative values, one could shift the values of all m^t by adding a sufficiently large constant to all of them. This does not change the optimal solution to ($\mathcal{P}$) and simply shifts the optimal objective value by a constant.

Note that we cannot solve problem ($\mathcal{P}$) because the joint distribution of $(X, \{Y^t\}_{t\in\mathcal{T}})$ is unknown. However, we have access to an observational dataset $\{(X_i, T_i, Y_i)\}_{i\in\mathcal{N}}$ of $n = |\mathcal{N}|$ i.i.d samples generated by a previously deployed policy, where $\mathcal{N}$ is an indexing set for the data samples, $T_i \in \mathcal{T}$ is the treatment received by individual i, and $Y_i = Y_i^{T_i}$ is the observed outcome for individual i under their received treatment T_i. Note that we do not observe the potential outcomes Y_i^t if $T_i \ne t$. We also assume that we do not know the historical assignment policy used to allocate treatments in our data.

In Section 3, we propose a simple and interpretable policy (referred to as dual-price queuing policy) assuming that we know the joint distribution of $(X, \{Y^t\}_{t\in\mathcal{T}})$ and show that this policy is optimal in problem ($\mathcal{P}$). In Section 4, we characterize a sample approximation of the dual-price queuing policy and show that this sample-based policy asymptotically attains the optimal value of ($\mathcal{P}$) as the number of data samples tends to infinity. We discuss in the following subsections the assumptions required for our results and elaborate on their generality.

2.1. Conditional Average Treatment Outcomes

In Section 3, we propose a dual-price queuing policy assuming full distributional knowledge and show that it is optimal in problem ($\mathcal{P}$) under the following technical assumption on conditional average treatment outcomes.

Assumption 1. *For any $t\in\mathcal{T}$, the following conditions hold.*

(i) The random variable $m^t(X)$ is continuously distributed and has a bounded support, i.e., there exists a $C \in \mathbb{R}_+$ such that $|m^t(x)| \leq C$ for all $x \in \mathcal{X}$.

(ii) For any $t' \in \mathcal{T}$ and $t' \neq t$, $m^t(X) - m^{t'}(X)$, the difference in conditional treatment effects between t and t', is continuously distributed and has a bounded support, i.e., there exists a $D \in \mathbb{R}_+$ such that $|m^t(x) - m^{t'}(x)| \leq D$ for all $x \in \mathcal{X}$ and $t' \in \mathcal{T}$.

Assumption 1 requires that the conditional average treatment outcomes $m^t(X)$ for all treatments $t \in \mathcal{T}$, and the treatment effect differences $m^t(X) - m^{t'}(X)$ between any two treatments t, t' are bounded and continuously distributed. Intuitively, this assumption holds if treatment outcomes and effects take on continuous values.

The following example illustrates a simple case where the conditional average treatment outcomes are linear functions of a continuous random variable and satisfy Assumption 1.

Example 1 (Assumptions 1 (i) and (ii) hold). Suppose that there exists a single covariate X that is continuously distributed on the support $[0,1]$. There are two potential treatments indexed by 0 and 1, and their respective conditional average treatment outcome functions are given by $m^0(x) = x$ and $m^1(x) = 2x$. In this case, the random variables $m^0(X)$, $m^1(X)$, $m^1(X) - m^0(X)$ and $m^0(X) - m^1(X)$ are all bounded and continuously distributed.

We also show in the following two examples wherein conditions (i) and (ii) in Assumption 1 are both needed, thereby demonstrating that one does not imply the other. These also give intuition as to when Assumption 1 does not hold.

Example 2 (Assumption 1 (i) does not imply Assumption 1 (ii)). Suppose again that there exists a single covariate X that is continuously distributed on the support $[0,1]$. There are two potential treatments indexed by 0 and 1, and their respective conditional average treatment outcome functions are given by $m^0(x) = x$ and $m^1(x) = x + 1$. The random variables $m^0(X)$ and $m^1(X)$ are therefore bounded and continuously distributed, which implies that Assumption 1 (i) holds. On the other hand, $m^1(x) - m^0(x) = 1$, and therefore $m^1(X) - m^0(X)$ is not continuously distributed. Thus, Assumption 1 (ii) does not hold.

Example 3 (Assumption 1 (ii) does not imply Assumption 1 (i)). Suppose again that there exists a single covariate X that is continuously distributed on the support $[0,1]$. There are two potential treatments indexed by 0 and 1, and their respective conditional average treatment outcome functions are given by $m^0(x) = 1$ and $m^1(x) = x$. Then, we have that $m^1(x) - m^0(x) = x - 1$ and $m^0(x) - m^1(x) = 1 - x$, and so both $m^1(X) - m^0(X)$ and $m^0(X) - m^1(X)$ are bounded and continuously distributed. However, $m^0(X)$ is constant valued in this case and not continuously distributed. Therefore, Assumption 1 (ii) does not imply Assumption 1 (i).

2.2. Sample-Based Estimates

In Section 4, we propose a sample-based dual-price queuing policy. In order to learn this policy from data, we first need to estimate the conditional treatment outcomes. Let $\hat{m}^t(\cdot|\{(X_i, T_i, Y_i)\}_{i\in\mathcal{N}}) \in \mathcal{L}_\infty(\mathcal{X}, \mathbb{R})$ denote an estimator of the conditional mean treatment response m^t from the data. For ease of notation, we suppress the dependency of $\hat{m}^t$ on the historical samples and simply write $\hat{m}^t_n$ to stress that we use a sample of size n. The following assumption formally states properties we need our estimators to have and how "good" they need to be asymptotically.

Assumption 2. *For any $t \in \mathcal{T}$, the following condition holds.*

(i) The estimate $\hat{m}^t_n$ is measurable and has bounded support, i.e., there exists a $\hat{C} \in \mathbb{R}_+$ such that $|\hat{m}^t_n(x)| \leq \hat{C}$ for all $x \in \mathcal{X}$ and $n \in \mathbb{N}$.

(ii) The estimate $\hat{m}^t_n$ converges uniformly and almost surely to m^t, i.e., $\lim_{n\to\infty} \sup_{x\in\mathcal{X}} |\hat{m}^t_n(x) - m^t(x)| = 0$.

Assumption 2 (i) is a technical assumption needed for our theoretical results and is standard within other policy learning works as well. Assumption 2 (ii), the uniform convergence of $\hat{m}^t_n$ to m^t for each treatment, is necessary for the asymptotic optimality of the sample-based dual-price policy; see Theorem 2. This is because we do not make any functional assumptions about the conditional expected outcomes m^t, $t \in \mathcal{T}$. In fact, if one has knowledge of the form of m^t, for example that it is a linear function with parameters θ, then Assumption 2 can be weakened to uniform convergence of the estimator parameters θ. In our general setting, the asymptotic convergence guarantees we require in Assumption 2 (ii) are satisfied by non-parametric estimators like nearest-neighbors (Liero 1989) if standard assumptions on observational data also hold. In the next subsection, we introduce the aforementioned assumptions on the observational data that enable estimating m^t with the asymptotic convergence guarantees in Assumption 2 (ii).

2.3. Observational Data and Identifiability

A fundamental challenge of policy learning from observational data is that, for any sample i, we observe the outcome $Y_i = Y_i^{T_i}$ associated with the treatment T_i assigned by the historical policy but we do not observe the outcome that would have been obtained if any other treatment $t \neq T_i$ had been assigned. Therefore, even evaluating the performance of a *counterfactual* policy π necessitates appropriate *identifiability* conditions to make causal quantities such as $\mathbb{E}[Y^t]$ and m^t possible to estimate from the available data. We make the following assumptions that are standard within the causal inference literature (Hernan and Robins 2023).

Assumption 3. *The following conditions hold.*

(i) Conditional Exchangeability: $Y^t \perp\!\!\!\perp T \mid X \quad \forall t \in \mathcal{T}$

(ii) Positivity: $\mathbb{P}(T=t \mid X=x)>0 \quad \forall t \in \mathcal{T}$ *and* $\forall x \in \mathcal{X}$

(iii) Consistency: $Y^T=Y$

Assumption 3 (i) says that the outcome Y^t of assigning treatment $t \in \mathcal{T}$ and the allocated treatment T are conditionally independent given the covariates $X=x$. Equivalently, the joint distribution of potential outcomes is the same for all individuals in each treatment group $t \in \mathcal{T}$ given they have the same value of x. This assumption is satisfied if we measure and condition on all covariates that are common causes of treatment assignment and outcomes. This implies that we can consider all individuals with the same x in each treatment group as exchangeable and we can calculate $m^t(x)$ as $\mathbb{E}[Y \mid T=t, X=x]$. In context of our motivating example, this means the VI-SPDAT assessment responses and HMIS database contain all the relevant information for individuals experiencing homelessness that case managers use to make their final decisions on housing assignments. In our case, this assumption is likely reasonable since the VI-SPDAT is used as a prioritization tool by case managers and contains a comprehensive overview of an individual's vulnerability factors.

Assumption 3 (ii) requires that the conditional probability of treatment assignment given features $X=x$ is positive for each treatment. Intuitively, this means that each individual receives any particular treatment with positive probability under the historical policy. Assumption 3 (iii) states that the observed outcome under an observed treatment T is actually the potential outcome. Practically, this means that our treatments are well specified or else if there were multiple versions of a particular treatment t, then Y^t is not well defined.

Although we do not explicitly refer to Assumption 3 in our theoretical results, we include it for the sake of completeness as it is closely related to Assumption 2 (ii). Specifically, under Assumption 3, the conditional average treatment outcomes m^t for $t \in \mathcal{T}$, which are *causal* quantities, can be expressed and estimated as *statistical* quantities that are solely a function of the observed data (Hernan and Robins 2023). When Assumption 3 holds and we can estimate a causal quantity as a statistical quantity from observed data, then an estimator such as nearest-neighbors will satisfy the uniform convergence property of Assumption 2 (ii) (Liero 1989).

3. Dual-Price Queuing Policy under Full Information

In this section, we temporarily assume that we know the joint distribution of the covariates X and the potential outcomes Y^t, $t \in \mathcal{T}$. Using a duality approach, we will characterize a policy (referred to as dual-price queuing policy) that is optimal in problem ($\mathcal{P}$) assuming full distributional knowledge. From this policy, we will characterize a sample-based policy learned from observational data in Section 4.

Before defining the dual-price queuing policy, we first showcase the duality approach that yields its characterization. The Lagrangian dual of problem $(\mathcal{P})$ is given by

$$\begin{aligned}\nu^\star &= \min_{\mu\in\mathbb{R}_+^{m+1}} \max_{\pi\in\Pi} \mathbb{E}\left[\sum_{t\in\mathcal{T}} \pi^t(X) m^t(X)\right] + \sum_{t\in\mathcal{T}} \mu^t \left(b^t - \mathbb{E}\left[\pi^t(X)\right]\right) \\ &= \min_{\mu\in\mathbb{R}_+^{m+1}} \max_{\pi\in\Pi} \mathbb{E}\left[\sum_{t\in\mathcal{T}} \pi^t(X)(m^t(X) - \mu^t)\right] + \sum_{t\in\mathcal{T}} \mu^t b^t,\end{aligned} \tag{1}$$

where μ collects the dual variables of the capacity constraints. By weak duality, we have $\nu^\star \geq z^\star$. For any $\mu \in \mathbb{R}_+^{m+1}$, the inner maximization problem is solved by the policy

$$\pi^t(x) = \begin{cases} 1 & \text{if} \quad t = \min\arg\max_{t'\in\mathcal{T}} \left(m^{t'}(x) - \mu^{t'}\right) \\ 0 & \text{otherwise} \end{cases} \qquad \forall t\in\mathcal{T}, \forall x\in\mathcal{X}. \tag{2}$$

We suppress the dependence of policy π on μ notationally in order to avoid clutter. We use a lexicographic tie-breaker in characterizing (2). We emphasize that our results do not rely on this particular tie-breaking rule. By substituting policy (2) into the dual problem (1), we obtain

$$\nu^\star = \min_{\mu\in\mathbb{R}_+^{m+1}} \mathbb{E}\left[\max_{t\in\mathcal{T}} (m_t(X) - \mu^t)\right] + \sum_{t\in\mathcal{T}} \mu^t b^t. \tag{$\mathcal{D}$}$$

In the remainder, we denote by $\mu^\star$ an optimal solution to $(\mathcal{D})$, which exists because we can restrict the feasible set of $(\mathcal{D})$ to a compact set without loss of generality; see Lemma EC.3 in the online appendix. We define a candidate policy $\pi^\star$ through

$$\pi^{\star,t}(x) = \begin{cases} 1 & \text{if} \quad t = \min\arg\max_{t'\in\mathcal{T}} (m^{t'}(x) - \mu^{\star,t'}) \\ 0 & \text{otherwise} \end{cases} \qquad \forall t\in\mathcal{T}, \forall x\in\mathcal{X}. \tag{3}$$

Note that policy $\pi^\star$ is obtained by choosing μ as $\mu^\star$ in (2).

Next, we introduce an online implementation of policy $\pi^\star$ that takes into account the fact that individuals arrive sequentially and that treatments become available over time. By construction, this policy satisfies the almost-surely capacity constraints of form (2) even in finite horizon, i.e., for fixed $|\mathcal{K}|$.

Definition 1 (Dual-price queuing policy). Establish m queues each associated with a treatment $t \in \mathcal{T}\setminus\{0\}$. At any time, individuals in queue t are waitlisted for treatment t. There are two events triggering an action: *(i)* arrival of an individual, and *(ii)* arrival of a treatment.

(i) When an individual with covariate vector x arrives, compute the treatment $t = \min\arg\max_{t'\in\mathcal{T}} m^{t'}(x) - \mu^{\star,t'}$, where t attains the maximum difference between the conditional mean treatment outcome and the associated dual variable. If $t \neq 0$, assign them to queue t, otherwise assign them no-treatment, meaning they will not wait for a particular treatment and will exit the system. If they are the only individual waiting in queue t, check whether there is an available treatment of type t. If there is, assign treatment t to the individual.

(ii) When a treatment of type t becomes available, assign this treatment to the first individual in queue t. If there is no one in queue t, keep the treatment until an individual is assigned to this queue.

In finite horizon, the dual-price queuing policy assigns $t = 0$ to individuals still waiting in queues at the end of the time horizon. Since in practice our experiments can only occur in the finite horizon setting, we will use a 'long' test horizon in the synthetic data experiments as a proxy for the asymptotic environment. We note that in general, $\mu^\star$, the optimal solution to $(\mathcal{D})$, may not be unique since the objective function of $(\mathcal{D})$ may not be *strictly* convex; this depends on the form of the joint distribution of $(X, \{Y^t\}_{t\in\mathcal{T}})$. This means that the candidate policy $\pi^\star$ defined in (3) and the dual-price queuing policy may not be uniquely specified.

The next theorem shows that the dual-price queuing policy satisfies the capacity constraints in problem $(\mathcal{P})$, and its asymptotic expected average outcome matches the optimal value $z^\star$ of $(\mathcal{P})$. It thus solves problem $(\mathcal{P})$.

Theorem 1. *Under Assumption 1, the dual-price queuing policy is optimal in problem* $(\mathcal{P})$.

The idea of the proof is that it is sufficient to show that $\pi^\star$ is feasible and optimal in $(\mathcal{P})$. Since strong duality holds for problem $(\mathcal{D})$, the optimal solutions of $(\mathcal{D})$ and its dual must satisfy the Karush-Kuhn-Tucker (KKT) conditions. The KKT conditions imply that $\pi^\star$, which is constructed from an optimal solution $\mu^\star$ of $(\mathcal{D})$, is a feasible solution of $(\mathcal{P})$ and yields an objective value equal to $v^\star = z^\star$.

The dual-price queuing policy in Definition 1 is unidentifiable because its definition relies on the joint distribution of $(X, \{Y^t\}_{t\in\mathcal{T}})$, which is unknown in reality. In the next section, we characterize a sample approximation of the dual-price queuing policy and show that it asymptotically attains the optimal value $z^\star$ of problem $(\mathcal{P})$ as the number of data samples tends to infinity. Our sample-based policy only depends on samples from an observational dataset, which are not necessarily realizations from the distribution $(X, \{Y^t\}_{t\in\mathcal{T}})$ since the observed samples depend on the historically deployed policy.

4. Sample-Based Dual-Price Queuing Policy

In reality, we do not know the joint distribution of $(X, \{Y^t\}_{t\in\mathcal{T}})$ nor do we have access to samples from this full distribution since our samples come from an observational dataset. Furthermore, we do not know the conditional average treatment outcome functions m^t for each $t \in \mathcal{T}$, or an optimal dual solution $\mu^\star$. Instead, we have access to a relevant historical dataset $\{(X_i, T_i, Y_i)\}_{i\in\mathcal{N}}$ of n i.i.d. samples that we can use to estimate these quantities. While the choice of estimator for m^t, $t \in \mathcal{T}$, is not the focus of our work, there exists a variety of methods within the causal inference literature for learning conditional average treatment effects and outcomes such as causal forests (Wager and Athey 2018). Algorithms such as metalearning, double machine learning, and doubly robust learning allow for arbitrary estimation methods to be used for modeling outcomes and treatment assignments and we can use non-parametric estimators that satisfy the theoretical asymptotic assumptions we make in Assumption 2 (Chernozhukov et al. 2018, Künzel et al. 2019, Foster and Syrgkanis 2019, Kennedy 2020). In this section, we characterize a sample approximation of the dual-price queuing policy, and we show that it asymptotically attains the optimal value $z^\star$ of problem $(\mathcal{P})$ as the number of data samples tends to infinity.

Consider the following sample approximation of the dual problem $(\mathcal{D})$, where $\hat{m}_n^t$ represents an estimate of m^t learned from the data

$$\hat{\nu}_n^\star = \min_{\mu\in\mathbb{R}_+^{m+1}} \frac{1}{n}\sum_{i=1}^n \max_{t\in\mathcal{T}} \left(\hat{m}_n^t(x_i) - \mu^t\right) + \sum_{t\in\mathcal{T}} \mu^t b^t. \tag{$\hat{\mathcal{D}}$}$$

Problem $(\hat{\mathcal{D}})$ is a convex optimization problem and can be reformulated as a finite linear program that can easily be solved by off-the-shelf solvers. Denote by $\hat{\mu}_n^\star$ an optimal solution to problem $(\hat{\mathcal{D}})$, which exists under Assumption 2 (i) because we can restrict the feasible set of $(\hat{\mathcal{D}})$ to a compact set without loss of generality; see Lemma EC.4. We now define a sample analogue of $\pi^\star$ in (3) through

$$\hat{\pi}_n^{\star,t}(x) = \begin{cases} 1 & \text{if} \quad t = \min\arg\max_{t'\in\mathcal{T}} \left(\hat{m}_n^{t'}(x) - \hat{\mu}_n^{\star,t'}\right) \\ 0 & \text{otherwise} \end{cases} \qquad \forall t\in\mathcal{T}, \forall x\in\mathcal{X}. \tag{4}$$

Next, we define a sample-based version of the dual-price queuing policy given in Definition 1 by replacing the conditional mean treatment outcome functions m^t for each $t\in\mathcal{T}$ by their respective estimates $\hat{m}_n^t$ and an optimal dual solution $\mu^\star$ by a sample-approximate solution $\hat{\mu}_n^\star$.

Definition 2 (Sample-based dual-price queuing policy). Establish m queues each associated with a treatment $t\in\mathcal{T}\setminus\{0\}$. At any time, individuals in queue t are waitlisted for treatment t. There are two events triggering an action: (1) arrival of an individual, (2) arrival of a treatment.

(1) When an individual with covariate vector x arrives, compute the treatment $t = \min\arg\max_{t'\in\mathcal{T}} (\hat{m}_n^{t'}(x) - \hat{\mu}_n^{\star,t'})$, where t attains the maximum difference between the estimated conditional mean treatment outcome and the associated sample optimal dual variable. If $t\neq 0$, assign them to queue t, otherwise assign them to no-treatment. If they are the only individual waiting in queue t, check whether there is an available treatment of type t. If there is, assign treatment t to the individual.

(2) When a treatment of type t becomes available, assign this treatment to the first individual in the queue t. If there is no one in the queue t, keep the treatment until an individual is assigned to this queue.

Similarly to the dual-price queuing policy, the sample-based dual-price queuing policy, when implemented in a finite horizon setting, assigns no-treatment to individuals who are still waiting in queues at the end of the horizon. We also note that in general, even if the historical samples are fixed, $\hat{\mu}_n^\star$, the optimal solution to problem $(\hat{\mathcal{D}})$, need not be unique since the objective function of $(\hat{\mathcal{D}})$ may not be strictly convex and therefore the sample-based dual-price queuing policy may not be uniquely specified.

In the following, we denote by $\hat{z}_n^{\mathrm{D}}$ the expected average outcome of the sample-based dual-price queuing policy, where the expectation is taken with respect to the distribution of covariates and outcomes of the individuals arriving during implementation and *not* with respect to the distribution of the historical samples. Note that $\hat{z}_n^{\mathrm{D}}$ is a random value because it relies on the historical samples, which are random. The next theorem shows that the sample-based dual-price queuing policy attains the optimal value $z^\star$ of problem $(\mathcal{P})$ as the number of data samples tends to infinity.

Theorem 2. *Under Assumptions 1 and 2, the expected average outcome $\hat{z}_n^{\mathrm{D}}$ of the sample-based dual-price queuing policy is almost surely asymptotically at least as high as the optimal value $z^\star$ of problem $(\mathcal{P})$ as $n \to \infty$, that is,*

$$\limsup_{n\to\infty} z^\star - \hat{z}_n^{\mathrm{D}} \leq 0.$$

The idea of the proof is that if the allocations of a sample-based dual-price queuing policy asymptotically match those of a dual-price queuing policy almost surely, then the sample-based dual-price queuing policy will asymptotically generate expected average outcomes at least as good as those of a dual-price queuing policy almost surely. Since the dual-price queuing policy is optimal in problem $(\mathcal{P})$, then asymptotically $\hat{z}_n^{\mathrm{D}}$ is least as high as $z^\star$ almost surely. We show in the online appendix that the optimal solutions to problem $(\hat{\mathcal{D}})$, $\hat{\mu}_n^\star$, will be arbitrarily close to an optimal solution $\mu^\star$ of problem $(\mathcal{D})$ as the number of samples increases, $n \to \infty$. This will imply that the allocations of a sample-based dual-price queuing policy and dual-price queuing policy match almost surely.

5. Fairness Extensions

In high-stakes settings, the allocation of scarce resources needed to satisfy basic needs raises issues of fairness with respect to some protected group(s). This is especially the case when minority groups historically discriminated against due to race, gender, or other protected characteristics disproportionately make up the population supported by the allocation system. In our motivating example, fairness and equity, or the lack thereof, are key concerns for policymakers and community members (see Section 1).

In this section, we showcase fairness-constrained extensions of the dual-price queuing policies introduced in Sections 3 and 4 based on two fairness notions of statistical parity in allocation and statistical parity in outcomes. In the following subsections, we introduce the aforementioned fairness notions and present the resulting fair policies. Details of the derivations of the respective policies are relegated to Electronic Companion 5. The proposed extensions and their derivations can be similarly derived for other well-defined linear fairness constraints such as conditional statistical parity, which, for example, could require that the percentage of housing resources allocated to individuals with low VI-SPDAT scores should not exceed a predefined threshold. Alternatively, one could incorporate a fairness constraint that enforces lower bounds on the percentage of resources allocated to specific protected groups, which, for example, could require that the percentage of housing allocated to minority groups should be at least as high as the percentage allocated historically. For a comprehensive overview of possible fairness notions and constraints for scarce resource allocation systems and impossibility results between these various fairness notions, see Jo et al. (2023).

