

Redesigning VolunteerMatch’s Ranking Algorithm: Toward More Equitable Access to Volunteers

(Authors’ names blinded for peer review)

In collaboration with VolunteerMatch (VM)—the world’s largest online platform for connecting volunteers with nonprofits—we designed and implemented a new display ranking algorithm. VM’s original ranking algorithm was intended to maximize efficiency (i.e., the total number of connections), but as a consequence it repeatedly displayed the same few opportunities at the top of its ranking, effectively limiting access to volunteers for the other opportunities. To incorporate VM’s desire for equity (defined as the weekly number of opportunities with at least one connection) along with efficiency, we propose a modeling framework for online display ranking in settings where it is important to manage the trade-off between the total number of connections and the equitable allocation of these connections. Taking an adversarial approach in evaluating the performance of online algorithms, we show that a simple class of algorithms that applies a *penalty* to opportunities after each connection provides a strong (and, in certain regimes, optimal) performance guarantee. Based on our theoretical development, we propose *SmartSort*, a simple online display ranking algorithm with a penalty term that we calibrated using VM’s data and simulation. We implemented SmartSort in two experiments, covering Dallas-Fort Worth and all of Southern California. Using a difference-in-differences analysis, we find that the implementation of SmartSort led to a 8-9% increase in the weekly average number of opportunities with at least one connection (consistent across both experiments) without any meaningful decrease in the total number of connections, implying a *Pareto improvement* for VM. Based on the success of our experiments, SmartSort has now been deployed nationwide. If SmartSort has a similar distributional effect on a national scale, every year, an additional 30,000 connections will go to opportunities that would have otherwise lacked access to volunteers.

Key words: volunteering, matching markets, online platforms, online experiments, online matching

1. Introduction

Every year, 75 million volunteers in the US contribute 7 billion hours of service, which represent roughly 167 billion dollars of economic activity.¹ Thousands of nonprofit organizations rely on these volunteer hours to operate, and nowadays, many nonprofits depend on volunteer recruiting platforms such as AmeriCorps, Idealist, and JustServe to help them find volunteers. At a high level, these platforms function as online labor markets for volunteering, where the platforms’ display ranking algorithms play a key role in connecting prospective volunteers with volunteering opportunities that suit their interests and availability. This paper describes a collaboration with

¹ <https://americorps.gov/about/our-impact/volunteering-civic-life>



Figure 1 (a-c): The usual path for prospective volunteers on VM. (d): An illustration of the sign-up distribution for opportunities in Dallas-Fort Worth from March 15-21, 2022. For aesthetic purposes, each bar represents a batch of opportunities. A bar's color (blue, green, or red) corresponds to the number of sign-ups an opportunity in that batch received (0, 1, or more than 1).

VolunteerMatch — the world’s largest online volunteer recruiting platform — to design and implement a new display ranking algorithm that improves equity in nonprofits’ access to volunteers.

VolunteerMatch (VM) receives tens of thousands of daily visitors, and hundreds of new volunteering opportunities — addressing a variety of important societal issues — are posted on the platform each day. Since its launch, VM has facilitated 18 million matches between volunteers and nonprofits. VM’s role in promoting civic engagement has been recognized by numerous awards² and highlighted by Presidents Clinton, Bush, and Obama.³

VM facilitates connections between prospective volunteers (*volunteers*, for short) and volunteering opportunities (*opportunities*, for short) by allowing volunteers to search for opportunities that suit their interests and availability. The search process of a volunteer is illustrated in Figure 1 and can be summarized as follows. When a volunteer visits VM, they are assigned a pre-populated search location. Upon selecting “Find Opportunities,” VM’s *display ranking algorithm* produces a ranked list of opportunities. After selecting an opportunity from this list, a volunteer may *sign up* for volunteering by clicking “I Want to Help.” VM refers to a sign-up as a *connection*. However, hereafter, we only use the term sign-up.

Typically, demand for volunteer labor far exceeds its supply, which means that many opportunities are left wanting more sign-ups. In addition to the general challenges posed by this volunteer shortage, our analysis of VM sign-up data shows significant inequity in the number of sign-ups that

² These awards include a MIT Sloan E-Commerce Award and two Webby Awards.

³ https://en.wikipedia.org/wiki/VolunteerMatch#Awards_and_presidential_support

opportunities receive. As an example, panel (d) of Figure 1 shows the distribution of sign-ups on VM for opportunities in the Dallas-Fort Worth geographic region available for sign-ups during the week of March 15-21, 2022. Observe that only 23.6% (174) of opportunities received a sign-up during this week, and 76.4% (563) did not receive *any* sign-ups. Moreover, some opportunities received a disproportionate number of sign-ups: 21.4% of the sign-ups went to just 10 opportunities.

Such an imbalanced distribution of sign-ups is problematic for many reasons. First, given its non-profit objectives, VM places a high priority on fulfilling the demand for volunteers in an egalitarian fashion across diverse nonprofits and causes. Second, even the nonprofits receiving many sign-ups may not benefit, as the value of the marginal volunteer may not exceed the cost of screening or training that volunteer. Third, most nonprofits have limited resources for volunteer management, so those receiving an excess of sign-ups — especially in a short period of time — may fail to respond to volunteers, which can be discouraging to them and may prevent them from volunteering in the future.

Prior to our collaboration, VM's display ranking algorithm likely exacerbated imbalances in the sign-up distribution. Roughly speaking, VM's display ranking was appeal-based, i.e., opportunities were ranked based on how appealing VM perceives them to be for the arriving volunteer. Appeal-based rankings are common in e-commerce and social media platforms, as they maximize efficiency (e.g., the number of purchases, clicks, or, in our case, sign-ups). However, appeal-based rankings also promote inequitable exposure by disproportionately displaying a few high-appeal options in top positions (Biega et al. 2018, Singh and Joachims 2018, Chen et al. 2022). In VM's context, inequitable exposure may result in some opportunities receiving an excessive amount of sign-ups, while others will effectively lack access to volunteers.

The shortcomings of VM's algorithm motivate our applied work: *how can we redesign VM's display ranking algorithm to allocate sign-ups more equitably—in particular, producing a more uniform sign-up distribution—without impacting the total number of sign-ups?*

To answer the above question, we: (i) developed a modeling framework to better understand how to design algorithms in settings where one cares about both the total number and the equitable allocation of sign-ups; (ii) translated these findings into an algorithm, SmartSort, that could be deployed by VM; (iii) designed and implemented two field experiments to measure the impact of SmartSort. We describe each contribution next.

Modeling and Algorithmic Contributions. We model the problem of online (i.e., real-time) display ranking that goes beyond the objective of *efficiency* (i.e., the total number of weekly sign-ups). Based on our conversations with VM managers, we define a natural *equity* objective that only rewards sign-ups up to an opportunity-specific threshold, thus discouraging excessive sign-ups

to any opportunity. Given the potential trade-off between the two objectives, we consider their convex combination as the platform's main objective. In designing and evaluating online display ranking algorithms for this bi-criteria objective, we take a robust (i.e., adversarial) approach which is frequently used when designing real-time algorithms for online platforms, as we further discuss in Section 1.1. That is, we assume no prior knowledge on the preferences or choice behavior of arriving users, and we evaluate the performance of an algorithm by its competitive ratio, i.e., its worst-case ratio when compared with a properly-defined clairvoyant solution.

To design algorithms for our setting, we focus on the natural class of balancing algorithms, which have been studied in the context of online budgeted allocation and several variants of the online assortment problem (see Section 1.1 for an overview). Algorithms in this class are defined by a penalty function that will determine the extent to which each opportunity's rank is diminished after each sign-up. In our context, a balancing algorithm with a properly-chosen penalty term allows us to manage the trade-off between efficiency and equity: in one extreme (where the platform only cares about efficiency), we can maximize the total number of sign-ups by applying no penalty to opportunities. In the other extreme (where the platform only cares about equity), we can maximally penalize any opportunity that has reached its sign-up threshold, thereby avoiding excessive sign-ups. For any particular balancing algorithm, we establish a parametric lower bound on its competitive ratio that depends on its penalty function and the platform's objective. In addition, for a fixed weight on efficiency, we show that a balancing algorithm with an appropriately-chosen penalty function is asymptotically optimal, meaning that it achieves the upper bound on the performance of any online algorithm in the limit when the opportunity-specific thresholds are all large. Our analysis adopts the LP-free framework of Goyal et al. (2020) and has connections to a related problem in Buchbinder et al. (2007), but our results do not follow from either. In fact, our achievable lower bounds compare favorably to corresponding results in Buchbinder et al. (2007).