Recall from Section 2 that X can include protected features $G \in \mathcal{G}$ that represent sensitive attributes (e.g., race, gender, sex). Since the fairness notions and constraints in the following subsections will depend on G, we now explicitly consider the protected feature G as an input into the policy π, i.e., $X = (X^{-G}, G)$, where $X^{-G} \in \mathcal{X}^{-\mathcal{G}}$ collects all features of X excluding G and $\mathcal{X}^{-\mathcal{G}}$ is the set of all features of $\mathcal{X}$ excluding the set of protected features, $\mathcal{G}$.

5.1. Statistical Parity in Allocation

Statistical parity in allocation, which requires that the treatment allocation probabilities are similar across protected groups, is an intuitive and commonly used notion by stakeholders for evaluating the fairness of high-stakes allocation systems (Jo et al. 2023). Formally, a policy π_{alloc} satisfies statistical parity in allocation if it satisfies

$$\mathbb{E}\left[\pi^t_{\text{alloc}}(X^{-G}, G) \mid G = g\right] - \mathbb{E}\left[\pi^t_{\text{alloc}}(X^{-G}, G) \mid G = g'\right] \leq \delta \quad \forall g, g' \in \mathcal{G},\ g \neq g',\ t \in \mathcal{T}, \tag{5}$$

which requires that the allocation probabilities of any treatment $t \in \mathcal{T}$ are approximately equal (up to a specified tolerance level δ) across different sensitive groups. Problem ($\mathcal{P}$) can be extended to include the additional constraint (5).

As in Section 3, let $\mu \in \mathbb{R}^{m+1}_+$ collect the dual variables of the capacity constraints in ($\mathcal{P}$). We now denote by $\lambda^t(g, g') \in \mathbb{R}_+$ the dual variable of the fairness constraint (5) for the pair (g, g') and for treatment t. Following similar steps to those in Section 3, we characterize the candidate policy $\pi^\star_{\text{alloc}}$, as a counterpart of $\pi^\star$ defined in (3), through

$$\pi^{\star,t}_{\text{alloc}}(x, g) = \begin{cases} 1 & \text{if } \ t = \min \arg\max_{t' \in \mathcal{T}} \left(m^{t'}(x, g) - \mu^{\star,t'} - \frac{\gamma^{\star,t'}(g)}{\mathbb{P}(G=g)} \right) \\ 0 & \text{otherwise} \end{cases} \quad \forall t \in \mathcal{T},\ x \in \mathcal{X}^{-\mathcal{G}},\ g \in \mathcal{G},$$

where for each $g \in \mathcal{G}$ and $t \in \mathcal{T}$, $\gamma^{\star,t}(g) = \sum_{g' \in \mathcal{G}, g \neq g'} (\lambda^{\star,t}(g,g') - \lambda^{\star,t}(g',g))$ is an aggregation of dual variables across all $g' \neq g$, and $\mu^{\star,t}$ and $\lambda^{\star,t}(g,g')$ are optimal solutions to the Lagrangian dual problem of $(\mathcal{P})$ with additional fairness constraints (5). Intuitively, if $\mu^{\star,t}$ is roughly interpreted as an opportunity cost/threshold for assigning treatment t, then our new policy uses $\gamma^{\star,t}(g)$ to adjust that threshold for each group to achieve allocation parity. We can see that the adjustment $\gamma^{\star,t}(g)$ is unrestricted in sign so that it can appropriately increase (to allocate less) or decrease (to allocate more) the 'opportunity cost' of each treatment for each group. Finally, our adjustment is also inversely weighted by the population proportion of each group g to appropriately account for the protected group distribution within the population.

We define a sample analogue $\hat{\pi}^\star_{\text{alloc}}$ of $\pi^\star_{\text{alloc}}$, as a counterpart of $\hat{\pi}^\star$ in (4), through

$$(\hat{\pi}^{\star,t}_n)_{\text{alloc}}(x,g) = \begin{cases} 1 & \text{if } \ t = \min \arg\max_{t' \in \mathcal{T}} \left(\hat{m}^{t'}_n(x,g) - \hat{\mu}^{\star,t'}_n - \frac{n}{I_g} \hat{\gamma}^{\star,t'}(g) \right) \\ 0 & \text{otherwise} \end{cases} \quad \forall t \in \mathcal{T}, \ x \in \mathcal{X}^{-\mathcal{G}}, \ g \in \mathcal{G}.$$

where for each $g \in \mathcal{G}$ and $t \in \mathcal{T}$, $\hat{\gamma}^{\star,t}(g)$ is a sample analogue of $\gamma^{\star,t}(g)$, i.e., $\hat{\gamma}^{\star,t}(g) = \sum_{g' \in \mathcal{G}, g \neq g'} (\hat{\lambda}^{\star,t}_n(g,g') - \hat{\lambda}^{\star,t}_n(g',g))$, and $\hat{\mu}^{\star,t}_n$ and $\hat{\lambda}^{\star,t}_n(g,g')$ are optimal solutions to the sample-based Lagrangian dual problem of $(\mathcal{P})$ with additional fairness constraints (5). The number of individuals in our sample who belong to the group $g \in \mathcal{G}$ is denoted by the random variable I_g and so the quantity $\frac{n}{I_g}$ is the sample approximation of $\frac{1}{\mathbb{P}(G=g)}$.

Using the definition of $\hat{\pi}^\star_{\text{alloc}}$, we can construct a fairness-constrained sample-based dual-price queuing policy similarly to the Section 4. The fairness-constrained sample-based dual-price queuing policy assigns an individual with covariates (x,g) to the queue for treatment $t = \min \arg\max_{t' \in \mathcal{T}} \left(\hat{m}^{t'}_n(x,g) - \hat{\mu}^{\star,t'}_n - \frac{n}{I_g} \hat{\gamma}^{\star,t}(g) \right)$.

In certain allocation settings, policymakers may want to prioritize minority groups historically disadvantaged by racial, gender, or other forms of discrimination for allocation as a solution to mitigate inequities. While statistical parity in allocation ensures equal opportunities across protected groups, it may be insufficient for repairing existing disparities. Whether this prioritization is appropriate again depends on the fairness goals policymakers want to achieve. In this case, we can define a set $\mathcal{G}^{\min}$ of minority groups and a set $\mathcal{G}^{\text{maj}}$ of majority groups such that $\mathcal{G}^{\min}$ and $\mathcal{G}^{\text{maj}}$ are disjoint and $\mathcal{G}^{\min} \cup \mathcal{G}^{\text{maj}} = \mathcal{G}$. The goal of prioritizing allocation to minority groups could be formulated through the constraint

$$\mathbb{E}\left[\pi^t_{\text{alloc}}(X^{-G}, G) \mid G = g\right] \leq \mathbb{E}\left[\pi^t_{\text{alloc}}(X^{-G}, G) \mid G = g'\right] \quad \forall g \in \mathcal{G}^{\text{maj}}, \ g' \in \mathcal{G}^{\min}, \ t \in \mathcal{T} \setminus \{0\}, \tag{6}$$

where the average treatment assignment probability for any minority group $g' \in \mathcal{G}^{\min}$ is at least as high as the average treatment assignment probability for any majority group $g \in \mathcal{G}^{\text{maj}}$. We see that constraint (6) is a variant of (5), where we have excluded the no-treatment option of $t = 0$, set $\delta = 0$, and we enforce the constraint only for the pairs (g, g') such that $g \in \mathcal{G}^{\text{maj}}$ and $g' \in \mathcal{G}^{\text{min}}$. In this special case, our minority prioritizing allocation policy for minority groups $g' \in \mathcal{G}^{\text{min}}$ becomes

$$\pi^{\star,t}_{\text{alloc}}(x, g') = \begin{cases} 1 & \text{if } \; t = \min \arg\max_{t' \in \mathcal{T}} \left(m^{t'}(x,g) - \mu^{\star,t'} + \frac{\sum_{g \in \mathcal{G}^{\text{maj}}} \lambda^{\star,t'}(g,g')}{\mathbb{P}(G=g')} \right) \\ 0 & \text{otherwise} \end{cases} \quad \forall t \in \mathcal{T},\ x \in \mathcal{X}^{-\mathcal{G}}, \ g' \in \mathcal{G}^{\text{min}},$$

and for majority groups $g \in \mathcal{G}^{\text{maj}}$, we have

$$\pi^{\star,t}_{\text{alloc}}(x, g) = \begin{cases} 1 & \text{if } \; t = \min \arg\max_{t' \in \mathcal{T}} \left(m^{t'}(x,g) - \mu^{\star,t'} - \frac{\sum_{g' \in \mathcal{G}^{\text{min}}} \lambda^{\star,t'}(g,g')}{\mathbb{P}(G=g)} \right) \\ 0 & \text{otherwise} \end{cases} \quad \forall t \in \mathcal{T},\ x \in \mathcal{X}^{-\mathcal{G}}, \ g \in \mathcal{G}^{\text{maj}}.$$

where in the case that $t = 0$, we define $\lambda^{\star,0}(g, g') = 0$ for any pair of g and g'. In this case we see that the adjustments for all minority groups $g' \in \mathcal{G}^{\text{min}}$, given by $\sum_{g \in \mathcal{G}^{\text{maj}}} \lambda^{\star,t'}(g, g')$, are non-negative, and serve only to decrease the 'threshold' of treatment t for group g, therefore leading to more assignments. On the other-hand, since $\sum_{g' \in \mathcal{G}^{\text{min}}} \lambda^{\star,t'}(g, g')$ is also non-negative but has opposite sign, it serves to increase the 'threshold' of treatment t for all $g \in \mathcal{G}^{\text{maj}}$, thereby leading to fewer assignments.

We can similarly define a sample analogue of the minority allocation prioritization policy by replacing $m^t(x, g)$ with $\hat{m}^t_n(x, g)$ for all $t \in \mathcal{T}$ and $x \in \mathcal{X}$, $\mu^{\star,t}$ with $\hat{\mu}^{\star,t}_n$ for all $t \in \mathcal{T}$, $\lambda^{\star,t}(g, g')$ with $\hat{\lambda}^{\star,t}_n(g, g')$ for all $t \in \mathcal{T}$ and $g, g' \in \mathcal{G}$, and $\mathbb{P}(G = g)$ with $\frac{I_g}{n}$ for all $g \in \mathcal{G}$. We can then construct a minority allocation prioritization sample-based dual-price queuing policy by assigning an individual with covariates (x, g) to the queue for treatment $t = \min \arg\max_{t' \in \mathcal{T}} \left(\hat{m}^{t'}_n(x, g) - \hat{\mu}^{\star,t'}_n + \sum_{g' \in \mathcal{G}^{\text{maj}}} \hat{\lambda}^{\star,t'}_n(g', g) \frac{n}{I_g} \right)$ if $g \in \mathcal{G}^{\text{min}}$, or the queue for treatment $t = \min \arg\max_{t' \in \mathcal{T}} \left(\hat{m}^{t'}_n(x, g) - \hat{\mu}^{\star,t'}_n - \sum_{g' \in \mathcal{G}^{\text{min}}} \hat{\lambda}^{\star,t'}_n(g, g') \frac{n}{I_g} \right)$ if $g \in \mathcal{G}^{\text{maj}}$.

5.2. Statistical Parity in Outcomes

While statistical parity in allocation ensures similar chances of receiving each treatment across different protected groups, it may not be adequate for achieving fair outcomes for the individuals. If minority groups are more vulnerable to homelessness due to discrimination, then statistical parity in allocation may not address existing outcome inequities. This is precisely the case and a major concern in our motivating example, where PSH is allocated at nearly equal rates to all racial groups, but Black individuals experience higher rates of returns to homelessness compared to other racial and ethnic groups (LAHSA 2018b). For this reason, we will consider an alternative fairness notion called *statistical parity in outcomes*, which requires that a policy π_{out} satisfies constraints of the form

$$\mathbb{E}\left[\sum_{t \in \mathcal{T}} \pi^t_{\text{out}}(X^{-G}, G) m^t(X) \mid G = g\right] - \mathbb{E}\left[\sum_{t \in \mathcal{T}} \pi^t_{\text{out}}(X^{-G}, G) m^t(X) \mid G = g'\right] \le \delta \quad \forall g, g' \in \mathcal{G}, \ g \neq g'. \tag{7}$$

Constraint (7) ensures that the expected outcomes under the allocation policy π_{out} are approximately equal across different protected groups. We now denote by $\lambda(g,g') \in \mathbb{R}_+$ the dual variable of the statistical parity in outcome constraint for the group pair g and g'. Similarly to before, we denote by $\mu^\star$ and $\lambda^\star$ the optimal solutions to the Langragian dual problem of $(\mathcal{P})$ with additional statistical parity in outcomes constraints (7). Then we can similarly define a modified candidate policy by

$$\pi^{\star,t}_{\text{out}}(x,g) = \begin{cases} 1 & \text{if } \ t = \min \arg\max_{t' \in \mathcal{T}} \left(m^{t'}(x,g) \left(1 - \frac{\gamma^\star(g)}{\mathbb{P}(G=g)} \right) - \mu^{\star,t'} \right) \\ 0 & \text{otherwise} \end{cases} \quad \forall t \in \mathcal{T}, \ x \in \mathcal{X}^{-\mathcal{G}}, \ g \in \mathcal{G}, \tag{8}$$

where $\gamma^\star(g) = \sum_{g' \in \mathcal{G}, g \neq g'} (\lambda^\star(g,g') - \lambda^\star(g',g))$. Compared to statistical parity in allocation, we see now that the adjustment is directly applied to the contribution value of the conditional mean treatment outcome function m^t. The adjustment $\gamma^\star(g)$, which is unrestricted in sign and inversely weighted by the population proportion of group g, will scale up or down $m^t(x)$ for a given treatment t in order for assignments to achieve parity in outcomes. If on average one group's outcomes are not as high, we can upweight their conditional mean treatment outcomes so that individuals of that group are more likely to 'pass' the 'threshold' value $\mu^{\star,t}$. This leads individuals of that group being more likely to receive treatments, and therefore higher group-wise outcomes. As in the case of statistical parity in allocation in Section 5.1, we can similarly define sample analogue and minority prioritization versions of (8) and the respective fairness-constrained sample-based dual-price queuing policies.

6. Empirical Results

In this section, we evaluate the empirical performance of our methodology in two sets of experiments: *(1)* a synthetic example, where the data generation models for covariates, treatment assignments, and potential outcomes are known; *(2)* an example based on real HMIS and VI-SPDAT data from our central motivating application of allocating scarce housing resources to people experiencing homelessness. In both cases, we split the available data into training and testing/test sets, where the training data is used to learn the dual prices and counterfactual outcomes and the testing data is used to evaluate policy performance.

6.1. Synthetic Data Experiments

We use a simple synthetic example to evaluate the performance of our approach in a setting where counterfactuals are known and to study the effects of bias from insufficient training data, large noise terms, and model misspecification on the out-of-sample performance of our methodology. For the synthetic experiments, our primary performance metric is the ratio of the test set outcomes of the sample-based dual-price queuing policy to the outcomes of the perfect foresight test set policy (that knows both the people and resources arriving and the counterfactual outcomes). Thus, a ratio of 1 corresponds to the best possible performance. We also study the convergence of the performance of our policy to that of the perfect foresight test set policy as the training sample size is increased.

6.1.1. Data Generation. For the synthetic data experiments, we adapt the data generation process of Jo et al. (2021) by increasing the number of treatments from two to three to test our methodology in the multiple treatments case. In this setting, there are three possible treatments, i.e., $\mathcal{T} = \{0,1,2\}$, and each individual is characterized by two covariates, each following a standard normal distribution, i.e., $X = [X^1, X^2]$ and $X^1, X^2 \sim \mathcal{N}(0,1)$. The potential outcomes Y^t, $t \in \mathcal{T}$, are expressible as

$$Y^t = \frac{1}{2}X^1 + X^2 + (2 \cdot \mathbb{1}[t=1] - 1)\,\frac{1}{4}X^1 + (2 \cdot \mathbb{1}[t=2] - 1)\,\frac{1}{4}X^2 + \epsilon_t,$$

where $\epsilon_t \sim \mathcal{N}(0,\sigma)$, $t \in \mathcal{T}$, are i.i.d. noise terms. In our experiments, we will vary σ in the set $\{0.1, 0.5, 0.8, 1.15, 1.5\}$ to study the impact on our method of increasing noise. Indeed, increasing the standard deviation of the noise impacting Y^t, $t \in \mathcal{T}$, is expected to affect the quality of the estimates of m^t and μ^t used by our sample-based dual-price queuing policy. From the above, in our notation of conditional mean treatment outcomes, we see that

$$m^0(X) = \frac{1}{4}X^1 + \frac{3}{4}X^2, \quad m^1(X) = \frac{3}{4}X^1 + \frac{3}{4}X^2, \quad \text{and} \quad m^2(X) = \frac{1}{4}X^1 + \frac{5}{4}X^2.$$

The historical/training set treatment assignments T come from a discrete distribution that, in line with our assumptions, depends on the covariates X only, i.e., $p(X) = [p^0(X), p^1(X), p^2(X)]$, where $p^t(x)$ is the assignment probability of treatment t given $X = x$. In particular, we let treatment assignment probabilities depend only on the treatment that is the best in expectation. This partitions the covariate space into three disjoint regions, each with associated values for $p(x)$: *(1)* when treatment 0 is the best (or at least as good) in expectation amongst the treatments (i.e., $0 = \min \arg\max_{t \in \mathcal{T}} m^t(x)$), we set $p(x) = [0.8, 0.1, 0.1]$; *(2)* when treatment 1 is the best (or at least as good) in expectation amongst the treatments (i.e., $1 = \min \arg\max_{t \in \mathcal{T}} m^t(x)$), we set $p(x) = [0.6, 0.3, 0.1]$; and *(3)* when treatment 2 is the best (or at least as good) in expectation amongst the treatments (i.e., $2 = \min \arg\max_{t \in \mathcal{T}} m^t(x)$), we set $p(x) = [0.6, 0.1, 0.3]$. This modeling choice mimics realistic scenarios where we anticipate that policy-/decision-makers frequently assign treatments “correctly.” In addition, for cases *(2)* and *(3)* where treatments 1 and 2 are the best in expectation and there is a higher probability of receiving those treatments relative to the probabilities in case *(1)*, treatment 0 still has the highest probability at 60%. This modeling choice mimics realistic scenarios where there is a default no-treatment scenario (treatment 0) and there are limited quantities of treatments 1 and 2. Finally, we assume that the treatment capacities are known and given by $b^0 = 1$, $b^1 = 0.1$, and $b^2 = 0.05$.

To evaluate the performance of our method on this data, we run 25 simulations. Each of these involves $360,000$ test set samples (individuals) as a proxy for the asymptotic setting, and a number of training samples that we vary from 500 to $220,000$, though we observe convergence much earlier with fewer training samples.

6.1.2. Choice of Outcome Estimators. Since the asymptotic optimality of our policy π depends on the uniform almost sure convergence of $\hat{m}^t$ to m^t (see Assumption 2 (ii)), we study the robustness of our method to Assumption 2 (ii) by using a variety of estimators for learning the mean treatment outcome models. If an incorrect parametric model is employed when learning m^t, $t \in \mathcal{T}$, then such estimation errors may cause suboptimal policy learning. From the data generation process, we know that the true m^t functions are linear, see Section 6.1.1. Thus, we can use decision trees and lasso regression as our misspecified models and k-NN as a non-parametric model with asymptotic consistency. Finally, we investigate if causal inference methods for ameliorating functional form bias in treatment effects estimation can improve policy learning in the presence of estimation errors of treatment outcomes. In particular, we use inverse propensity score weighted (IPW) and doubly robust (DR) variations of our misspecified models (decision trees and lasso regression) to see if they can mitigate bias from estimation errors. Since IPW and DR require estimating propensity scores, we use decision trees and logistic regression as the model classes for propensity score estimation. For an introduction to these methods, please see Battocchi et al. (2019).

6.1.3. Performance and Discussion. Our main results are summarized in Figure 1, where we plot our main performance metric under policies using various estimators of treatment outcomes. The figure shows our policy performance metric for each estimator in dependence of training sample size, with the left (resp. right) figure focused on the low (resp. high) variance case where $\sigma = 0.1$ (resp. $\sigma = 1.5$). For brevity, in the rest of this section we refer to various dual-price queuing policies using different estimators of the conditional mean treatment outcomes by just the estimator name. For example, the linear, lasso, decision tree policies in Figure 1 are the dual-price queuing policies based on 'directly' estimating the conditional mean outcomes using linear, lasso, and decision tree models, where linear is the correct model form while decision trees and lasso are misspecified models. The policies that start with 'IPW' and 'DoublyRobust' in their names are based on estimators of treatment outcomes that attempt to correct the bias of the lasso based policy by IPW or DR estimates of treatment outcomes. Finally, we have omitted from these results the performance of a 'random' policy that randomly assigns treatments as they arrive to waitlisted

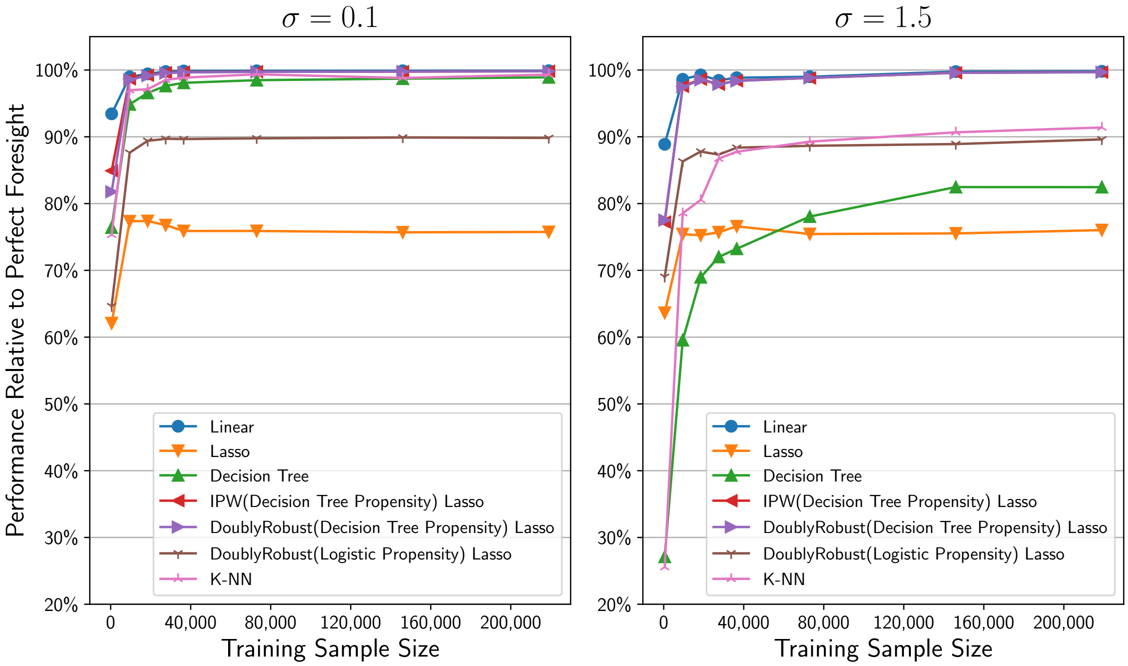


Figure 1 Companion figure to the synthetic data results for $\sigma = 0.1$ and $\sigma = 1.5$. Both subfigures show the ratio of out-of-sample performance of the sample-based queuing policy to that of the perfect foresight policy in dependence of training set size. Each line corresponds to a policy using a different treatment outcome estimator.

individuals (which performed near 0% irrespective of training sample size). Additional results for σ values of $\{0.5, 0.8, 1.15\}$ can be found in Electronic Companion EC.3.1.

From Figure 1 (left), we see that, in the low variance setting, policy performance converges relatively fast (with under $9,000$ training samples). This is expected: since observed outcomes affect the estimation of counterfactuals, which in turn affect the estimation of policy parameters $\hat{\mu}^{\star}$, low noise in the observed outcomes leads to faster convergence of policy performance. As we would expect, the linear estimator policy attains close to perfect performance since it correctly estimates the true conditional mean treatment outcomes. On the other hand, the various modified lasso policies appear to mitigate the estimation bias of the direct lasso estimates and improve policy performance. However, these results also show that the choice of propensity score estimates used to modify the direct treatment outcome estimates can impact policy performance. We see that policies using decision tree estimated propensity scores resulted in better convergence of policy performance (close to the linear policy performance) compared to policies using logistic regression estimated propensity scores. The worse performance of policies using logistic regression estimated propensity scores, compared to that of policies using decision tree estimated propensity scores, may result from the non-linear relation between treatment assignment probabilities and covariates. We emphasize that selecting an accurate propensity score model (even when the historical treatment assignment policy is unknown) is usually feasible in practice with modern machine learning approaches. In this low noise setting, non-parametric estimator based policies such as k-NN can effectively estimate the counterfactuals well enough to achieve close to perfect performance. Thus, as expected, in the direct estimator case, our method's performance converges to 1 if the model employed is correct. On the otherhand, in policies using IPW or DR adjustments, the choice of propensity score model and estimates impacts the convergence of policy performance. With a nonparametric estimator, convergence to 1 also occurs.

From Figure 1 (right), we see that in the high variance setting, the policies based on the non-parametric estimators of m_t all have a deterioration in performance relative to the low variance setting. In particular, the policy using k-NN estimators exhibits slower performance convergence as the high variance noise terms causes the k-NN mean outcome estimation error to decrease at a slower rate. On the other hand, compared to the low variance setting, the simpler parametric lasso model maintains a similar, though suboptimal, performance. The lasso model is more 'robust', in some sense, to more noise as both its mean treatment outcome estimation error and overall policy performance do not deteriorate much from the low to high variance setting. Finally, we see similar behavior in terms of improvement in policy performance from the IPW and DR correction methods to the lasso policy, though the degree of bias correction of IPW and DR methods depend on the estimated propensity scores used. These simple simulation results highlight the importance of model specification for counterfactual estimation and impact on model performance, and potential bias improvements we can make from reweighting or doubly robust estimation of treatment outcomes.

6.2. Real Data Experiments: Allocating PSH and RRH Housing in LA

We next showcase the performance of our methodology on real data to design policies for allocating scarce resources to individuals experiencing homelessness in LA. We obtained our data from LAHSA through a Data Use Agreement and our data comes from the LA County Homeless Management Information System (HMIS) database (LAHSA 2015) and administered VI-SPDAT assessments (LAHSA 2017). We obtained Institutional Review Board approval for our data analysis. Additional details on the data sources and preprocessing can be found in Electronic Companion EC.3.2.

6.2.1. HMIS and VI-SPDAT Data Description. The HMIS data contains deanonymized data of individuals experiencing homelessness and their interactions with the system, which we refer to as *enrollments*. Each enrollment represents an instance in which an individual is assigned a specific resource, where the resource types (introduced in Section 1) are `PSH`, `RRH`, and `SO`. Note that `SO` is a broad category that includes various services that are not considered permanent housing such as street outreach, homelessness prevention, emergency shelter, supportive services like childcare, and more. We also further breakout `PSH` into two sub-types `PSH Tenant-Based` and `PSH`

Site-Based that differ, broadly speaking, in terms of supportive services available on-site and process of actually receiving housing. Further details can be found in in Electronic Companion EC.3.2. Since an individual can have multiple enrollments, we observe for each individual and enrollment pair the relevant date, and assigned resource type. The administered VI-SPDAT surveys, which are used for assessing individual vulnerability and resource prioritization, contain individual covariates, such as disabilities and prior housing history at the time of assessment.

We include all individuals with an assessment between the time period of 1/12/2015 (roughly when VI-SPDAT started being administered) to 12/31/2019 (to avoid idiosyncratic effects of Covid-19 during 2020) for a total of $63,764$ samples. We use all individuals assessed between 1/12/2015 to 12/31/2017 as our training set to learn treatment outcomes and construct our policy, and then evaluate on individuals assessed between 1/1/2018 to 12/31/2019 to measure out-of-sample performance. We have a total of 4 possible resources, i.e., $\mathcal{T} = \{0, 1, 2, 3\}$, which represent no-treatment, RRH, PSH Tenant-Based, and PSH Site-Based. To solve problem ($\hat{\mathcal{D}}$) and derive a sample-based queuing policy, we use the training set capacities for no-treatment, RRH, PSH Tenant-Based, and PSH Site-Based as the estimates of capacity per person, which are 100%, 14.2%, 4.1%, and 4.6%, respectively. Since racial equality and equity are important concerns in our motivating example in Section 1, we use race as the protected characteristic for evaluating allocation and outcome fairness. The protected features set is $\mathcal{G} = \{\texttt{BlackAfAmerican}, \texttt{Hispanic}, \texttt{Other}, \texttt{White}\}$, where BlackAfAmerican stands for Black or African American, and Other includes all other individuals who identify as American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, more than one racial group, or answered ‘Doesn’t know’. In this example, we consider non-White individuals to be minority groups that policymakers may want to prioritize due to historical discrimination.

Outcome Definition. According to the U.S. Department of Housing and Urban Development (HUD), the LA CoC should develop “action steps to end homelessness and prevent a return to homelessness” (LAHSA 2023). In line with this goal, we will use an individual’s return to homelessness after an intervention as the outcome to measure the performance of a policy. This outcome will serve as an input to our optimization model for designing a dual-price policy for allocating scarce housing resources effectively. Specifically, we focus on returns to homelessness within a two-year window following intervention for our real data experiments. However ‘returns to homelessness’ for individuals are not explicitly tracked. Instead, we construct a proxy outcome variable where an individual ‘returned to homelessness’ following intervention if we observe a subsequent enrollment into ‘emergency shelter’, ‘safe haven’, or ‘street outreach’. We chose these enrollment categories based on discussions with domain experts like matchers and case managers working within the LA CES since these types of interactions with the system indicate a need for homeless services and suggest an individual returned to homelessness again. For individuals without a treatment enrollment into RRH or PSH, we treat their first enrollment into a SO resource as the start of a no-treatment 'intervention'. Therefore, we define a positive outcome, $Y=1$, as *not* observing a subsequent return to homelessness within a two-year window of receiving intervention, and $Y=0$ otherwise. While other observation window lengths could be chosen, there is a trade-off in terms of window length. Shorter windows are potentially biased proxies since there is not enough time to observe post-intervention outcomes while longer windows result in less individuals for whom we can observe the entire window in our dataset.

Policy Evaluation. We learn three different versions of our proposed policy from the training set and evaluate their performance by their overall proportion of positive outcomes as the 'effectiveness' metric and their fairness properties in terms of statistical parity in allocation (5) and outcomes (7) on the test set. To investigate the fairness properties, we look at discrepancies in the percentage receiving housing resources and discrepancies in the proportion of positive outcomes between racial groups in the out-of-sample allocation. In our results, Base refers to the policy without fairness constraints, while the other two refer to fairness-constrained variations defined in Section 5. Based on our motivating example, we chose to focus on fairness extensions that prioritized minority groups in terms of either allocation, Alloc Min Priority, or outcomes, Outcome Min Priority. To benchmark performance, we compare the three policies to the Historical policy, which uses the same assignments originally found in the data, in terms of overall proportion of positive outcomes and the two statistical parity metrics. This way we can see if our proposed policies can improve on the 'effectiveness' and fairness of the currently deployed policy. Finally we also compare performance of overall outcomes to a Perfect Foresight policy that implements the sample-based dual-price queuing policy while knowing both the test set distribution of arriving individuals and their counterfactuals.

6.2.2. Counterfactuals and Outcome Estimators. Unlike the synthetic experiment, we do not have access to counterfactuals for each individual under treatments other than the one received. Therefore, we take a semi-synthetic approach and use model generated counterfactuals for the test set by choosing models that are well calibrated to the entire data, therefore including data not used by the training estimators. We choose calibration as the measure of fit since we are evaluating based on expected outcomes of a binary variable, and therefore we want our predicted probabilities of a positive outcome to match the observed data. Our choice of models include logistic regression (with and without regularization), decision trees, random forests, gradient boosted trees (Friedman 2001), and XGBoost trees (Chen and Guestrin 2016). For model selection, we used standard cross-validation to find the best hyperparameters for each model class that minimizes log-loss and then select the final model based on calibration curves on a held-out validation set. An additional concern of our model generated counterfactuals was potential bias by race since any calibration bias by race would mean our evaluation results could be potentially biased by race. Though this was not a part of our model selection process, we investigated potential racial bias by comparing calibration curves for each racial group across treatments using our selected counterfactual generating models. For the most part, the calibration curves across racial groups are similar except for the 'Other' racial group, likely because this group contained too few samples to get accurate estimates. While the above addresses generating counterfactuals for evaluation, we take a similar approach for actually learning $\hat{m}^t$ functions but *only* use the training data. For more details on counterfactual generation, model calibration, and racial fairness of the estimates, please see the Electronic Companions EC.3.2.2 and EC.3.2.3.

6.2.3. Experimental Results. Our results are summarized in Table 1, which shows the performance of the proposed policies by their out-of-sample expected outcomes. Figure 2 illustrates the allocation fairness of each policy by showing the proportion of each racial group receiving `RRH` and `PSH` and the outcome fairness of each policy by showing the expected outcome within each racial group.

In terms of the out-of-sample expected outcomes, we see in Table 1 that our base model achieves an improvement of 1.9% above the `Historical` policy while having a suboptimality gap of 1.05% compared to the `Perfect Foresight` policy. We found that the performance gap comes from two sources of error: *(i)* estimation error of treatment outcomes $\hat{m}^t$, which are plug-ins into the assignment policy, and *(ii)* shifts in the distribution of individuals we observe and capacities per person for each treatment from training to testing, which affects the optimal value of $\mu^\star$ and any estimates $\hat{\mu}^\star$ of it. These sources of error could come from insufficient training samples for estimating outcomes or observing the distribution of individuals, or from violations of assumptions needed such as stationary distribution from training to testing, or positivity assumption for causal inference. Practically speaking, we see that the `Historical` policy already performs rather well compared to the `Perfect Foresight` policy given the scarce amount of resources. An additional improvement of 1.9% above `Historical`, on the other hand, would lead to hundreds more individuals exiting homelessness. This shows our proposed policy can be a *systematic* method to assist and supplement the experience and knowledge of domain experts.

While `Base` improves upon the expected outcomes of `Historical`, `Base` has worse allocation fairness properties than `Historical` when considering the proportion of minority groups receiving `RRH` and `PSH` as displayed in Figure 2. However, by adding appropriate fairness constraints, the `Alloc Min Priority` policy is able to mitigate this problem and achieve better prioritization of

Table 1 **Companion table to the real data experiment showing out-of-sample proportion of positive outcomes under each policy.**

Policy	Proportion of Positive Outcomes
`Historical`	60.35%
`Base`	62.25%
`Alloc Min Priority`	62.24%
`Outcome Min Priority`	62.25%
`Perfect Foresight`	63.30%

minority groups for allocations. The first subplot of Figure 2 shows the proportion of each racial group receiving `RRH` under each policy. We see that the `Base` and `Outcome Min Priority` policies actually worsen the `RRH` allocation disparity gap between `White` vs `Hispanic` and `Other` groups compared to the `Historical`. While `Alloc Min Priority` does not satisfy the minority allocation prioritization fairness constraint presented in Section 5.1 out-of-sample either, qualitatively the disparity between `White` vs `Hispanic` and `Other` groups is similar to that of `Historical` and much smaller than `Base`. Looking at `PSH` allocations in the second subplot of Figure 2, we see that the `Historical` policy does not prioritize the `Hispanic` and `Other` racial groups to be at least as likely to receive `PSH` as the majority group. Again, our `Base` and `Outcome Min Priority` policies allocate even less to `Hispanic` and `Other` individuals relative to `White` individuals. However, if we explicitly add allocation fairness constraints to our sample-based problem to prioritize minority groups for receiving `PSH` as shown in Section 5.1, then our `Alloc Min Priority` is able to achieve our desired fairness goal.

In the last subplot of Figure 2, we show the proportion of positive outcomes within each racial group for each policy and focus on the prioritization of minority groups to have groupwise outcomes that are at least as good as the majority group. In this case, the `Historical` policy already achieves this notion of fairness and `Base` does not create outcome disparities between minority groups compared to the majority group. The `Outcome Min Priority` policy, which explicitly includes outcome fairness constraints in the sample problem, also achieves the desired fairness goal while also improving `BlackAfAmerican`, `Hispanic`, and `White` individual outcomes relative to `Historical` (this is also true for `Base`). We also see in these results that `Alloc Min Priority` has favorable fairness properties by achieving or almost achieving both fairness notions of prioritization of minority groups for allocation and outcomes, though this need not hold in other problems. In general, allocation statistical parity based fairness notions and constraints are incompatible with outcome

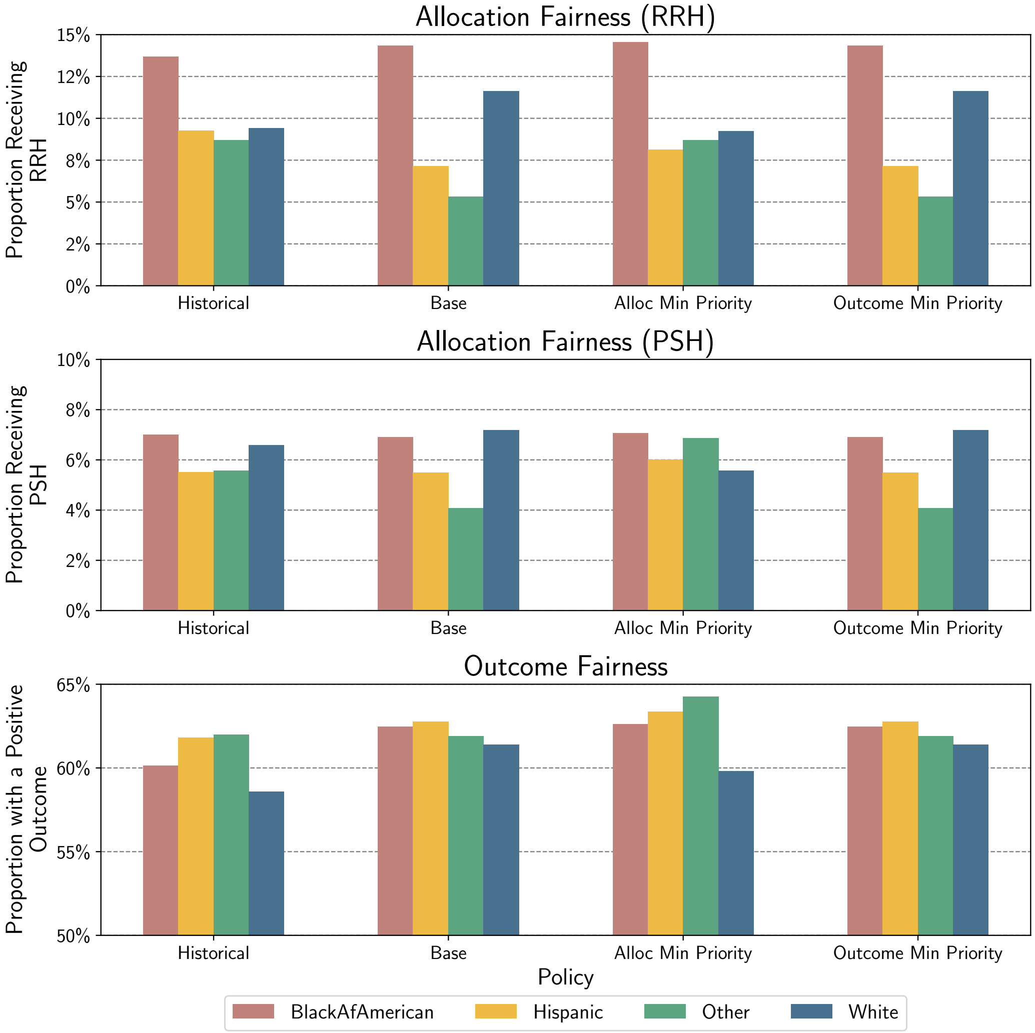


Figure 2 **Companion figure to the real data experiments: Out-of-sample fairness results showing the proportion of each racial group receiving** `RRH`**,** `PSH` **(Allocation Fairness) and the proportion of each racial group with a positive outcome (Outcome Fairness).**

statistical parity based fairness (Jo et al. 2023). Therefore, depending on the desired fairness goals, policymakers may need to choose between alternative fairness notions, or choose a combination of them to achieve a compromise.

From Figure 2, we can conclude that fairness-constrained policies such as `Alloc Min Priority` and `Outcome Min Priority` can improve the fairness properties of the `Base` policy, which actually worsened the allocation disparities of `Historical`. More interestingly, we see in Table 1 that our fairness-constrained policies suffer almost no performance gap in terms of expected outcomes compared to `Base`, suggesting almost no ‘price of fairness’. Intuitively, we would expect some loss of performance by enforcing certain constraints on allocations. This may either be due to this application or the flexibility of our methodology to find well-performing policies that satisfy various fairness goals.

In summary, our proposed policy is able to achieve an improvement above the historical policy, while also being flexible enough to achieve a desired fairness goal in allocation or outcome without suffering much in performance. Compared to the historical allocation, which depended on varying individual decision making, our proposed policy can serve as a standardized allocation process or decision making tool.

Disclaimer

The views and opinions of the data presented belong to the authors and do not represent the views or opinions of LAHSA.

Acknowledgments

P. Vayanos is funded in part by the National Science Foundation under CAREER award number 2046230. B. Tang is funded in part by the National Science Foundation Graduate Research Fellowship Program (GRFP). E. Rice, P. Vayanos, and B. Tang are funded in part from the Hilton C. Foundation, the Homeless Policy Research Institute, and the Home for Good foundation under the "C.E.S. Triage Tool Research & Refinement" grant. The authors are grateful for the support.

References

Adelman D (2007) Dynamic bid prices in revenue management. *Operations Research* 55(4):647–661.

Athanassoglou S, Sethuraman J (2011) House allocation with fractional endowments. *International Journal of Game Theory* 40:481–513.

Athey S, Wager S (2021) Policy learning with observational data. *Econometrica* 89(1):133–161.

Azizi MJ, Vayanos P, Wilder B, Rice E, Tambe M (2018) Designing fair, efficient, and interpretable policies for prioritizing homeless youth for housing resources. *Integration of Constraint Programming, Artificial Intelligence, and Operations Research: 15th International Conference*, 35–51.

Battocchi K, Dillon E, Hei M, Lewis G, Oka P, Oprescu M, Syrgkanis V (2019) EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation. Https://github.com/py-why/EconML.

Bertsekas D (2009) *Convex Optimization Theory*. Athena Scientific Optimization and Computation Series (Athena Scientific), ISBN 9781886529311.

Bertsimas D, Dunn J, Mundru N (2019) Optimal prescriptive trees. *INFORMS Journal on Optimization* 1(2):164–183.

Bertsimas D, Farias VF, Trichakis N (2013) Fairness, efficiency, and flexibility in organ allocation for kidney transplantation. *Operations Research* 61(1):73–87.

Bhattacharya D, Dupas P (2012) Inferring welfare maximizing treatment assignment under budget constraints. *Journal of Econometrics* 167(1):168–196.

Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, Robins J (2018) Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal* 21(1):C1–C68, ISSN 1368-4221.

Corbett-Davies S, Pierson E, Feller A, Goel S, Huq A (2017) Algorithmic decision making and the cost of fairness. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 797–806.

Dickerson J, Sandholm T (2015) Futurematch: Combining human value judgments and machine learning to match in dynamic environments. *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, 622–628.

Dudík M, Langford J, Li L (2011) Doubly robust policy evaluation and learning. *Proceedings of the 28th International Conference on Machine Learning*, 1097–1104.

Foster DJ, Syrgkanis V (2019) Orthogonal statistical learning. *arXiv preprint arXiv:1901.09036* .

Freeman R, Shah N, Vaish R (2020) Best of both worlds: Ex-ante and ex-post fairness in resource allocation. *Proceedings of the 21st ACM Conference on Economics and Computation*, 21–22.

Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Annals of Statistics* 29(5):1189–1232.

Gneiting T, Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* 102(477):359–378.

Grand-Clément J, Chan CW, Goyal V, Escobar G (2022) Robustness of proactive intensive care unit transfer policies. *Operations Research* (In press).

Hernan M, Robins J (2023) *Causal Inference: What If*. Boca Raton: Chapman & Hall (CRC Press).

Jo N, Aghaei S, Gómez A, Vayanos P (2021) Learning optimal prescriptive trees from observational data. *arXiv preprint arXiv:2108.13628* .

Jo N, Tang B, Dullerud K, Aghaei S, Rice E, Vayanos P (2023) Fairness in contextual resource allocation systems: Metrics and incompatibility results. *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 11837–11846.

Johnston CM, Blessenohl S, Vayanos P (2020) Preference elicitation and aggregation to aid with patient triage during the covid-19 pandemic. *Workshop on Participatory Approaches to Machine Learning.*

Kallus N (2017) Recursive partitioning for personalization using observational data. *Proceedings of the 34th International Conference on Machine Learning*, 1789–1798.

Kaya YB, Maass KL, Dimas GL, Konrad R, Trapp AC, Dank M (2022) Improving access to housing and supportive services for runaway and homeless youth: Reducing vulnerability to human trafficking in new york city. *IISE Transactions* 1–15.

Kennedy EH (2020) Towards optimal doubly robust estimation of heterogeneous causal effects. *arXiv preprint arXiv:2004.14497* .

Kitagawa T, Tetenov A (2018) Who should be treated? Empirical welfare maximization methods for treatment choice. *Econometrica* 86(2):591–616.

Kube A, Das S, Fowler PJ (2019) Allocating interventions based on predicted outcomes: A case study on homelessness services. *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*, volume 33, 622–629.

Kunnumkal S, Talluri K (2016) On a piecewise-linear approximation for network revenue management. *Mathematics of Operations Research* 41(1):72–91.

Künzel SR, Sekhon JS, Bickel PJ, Yu B (2019) Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences* 116(10):4156–4165.

Li D, Pang Z, Qian L (2023) Bid price controls for car rental network revenue management. *Production and Operations Management* 32(1):261–282.

Liero H (1989) Strong uniform consistency of nonparametric regression function estimates. *Probability Theory and Related Fields* 82(4):587–614.

Los Angeles Homeless Services Authority (2015) LA HMIS Poplices and Procedures. `https://www.lahsa.org/documents?id=1128-la-hmis-policies-and-procedures.pdf&ref=hmis`.

Los Angeles Homeless Services Authority (2017) CES Survey for Individuals Survey Packet. `https://www.lahsa.org/documents?id=1306-form-1306-ces-survey-for-individuals-survey-packet.pdf`.

Los Angeles Homeless Services Authority (2018a) An Introduction to the Coordinated Entry System & How to Conduct the CES Triage Tools. `http://publichealth.lacounty.gov/sapc/docs/providers/special-programs/Coordinated%20Entry%20HMIS%20Training.pdf`.

Los Angeles Homeless Services Authority (2018b) Report and Recommendations of the Ad Hoc Committee on Black People Experiencing Homelessness. `https://www.lahsa.org/documents?id=2823-report-and-recommendations-of-the-ad-hoc-committee-on-black-people-experiencing-homelessness.pdf`.

Los Angeles Homeless Services Authority (2022a) 2022 Greater Los Angeles Homeless Count. `https://www.lahsa.org/documents?id=6545-2022-greater-los-angeles-homeless-count-deck.pdf`.