Designing an algorithm for VM. While the framework above allows us to formally understand how to develop algorithms given dual objectives of efficiency and equity, we must still make several design choices to deploy an algorithm in the field. In collaboration with VM, we first determine appropriate metrics for evaluating efficiency and equity on their platform and define objectives that reflect their desired outcomes. Then, inspired by our theoretical framework and taking into account various practical considerations on VM, we propose a new display ranking algorithm called *SmartSort*. Like the balancing algorithms we analyzed in our modeling framework, SmartSort ranks opportunities based on both their appeal to the current volunteer and a penalty term that depends on the opportunity's sign-up history. While appeal cannot be directly observed with the available data, SmartSort proxies for it based on the distance of the opportunity to a volunteer and how recently it was updated; this proxy for appeal is consistent with the display

ranking algorithm previously used by VM. Toward specifying the penalty term, we first use simulations to explore the potential trade-off between the two objectives (efficiency and equity) under SmartSort; we find that the extent of the trade-off crucially depends on penalty parameters and on characteristics of the underlying marketplace. Our calibration of the penalty parameter was informed by our simulation study as well as conversations with VM managers.

Field Experiments. To measure the impact of Smartsort, we designed and implemented two field experiments in the Dallas-Fort Worth and Southern California geographic regions (see Figure 2), which jointly comprise around 14% of VM's user base. We use a difference-in-differences analysis to rigorously identify the causal impacts of SmartSort on VM's two primary metrics: weekly number of sign-ups (efficiency) and weekly number of opportunities with a sign-up (equity). Our analyses suggest that SmartSort creates a Pareto improvement for VM: equity significantly increases by around 8% without any significant cost to efficiency. This is illustrated in Figure 2, where the bars correspond to the estimated percentage change in our two outcome metrics (i.e., efficiency and equity). The consistency of these results across regions is striking, as SmartSort was implemented in Dallas-Fort Worth in Summer '22 while it was implemented in Southern California in Winter '22-'23. In total, across the two experimental areas, SmartSort increased the number of opportunities with at least one sign-up (equity) by an average of roughly 80 per week.

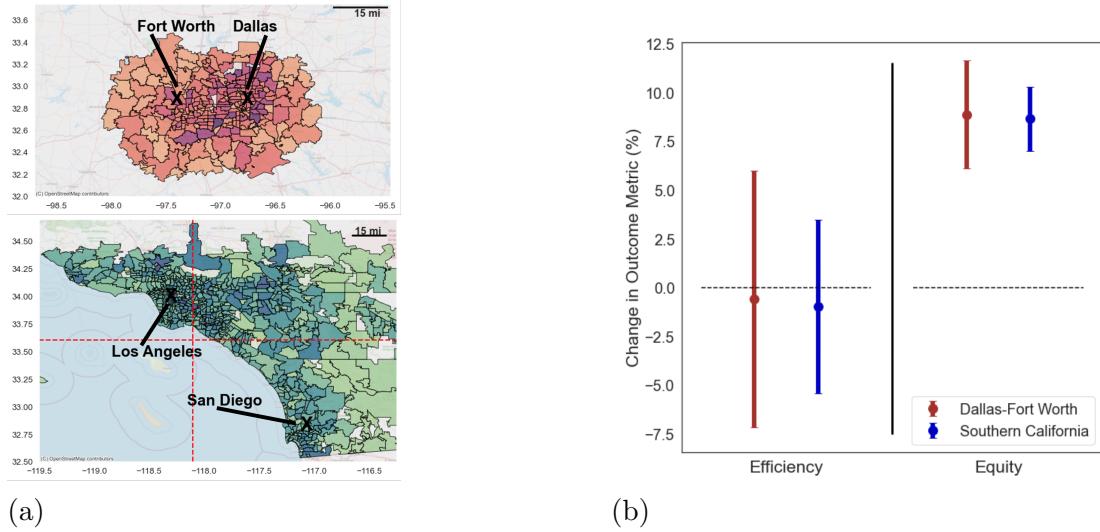


Figure 2 (a): The population density of the experimental areas. (b): The difference-in-differences effects of SmartSort on efficiency and equity in Dallas-Fort Worth and Southern California.

Further Impact. Encouraged by these findings, starting on April 27, 2023, VM began using SmartSort as the default display ranking algorithm nationwide. Extrapolating our estimated impact to a national scale, we expect that SmartSort will provide an additional 30,000 sign-ups every year to opportunities that would have otherwise lacked access to volunteers.