Los Angeles Homeless Services Authority (2022b) 2022 Housing Inventory Count. `https://www.lahsa.org/documents?id=6544-2022-housing-inventory-count.xlsx`.

Los Angeles Homeless Services Authority (2023) Los Angeles Continuum of Care. `https://www.lahsa.org/coc/`.

Manshadi V, Niazadeh R, Rodilitz S (2021) Fair dynamic rationing. *Proceedings of the 22nd ACM Conference on Economics and Computation*, 694–695.

Mashiat T, Gitiaux X, Rangwala H, Fowler P, Das S (2022) Trade-offs between group fairness metrics in societal resource allocation. *ACM Conference on Fairness, Accountability, and Transparency*, 1095–1105.

Milburn NG, Edwards E, Obermark D, Rountree J (2021) Inequity in the permanent supportive housing system in los angeles: Scale, scope and reasons for black residents' returns to homelessness. `https://www.capolicylab.org/wp-content/uploads/2021/10/Inequity-in-the-PSH-System-in-Los-Angeles.pdf`.

Nguyen Q, Das S, Garnett R (2021) Scarce societal resource allocation and the price of (local) justice. *Proceedings of the AAAI Conference on Artificial Intelligence*, 5628–5636.

OrgCode (2015) Vulnerability index - service prioritization decision assistance tool (vi-spdat): Prescreen triage tool for single adults. `https://www.havenforhope.org/wp-content/uploads/2018/11/VI-SPDAT-v2.01-Family-US-Fillable.pdf`.

OrgCode (2020) The Time Seems Right: Let's Begin the End of the VI-SPDAT. `https://www.orgcode.com/blog/the-time-seems-right-lets-begin-the-end-of-the-vi-spdat`.

OrgCode Consulting, Inc and Community Solutions (2015) Changes in the VI-SPDAT v2.0. `https://www.ncceh.org/media/files/page/359d0ab6/VI-SPDAT_v2_Backgrounder.pdf`.

Qian M, Murphy SA (2011) Performance guarantees for individualized treatment rules. *Annals of Statistics* 39(2):1180.

Radovanovic D, Pini S, Franceschi E, Pecis M, Airoldi A, Rizzi M, Santus P (2021) Characteristics and outcomes in hospitalized covid-19 patients during the first 28 days of the spring and autumn pandemic waves in milan: an observational prospective study. *Respiratory Medicine* 178:106323.

Rahmattalabi A, Jabbari S, Lakkaraju H, Vayanos P, Izenberg M, Brown R, Rice E, Tambe M (2021) Fair influence maximization: a welfare optimization approach. *Proceedings of the AAAI Conference on Artificial Intelligence*, 11630–11638.

Rahmattalabi A, Vayanos P, Dullerud K, Rice E (2022) Learning resource allocation policies from observational data with an application to homeless services delivery. *ACM Conference on Fairness, Accountability, and Transparency*, 1240–1256.

Shapiro A, Dentcheva D, Ruszczynski A (2014) *Lectures on Stochastic Programming: Modeling and Theory, Second Edition* (SIAM).

Swaminathan A, Joachims T (2015) Batch learning from logged bandit feedback through counterfactual risk minimization. *The Journal of Machine Learning Research* 16(1):1731–1755.

Talluri K, Van Ryzin G (1998) An analysis of bid-price controls for network revenue management. *Management Science* 44(11-part-1):1577–1593.

Talluri KT, Van Ryzin G (2004) *The Theory and Practice of Revenue Management*, volume 1 (Springer).

US Department of Housing and Urban Development (2021) FY 2022 HMIS Data Standards Data Dictionary. `https://files.hudexchange.info/resources/documents/HMIS-Data-Dictionary.pdf`.

US Department of Housing and Urban Development Office of Policy Development and Research (2007) The Applicability of Housing First Models to Homeless Persons with Serious Mental Illness. `https://www.huduser.gov/portal/publications/hsgfirst.pdf`.

Wager S, Athey S (2018) Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association* 113(523):1228–1242.

Zhang D, Weatherford L (2017) Dynamic pricing for network revenue management: A new approach and application in the hotel industry. *INFORMS Journal on Computing* 29(1):18–35.

Zhou Z, Athey S, Wager S (2022) Offline multi-action policy learning: Generalization and optimization. *Operations Research* 71(1):148–183.

Online Appendix

EC.1. Proofs

In this section, we provide the proofs of the results in Sections 3 and 4.

EC.1.1. Proofs for Section 3

The proof of Theorem 1 requires the following two technical lemmas.

LEMMA EC.1. *For any $\mu \in \mathbb{R}_+^{m+1}$, denote by μ^{-t} its subvector $(\mu^0, \ldots, \mu^{t-1}, \mu^{t+1}, \ldots, \mu^m) \in \mathbb{R}_+^m$ excluding μ^t, and define the random variable $Z^{\mu^{-t}} = m^t(X) - \max_{t' \in \mathcal{T} \setminus \{t\}} (m^{t'}(X) - \mu^{t'})$ for all $t \in \mathcal{T}$. Under Assumption 1, $Z^{\mu^{-t}}$ is continuously distributed for any $t \in \mathcal{T}$ and $\mu^{-t} \in \mathbb{R}_+^m$.*

Proof. We will show that $\mathbb{P}(Z^{\mu^{-t}} = z) = 0$ for all $t \in \mathcal{T}$, $\mu^{-t} \in \mathbb{R}_+^m$, and $z \in \mathbb{R}$. To this end, fix arbitrary $t \in \mathcal{T}$, $\mu^{-t} \in \mathbb{R}_+^m$, and $z \in \mathbb{R}$. We have $\mathbb{P}(Z^{\mu^{-t}} = z) \geq 0$ so we only need to show that $\mathbb{P}(Z^{\mu^{-t}} = z) \leq 0$. We have

$$\begin{aligned}\mathbb{P}(Z^{\mu^{-t}} = z) &= \mathbb{P}(m^t(X) - \max\nolimits_{t' \in \mathcal{T} \setminus \{t\}} (m^{t'}(X) - \mu^{t'}) = z) \\ &= \mathbb{P}(\min\nolimits_{t' \in \mathcal{T} \setminus \{t\}} (m^t(X) - m^{t'}(X) + \mu^{t'}) = z) \\ &\leq \mathbb{P}(\cup_{t' \in \mathcal{T} \setminus \{t\}} \{m^t(X) - m^{t'}(X) + \mu^{t'} = z\}) \\ &\leq \sum_{t' \in \mathcal{T} \setminus \{t\}} \mathbb{P}(m^t(X) - m^{t'}(X) + \mu^{t'} = z) = 0,\end{aligned}$$

where the first inequality holds because the probability of $\min_{t' \in \mathcal{T} \setminus \{t\}} (m^t(X) - m^{t'}(X) + \mu^{t'}) = z$ is less than or equal to the probability that $m^t(X) - m^{t'}(X) + \mu^{t'} = z$ for at least one $t' \in \mathcal{T}$, the second inequality follows from the union bound, and the last equality holds because $m^t(X) - m^{t'}(X)$ is continuously distributed for all t, $t' \in \mathcal{T}$ by Assumption 1 (ii). □

In the remainder, we denote by $\mathbb{P}_{\mathrm{x}}$ the marginal distribution of X and by $\mathbb{P}_{\mathrm{z}^{\mu^{-t}}}$ the marginal distribution of $Z^{\mu^{-t}}$ defined in Lemma EC.1 when necessary for clarity. The next lemma shows that the objective function of the dual problem ($\mathcal{D}$) is differentiable under Assumption 1, and this result will later allow us to establish the optimality conditions for problem ($\mathcal{D}$).

LEMMA EC.2. *Define the auxiliary function $f : \mathbb{R}^{m+1} \to \mathbb{R}$ through*

$$f(\mu) = \mathbb{E}\left[\max_{t' \in \mathcal{T}} \left(m^{t'}(X) - \mu^{t'}\right)\right] = \int_{\mathcal{X}} \max_{t' \in \mathcal{T}} \left(m^{t'}(x) - \mu^{t'}\right) \mathrm{d}\mathbb{P}_{\mathrm{x}}(x), \qquad \text{(EC.1)}$$

which appears in the objective function of the dual problem ($\mathcal{D}$). Under Assumption 1, the partial derivative of function f with respect to μ^t, $t \in \mathcal{T}$, is given by

$$\frac{\partial}{\partial \mu^t} f(\mu) = -\mathbb{P}_{\mathrm{x}}\left(m^t(X) - \mu^t \geq \max_{t' \in \mathcal{T} \setminus \{t\}} \left(m^{t'}(X) - \mu^{t'}\right)\right).$$

Furthermore, each partial derivative is continuous in μ, and f is therefore differentiable.

Proof. We first derive the partial derivative of f with respect to each μ^t and then show that each partial derivative is continuous in $\mu \in \mathbb{R}_+^{m+1}$. Given a real number a, we use a^+ to denote the positive part function $\max(a, 0)$. For any $t \in \mathcal{T}$, the partial derivative of f with respect to μ^t is given by

$$\begin{aligned}
\frac{\partial}{\partial \mu^t} f(\mu) \quad &= \frac{\partial}{\partial \mu^t} \int_{\mathcal{X}} \max_{t' \in \mathcal{T}} \left(m^{t'}(x) - \mu^{t'} \right) \mathrm{d}\mathbb{P}_{\mathrm{x}}(x) \\
&= \frac{\partial}{\partial \mu^t} \int_{\mathcal{X}} \max \left\{ m^t(x) - \mu^t, \max_{t' \in \mathcal{T} \setminus \{t\}} \left(m^{t'}(x) - \mu^{t'} \right) \right\} \mathrm{d}\mathbb{P}_{\mathrm{x}}(x) \\
&= \frac{\partial}{\partial \mu^t} \int_{\mathcal{X}} \left(m^t(x) - \max_{t' \in \mathcal{T} \setminus \{t\}} \left(m^{t'}(x) - \mu^{t'} \right) - \mu^t \right)^+ + \max_{t' \in \mathcal{T} \setminus \{t\}} \left(m^{t'}(x) - \mu^{t'} \right) \mathrm{d}\mathbb{P}_{\mathrm{x}}(x) \\
&= \frac{\partial}{\partial \mu^t} \int_{-\infty}^{\infty} (z - \mu^t)^+ \, \mathrm{d}\mathbb{P}_{\mathrm{z}^{\mu^{-t}}}(z) \\
&= \frac{\partial}{\partial \mu^t} \int_{\mu^t}^{\infty} (z - \mu^t) \, \mathrm{d}\mathbb{P}_{\mathrm{z}^{\mu^{-t}}}(z) \\
&= \int_{\mu^t}^{\infty} \frac{\partial}{\partial \mu^t} (z - \mu^t) \, \mathrm{d}\mathbb{P}_{\mathrm{z}^{\mu^{-t}}}(z),
\end{aligned}$$

where the fourth equality follows from a variable substitution by using the definition of $Z^{\mu^{-t}}$ from Lemma EC.1 and that $\frac{\partial}{\partial \mu^t} \int_{\mathcal{X}} \max_{t' \in \mathcal{T} \setminus \{t\}} \left(m^{t'}(x) - \mu^{t'} \right) \mathrm{d}\mathbb{P}_{\mathrm{x}}(x) = 0$, and the last equality follows from the Leibniz integral rule, which applies because for all $z \in \mathbb{R}$, $\frac{\partial}{\partial \mu^t}(z - \mu^t) = -1$ exists and is continuous for all $\mu^t \in \mathbb{R}_+$, and it is also uniformly bounded by the constant 1. We thus have

$$\begin{aligned}
\frac{\partial}{\partial \mu^t} f(\mu) &= \int_{\mu^t}^{\infty} \frac{\partial}{\partial \mu^t} (z - \mu^t) \, \mathrm{d}\mathbb{P}_{\mathrm{z}^{\mu^{-t}}}(z) \\
&= -\mathbb{P}_{\mathrm{z}^{\mu^{-t}}}(Z^{\mu^{-t}} \geq \mu^t) \qquad \text{(EC.2)} \\
&= -\mathbb{P}_{\mathrm{x}} \left(m^t(X) - \mu^t \geq \max_{t' \in \mathcal{T} \setminus \{t\}} \left(m^{t'}(X) - \mu^{t'} \right) \right),
\end{aligned}$$

where the last equality follows again from the definition of $Z^{\mu^{-t}}$.

Next, we show that $\frac{\partial}{\partial \mu^t} f(\mu)$ is continuous in μ, i.e., $\lim_{\mu \to \mu_0} \frac{\partial}{\partial \mu^t} f(\mu) = \frac{\partial}{\partial \mu^t} f(\mu_0)$ for every $\mu_0 \in \mathbb{R}_+^{m+1}$. To this end, consider an arbitrary $\mu_0 \in \mathbb{R}_+^{m+1}$, and note that we have

$$\lim_{\mu \to \mu_0} \frac{\partial}{\partial \mu^t} f(\mu) \quad = \lim_{\mu \to \mu_0} - \int_{\mathcal{X}} \mathbb{1}[h^t(x, \mu) > 0] \, \mathrm{d}\mathbb{P}_{\mathrm{x}}(x),$$

where $h^t(x, \mu) = m^t(x) - \max_{t' \in \mathcal{T} \setminus \{t\}} \left(m^{t'}(X) - \mu^{t'} \right) - \mu^t = Z^{\mu^{-t}} - \mu^t$, and the equality follows from (EC.2). The indicator function $\mathbb{1}[h^t(x, \mu) > 0]$ is uniformly bounded by the constant 1. By Lemma EC.1 , $h^t(x, \mu)$ is continuously distributed for every $\mu \in \mathbb{R}_+^{m+1}$ since $Z^{\mu^{-t}}$ is continuously distributed for any $\mu^{-t} \in \mathbb{R}_+^m$, and therefore $\mathbb{P}_{\mathrm{x}}(h^t(x, \mu) = 0) = 0$. As the function $h^t(x, \mu)$ is continuous in μ for every $x \in \mathcal{X}$, the indicator function $\mathbb{1}[h^t(x, \mu) > 0]$ is discontinuous at μ_0 only if $h(x,\mu_0)=0$. As $\mathbb{P}_{\mathrm{x}}(h(X,\mu_0)=0)=0$, we have $\lim_{\mu\to\mu_0}\mathbb{1}[h^t(X,\mu)>0]=\mathbb{1}[h(X,\mu_0)>0]$ $\mathbb{P}_{\mathrm{x}}$-almost surely. We thus obtain

$$\begin{aligned}\lim_{\mu\to\mu_0}\frac{\partial}{\partial\mu^t}f(\mu)&=\lim_{\mu\to\mu_0}-\int_{\mathcal{X}}\mathbb{1}[h^t(x,\mu)>0]\,\mathrm{d}\mathbb{P}_{\mathrm{x}}(x)\\&=-\int_{\mathcal{X}}\lim_{\mu\to\mu_0}\mathbb{1}[h^t(x,\mu)>0]\,\mathrm{d}\mathbb{P}_{\mathrm{x}}(x)\\&=-\int_{\mathcal{X}}\mathbb{1}[h(x,\mu_0)>0]\,\mathrm{d}\mathbb{P}_{\mathrm{x}}(x)\\&=\frac{\partial}{\partial\mu^t}f(\mu_0),\end{aligned}$$

where the second equality follows from the Dominated Convergence Theorem. This proves that the partial derivative $\frac{\partial}{\partial\mu^t}f(\mu)$ is continuous in μ. Function f is thus differentiable. □

Lemma EC.2 implies that the objective function of problem $(\mathcal{D})$ is differentiable. As problem $(\mathcal{D})$ is a convex optimization problem and satisfies Slater's condition, strong duality holds, i.e., the optimal value of $(\mathcal{D})$ coincides with the one of its dual problem. Denote by λ the dual variable of the constraint $\mu\geq 0$ in problem $(\mathcal{D})$. If $\mu^\star$ and $\lambda^\star$ are solutions to problem $(\mathcal{D})$ and its dual, respectively, then $\mu^\star,\lambda^\star$ satisfy the Karush–Kuhn–Tucker (KKT) conditions

$$\begin{aligned}\mathbb{P}\left(m^t(X)-\mu^{\star,t}\geq\max_{t'\in\mathcal{T}\setminus\{t\}}\left(m^{t'}(X)-\mu^{\star,t'}\right)\right)+\lambda^{\star,t}=b^t&\qquad\text{(Stationarity Condition)},\\\lambda^{\star,t}\mu^{\star,t}=0&\qquad\text{(Complementary Slackness)},\\\mu^{\star,t}\geq 0,\quad\lambda^{\star,t}\geq 0&\qquad\text{(Primal \& Dual Feasibility)},\end{aligned}\tag{EC.3}$$

for all $t\in\mathcal{T}$. Intuitively, the KKT conditions imply that policy $\pi^\star$ defined in (3) satisfies the capacity constraints in expectation. To see this, note that $\mathbb{P}\left(m^t(X)-\mu^{\star,t}\geq\max_{t'\in\mathcal{T}\setminus\{t\}}\left(m^{t'}(X)-\mu^{\star,t'}\right)\right)$ is the expected number of individuals assigned treatment t under policy $\pi^\star$. As $\lambda^{\star,t}\geq 0$ by dual feasibility, this quantity cannot exceed the available capacity b_t by the stationarity condition.

We are now ready to prove Theorem 1.

Proof of Theorem 1. For any $\mu^\star$ optimal in $(\mathcal{D})$, the treatment assignment probabilities of $\pi^\star$ defined in (3) and that of the dual-price queuing policy defined in Definition 1 coincide. Indeed, the dual-price queuing policy is nothing more than an online implementation of $\pi^\star$ since both make assignments using the same treatment assignment criteria. The difference is that the dual-price queuing policy waitlists individuals for treatments by assigning them to queues since there may be a mismatch in timing between the arrival of individuals and resources. It is thus sufficient to show that $\pi^\star$ is optimal in $(\mathcal{P})$. We first prove that $\pi^\star$ is feasible in $(\mathcal{P})$ and then show that its expected average outcome matches the optimal value $z^\star$ of $(\mathcal{P})$.

Policy $\pi^\star$ is feasible in $(\mathcal{P})$ because

$$\begin{aligned}\mathbb{E}[\pi^{\star,t}(X)]&=\mathbb{P}\left(t=\min\operatorname{argmax}_{t'\in\mathcal{T}}m^{t'}(X)-\mu^{\star,t'}\right)\\&\leq\mathbb{P}\left(m^t(X)-\mu^{\star,t}\geq\max_{t'\in\mathcal{T}\setminus\{t\}}\left(m^{t'}(X)-\mu^{\star,t'}\right)\right)\leq b^t\quad\forall t\in\mathcal{T},\end{aligned}$$

where the first inequality follows from the fact that if the event $t = \min \operatorname{argmax}_{t' \in \mathcal{T}} m^{t'}(X) - \mu^{\star,t'}$ occurs, then the inequality $m^t(X) - \mu^{\star,t} \geq \max_{t' \in \mathcal{T} \setminus \{t\}} (m^{t'}(X) - \mu^{\star,t'})$ must hold, and the second inequality follows from the stationarity and dual feasibility conditions in (EC.3).

As $\pi^\star$ is feasible in ($\mathcal{P}$), its expected average outcome cannot exceed $z^\star$. We next show that the expected average outcome of $\pi^\star$ is at least as high as $z^\star$, which implies that the two values coincide. The expected average outcome of $\pi^\star$ amounts to

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t\in\mathcal{T}} \mathbb{1}\left[t = \min \operatorname{argmax}_{t' \in \mathcal{T}} m^{t'}(X) - \mu^{\star,t'}\right] m^t(X)\right] \\
&= \mathbb{E}\left[\max_{t\in\mathcal{T}} (m^t(X) - \mu^{\star,t}) + \sum_{t\in\mathcal{T}} \mathbb{1}\left[t = \min \operatorname{argmax}_{t' \in \mathcal{T}} m^{t'}(X) - \mu^{\star,t'}\right] \mu^{\star,t}\right] \\
&= \mathbb{E}\left[\max_{t\in\mathcal{T}} (m^t(X) - \mu^{\star,t})\right] + \sum_{t\in\mathcal{T}} \mathbb{P}\left(t = \min \operatorname{argmax}_{t' \in \mathcal{T}} m^{t'}(X) - \mu^{\star,t'}\right) \mu^{\star,t} \\
&= \mathbb{E}\left[\max_{t\in\mathcal{T}} (m^t(X) - \mu^{\star,t})\right] + \sum_{t\in\mathcal{T}} b^t \mu^{\star,t} \\
&= \nu^\star \geq z^\star.
\end{aligned} \tag{EC.4}$$

The third equality above follows from the KKT conditions (EC.3), which imply that

$$\begin{aligned}
\mu^{\star,t} b^t &= \mu^{\star,t}\left[\mathbb{P}\left(m^t(X) - \mu^{\star,t} \geq \max_{t' \in \mathcal{T} \setminus \{t\}} (m^{t'}(X) - \mu^{\star,t'})\right) + \lambda^{\star,t}\right] \\
&= \mu^{\star,t}\,\mathbb{P}\left(m^t(X) - \mu^{\star,t} \geq \max_{t' \in \mathcal{T} \setminus \{t\}} (m^{t'}(X) - \mu^{\star,t'})\right) \\
&= \mu^{\star,t}\,\mathbb{P}\left(t = \min \operatorname{argmax}_{t' \in \mathcal{T}} m^{t'}(X) - \mu^{\star,t'}\right),
\end{aligned}$$

where the first and second equalities follows from the stationarity and complementary slackness conditions, respectively, and the third equality holds because $\operatorname{argmax}_{t' \in \mathcal{T}} m^{t'}(X) - \mu^{\star,t}$ is a singleton almost surely. The last equality in (EC.4) holds because $\mu^\star$ is optimal in ($\mathcal{D}$), and the inequality follows from weak duality. The claim thus follows. □

EC.1.2. Proofs for Section 4

The proof of Theorem 2 requires the following technical lemmas: Lemmas EC.3 – EC.7.

We first define the following functions $\nu : \mathbb{R}^{m+1} \to \mathbb{R}$ and $\hat{\nu}_n : \mathbb{R}^{m+1} \to \mathbb{R}$ for denoting the objective functions of the dual problem ($\mathcal{D}$) and its sample approximation ($\hat{\mathcal{D}}$) as functions of μ. These definitions will be used throughout Section EC.1.2.

$$\begin{aligned}
\nu(\mu) &= \mathbb{E}\left[\max_{t\in\mathcal{T}} (m^t(X) - \mu^t)\right] + \sum_{t\in\mathcal{T}} \mu^t b^t \\
\hat{\nu}_n(\mu) &= \frac{1}{n}\sum_{i=1}^n \max_{t\in\mathcal{T}} (\hat{m}_n^t(x_i) - \mu^t) + \sum_{t\in\mathcal{T}} \mu^t b^t.
\end{aligned} \tag{EC.5}$$

By definition, we have $\nu^\star = \min_{\mu \in \mathbb{R}_+^{m+1}} \nu(\mu)$ and $\hat{\nu}_n^\star = \min_{\mu \in \mathbb{R}_+^{m+1}} \hat{\nu}_n(\mu)$. Denote by $\mathcal{S}^\star$ and $\hat{\mathcal{S}}_n^\star$ the sets of optimal solutions to problems ($\mathcal{D}$) and ($\hat{\mathcal{D}}$), respectively. The next lemma shows that $\mathcal{S}^\star$ is contained in a compact set as long as the function m^t is uniformly bounded for all $t \in \mathcal{T}$, which is guaranteed under Assumption 1.

Lemma EC.3. *Under Assumption 1, the set $\mathcal{S}^\star$ of optimal solutions to problem $(\mathcal{D})$ is non-empty and contained in the compact set $[0,C]^{m+1} \subset \mathbb{R}_+^{m+1}$, where C is defined as in Assumption 1(i), i.e., $|m^t(x)| \le C$ for all $x \in \mathcal{X}$ and $t \in \mathcal{T}$.*

Proof. We will first show that the optimal solution set $\mathcal{S}^\star$ is contained in the compact set $[0,C]^{m+1}$ and then show that the set $\mathcal{S}^\star$ is non-empty.

Suppose for contradiction that $\mathcal{S}^\star \nsubseteq [0,C]^{m+1}$, i.e., there exists a $\mu \in \mathcal{S}^\star$ such that $\mu \notin [0,C]^{m+1}$. As $\mu \notin [0,C]^{m+1}$, there exists at least one $t \in \mathcal{T}$ such that $\mu^t > C$. Denote by $\mathcal{T}' = \{t \in \mathcal{T} : \mu^t > C\}$ the nonempty set of treatments t for which $\mu^t > C$. We construct another feasible solution $\mu' \in [0,C]^{m+1}$ via $(\mu')^t = \mu^t$ for all $t \in \mathcal{T} \setminus \mathcal{T}'$ (all t's where μ^t is in the interval $[0,C]$), and $(\mu')^t = C$ for all $t \in \mathcal{T}'$ (all t's where μ^t is not in the interval $[0,C]$). We will show that the objective value of μ' in problem $(\mathcal{D})$ is strictly lower than that of μ. This implies that μ cannot be optimal and results in a contradiction.

We will first show that $\max_{t\in\mathcal{T}}(m^t(x) - \mu^t) = \max_{t\in\mathcal{T}}(m^t(x) - (\mu')^t)$ for all $x \in \mathcal{X}$. To this end, recall that treatment 0 represents the no-treatment option, and $b^0 = 1$, which makes the capacity constraint corresponding to this treatment redundant. This implies that we can assume without loss of generality $\mu^0 = 0$, and therefore $0 \notin \mathcal{T}'$. For any $t \in \mathcal{T}'$ and $x \in \mathcal{X}$, we have

$$m^t(x) - \mu^t < m^t(x) - C = m^t(x) - (\mu')^t \le m^0(x) = m^0(x) - \mu^0 = m^0(x) - (\mu')^0, \tag{EC.6}$$

where the first inequality follows from the definition of $\mathcal{T}'$, the first equality follows from the definition of μ', and the second inequality holds because $m^t(x) \le C$ and $m^0(x)$ is non-negative. The last two equalities hold because $\mu^0 = (\mu')^0 = 0$. By equation (EC.6), we have $m^t(x) - \mu^t < m^0(x) - \mu^0$ and $m^t(x) - (\mu')^t \le m^0(x) - (\mu')^0$ for all $t \in \mathcal{T}'$. This implies that for any $x \in \mathcal{X}$, we have

$$\max_{t\in\mathcal{T}}(m^t(x) - \mu^t) = \max_{t\in\mathcal{T}\setminus\mathcal{T}'}(m^t(x) - \mu^t) = \max_{t\in\mathcal{T}\setminus\mathcal{T}'}(m^t(x) - (\mu')^t) = \max_{t\in\mathcal{T}}(m^t(x) - (\mu')^t),$$

where the first and third equality hold because no $t \in \mathcal{T}'$ can attain the maximum in $\max_{t\in\mathcal{T}}(m^t(x) - \mu^t)$ and $\max_{t\in\mathcal{T}}(m^t(x) - (\mu')^t)$ by (EC.6), and the second equality holds because $(\mu')^t = \mu^t$ for all $t \in \mathcal{T} \setminus \mathcal{T}'$.

The objective value $\nu(\mu)$ of μ thus exceeds the objective value $\nu(\mu')$ of μ' as

$$\begin{aligned}\nu(\mu) = \mathbb{E}\left[\max_{t\in\mathcal{T}}(m^t(X) - \mu^t)\right] + \sum_{t\in\mathcal{T}} \mu^t b^t &= \mathbb{E}\left[\max_{t\in\mathcal{T}}(m^t(X) - (\mu')^t)\right] + \sum_{t\in\mathcal{T}} \mu^t b^t \\ &> \mathbb{E}\left[\max_{t\in\mathcal{T}}(m^t(X) - (\mu')^t)\right] + \sum_{t\in\mathcal{T}} (\mu')^t b^t \\ &= \nu(\mu'),\end{aligned}$$

where the second equality holds because we showed that $\max_{t\in\mathcal{T}}(m^t(x)-\mu^t) = \max_{t\in\mathcal{T}}(m^t(x)-(\mu')^t)$ for all $x\in\mathcal{X}$, and the inequality holds because $(\mu')^t \le \mu^t$ for all $t\in\mathcal{T}$ and the inequality is strict for all $t\in\mathcal{T}'$ by definition of μ'. This implies that μ cannot be optimal and results in a contradiction. Thus, we must have $\mathcal{S}^\star \subseteq [0,C]^{m+1}$.

We now show that $\mathcal{S}^\star$ is always non-empty. As $\mathcal{S}^\star \subseteq [0,C]^{m+1}$, the dual problem $(\mathcal{D})$ is equivalent to $\min_{\mu\in[0,C]^{m+1}} \nu(\mu)$, where we only minimize over the compact set $[0,C]^{m+1}$. Since $\nu(\mu)$ is a convex function, and hence continuous, in $\mathbb{R}^n$, it attains a minimum over any compact set (Proposition A.2.7 of Bertsekas (2009)), which implies $\mathcal{S}^\star$ is non-empty. □

Similar to Lemma EC.3, the next lemma shows that the optimal solution set $\hat{\mathcal{S}}_n^\star$ must also lie within a compact set.

LEMMA EC.4. *Under Assumption 2, for any $n\in\mathbb{N}$, the set $\hat{\mathcal{S}}_n^\star$ is non-empty and contained in the compact set $[0,\hat{C}]^{m+1}\subset\mathbb{R}_+^{m+1}$, where $\hat{C}$ is defined as in Assumption 2(i), i.e., $|\hat{m}_n^t(x)|\le\hat{C}$ for all $x\in\mathcal{X}$, $t\in\mathcal{T}$, and $n\in\mathbb{N}$.*

Proof. Fix an arbitrary $\omega\in\Omega$, *i.e*, fix historical samples. By Assumption 2(i), $\hat{C}$ is a uniform upper bound on $|\hat{m}_n^t(x)|$ for all $x\in\mathcal{X}$, $t\in\mathcal{T}$, and $n\in\mathbb{N}$. Following the same arguments used in the proof of Lemma EC.3, we can show that any solution $\mu\notin[0,\hat{C}]^{m+1}$ will be sub-optimal and there exists a solution in $[0,\hat{C}]^{m+1}$ with strictly lower objective value. Therefore, the feasible set of problem $(\hat{\mathcal{D}})$ can be restricted to the compact set $[0,\hat{C}]^{m+1}$. For all $n\in\mathbb{N}$, $\hat{\nu}_n(\mu)$ is a convex function and hence attains a minimum over any compact set. Since the choice of ω was arbitrary, for any $\omega\in\Omega$, $\hat{\mathcal{S}}_n^\star$ is non-empty for all $n\in\mathbb{N}$ and contained in $[0,\hat{C}]^{m+1}$. □

The next lemma shows that the objective function $\hat{\nu}_n(\mu)$ of the sample-approximate dual problem $(\hat{\mathcal{D}})$ converges almost surely to the objective function $\nu(\mu)$ of the true dual problem $(\mathcal{D})$ uniformly in μ.

LEMMA EC.5. *Under Assumptions 1 and 2, the objective function $\hat{\nu}_n(\mu)$ of problem $(\hat{\mathcal{D}})$ converges almost surely to the objective function $\nu(\mu)$ of problem $(\mathcal{D})$ uniformly on any nonempty compact set $\mathcal{S}\subseteq\mathbb{R}_+^{m+1}$ as n tends to infinity, that is,*

$$\lim_{n\to\infty}\sup_{\mu\in\mathcal{S}}|\nu(\mu)-\hat{\nu}_n(\mu)| = 0.$$

Proof. Consider an arbitrary nonempty compact set $\mathcal{S}\subseteq\mathbb{R}_+^{m+1}$. By triangle inequality, we have

$$\begin{aligned}\sup_{\mu\in\mathcal{S}}|\nu(\mu)-\hat{\nu}_n(\mu)| &\le \sup_{\mu\in\mathcal{S}}\Big(|\nu(\mu)-\nu_n'(\mu)| + |\nu_n'(\mu)-\hat{\nu}_n(\mu)|\Big)\\ &\le \sup_{\mu\in\mathcal{S}}|\nu(\mu)-\nu_n'(\mu)| + \sup_{\mu\in\mathcal{S}}|\nu_n'(\mu)-\hat{\nu}_n(\mu)|,\end{aligned}$$

where ν'_n is an sample average approximation of ν using the true m^t functions, i.e.,

$$\nu'_n(\mu) \quad = \frac{1}{n}\sum_{i=1}^{n} \max_{t\in\mathcal{T}} \left(m^t(x_i) - \mu^t\right) + \sum_{t\in\mathcal{T}} \mu^t b^t.$$

In the following, we investigate the terms $\sup_{\mu\in\mathcal{S}} |\nu(\mu) - \nu'_n(\mu)|$ and $\sup_{\mu\in\mathcal{S}} |\nu'_n(\mu) - \hat{\nu}_n(\mu)|$ one by one.

We first consider the term $\sup_{\mu\in\mathcal{S}} |\nu(\mu) - \nu'_n(\mu)|$ and show that

$$\lim_{n\to\infty} \sup_{\mu\in\mathcal{S}} |\nu(\mu) - \nu'_n(\mu)| = 0. \tag{EC.7}$$

To this end, let $F(x,\mu) = \max_{t\in\mathcal{T}} \left(m^t(x) - \mu^t\right)$, and note that, for any $x \in \mathcal{X}$, $F(x,\mu)$ is continuous in μ as it is a piecewise-linear function of μ. For any $\mu \in \mathcal{S}$, the function $F(x,\mu)$ is dominated by the integrable function $g(x) = C$ as C is a uniform bound on m^t for each $t \in \mathcal{T}$ by Assumption 1(i). We can thus invoke Theorem 7.53 of Shapiro et al. (2014) and conclude that $\frac{1}{n}\sum_{i=1}^{n} F(X_i,\mu)$ converges uniformly on $\mathcal{S}$ and almost surely to $\mathbb{E}\left[F(X,\mu)\right]$ as n tends to infinity. This implies that $\nu'_n(\mu)$ converges uniformly on $\mathcal{S}$ and almost surely to $\nu(\mu)$. We thus proved (EC.7).

Next, we consider the term $\sup_{\mu\in\mathcal{S}} |\nu'_n(\mu) - \hat{\nu}_n(\mu)|$ and similarly show that

$$\lim_{n\to\infty} \sup_{\mu\in\mathcal{S}} |\nu'_n(\mu) - \hat{\nu}_n(\mu)| = 0. \tag{EC.8}$$

For any $\mu \in \mathcal{S}$, we have

$$\begin{aligned}
|\nu'_n(\mu) - \hat{\nu}_n(\mu)| &= \left| \frac{1}{n}\sum_{i=1}^{n} \max_{t\in\mathcal{T}} \left(m^t(x_i) - \mu^t\right) - \max_{t\in\mathcal{T}} \left(\hat{m}^t_n(x_i) - \mu^t\right) \right| \\
&\le \frac{1}{n}\sum_{i=1}^{n} \left| \max_{t\in\mathcal{T}} \left(m^t(x_i) - \mu^t\right) - \max_{t\in\mathcal{T}} \left(\hat{m}^t_n(x_i) - \mu^t\right) \right| \\
&\le \frac{1}{n}\sum_{i=1}^{n} \left| \max_{t\in\mathcal{T}} \left(m^t(x_i) - \hat{m}^t_n(x_i)\right) \right| \\
&\le \frac{1}{n}\sum_{i=1}^{n} \max_{t\in\mathcal{T}} \left| m^t(x_i) - \hat{m}^t_n(x_i) \right| \\
&\le \sup_{x\in\mathcal{X}} \max_{t\in\mathcal{T}} \left| m^t(x) - \hat{m}^t_n(x) \right|.
\end{aligned} \tag{EC.9}$$

The first inequality follows by triangle inequality. To see that the second inequality holds, fix an arbitrary $x \in \mathcal{X}$ and $\mu \in \mathcal{S}$ and let $t^\star = \min\arg\max_{t\in\mathcal{T}} \left(m^t(x) - \mu^t\right)$. We then have

$$\begin{aligned}
\max_{t\in\mathcal{T}} \left(m^t(x) - \mu^t\right) - \max_{t\in\mathcal{T}} \left(\hat{m}^t_n(x) - \mu^t\right) &\le \left(m^{t^\star}(x) - \mu^{t^\star}\right) - \left(\hat{m}^{t^\star}_n(x) - \mu^{t^\star}\right) \\
&= m^{t^\star}(x) - \hat{m}^{t^\star}_n(x) \\
&\le \max_{t\in\mathcal{T}} \left[m^t(x) - \hat{m}^t_n(x)\right],
\end{aligned}$$

where the first inequality holds by definition of $t^\star$ and $\hat{m}_n^{t^\star}(x) - \mu^{t^\star} \leq \max_{t\in\mathcal{T}} \ (\hat{m}_n^t(x) - \mu^t)$. Since the choice of x and μ were arbitrary, the above holds for all $x \in \mathcal{X}$ and $\mu \in \mathcal{S}$, thereby proving the second inequality in (EC.9). Finally, by Assumption 2(ii), we have $\lim_{n\to\infty} \sup_{x\in\mathcal{X}} \max_{t\in\mathcal{T}} \ |m^t(x) - \hat{m}_n^t(x)| = 0$. As this claim holds irrespective of the value of $\mu \in \mathcal{S}$ and in view of (EC.9), we conclude that (EC.8) holds.

The findings above imply that $\lim_{n\to\infty} \sup_{\mu\in\mathcal{S}} |\nu(\mu) - \hat{\nu}_n(\mu)| = 0$. The claim thus follows. □

Next, we use Lemmas EC.3–EC.5 to show that the optimal value $\hat{\nu}_n^\star$ of the sample-approximate dual problem ($\hat{\mathcal{D}}$) converges almost surely to the optimal value $\nu^\star$ of the true dual problem ($\mathcal{D}$). We will also prove a convergence result in terms of their respective optimal solution sets. The following definitions will be relevant for this result. We use $\|\cdot\|$ to denote the Euclidean norm and define

$$\mathrm{dist}(x, \mathcal{A}) = \inf_{x'\in\mathcal{A}} \|x - x'\|, \quad \mathbb{D}(\mathcal{A}, \mathcal{B}) = \sup_{x\in\mathcal{A}} \mathrm{dist}(x, \mathcal{B}),$$

where $\mathrm{dist}(x, \mathcal{A})$ denotes the distance between a point $x \in \mathbb{R}^d$ and a set $\mathcal{A} \subseteq \mathbb{R}^d$, and $\mathbb{D}(\mathcal{A}, \mathcal{B})$ denotes the deviation between two sets $\mathcal{A} \subseteq \mathbb{R}^d$ and $\mathcal{B} \subseteq \mathbb{R}^d$.

Lemma EC.6. *Under Assumptions 1 and 2, the optimal value $\hat{\nu}_n^\star$ of problem ($\hat{\mathcal{D}}$) converges almost surely to the optimal value $\nu^\star$ of problem ($\mathcal{D}$), and the deviation $\mathbb{D}(\hat{\mathcal{S}}_n^\star, \mathcal{S}^\star)$ between their respective optimal solution sets $\hat{\mathcal{S}}_n^\star$ and $\mathcal{S}^\star$ converges almost surely to zero as n tends to infinity, that is,* $\lim_{n\to\infty} |\nu^\star - \hat{\nu}_n^\star| = 0$ *and* $\lim_{n\to\infty} \mathbb{D}(\hat{\mathcal{S}}_n^\star, \mathcal{S}^\star) = 0$.

Proof. Consider the compact set $[0, \max(C, \hat{C})]^{m+1} \subset \mathbb{R}_+^{m+1}$, and note that: (i) the optimal solution set $\mathcal{S}^\star$ is nonempty and contained in $[0, \max(C, \hat{C})]^{m+1}$ by Lemma EC.3; (ii) the function $\nu(\mu)$ is finite valued by Assumption 1(i) and continuous on $[0, \max(C, \hat{C})]^{m+1}$; (iii) $\hat{\nu}_n(\mu)$ converges uniformly on $[0, \max(C, \hat{C})]^{m+1}$ and almost surely to $\nu(\mu)$ by Lemma EC.5; and, (iv) $\hat{\mathcal{S}}_n^\star$ is nonempty and contained in $[0, \max(C, \hat{C})]^{m+1}$ by Lemma EC.4. We can thus invoke Theorem 5.3 of Shapiro et al. (2014) and conclude that $\hat{\nu}_n^\star$ converges almost surely to $\nu^\star$, and $\mathbb{D}(\hat{\mathcal{S}}_n^\star, \mathcal{S}^\star)$ converges almost surely to zero. □

The proof of Theorem 2 will require comparing the treatment assignments made by a dual-price queuing policy and its sample approximation defined in Definitions 1 and 2, respectively. To that end, we extend the sample solution set convergence results of Lemma EC.6 by showing that for n large enough, any optimal sample-based solution $\hat{\mu}_n^\star$ will be arbitrarily close to some element within the optimal solution set $\mathcal{S}^\star$. This will help ensure that almost surely for n large enough, a sample approximation policy will make the same treatment assignments as a dual-price queuing policy.

Lemma EC.7. *Under Assumptions 1 and 2, there exists $\Omega' \subseteq \Omega$, where $\mathbb{P}(\Omega') = 1$, with the following property. For any $\omega \in \Omega'$ and $\epsilon > 0$, there exists $N(\omega) \in \mathbb{N}$ such that*

$$\forall \hat{\mu}_n^\star(\omega) \in \hat{\mathcal{S}}_n^\star(\omega), \exists \mu^\star \in \mathcal{S}^\star : \|\hat{\mu}_n^\star(\omega) - \mu^\star\| < \epsilon \quad \forall n \geq N(\omega).$$

Proof. By Lemma EC.6, $\mathbb{D}(\hat{\mathcal{S}}_n^\star, \mathcal{S}^\star)$ converges almost surely to zero as $n \to \infty$. In other words, there exists $\Omega' \subseteq \Omega$, where $\mathbb{P}(\Omega') = 1$, with the following property. For any $\omega \in \Omega'$ and $\epsilon > 0$, there exists $N(\omega) \in \mathbb{N}$ such that

$$\sup_{\hat{\mu}_n(\omega) \in \hat{\mathcal{S}}_n^\star(\omega)} \inf_{\mu \in \mathcal{S}^\star} \|\hat{\mu}_n(\omega) - \mu\| < \epsilon \quad \forall n \geq N(\omega),$$

which follows by the definition of $\mathbb{D}(A, B)$. The above implies that

$$\forall \hat{\mu}_n(\omega) \in \hat{\mathcal{S}}_n^\star(\omega) : \inf_{\mu \in \mathcal{S}^\star} \|\hat{\mu}_n(\omega) - \mu\| < \epsilon \quad \forall n \geq N(\omega). \tag{EC.10}$$

Next, we show that for every $\omega \in \Omega$ (i.e., for fixed historical sample), $\inf_{\mu \in \mathcal{S}^\star} \|\hat{\mu}_n(\omega) - \mu\|$ has an optimal solution for all $\hat{\mu}_n(\omega) \in \hat{\mathcal{S}}_n^\star(\omega)$. To this end, we first show that $\mathcal{S}^\star$ is a compact set. Recall that $\mathcal{S}^\star$ denotes the optimal solution set to problem ($\mathcal{D}$). The set $\mathcal{S}^\star$ is bounded as it is contained in the compact set $[0, C]^{m+1}$ by Lemma EC.3, where C is defined as in Assumption 1(i). The objective function $\nu(\mu)$ of problem ($\mathcal{D}$) is convex and thus continuous on $\mathbb{R}^{m+1}$. This implies that it has closed sublevel sets of the form $\{\mu \in \mathbb{R}_+^{m+1} \mid \nu(\mu) \leq c\}$, where c is any scalar. We now have $\mathcal{S}^\star = [0, C]^{m+1} \cap \{\mu \in \mathbb{R}_+^{m+1} \mid \nu(\mu) \leq \nu^\star\}$, which is closed as it is an intersection of closed sets. The set $\mathcal{S}^\star$ is thus compact.

Since all norms are continuous on $\mathbb{R}_+^{m+1}$ and $\mathcal{S}^\star$ is compact, by Weierstrass' Extreme Value Theorem (Proposition A.2.7 of Bertsekas (2009)), $\inf_{\mu \in \mathcal{S}^\star} \|\hat{\mu}_n(\omega) - \mu\|$ has an optimal solution for any input $\hat{\mu}_n(\omega) \in \hat{\mathcal{S}}_n^\star(\omega)$. This together with (EC.10) implies that for any $\omega \in \Omega'$ and $\epsilon > 0$, there exists $N(\omega) \in \mathbb{N}$ such that

$$\forall \hat{\mu}_n^\star(\omega) \in \hat{\mathcal{S}}_n^\star(\omega), \exists \mu^\star \in \mathcal{S}^\star : \|\hat{\mu}_n^\star - \mu^\star\| < \epsilon \quad \forall n \geq N(\omega).$$

The claim thus follows as $\mathbb{P}(\Omega') = 1$. □

We are now ready to prove Theorem 2.

Proof of Theorem 2. We use the following notation throughout the proof. Define the function h^t through $h^t(x, \mu) = m^t(x) - \mu^t$ for all $x \in \mathcal{X}$, $\mu \in \mathbb{R}_+^{m+1}$, and $t \in \mathcal{T}$. For any $\omega \in \Omega$ (i.e., fixed historical samples), define the function $\hat{h}_n^t(\cdot, \omega)$ through $\hat{h}_n^t(x, \mu, \omega) = \hat{m}_n^t(x, \omega) - \mu^t$ for all $x \in \mathcal{X}$, $\mu \in \mathbb{R}_+^{m+1}$, $t \in \mathcal{T}$, and $n \in \mathbb{N}$. Denote by z^{D} the expected average outcome of a dual-price queuing policy. Recall that $\hat{z}_n^{\mathrm{D}}$ and $z^\star$ denote the expected average outcome of a sample-based dual-price queuing policy, where the expectation is taken with respect to the distribution of covariates and outcomes of the individuals arriving during implementation and *not* with respect to the distribution of the historical samples, and the optimal value of problem ($\mathcal{P}$), respectively. Note that, by definition, $\hat{z}_n^{\mathrm{D}}$ is a random variable as it depends on the historical samples observed.

The proof is divided into two steps. In Step 1, we show that the allocations of a sample-based dual-price queuing policy match almost surely those of a dual-price queuing policy as $n \to \infty$. In

Step 2, we then use this fact to show that a sample-based dual-price queuing policy asymptotically generates at least as high expected average outcome as that of a dual-price queuing policy as $n \to \infty$ almost surely. As the expected average outcome of any dual-price queuing policy is equal to $z^\star$ by Theorem 1, we will thus conclude that the claim holds.