The rest of the paper is organized as follows. In Section 1.1 we discuss further related work. Section 2 introduces our modeling and algorithmic framework and characterizes the achievable performance guarantees in such a setting. Section 3 provides more background on VM and introduces SmartSort, our new ranking algorithm. In Section 4 we present our empirical analysis of SmartSort's impact before concluding in Section 5.

1.1. Further Related Work

Our work relates to the growing line of research on how to design platforms to promote social good. These studies span a wide range of applications, including recently-developed platforms in the domains of agriculture (de Zegher and Lo 2020, Levi et al. 2020, Adebola et al. 2022), healthcare (McElfresh et al. 2020, Anunrojwong et al. 2022), and volunteer management (Lo et al. 2021, Manshadi and Rodilitz 2022, Manshadi et al. 2022b). Our work contributes to this last stream of papers, which explore how online platforms can improve the effectiveness, engagement, and retention of volunteers.

Our work also contributes to the burgeoning literature on fairness in operations. Motivated by societal and ethical considerations, several recent papers study how to incorporate fairness in operational decisions such as pricing (Cohen et al. 2021, 2022), online learning (Baek and Farias 2021, Gupta and Kamble 2021), and dynamic allocation and selection (Correa et al. 2021, Arsenis and Kleinberg 2022, Manshadi et al. 2022a, Aminian et al. 2023). In this line of research, our work is most closely related to the papers that focus on fairness issues that arise in two-sided online platforms (Biega et al. 2018, Singh and Joachims 2018, Athey et al. 2022, Chen et al. 2022, Freund et al. 2023).

While the aforementioned papers take either a theoretical or empirical approach, our work bridges theory and practice by implementing a new ranking algorithm to improve equity based on well-grounded theory. On the theory side, our model of VM's problem can be viewed as a variant of online assortment optimization (see, e.g., Golrezaei et al. 2014, Ma and Simchi-Levi 2020, Rusmevichientong et al. 2020, Gong et al. 2021, Feng et al. 2022); however, those papers typically consider maximizing matches or revenue without considering equity. Our framework is also related to the literature on online matching with concave rewards (Buchbinder et al. 2007, Devanur and Jain 2012), though we consider a more general assortment setting and establish tight bounds for a particular class of objective functions. For our algorithmic analysis, we focus on the class of balancing algorithms that has been extensively studied in the contexts of revenue maximization for online budgeted allocation (e.g., Mehta et al. 2007a), online assortment for retail (e.g., Golrezaei et al. 2014), and two-sided matching markets (e.g., Aouad and Saban 2023). This class of algorithms has also been successfully used in practice (e.g., Vazirani (2022)). Our analysis leverages the LP-free framework pioneered in Goyal et al. (2020).

In terms of implementation, our work is related to other collaborative projects that re-design matching markets. These types of projects usually focus on improving some measure of efficiency in the outcomes, such as improving productivity in rideshare marketplaces (Ong et al. 2021, Krishnan et al. 2022) or maximizing match rates in dating markets (Rios et al. 2022). Some of these projects also involve public sector applications, e.g., food donations (Prendergast 2022), refugee placement (Ahani et al. 2021b,a), kidney donation (Roth et al. 2005), and school choice (Baswana et al. 2019, Correa et al. 2022). Like these papers, our work aims to re-design a matching market with an eye toward improving social welfare. Closest to our work are Allman et al. (2022) and Bansak and Paulson (2022), which also consider distributional objectives in the resultant matching, albeit in school choice and refugee matching contexts, respectively.

2. Equitable and Efficient Online Display Ranking

This section studies the problem of online display ranking for matching platforms such as VM that have dual objectives of efficiency and equity. Section 2.1 defines the formal model, the objective, and the performance metric. Section 2.2 provides an upper bound on the performance of any online algorithm as well as lower bounds on the performance of a simple and intuitive class of algorithms.

2.1. Model

To better understand how platforms like VM can balance competing objectives of efficiency and equity, we introduce a stylized model for a platform's display rank optimization problem in the presence of equity considerations. We first describe the dynamics of the model, which is a generalization of online matching. Then we define the platform's bi-criteria objective as well as the competitive ratio, the metric that we will use to evaluate the performance of an online algorithm.

Throughout the paper, we follow the convention of denoting vectors with bold letters. Furthermore, for any $x \in \mathbb{N}$, we define $[x] := \{1, 2, \dots, x\}$.