Step 1. For any $\omega \in \Omega$, $n \in \mathbb{N}$, $\hat{\mu}_n^\star(\omega) \in \hat{\mathcal{S}}_n^\star(\omega)$, and $\mu^\star \in \mathcal{S}^\star$, by triangle inequality we have

$$\begin{aligned}\sup_{t\in\mathcal{T}}\sup_{x\in\mathcal{X}}|\hat{h}_n^t(x,\hat{\mu}_n^\star(\omega),\omega) - h^t(x,\mu^\star)| &= \sup_{t\in\mathcal{T}}\sup_{x\in\mathcal{X}}|\hat{m}_n^t(x,\omega) - \hat{\mu}_n^{\star,t}(\omega) - m^t(x) + \mu^{\star,t}| \\ &\leq \sup_{t\in\mathcal{T}}\sup_{x\in\mathcal{X}}|\hat{m}_n^t(x,\omega) - m^t(x)| + \sup_{t\in\mathcal{T}}|\hat{\mu}_n^{\star,t}(\omega) - \mu^{\star,t}|.\end{aligned} \tag{EC.11}$$

For the first term in the last line of (EC.11), by Assumption 2(ii), there exists $\Omega' \subseteq \Omega$, where $\mathbb{P}(\Omega') = 1$, with the following property. For every $\omega \in \Omega'$ and $\epsilon > 0$, there exists $N(\omega) \in \mathbb{N}$ such that

$$\sup_{t\in\mathcal{T}}\sup_{x\in\mathcal{X}}|\hat{m}_n^t(x,\omega) - m^t(x)| < \epsilon \quad \forall n \geq N(\omega).$$

For the second term in the last line of (EC.11), Lemma EC.7 and the property $\sup_{t\in\mathcal{T}}|\mu^t| \leq \|\mu\|$ imply that there exists $\Omega'' \subseteq \Omega$, where $\mathbb{P}(\Omega'') = 1$, with the following property. For every $\omega \in \Omega''$ and $\epsilon > 0$, there exists $N(\omega) \in \mathbb{N}$ such that

$$\forall \hat{\mu}_n^\star(\omega) \in \hat{\mathcal{S}}_n^\star(\omega), \exists \mu^\star \in \mathcal{S}^\star : \sup_{t\in\mathcal{T}}|\hat{\mu}_n^{\star,t}(\omega) - \mu^{\star,t}| < \epsilon \quad \forall n \geq N(\omega),$$

Let $\hat{\Omega} = \Omega' \cap \Omega''$ and note that $\mathbb{P}(\hat{\Omega}) = 1$ because

$$\mathbb{P}(\hat{\Omega}) \geq 1 - \mathbb{P}((\Omega \setminus \Omega') \cup (\Omega \setminus \Omega'')) \geq 1 - \mathbb{P}((\Omega \setminus \Omega')) - \mathbb{P}((\Omega \setminus \Omega'')) = 1,$$

where the second inequality follows from the union bound, and the equality holds because $\mathbb{P}(\Omega') = 1$ and $\mathbb{P}(\Omega'') = 1$. Assumption 2(ii) and Lemma EC.7 thus imply that for every $\omega \in \hat{\Omega}$ and $\epsilon > 0$, there exists $N(\omega) \in \mathbb{N}$ such that

$$\forall \hat{\mu}_n^\star(\omega) \in \hat{\mathcal{S}}_n^\star(\omega), \exists \mu^\star \in \mathcal{S}^\star : \sup_{t\in\mathcal{T}}\sup_{x\in\mathcal{X}}|\hat{m}_n^t(x,\omega) - m^t(x)| + \sup_{t\in\mathcal{T}}|\hat{\mu}_n^{\star,t}(\omega) - \mu^{\star,t}| < \epsilon \quad \forall n \geq N(\omega).$$

The above expression, in conjunction with equation (EC.11), imply that for every $\omega \in \hat{\Omega}$ and $\epsilon > 0$, there exists $N(\omega) \in \mathbb{N}$ such that

$$\forall \hat{\mu}_n^\star(\omega) \in \hat{\mathcal{S}}_n^\star(\omega), \exists \mu^\star \in \mathcal{S}^\star : \sup_{t\in\mathcal{T}}\sup_{x\in\mathcal{X}}|\hat{h}_n^t(x,\hat{\mu}_n^\star(\omega),\omega) - h^t(x,\mu^\star)| < \epsilon \quad \forall n \geq N(\omega). \tag{EC.12}$$

Decomposing the expression $\sup_{t\in\mathcal{T}}\sup_{x\in\mathcal{X}}|\hat{h}_n^t(x,\hat{\mu}_n^\star(\omega),\omega) - h^t(x,\mu^\star)| < \epsilon$ in Equation (EC.12), we have that for every $\omega \in \hat{\Omega}$ and $\epsilon > 0$, there exists $N(\omega) \in \mathbb{N}$ such that

$$\begin{aligned}\forall \hat{\mu}_n^\star(\omega) \in \hat{\mathcal{S}}_n^\star(\omega), \exists \mu^\star \in \mathcal{S}^\star : \hat{h}_n^t(x,\hat{\mu}_n^\star(\omega),\omega) - h^t(x,\mu^\star) < \epsilon,\ h^t(x,\mu^\star) - \hat{h}_n^t(x,\hat{\mu}_n^\star(\omega),\omega) < \epsilon \\ \forall t \in \mathcal{T}, \forall x \in \mathcal{X}, \forall n \geq N(\omega).\end{aligned} \tag{EC.13}$$

For any $t, t' \in \mathcal{T}$, by summing the inequalities from above, i.e.,

$$h^t(x,\mu^\star) - \hat{h}_n^t(x,\hat{\mu}_n^\star(\omega),\omega) < \epsilon \quad \text{and} \quad \hat{h}_n^{t'}(x,\hat{\mu}_n^\star(\omega),\omega) - h^{t'}(x,\mu^\star) < \epsilon,$$

we obtain

$$h^t(x,\mu^\star) - \hat{h}_n^t(x,\hat{\mu}_n^\star(\omega),\omega) + \hat{h}_n^{t'}(x,\hat{\mu}_n^\star(\omega),\omega) - h^{t'}(x,\mu^\star) < 2\epsilon. \tag{EC.14}$$

Using (EC.13) and (EC.14), we have that for every $\omega \in \hat{\Omega}$ and $\epsilon > 0$, there exists $N(\omega) \in \mathbb{N}$ such that

$$\begin{aligned} \forall \hat{\mu}_n^\star(\omega) \in \hat{\mathcal{S}}_n^\star(\omega), \exists \mu^\star \in \mathcal{S}^\star : h^t(x,\mu^\star) - \hat{h}_n^t(x,\hat{\mu}_n^\star(\omega),\omega) + \hat{h}_n^{t'}(x,\hat{\mu}_n^\star(\omega),\omega) - h^{t'}(x,\mu^\star) < 2\epsilon \\ \forall t,t' \in \mathcal{T}, \forall x \in \mathcal{X}, \forall n \geq N(\omega). \end{aligned} \tag{EC.15}$$

Let $t(x,\mu) = \min \arg\max_{t\in\mathcal{T}} h^t(x,\mu)$, and define $\delta(x,\mu) = h^{t(x,\mu)}(x,\mu) - \max_{t \neq t(x,\mu)} h^t(x,\mu)$ for all $x \in \mathcal{X}$ and $\mu \in \mathbb{R}_+^{m+1}$. Note that $\delta(x,\mu)$ represents the difference between the highest and second highest value of the set $\{h^t(x,\mu)\}_{t\in\mathcal{T}}$ and that we have $\delta(x,\mu) > 0$ if and only if $\arg\max_{t\in\mathcal{T}} h^t(x,\mu) = \{t(x,\mu)\}$, i.e., $t(x,\mu)$ is the unique treatment that attains the maximum. Next, we write (EC.15) not for all treatment pairs $t,t' \in \mathcal{T}$ but only for the treatment pairs of the form $(t(x,\mu^\star),t)$, where $t \in \mathcal{T} \setminus \{t(x,\mu^\star)\}$. Using the new notation we introduced and (EC.15), we have that for every $\omega \in \hat{\Omega}$ and $\epsilon > 0$, there exists $N(\omega) \in \mathbb{N}$

$$\begin{aligned} \forall \hat{\mu}_n^\star(\omega) \in \hat{\mathcal{S}}_n^\star(\omega), \exists \mu^\star \in \mathcal{S}^\star : \hat{h}_n^t(x,\hat{\mu}_n^\star(\omega),\omega) - \hat{h}_n^{t(x,\mu^\star)}(x,\hat{\mu}_n^\star(\omega),\omega) < 2\epsilon - \delta(x,\mu^\star) \\ \forall t \in \mathcal{T} \setminus \{t(x,\mu^\star)\}, \forall x \in \mathcal{X}, \forall n \geq N(\omega), \end{aligned} \tag{EC.16}$$

where we used the fact that

$$\delta(x,\mu^\star) = h^{t(x,\mu)}(x,\mu^\star) - \max_{t \neq t(x,\mu^\star)} h^t(x,\mu^\star) \leq h^{t(x,\mu^\star)}(x,\mu^\star) - h^t(x,\mu^\star) \quad \forall t \in \mathcal{T} \setminus \{t(x,\mu^\star)\}, \forall x \in \mathcal{X}.$$

Equation (EC.16) implies that for a fixed $\omega \in \hat{\Omega}$ and $x \in \mathcal{X}$, if $\delta(x,\mu^\star) > 0$, then for $\epsilon > 0$ small enough (an therefore n large enough), $2\epsilon - \delta(x,\mu^\star) < 0$. This in turn implies $\hat{h}_n^t(x,\hat{\mu}_n^\star(\omega),\omega) - \hat{h}_n^{t(x,\mu^\star)}(x,\hat{\mu}_n^\star(\omega),\omega) < 0$ for all $\forall t \in \mathcal{T} \setminus \{t(x,\mu^\star)\}$. Since $t(x,\mu^\star)$ is also the unique maximizer of $\arg\max_{t\in\mathcal{T}} h^t(x,\mu)$ when $\delta(x,\mu^\star) > 0$, this implies that the allocation of a sample-based dual-price queuing policy for an individual with covariates x match that of a dual-price queuing policy for n large enough for any $\omega \in \hat{\Omega}$, where $\mathbb{P}(\hat{\Omega}) = 1$. By Assumption 1 and Lemma EC.1, we have $\delta(X,\mu^\star) > 0$ almost surely. We will use this observation in Step 2 to prove that the expected average outcome of a sample-based dual-price policy almost surely will be as high as that of a true dual-price policy as $n \to \infty$.

Step 2. By Step 1, for any $\omega \in \hat{\Omega}$ and sequence $(\hat{\mu}_n^\star(\omega))_{n\in\mathbb{N}}$ of sample-based solutions, where $\hat{\mu}_n^\star(\omega) \in \hat{\mathcal{S}}_n^\star(\omega)$ for all $n \in \mathbb{N}$, there exists a sequence $(\mu_n^\star(\omega))_{n\in\mathbb{N}}$ of solutions, where $\mu_n^\star(\omega) \in \mathcal{S}^\star$ for all $n \in \mathbb{N}$, with the following property. For every $\epsilon > 0$, there exists $N(\omega) \in \mathbb{N}$ such that

$$\begin{aligned} \hat{h}_n^t(x,\hat{\mu}_n^\star(\omega),\omega) - \hat{h}_n^{t(x,\mu_n^\star(\omega))}(x,\hat{\mu}_n^\star(\omega),\omega) < 2\epsilon - \delta(x,\mu_n^\star(\omega)) \\ \forall t \in \mathcal{T} \setminus \{t(x,\mu_n^\star(\omega))\}, \forall x \in \mathcal{X}, \forall n \geq N(\omega). \end{aligned} \tag{EC.17}$$

Fix a $\omega \in \hat{\Omega}$ and sequence couple $(\hat{\mu}_n^\star(\omega))_{n\in\mathbb{N}}$ and $(\mu_n^\star(\omega))_{n\in\mathbb{N}}$ such that (EC.17) holds. From now on, we slightly abuse the notation and drop ω notation. In the following, intuitively, we fix a path of historical samples, which implies that our estimates $\hat{\mu}_n^\star$, $\hat{m}_n^t$, $\hat{h}_n^t$ and $\hat{z}_n^{\mathrm{D}}$ are no longer random. We still perceive the covariates and outcomes of the individuals arriving during the implementation as random and will denote by $\mathbb{E}_{\mathrm{x}}$ the expectation taken with respect to the distribution of these individuals' covariates. Recalling that $z^{\mathrm{D}} = z^\star$ by Theorem 1, we have

$$\begin{aligned}
&\limsup_{n\to\infty}\ z^\star - \hat{z}_n^{\mathrm{D}} = \limsup_{n\to\infty}\ z^{\mathrm{D}} - \hat{z}_n^{\mathrm{D}} \\
&\le \limsup_{n\to\infty}\ \mathbb{E}_{\mathrm{x}}\left[C\mathbb{1}\left\{\min\arg\max_{t\in\mathcal{T}} h^t(X,\mu_n^\star) \neq \min\arg\max_{t\in\mathcal{T}} \hat{h}_n^t(X,\hat{\mu}_n^\star)\right\}\right] \\
&\le \mathbb{E}_{\mathrm{x}}\left[\limsup_{n\to\infty} C\mathbb{1}\left\{\min\arg\max_{t\in\mathcal{T}} h^t(X,\mu_n^\star) \neq \min\arg\max_{t\in\mathcal{T}} \hat{h}_n^t(X,\hat{\mu}_n^\star)\right\}\right] \\
&\le \mathbb{E}_{\mathrm{x}}\left[\limsup_{n\to\infty} C\mathbb{1}\left\{\hat{h}_n^t(X,\hat{\mu}_n^\star) - \hat{h}_n^{t(X,\mu_n^\star)}(X,\hat{\mu}_n^\star) \ge 0 \text{ for some } t\in\mathcal{T}\setminus\{t(X,\mu_n^\star)\}\right\}\right] \\
&\le \mathbb{E}_{\mathrm{x}}\left[\limsup_{n\to\infty} C\mathbb{1}\left\{\hat{h}_n^t(X,\hat{\mu}_n^\star) - \hat{h}_n^{t(X,\mu_n^\star)}(X,\hat{\mu}_n^\star) + \delta(X,\mu_n^\star) > 0 \text{ for some } t\in\mathcal{T}\setminus\{t(X,\mu_n^\star)\}\right\}\right],
\end{aligned} \tag{EC.18}$$

where the first inequality holds because a gap between z^{D} and z_n^{D} occurs only when the dual-price queuing policy and its sample-based variant assign different treatments to an individual with covariates X and because m^t is uniformly bounded by C by Assumption 1 (i), which implies that the gap is also bounded by the same value. The second inequality follows from the reverse Fatou's lemma, which applies because the indicator function is bounded above by 1. The third inequality holds because for every $x \in \mathcal{X}$, $\min\arg\max_{t\in\mathcal{T}} h^t(x,\mu_n^\star) \neq \min\arg\max_{t\in\mathcal{T}} \hat{h}_n^t(x,\hat{\mu}_n^\star)$ implies that there exists a $t\in\mathcal{T}\setminus\{t(x,\mu_n^\star)\}$ such that $\hat{h}_n^t(x,\hat{\mu}_n^\star) - \hat{h}_n^{t(x,\mu_n^\star)}(x,\hat{\mu}_n^\star) \ge 0$, where $t(x,\mu_n^\star)$ is defined as before, i.e., $t(x,\mu_n^\star) = \min\arg\max_{t\in\mathcal{T}} h^t(x,\mu_n^\star)$. To see this, for a fixed $x\in\mathcal{X}$, if $t(x,\mu_n^\star) \notin \arg\max_{t\in\mathcal{T}} \hat{h}_n^t(x,\hat{\mu}_n^\star)$, then there exists $t\in\mathcal{T}\setminus\{t(x,\mu_n^\star)\}$ such that $\hat{h}_n^t(x,\hat{\mu}_n^\star) - \hat{h}_n^{t(x,\mu_n^\star)}(x,\hat{\mu}_n^\star) > 0$. On the otherhand, if $t(x,\mu_n^\star) \in \arg\max_{t\in\mathcal{T}} \hat{h}_n^t(x,\hat{\mu}_n^\star)$ but is not the minimum index treatment, then there exists $t\in\mathcal{T}\setminus\{t(x,\mu_n^\star)\}$ such that $\hat{h}_n^t(x,\hat{\mu}_n^\star) - \hat{h}_n^{t(x,\mu_n^\star)}(x,\hat{\mu}_n^\star) = 0$. Finally, the last inequality holds because $\hat{h}_n^t(X,\hat{\mu}_n^\star) - \hat{h}_n^{t(X,\mu_n^\star)}(X,\hat{\mu}_n^\star) \ge 0$ implies that $\hat{h}_n^t(X,\hat{\mu}_n^\star) - \hat{h}_n^{t(X,\mu_n^\star)}(X,\hat{\mu}_n^\star) + \delta(X,\mu_n^\star) > 0$ almost surely as $\delta(X,\mu_n^\star) > 0$ almost surely by Assumption 1 and Lemma EC.1.

We will next show that the last expression in (EC.18) equals zero. By (EC.17), we have

$$\limsup_{n\to\infty} \hat{h}_n^t(x,\hat{\mu}_n^\star) - \hat{h}_n^t(x,\hat{\mu}_n^\star) + \delta(x,\mu_n^\star) \le 0 \quad \forall t\in\mathcal{T}\setminus\{t(x,\mu_n^\star)\}, x\in\mathcal{X}. \tag{EC.19}$$

Equation (EC.19) implies that for any $x\in\mathcal{X}$, we also have

$$\liminf_{n\to\infty} \mathbb{1}\left\{\hat{h}_n^t(x,\hat{\mu}_n^\star) - \hat{h}_n^t(x,\hat{\mu}_n^\star) + \delta(x,\mu_n^\star) \le 0 \quad \forall t\in\mathcal{T}\setminus\{t(x,\mu_n^\star)\}\right\} = 1.$$

We then have

$$
\begin{aligned}
&\limsup_{n\to\infty} \mathbb{1}\left\{\hat{h}_n^t(x,\hat{\mu}_n^\star) - \hat{h}_n^{t(x,\mu_n^\star)}(x,\hat{\mu}_n^\star) + \delta(x,\mu_n^\star) > 0 \text{ for some } t\in\mathcal{T}\setminus\{t(x,\mu_n^\star)\}\right\}\\
=&\,1-\liminf_{n\to\infty} \mathbb{1}\left\{\hat{h}_n^t(x,\hat{\mu}_n^\star) - \hat{h}_n^{t}(x,\hat{\mu}_n^\star) + \delta(x,\mu_n^\star) \le 0 \quad \forall t\in\mathcal{T}\setminus\{t(x,\mu_n^\star)\}\right\}\\
=&\,0.
\end{aligned}
$$

As the above equality holds true for all $x\in\mathcal{X}$, the last expression in (EC.18) therefore amounts to zero. By (EC.18), we thus have $\limsup_{n\to\infty} z^\star - \hat{z}_n^{\mathrm{D}} \le 0$ for the fixed path of historical samples. The claim now follows from the observation that our choice of sequence couple $(\hat{\mu}_n^\star)_{n\in\mathbb{N}}$ and $(\mu_n^\star)_{n\in\mathbb{N}}$ was arbitrary and that $\limsup_{n\to\infty} z^\star - \hat{z}_n^{\mathrm{D}} \le 0$ holds for every sequence couple $(\hat{\mu}_n^\star)_{n\in\mathbb{N}} = (\hat{\mu}_n^\star)_{n\in\mathbb{N}}(\omega)$ and $(\mu_n^\star)_{n\in\mathbb{N}} = (\mu_n^\star)_{n\in\mathbb{N}}(\omega)$, where $\omega\in\hat{\Omega}$ and $\mathbb{P}(\hat{\Omega})=1$. In other words, we have $\limsup_{n\to\infty} z^\star - \hat{z}_n^{\mathrm{D}} \le 0$ almost surely. □

EC.2. Derivations of Fairness-Constrained Policies for Section 5

EC.2.1. Allocation-Parity-Constrained Policies

We derive the policy $\pi^\star_{\text{alloc}}$ of Section 5.1, with the sample analogue being similar in derivation and the minority prioritization being a special case. Recall that we now explicitly consider the protected feature G as an input into the policy π, i.e., $X=(X^{-G},G)$, where $X^{-G}\in\mathcal{X}^{-\mathcal{G}}$ collects all features of X excluding G. The allocation-parity-constrained policy design problem is given by

$$
\begin{aligned}
z^\star_{\text{alloc}} = \max_{\pi\in\Pi}\quad & \mathbb{E}\left[\sum_{t\in\mathcal{T}} \pi^t(X^{-G},G) m^t(X)\right]\\
\text{s.t.}\quad & \mathbb{E}\left[\pi^t(X^{-G},G)\right] \le b^t \quad \forall t\in\mathcal{T}\\
& \mathbb{E}\left[\pi^t(X^{-G},G) \mid G=g\right] - \mathbb{E}\left[\pi^t(X^{-G},G)\mid G=g'\right] \le \delta\\
& \qquad\qquad\qquad\qquad \forall g,g'\in\mathcal{G},\;\; g\neq g',\;\; t\in\mathcal{T}.
\end{aligned}
\tag{EC.20}
$$

As in Section 3, let $\mu\in\mathbb{R}_+^{m+1}$ collect the dual variables of the capacity constraints and denote by $\lambda^t(g,g')\in\mathbb{R}_+$ the dual variable of the fairness constraint (5) for the pair (g,g') and for treatment t. The Lagrangian dual of problem (EC.20) is given by

$$
v^\star_{\text{alloc}} = \min_{\substack{\mu\in\mathbb{R}_+^{m+1}\\ \lambda\in\mathcal{L}_\infty(\mathcal{G}\times\mathcal{G},\mathbb{R}_+^{m+1})}} \max_{\pi\in\Pi} L_{\text{alloc}}(\pi,\mu,\lambda),
\tag{EC.21}
$$

where

$$
\begin{aligned}
L_{\text{alloc}}(\pi,\mu,\lambda) =& \mathbb{E}\left[\sum_{t\in\mathcal{T}}\pi^t(X^{-G},G)m^t(X)\right] + \sum_{t\in\mathcal{T}}\mu^t\Big(b^t - \mathbb{E}\left[\pi^t(X^{-G},G)\right]\Big)\\
&+\sum_{t\in\mathcal{T}}\sum_{g\in\mathcal{G}}\sum_{g'\in\mathcal{G},g\neq g'} \lambda^t(g,g')\delta\\
&-\sum_{t\in\mathcal{T}}\sum_{g\in\mathcal{G}}\sum_{g'\in\mathcal{G},g\neq g'} \lambda^t(g,g')\Big(\mathbb{E}\left[\pi^t(X^{-G},G)\mid G=g\right] - \mathbb{E}\left[\pi^t(X^{-G},G)\mid G=g'\right]\Big).
\end{aligned}
$$

Looking at the last term above, note that for each $t \in \mathcal{T}$, each $g \in \mathcal{G}$ appears $n-1$ times in the summation as $\lambda^t(g,g')\mathbb{E}\left[\pi^t(X^{-G},G) \mid G=g\right]$ (once for each possible (g,g') pairing) and $n-1$ times as $-\lambda^t(g',g)\mathbb{E}\left[\pi^t(X^{-G},G) \mid G=g\right]$ (once for each (g',g) pairing). Therefore we can rewrite this last term as

$$\sum_{t\in\mathcal{T}}\sum_{g\in\mathcal{G}}\sum_{g'\in\mathcal{G},g\neq g'}\lambda^t(g,g')\Big(\mathbb{E}\left[\pi^t(X^{-G},G)\mid G=g\right]-\mathbb{E}\left[\pi^t(X^{-G},G)\mid G=g'\right]\Big)=$$

$$\sum_{t\in\mathcal{T}}\sum_{g\in\mathcal{G}}\mathbb{E}\left[\pi^t(X^{-G},G)\mid G=g\right]\left[\sum_{g'\in\mathcal{G},g\neq g'}(\lambda^t(g,g')-\lambda^t(g',g))\right].$$

For ease of notation, we let $\gamma^t(g)=\sum_{g'\in\mathcal{G},g\neq g'}(\lambda^t(g,g')-\lambda^t(g',g))$. We now have

$$\begin{aligned}
L_{\text{alloc}}(\pi,\mu,\lambda)= &\ \mathbb{E}\left[\sum_{t\in\mathcal{T}}\pi^t(X^{-G},G)(m^t(X)-\mu^t)\right]+\sum_{t\in\mathcal{T}}\mu^t b^t\\
&+\sum_{t\in\mathcal{T}}\sum_{g\in\mathcal{G}}\sum_{g'\in\mathcal{G},g\neq g'}\lambda^t(g,g')\delta-\sum_{t\in\mathcal{T}}\sum_{g\in\mathcal{G}}\mathbb{E}\left[\pi^t(X^{-G},G)\mid G=g\right]\gamma^t(g)\\
= &\ \sum_{g\in\mathcal{G}}\mathbb{E}\left[\sum_{t\in\mathcal{T}}\pi^t(X^{-G},G)(m^t(X)-\mu^t)\mid G=g\right]\mathbb{P}(G=g)\\
&-\sum_{g\in\mathcal{G}}\mathbb{E}\left[\sum_{t\in\mathcal{T}}\pi^t(X^{-G},G)\gamma^t(g)\mid G=g\right]+\sum_{t\in\mathcal{T}}\mu^t b^t+\sum_{t\in\mathcal{T}}\sum_{g\in\mathcal{G}}\sum_{g'\in\mathcal{G},g\neq g'}\lambda^t(g,g')\delta\\
= &\ \sum_{g\in\mathcal{G}}\mathbb{P}(G=g)\mathbb{E}\left[\sum_{t\in\mathcal{T}}\pi^t(X^{-G},G)\left(m^t(X)-\mu^t-\frac{\gamma^t(g)}{\mathbb{P}(G=g)}\right)\mid G=g\right]\\
&+\sum_{t\in\mathcal{T}}\mu^t b^t+\sum_{t\in\mathcal{T}}\sum_{g\in\mathcal{G}}\sum_{g'\in\mathcal{G},g\neq g'}\lambda^t(g,g')\delta.
\end{aligned}$$

For any $\mu\in\mathbb{R}_+^{m+1}$ and $\lambda\in\mathcal{L}_\infty(\mathcal{G}\times\mathcal{G},\mathbb{R}_+^{m+1})$, the inner maximization problem of (EC.21) is solved by

$$\pi^t_{\text{alloc}}(x,g)=\begin{cases}1 & \text{if } \ t=\min\arg\max_{t'\in\mathcal{T}}\left(m^{t'}(x,g)-\mu^{t'}-\frac{\gamma^{t'}(g)}{\mathbb{P}(G=g)}\right)\\ 0 & \text{otherwise}\end{cases}\quad \forall t\in\mathcal{T},\ x\in\mathcal{X}^{-\mathcal{G}},\ g\in\mathcal{G}.$$

We suppress the dependence of policy π^t_{alloc} on μ,λ notationally in order to avoid clutter and tie-break using a lexicographic tie-breaker. By substituting the above policy into the dual problem (EC.21), we obtain the convex program

$$v^\star_{\text{alloc}}=\min_{\substack{\mu\in\mathbb{R}_+^{m+1};\\ \lambda\in\mathcal{L}_\infty(\mathcal{G}\times\mathcal{G},\mathbb{R}_+^{m+1})}}\overline{L}_{\text{alloc}}(\mu,\lambda),$$

where

$$\begin{aligned}
\overline{L}_{\text{alloc}}(\mu,\lambda)=&\sum_{g\in\mathcal{G}}\mathbb{P}(G=g)\mathbb{E}\left[\max_{t\in\mathcal{T}}\left(m^t(X)-\mu^t-\frac{\gamma^t(g)}{\mathbb{P}(G=g)}\right)\mid G=g\right]+\sum_{t\in\mathcal{T}}\mu^t b^t\\
&+\sum_{t\in\mathcal{T}}\sum_{g\in\mathcal{G}}\sum_{g'\in\mathcal{G},g\neq g'}\lambda^t(g,g')\delta.
\end{aligned}$$

We denote by $\mu^\star, \lambda^\star$ an optimal solution to this convex program. We now define our allocation-parity-constrained policy $\pi^\star_{\text{alloc}}$ through

$$\pi^{\star,t}_{\text{alloc}}(x,g) = \begin{cases} 1 & \text{if } t = \min \arg\max_{t' \in \mathcal{T}} \left(m^{t'}(x,g) - \mu^{\star,t'} - \frac{\gamma^{\star,t'}(g)}{\mathbb{P}(G=g)} \right) \\ 0 & \text{otherwise,} \end{cases} \quad \forall t \in \mathcal{T},\ x \in \mathcal{X}^{-\mathcal{G}},\ g \in \mathcal{G},$$

where $\gamma^{\star,t}(g) = \sum_{g' \in \mathcal{G}, g \neq g'} (\lambda^{\star,t}(g,g') - \lambda^{\star,t}(g',g))$.

EC.2.2. Outcome-Parity-Constrained Policies

We now derive the policy $\pi^\star_{\text{out}}$ of Section 5.2, with the sample analogue similar in derivation and the minority prioritization being a special case. The outcome-parity-constrained policy design problem is given by

$$\begin{aligned} z^\star_{\text{out}} = \max_{\pi \in \Pi} \quad & \mathbb{E}\left[\sum_{t \in \mathcal{T}} \pi^t(X^{-G}, G) m^t(X)\right] \\ \text{s.t.} \quad & \mathbb{E}\left[\pi^t(X^{-G}, G)\right] \leq b^t \quad \forall t \in \mathcal{T} \\ & \mathbb{E}\left[\sum_{t \in \mathcal{T}} \pi^t(X^{-G}, G) m^t(X) \mid G = g\right] - \mathbb{E}\left[\sum_{t \in \mathcal{T}} \pi^t(X^{-G}, G) m^t(X) \mid G = g'\right] \leq \delta \\ & \qquad \forall g, g' \in \mathcal{G}, g \neq g'. \end{aligned} \tag{EC.22}$$

Again let $\mu \in \mathbb{R}^{m+1}_+$ be the dual variable for our budget constraints, and denote by $\lambda(g,g') \in \mathbb{R}_+$ the dual variable of the statistical parity in outcome constraint (7) for the group pair g and g'. The Lagrangian relaxation of problem (EC.22) is given by

$$v^\star_{\text{out}} = \min_{\substack{\mu \in \mathbb{R}^{m+1}_+ \\ \lambda \in \mathcal{L}_\infty(\mathcal{G} \times \mathcal{G}, \mathbb{R}_+)}} \max_{\pi \in \Pi} L_{\text{out}}(\pi, \mu, \lambda), \tag{EC.23}$$

where

$$\begin{aligned} L_{\text{out}}(\pi, \mu, \lambda) = & \mathbb{E}\left[\sum_{t \in \mathcal{T}} \pi^t(X, G) m^t(X)\right] + \sum_{t \in \mathcal{T}} \mu^t \left(b^t - \mathbb{E}\left[\pi^t(X, G)\right]\right) + \sum_{g \in \mathcal{G}} \sum_{g' \in \mathcal{G}, g \neq g'} \lambda(g,g') \delta \\ & - \sum_{g \in \mathcal{G}} \sum_{g' \in \mathcal{G}, g \neq g'} \lambda(g,g') \left(\mathbb{E}\left[\sum_{t \in \mathcal{T}} \pi^t(X^{-G}, G) m^t(X) \mid G = g\right] - \right. \\ & \left. \mathbb{E}\left[\sum_{t \in \mathcal{T}} \pi^t(X^{-G}, G) m^t(X) \mid G = g'\right] \right). \end{aligned}$$

Looking at the last term above, note that each $g \in \mathcal{G}$ appears $n-1$ times in the summation as $\lambda(g,g')\mathbb{E}\left[\sum_{t \in \mathcal{T}} \pi^t(X^{-G}, G) m^t(X) \mid G = g\right]$ (once for each possible (g,g') pairing) and $n-1$ times as $-\lambda(g',g)\mathbb{E}\left[\sum_{t \in \mathcal{T}} \pi^t(X^{-G}, G) m^t(X) \mid G = g\right]$ (once for each (g',g) pairing). Therefore we can rewrite this last term as

$$\sum_{g \in \mathcal{G}} \sum_{g' \in \mathcal{G}, g \neq g'} \lambda(g,g') \left(\mathbb{E}\left[\sum_{t \in \mathcal{T}} \pi^t(X^{-G}, G) m^t(X) \mid G = g\right] - \mathbb{E}\left[\sum_{t \in \mathcal{T}} \pi^t(X^{-G}, G) m^t(X) \mid G = g'\right] \right) =$$

$$\sum_{g\in\mathcal{G}}\mathbb{E}\left[\sum_{t\in\mathcal{T}}\pi^t(X^{-G},G)m^t(X)\mid G=g\right]\left[\sum_{g'\in\mathcal{G},g\neq g'}(\lambda(g,g')-\lambda(g',g))\right].$$

For ease of notation, we let $\gamma(g)=\sum_{g'\in\mathcal{G},g'\neq g}(\lambda(g,g')-\lambda(g',g))$. We now have

$$\begin{aligned}L_{\text{out}}(\pi,\mu,\lambda)= &\ \mathbb{E}\left[\sum_{t\in\mathcal{T}}\pi^t(X^{-G},G)(m^t(X)-\mu^t)\right]+\sum_{t\in\mathcal{T}}\mu^t b^t\\ &+\sum_{g\in\mathcal{G}}\sum_{g'\in\mathcal{G},g\neq g'}\lambda(g,g')\delta-\sum_{g\in\mathcal{G}}\mathbb{E}\left[\sum_{t\in\mathcal{T}}\pi^t(X^{-G},G)m^t(X)\mid G=g\right]\gamma(g)\\ = &\ \sum_{g\in\mathcal{G}}\mathbb{E}\left[\sum_{t\in\mathcal{T}}\pi^t(X^{-G},G)(m^t(X)-\mu^t)\mid G=g\right]\mathbb{P}(G=g)\\ &-\sum_{g\in\mathcal{G}}\mathbb{E}\left[\sum_{t\in\mathcal{T}}\pi^t(X^{-G},G)m^t(X)\gamma(g)\mid G=g\right]+\sum_{t\in\mathcal{T}}\mu^t b^t+\sum_{g\in\mathcal{G}}\sum_{g'\in\mathcal{G},g\neq g'}\lambda(g,g')\delta\\ = &\ \sum_{g\in\mathcal{G}}\mathbb{P}(G=g)\mathbb{E}\left[\sum_{t\in\mathcal{T}}\pi^t(X^{-G},G)\left(m^t(X)\left(1-\frac{\gamma(g)}{\mathbb{P}(G=g)}\right)-\mu^t\right)\mid G=g\right]\\ &+\sum_{t\in\mathcal{T}}\mu^t b^t+\sum_{g\in\mathcal{G}}\sum_{g'\in\mathcal{G},g\neq g'}\lambda(g,g')\delta.\end{aligned}$$

For any $\mu\in\mathbb{R}_+^{m+1}$ and $\lambda\in\mathcal{L}_\infty(\mathcal{G}\times\mathcal{G},\mathbb{R}_+)$, the inner maximization problem of (EC.23) is thus solved by

$$\pi^t_{\text{out}}(x,g)=\begin{cases}1 & \text{if } t=\min\arg\max_{t'\in\mathcal{T}}\left(m^{t'}(x,g)\left(1-\frac{\gamma(g)}{\mathbb{P}(G=g)}\right)-\mu^{t'}\right)\\ 0 & \text{otherwise}\end{cases}\quad\forall t\in\mathcal{T},\ x\in\mathcal{X}^{-\mathcal{G}},\ g\in\mathcal{G}.$$

We suppress the dependence of policy π^t_{out} on μ,λ notationally and tie-break using a lexicographic tie-breaker. By substituting the above policy into the dual problem (EC.23), we obtain the convex program

$$v^\star_{\text{out}}=\min_{\substack{\mu\in\mathbb{R}_+^{m+1};\\ \lambda\in\mathcal{L}_\infty(\mathcal{G}\times\mathcal{G},\mathbb{R}_+)}}\overline{L}_{\text{out}}(\mu,\lambda),$$

where

$$\begin{aligned}\overline{L}_{\text{out}}(\mu,\lambda)=&\sum_{g\in\mathcal{G}}\mathbb{P}(G=g)\mathbb{E}\left[\max_{t\in\mathcal{T}}\left(m^t(X)\left(1-\frac{\gamma(g)}{\mathbb{P}(G=g)}\right)-\mu^t\right)\mid G=g\right]+\sum_{t\in\mathcal{T}}\mu^t b^t\\ &+\sum_{t\in\mathcal{T}}\sum_{g\in\mathcal{G}}\sum_{g'\in\mathcal{G},g\neq g'}\lambda(g,g')\delta.\end{aligned}$$

We denote by $\mu^\star,\lambda^\star$ an optimal solution to this convex program. We now define outcome-parity-constrained policy $\pi^\star_{\text{out}}$ through

$$\pi^{\star,t}_{\text{out}}(x,g)=\begin{cases}1 & \text{if } t=\min\arg\max_{t'\in\mathcal{T}}\left(m^{t'}(x,g)\left(1-\frac{\gamma^\star(g)}{\mathbb{P}(G=g)}\right)-\mu^{\star,t'}\right)\\ 0 & \text{otherwise,}\end{cases}\quad\forall t\in\mathcal{T},\ x\in\mathcal{X}^{-\mathcal{G}},\ g\in\mathcal{G},$$

where $\gamma^\star(g)=\sum_{g'\in\mathcal{G},g\neq g'}(\lambda^\star(g,g')-\lambda^\star(g',g))$.

EC.3. Empirical Results Details

EC.3.1. Simulated Data

In Figure EC.1, we show additional simulation results for values of σ from the set $\{0.5, 0.8, 1.15\}$ for the noise term. The qualitative takeaways are the same as those in Section 6.1.3 and serve primarily to show the degradation of non-parametric methods of decision tree and KNN from added noise in the data.

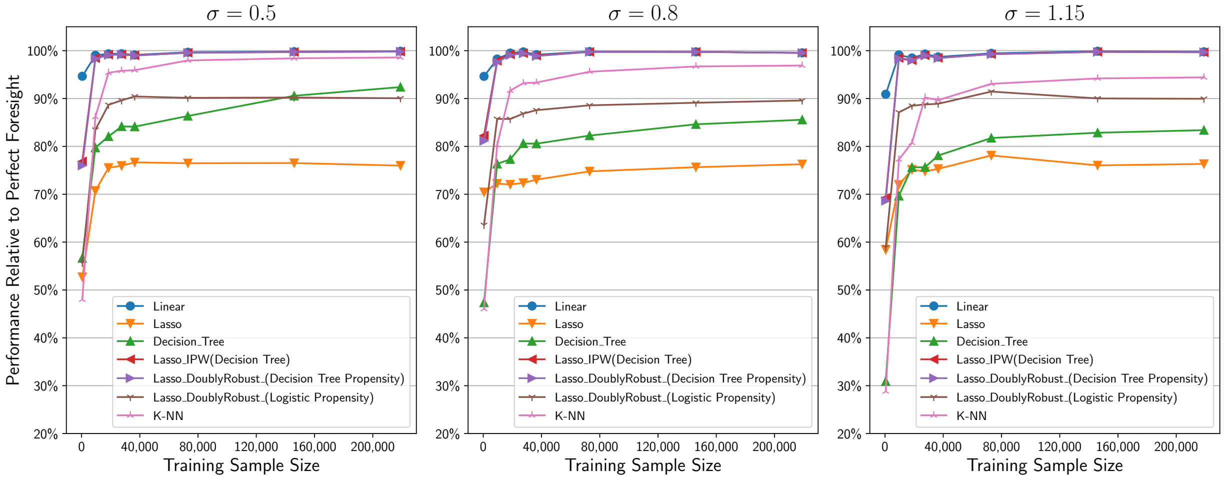


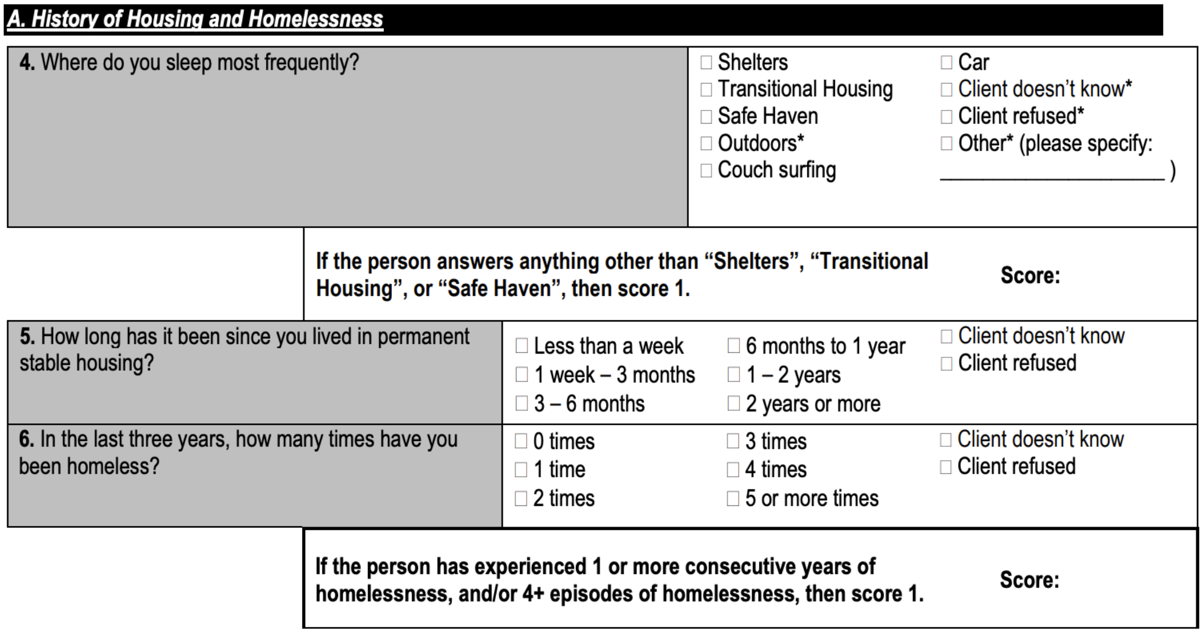
Figure EC.1 Synthetic data results for σ values $\{0.5, 0.8, 1.15\}$. All subfigures show the ratio of out-of-sample performance of the sample-based queuing policy to that of the perfect foresight policy in dependence of training set size. Each line corresponds to a policy using a different outcome estimator.

EC.3.2. HMIS and VI-SPDAT Data

EC.3.2.1. Data Details and Preparation

In this subsection, we provide further details on the current system, raw data, and how we defined outcomes for our experimental results.

When an individual arrives to the system, they are assessed for vulnerability with the VI-SPDAT tool, which contains a series of questions such as prior length of and number of experiences of homelessness. Adverse answers, such as sleeping situations indicative of homelessness, are given a weight of 1 so that an individual's vulnerability score ranges from 0 to 17, where a higher score indicates more vulnerability. The question and answers from the tool are used as features in our estimation methods and a sample of the tool is shown in Figure EC.2 (LAHSA 2017). To determine treatment assignments and outcomes after individuals receive their VI-SPDAT assessments, we can match them with their enrollment history from the raw HMIS data, which details their interactions with the system. Full details on the HMIS data and variables can be found in the HMIS Data Standards Data Dictionary (U.S. HUD 2021). For a given individual, we observe a sequence of enrollments, each characterized by a type of service and the corresponding date information.



A. History of Housing and Homelessness

4. Where do you sleep most frequently?	□ Shelters □ Transitional Housing □ Safe Haven □ Outdoors* □ Couch surfing	□ Car □ Client doesn't know* □ Client refused* □ Other* (please specify: ____________________)

If the person answers anything other than "Shelters", "Transitional Housing", or "Safe Haven", then score 1. **Score:**

5. How long has it been since you lived in permanent stable housing?	□ Less than a week □ 1 week – 3 months □ 3 – 6 months	□ 6 months to 1 year □ 1 – 2 years □ 2 years or more	□ Client doesn't know □ Client refused
6. In the last three years, how many times have you been homeless?	□ 0 times □ 1 time □ 2 times	□ 3 times □ 4 times □ 5 or more times	□ Client doesn't know □ Client refused

If the person has experienced 1 or more consecutive years of homelessness, and/or 4+ episodes of homelessness, then score 1. **Score:**

Figure EC.2 VI-SPDAT sample

Finally note that there is not a one-to-one correspondence between individuals found in VI-SPDAT assessment data and the HMIS data. After merging the two datasets, and removing missing data issues, we obtain our final dataset that consists of $63,764$ individuals assessed between 1/12/2015 and 12/31/2019, their assessment information, treatment received, and observed outcome. Recall that we use all individuals assessed between 1/12/2015 to 12/31/2017 as our training set to learn treatment outcomes and construct our policy, and then evaluate on individuals assessed between 1/1/2018 to 12/31/2019 to measure out-of-sample performance.

Historical Treatment Assignment Definition. Since the original HMIS data does not explicitly contain "Treatment", we use the enrollment history to create these variables. In Table EC.1 below, we breakout the 11 types of enrollments found within HMIS data, where the enrollments 'PH-PSH' and 'PH-RRH' are considered permanent interventions or treatments, while the enrollments 'Street Outreach', 'Emergency Shelter', and 'Safe Haven' are considered to be indicators of a return to homelessness. The remaining enrollments are considered no-treatment, or 'Services Only', either due to their temporary nature, shelter type, or nature of service provided. We also further breakout 'PH-PSH' into 'tenant' and 'site' based on an additional column, where the two sub-types differ, broadly speaking, in terms of supportive services available on-site and process of actually receiving housing. This categorization of the enrollment types was chosen based on discussions with matchers working within the LA CES system.

Outcome Definition. Recall from Section 6.2.1 that we focus on measuring returns to homelessness after an intervention as the outcome variable, in line with the goals listed by HUD. Since such

Table EC.1 HMIS enrollment types grouped by enrollments considered returns to homelessness, permanent interventions, and no treatments.

Treatment Type	Enrollment Type	Proportion of All Enrollments
Return to Homelessness	Street Outreach	34.31%
Return to Homelessness	Emergency Shelter	32.91%
Return to Homelessness	Safe Haven	0.13%
Permanent Intervention	PH - RRH	8.31%
Permanent Intervention	PH - PSH	3.55%
No Treatment	Services	14.33%
No Treatment	Transitional Housing	3.81%
No Treatment	Homelessness Prevention	1.99%
No Treatment	PH - Other	0.41%
No Treatment	Day Shelter	0.15%
No Treatment	Other	0.10%

an outcome is not explicitly tracked within the data, we construct a proxy variable as follows. To determine an individual's outcome, we first define an two-year observation window depending on their treatment received. For those receiving 'PH-RRH', we also observe a 'Move-In Date' column signifying when an individual was recorded to have officially received the resource. We define the 'PH-RRH' window to start from the 'Move-In Date'. Similarly for 'PH-PSH', we use the 'Move-In Date' to determine the starting point, but add a 100 day lag to the 'Move-In Date' because we concluded the recorded 'Move-In Date' within HMIS is likely not accurate. Through discussions with matchers within the LA CES system, we learned that the 'Move-In Date' within HMIS is an initial date inputted by case managers, but often does not get updated after an individual eventually moves in. Furthermore, we analyzed an auxiliary dataset, the Resource Management System (RMS) data, that supports this conclusion. For a subset of PSH units found within the HMIS data, the RMS data contains specific information such as their eligibility requirements and the updated move-in dates. For units found within both the HMIS and RMS data, we found discrepancies between the reported move-in dates, where the RMS 'Move-In Date' occurs often months after the corresponding HMIS 'Move-In Date', indicating that the HMIS 'Move-In Date' column for 'PH-PSH' have inaccuracies. This presents a problem where we often found individuals receiving PSH within the HMIS data to subsequently enroll in enrollments indicating an individual was experiencing homelessness ('Street Outreach', 'Emergency Shelter', and 'Safe Haven') shortly after their reported HMIS PSH 'Move-In Date'. If the HMIS 'Move-In Date' was incorrect and those individuals had actually not yet moved-in yet to their PSH units, then we would incorrectly conclude they experienced an adverse outcome after receiving PSH.