Model Dynamics. Consider a platform with n static opportunities and a sequence of T (prospective) volunteers arriving one-at-a-time, in discrete periods that we index according their arrival order. Each arriving volunteer $t \in [T]$ has a *choice function* that determines the probability that they sign up for an opportunity when presented with a particular ranking of opportunities. Specifically, we use $\phi_{i,t}(\vec{S})$ to denote the probability that volunteer t signs up for opportunity i when presented with a ranking \vec{S} . (Formally, \vec{S} is a permutation of $[n]$, and we denote the set of all permutations by \mathcal{S} .) We make no assumptions about the arrival sequence of the volunteers nor do we require knowledge of T ; in other words, we allow for an adversarial arrival sequence.

Note that the choice function can capture cases where an opportunity i arrives and/or departs during the horizon (by setting the choice probability of i to be zero under any ranking for all

volunteers arriving when opportunity i is not present), as well as cases where the volunteer only considers the k top-ranked opportunities (by setting the choice probability to 0 for any opportunity after position k in the display ranking).

We make the following assumption about the volunteer choice functions that we keep throughout.

Assumption 1 *The platform can efficiently solve for the ranking \vec{S} that maximizes $\sum_{i \in \vec{S}} w_i \phi_{i,t}(\vec{S})$ for any non-negative vector of weights $\mathbf{w} \in \mathbb{R}_+^n$.*⁴

Upon the arrival of volunteer t , the platform observes the volunteer's choice function and must immediately present a ranking $\vec{S} \in \mathcal{S}$. Faced with ranking \vec{S} , volunteer t signs up for opportunity i with probability $\phi_{i,t}(\vec{S})$ or exits without signing up with probability $1 - \sum_{i=1}^n \phi_{i,t}(\vec{S})$. We introduce the binary random variable $X_{i,t}(\vec{S})$ which is equal to 1 if volunteer t signs up for opportunity i when presented with ranking \vec{S} , and 0 otherwise.

Platform Objective. To model the platform's dual goals of efficiency and equity, we consider a *bi-criteria objective* function with both an *efficiency-oriented component* and an *equity-oriented component*. For the efficiency-oriented component, we use the total number of sign-ups (also known as market share in private-sector contexts). For the equity-oriented component, we only reward sign-ups for an opportunity i up to an opportunity-specific *equity threshold* denoted $m_i \in \mathbb{N}$, which incentivizes equitable access (or exposure). Any sign-ups for opportunity i beyond m_i can be thought of as an excessive and inequitable allocation of volunteers, and thus they do not contribute toward the equity component of the objective. To capture the heterogeneous needs across opportunities, we let m_i vary, and we use \mathbf{m} to denote the vector of equity thresholds. We consider the convex combination of these two components with a weight of $\gamma \in [0, 1]$ on the efficiency component.

This notion of equity is motivated by our collaboration with VM and is aligned with VM's desire to have a more uniform sign-up distribution (as discussed in Section 1). To see why, as a thought exercise, suppose $m_i = m$ universally for $i \in [n]$. Then, for a fixed number of sign-ups (i.e., fixing efficiency), a uniform distribution of sign-ups maximizes equity. We note that this is just one choice of objectives that incentivizes a uniform sign-up distribution.⁵

To help formally state the goal of a platform under this bi-criteria objective, for any $t \in [T]$, we define $Z_{i,t} \in \mathbb{N}$ to be the number of sign-ups for opportunity i at the end of period t . In other

⁴ This assumption is usually made in the literature (e.g. Golrezaei et al. 2014, Feng et al. 2022) and is satisfied by some commonly-used choice functions (e.g., Kempe and Mahdian 2008, Sumida et al. 2021). If the platform can only solve the above maximization up to a multiplicative α -factor for $\alpha \in (0, 1)$, our performance guarantee will be degraded by the same factor α .

⁵ In settings where $m = 1$ and the number of opportunities exceeds the number of potential sign-ups within the considered time horizon, then many reasonable equity-oriented objectives coincide (e.g., total variation distance), as they are proportional to the number of opportunities with at least one sign up. This is the setting for VM, as we discuss in Section 3.1.