To address the problems with the HMIS PSH 'Move-In Date' column, we add a 100 day lag to the 'Move-In Date' to determine the starting point for the 'PH-PSH' observation window. We determined this 100 day lag based on the distribution of positive differences between the RMS and HMIS 'Move-In Date'. Finally for those receiving `no-treatment`, we consider their first enrollment date of any type to be the start of their observation window. Our choice of a two-year observation window for determining if an individual returned to homelessness after an intervention involved a trade-off where shorter windows resulted in insufficient time to observe post-intervention outcomes while longer windows resulted in fewer individuals for whom we can observe their full outcome window. If we observe any occurrence of the three proxies of homelessness enrollments ('Street Outreach', 'Emergency Shelter', or 'Safe Haven') within that observation window, then we consider an individual to have experienced a negative outcome. Otherwise we consider them to have a positive otherwise. We made the above choices of treatment types, window length, start dates, and more considering the trade-offs in terms of number of samples, 'bias' of the proxies, and data quality. For individuals for whom we do not fully observe their outcome window, i.e., the outcome window ends after the last observation date in our data, we remove them from our analysis. Keeping those indivisuals would bias the data towards positive outcomes since individuals may experience negative outcomes after the observable time in our data. For a summary of our outcome column construction, see Figure EC.3.

EC.3.2.2. Generating Counterfactuals and Fitting Training Set Estimators To obtain counterfactuals for evaluating performance within the held out test set, we took a semi-synthetic approach and used model generated outcomes based on the entire dataset. For outcome model selection for each treatment, we want our model output probabilities to be as well-calibrated to the observed outcome distribution as possible. Since our outcome is binary, the probability of positive outcome for individual with features X under treatment t is exactly the function $m^t(X)$. Therefore, requiring calibrated probabilities under each treatment is the same as requiring $\hat{m}^t(x) \sim m^t(x)$ for all $x \in \mathcal{X}$ for our chosen treatment outcome generating model $\hat{m}^t$. Therefore, we generate counterfactuals from models that empirically match the outcome distribution of each treatment as well as possible. To ensure generalization as well, we take a standard cross-validation approach to select the best model for each treatment group. Finally, we noticed individuals without any indication of a disability in the data were very unlikely to receive `PSH` and, generally speaking, `PSH` units required some form of disability to be eligible. Therefore modeling the `PSH` outcomes of non-disabled individuals would possibly violate the positivity assumption in Assumption 3 and

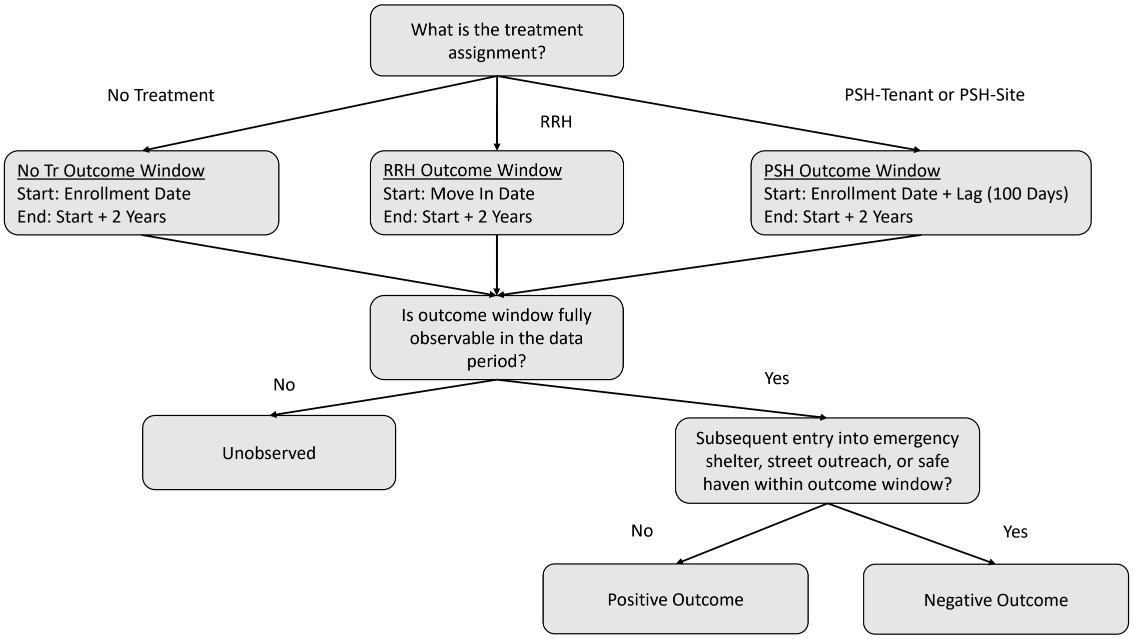


Figure EC.3 **Construction of treatment outcomes.**

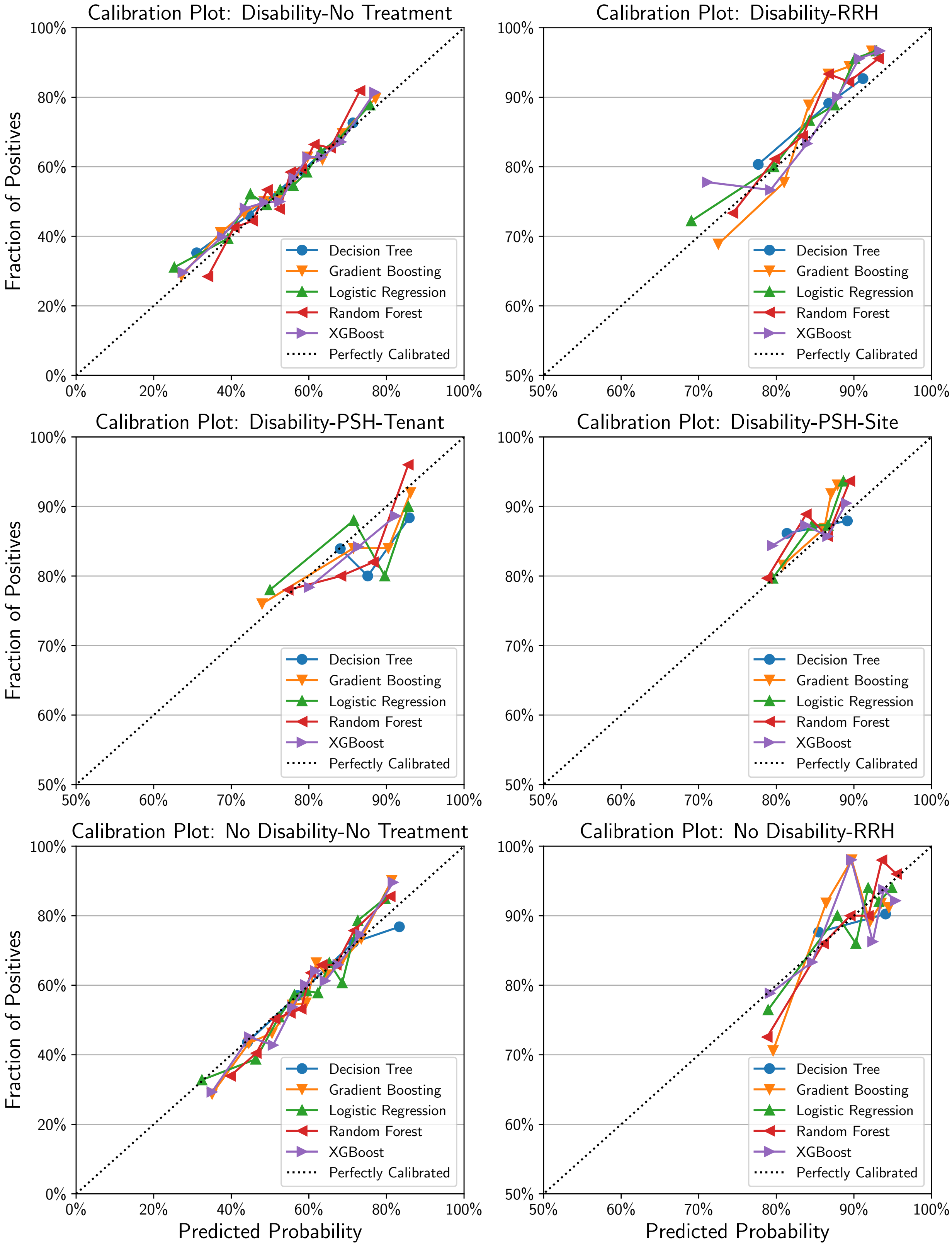
practically we have insufficient samples. Therefore, we model disabled and non-disabled individuals separately and did not model non-disabled individual outcomes under `PSH` treatment.

Note that the procedure outlined in this section was applied to the entire dataset and does not relate to the time-based splitting of training (1/12/2015 to 12/31/2017 assessments) and testing sets (1/1/2018 to 12/31/2019 assessments) used to evaluate performance in our experiment. Here, we are simply trying to find models that empirically fit our data and outcome distributions well enough to serve as counterfactuals.

Cross Validation. Since we want to have well-calibrated probabilistic predictions, we use log-loss during the cross validation phase as the evaluation metric for tuning hyper-parameters of each class. We choose to minimize log-loss since log-likelihood is a strictly proper scoring rule, i.e., optimizing this metric will yield well-calibrated probability predictions (Gneiting and Raftery 2007). For each treatment and model class we considered (logistic regression with and without regularization, decision trees, random forests, gradient boosted trees, and XGBoost trees), we use 5-folds and selected the hyper-parameters with the best average log-loss across all folds.

Validation Set of Calibration. For each treatment group and hyper-parameter tuned models, we generate a probability calibration to visually pick the model with the best calibration. In Figure EC.4, we plot the calibration curves from a held-out validation set. In each of the subplots of Figure EC.4, the dotted diagonal line represents probability outputs of a perfectly calibrated model while the colored solid lines with markers are the calibration curves of each model class with the tuned hyper-parameters. To generate each calibration curve, we split the probability predictions for a given model into equal sized buckets and find the average predicted probability (x-axis) and plot it against the average proportion of positive outcomes (y-axis). Note that because we only observe outcomes for those who received that treatment, we can only plot the calibration for those who were treated. Implicitly we are assuming that we satisfy the positivity assumption in 3 so that being well-calibrated on those who were treated will generalize to the entire population. Also, because the treatment groups `RRH`, `PSH`-Tenant, and `PSH`-Site had much smaller numbers of samples relative to the `no-treatment` group, we used a smaller number of buckets to evaluate calibration to avoid each bucket being too small in size and resulting in noise. Generally speaking, logistic regression, gradient boosting, XGboost produced the best calibration plots across all treatment groups, and where calibration appears similar, we defaulted to selecting the simpler model class of logistic regression.

After selecting the appropriate model class for each treatment group of disability and non-disability groups, we refitted the chosen models on the training and validation sets and checked for out-of-sample generalization on the held-out test set. In Figure EC.5. we plot the final test-set calibration curves for each treatment group for individuals with and without disabilities. In general, the final chosen models for each treatment group appear well-calibrated visually with some slight deviations from the diagonal line. Finally, while the above procedure describes how we chose the counterfactual generating models, we still need to generate predictions of treatment outcomes for the training set only (all individuals assessed 1/12/2015 to 12/31/2017) to learn a sample-based dual-price queuing policy. We repeat the above procedure but only on the training data to generate $\hat{m}^t$.



Calibration Plot: Disability-No Treatment

Calibration Plot: Disability-RRH

Calibration Plot: Disability-PSH-Tenant

Calibration Plot: Disability-PSH-Site

Calibration Plot: No Disability-No Treatment

Calibration Plot: No Disability-RRH

Figure EC.4 Calibration plots on validation set for each model class and different combinations of disability and treatment group.

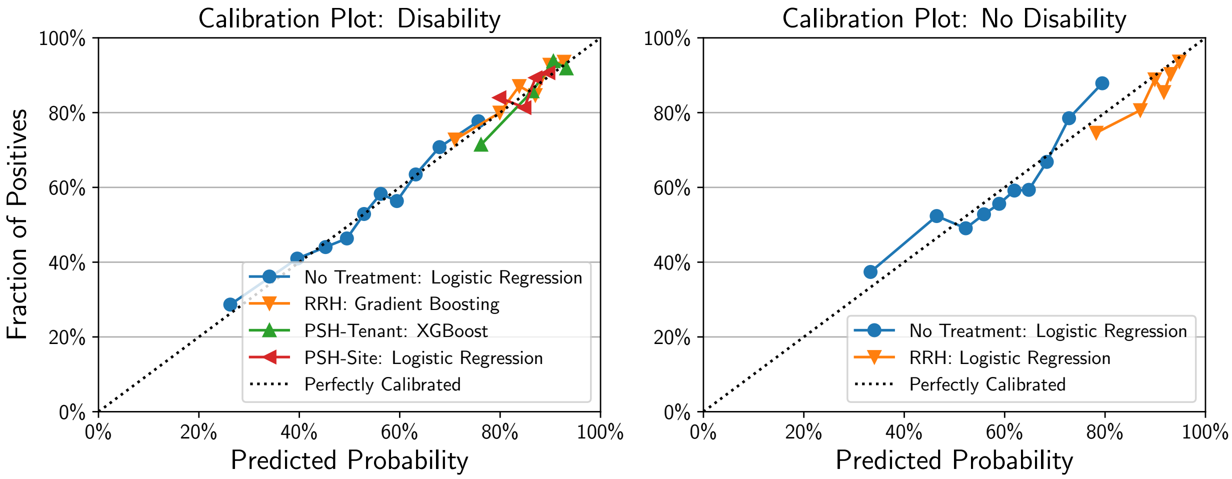


Figure EC.5 Out-of-sample calibration plots of final counterfactual generating models chosen for each treatment group and for individuals with and without disabilities.

EC.3.2.3. Fairness of Estimates An additional concern with our counterfactual models for evaluation was potential bias by race. Since we want to use our generated counterfactuals as reflections of the true generating process, any calibration bias would mean our evaluation results would be potentially biased by race. In Figure EC.6, we plot and compare calibration curves for each racial group across treatments after we selected our counterfactual generating models and fitted to the entire dataset. For the most part, the calibration curves across racial groups are similar for each treatment, with the exception of the ‘Other’ racial group. This is likely because the ‘Other’ group is only a small portion of the sample, representing only $\sim 7\%$ of the sample. Furthermore, the scarce PSH-Tenant and PSH-Site treatment groups also have worse calibration than RRH, again likely due to insufficient samples from this small treatment group.

EC.3.2.4. Full Experimental Test Results In this subsection, we present further results on two additional fairness-constrained policies `Allocation SP` and `Outcome SP` policies, which seek to achieve statistical parity in allocation (constraint (5)) and statistical parity in outcomes (constraint (7)) by race, respectively. The out-of-sample performance in terms of expected outcomes are summarized in Table EC.2 and allocation and outcome fairness properties are presented in Figure EC.7. We use a δ of 0.01 for the fairness constraints in the sample-based problem to learn the `Allocation SP` and `Outcome SP` policies.

For `RRH`, we see that `Allocation SP` results in the smallest differences in percentage receiving `RRH` between all racial groups compared to all other policies. Amongst all policies, `Allocation SP` is the closest towards satisfying statistical parity in allocation of `RRH` across all racial groups. On the other hand, it allocates fewer resources to Black individuals compared to `Historical` or any

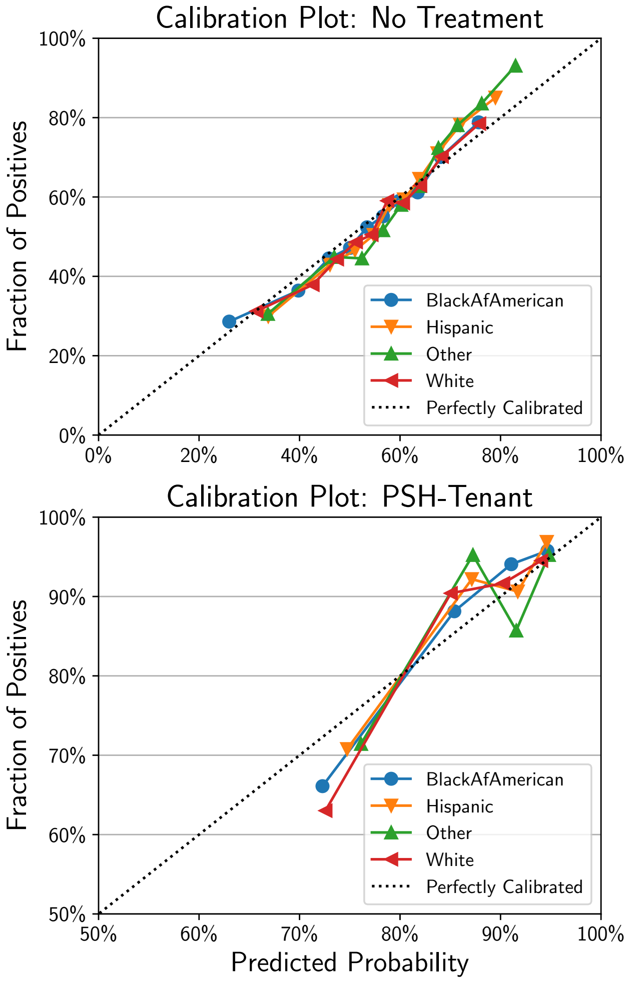


Figure EC.6 **Calibration plots for different racial groups and treatments using model outputs from the final chosen counterfactual generating models.**

other policy, which may be undesirable when policymakers want to prioritize minority groups. In this instance, policymakers need to choose between statistical parity in allocation or prioritizing minority groups for allocation of `RRH` but our proposed policy can incorporate either notions into consideration or possibly some combination of the two. For `PSH`, `Allocation SP` also results in similar maximum differences in percentage receiving `PSH` between all racial groups compared to `Historical` and `Alloc Min Priority`, while generally having smaller differences between all racial groups compared to the remaining policies. However, in simply trying to equalize allocation rates of `PSH` across all racial groups, `Allocation SP` allocated to Black individuals at the lowest rate relative to all other policies. Also, of note in the allocation fairness results is that the `Outcome SP` policy allocates `RRH` and `PSH` to the `Other` group at far lower rates compared to other racial groups. This is because in our predictions, individuals of the `Other` group had predicted outcomes under `no-treatment` that skewed more positive relative to other racial groups. Therefore, the `Outcome`

`SP` policy allocated less to the `Other` group in an attempt to achieve outcome parity across all groups because their predicted outcomes under `no-treatment` were already higher compared to the other racial groups. In terms of statistical parity in outcomes, we see that `Outcome SP`, in addition to `Base` and `Outcome Min Priority`, resulted in the smallest differences in proportion of positive outcomes between all racial groups out-of-sample.

Finally, we see in Table EC.2 that our fairness-constrained policies `Allocation SP` and `Outcome SP` suffer almost no performance gap in terms of expected outcomes compared to `Base` while potentially improving the fairness properties of the `Base` policy. Again, these further results suggest almost no ‘price of fairness’.

Table EC.2 Out-of-sample average expected outcome under each policy.

Policy	**Proportion of Positive Outcomes**
`Historical`	60.35%
`Base`	62.25%
`Allocation SP`	62.20%
`Alloc Min Priority`	62.24%
`Outcome SP`	62.25%
`Outcome Min Priority`	62.25%
`Perfect Foresight`	63.30%

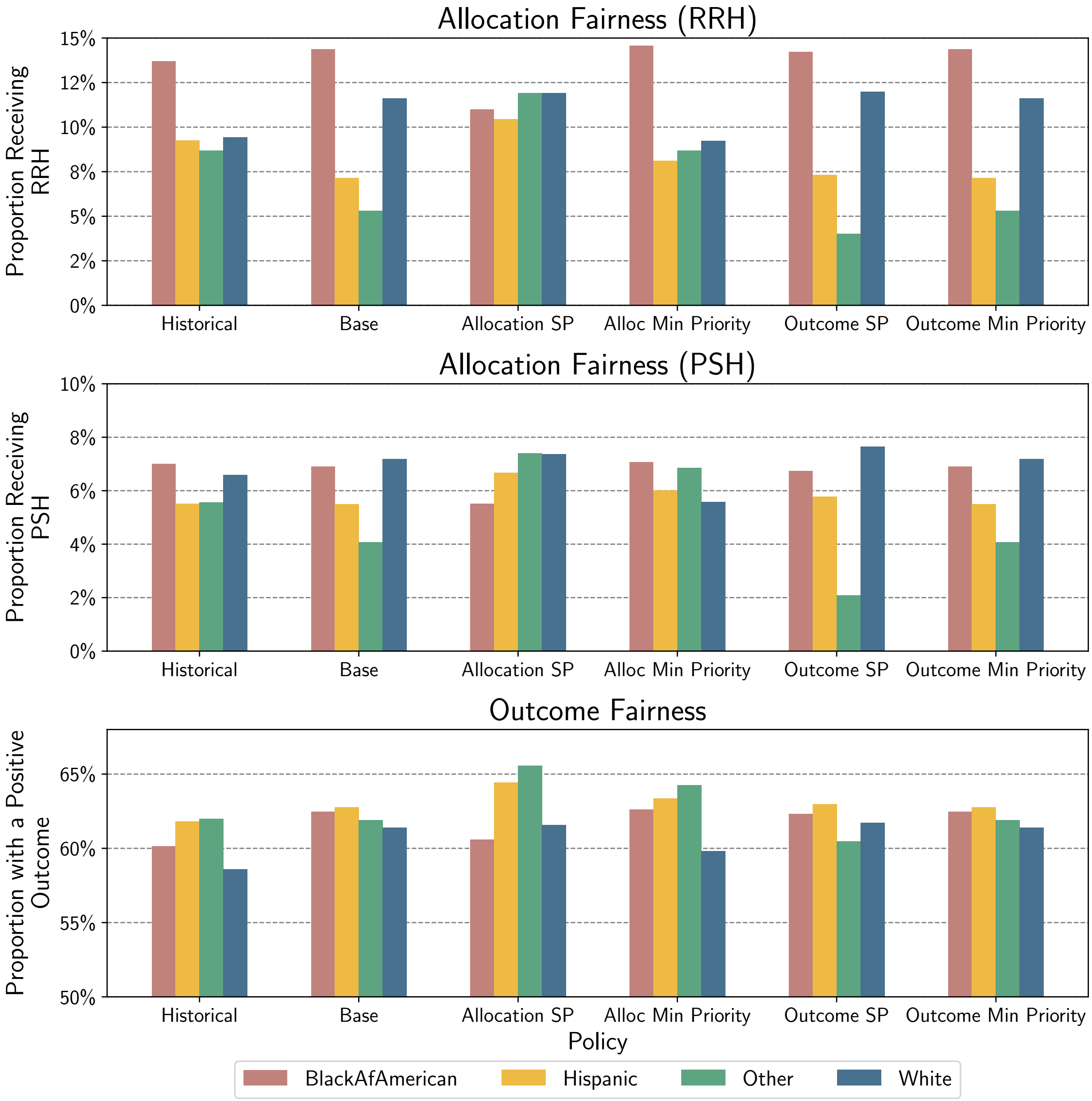


Figure EC.7 Out-of-sample fairness results showing proportion of each racial group receiving RRH, PSH (Allocation Fairness) and the proportion of positive outcomes for each racial group (Outcome Fairness) under different policies.