words, if a ranking \vec{S}_τ is the ranking shown in period τ for each $\tau \in [T]$, then $Z_{i,t} = \sum_{\tau=1}^t X_{i,\tau}(\vec{S}_\tau)$. The platform's objective is to choose a ranking algorithm π that maximizes

$$\max_{\pi} \quad \mathbb{E} \left[\gamma \sum_{i=1}^n \sum_{t=1}^T X_{i,t}(\vec{S}^\pi) + (1 - \gamma) \sum_{i=1}^n \sum_{t=1}^T \mathbb{1}\{Z_{i,t} \leq m_i\} X_{i,t}(\vec{S}^\pi) \right], \quad (\text{OBJ})$$

where the expectation is taken with respect to randomness in volunteer choices and (possibly) to randomness in the algorithm itself. This objective is equivalent to a setting where the platform obtains a particular form of concave rewards from each additional sign-up: the platform generates a reward of 1 for the first m_i sign-ups for opportunity i and a reward of γ for each subsequent sign-up. In other words, we can re-write (OBJ) as

$$\max_{\pi} \quad \mathbb{E} \left[\sum_{i=1}^n \sum_{t=1}^T \left(\mathbb{1}\{Z_{i,t} \leq m_i\} X_{i,t}(\vec{S}^\pi) + \gamma \mathbb{1}\{Z_{i,t} > m_i\} X_{i,t}(\vec{S}^\pi) \right) \right]. \quad (1)$$

Performance Metric. As mentioned above, we take an adversarial approach—thus assuming no prior knowledge on the arrival sequence of volunteers—and evaluate the performance of an online algorithm by its competitive ratio, i.e., the worst-case ratio it achieves when compared to a benchmark solution. We now formally define our benchmark and the competitive ratio metric.

Consider any *instance* of the problem \mathcal{I} , which consists of (i) a set of opportunities $[n]$, (ii) a sequence of volunteer choice functions $\{\phi_{i,t}(\vec{S}) : i \in [n], t \in [T], \vec{S} \in \mathcal{S}\}$, and (iii) an objective function parameterized by an efficiency weight γ , and (iv) a vector of equity thresholds \mathbf{m} . As is common in the literature (see, e.g., [Gong et al. 2021](#)), our benchmark is an optimal clairvoyant algorithm that knows the entire sequence of arrivals in advance but does not know the realization of volunteer choices in advance, as formally defined below.

Definition 1 (Optimal Clairvoyant Benchmark) *The optimal clairvoyant algorithm is the solution to an exponential-sized dynamic program for the problem instance $\mathcal{I} := ([n]; \{\phi_{i,t}(\vec{S}) : i \in [n], t \in [T], \vec{S} \in \mathcal{S}\}; \gamma; \mathbf{m})$. Fixing this instance, for each arriving volunteer $t \in [T]$, the optimal clairvoyant algorithm presents a ranking \vec{S}_t^{OPT} that maximizes the expected objective value at the end of the time horizon given the instance and the full history of the first $t - 1$ volunteer choices.*

For a fixed problem instance, we denote the expected value of the benchmark as $\mathbb{E}[\text{OPT}(\mathcal{I})]$ and the expected value of an online algorithm π as $\mathbb{E}[\pi(\mathcal{I})]$, where the randomness is taken over the volunteers' choices and (potentially) the randomization within the algorithm. As our objective function is parameterized by γ and \mathbf{m} , we similarly parameterize the competitive ratio, as defined below.

Definition 2 (Competitive Ratio) For parameters $\gamma \in [0, 1]$ and $\underline{m} \in \mathbb{Z}_+$, the competitive ratio of an online algorithm π is given by

$$CR(\pi, \gamma, \underline{m}) = \min_{\mathcal{I} \in \mathcal{I}_{\gamma, \underline{m}}} \frac{\mathbb{E}[\pi(\mathcal{I})]}{\mathbb{E}[\text{OPT}(\mathcal{I})]},$$

where $\mathcal{I}_{\gamma, \underline{m}}$ is the set of instances with efficiency weight γ and equity thresholds all at least \underline{m} , i.e., $m_i \geq \underline{m}$ for all $i \in [n]$.

For the above setting, we reiterate that the platform's goal is to design an online algorithm that achieves a strong competitive ratio given its objective function.

Connection to Retail Settings. We conclude by positioning our modeling framework in the broader literature of online matching and assortment in retail settings.

The particular form of the equity-oriented component of our objective function enables us to connect our model to more traditional online matching and assortment problems, often studied in the context of retail. Typically in retail settings, there are hard “inventory” or capacity constraints, meaning that there is a limited quantity of each resource. On VM’s platform, opportunities can receive an arbitrary number of sign-ups but, in a way, our equity metric plays a similar role to a soft inventory constraints.

To see this, note that when $\gamma = 0$, our problem can equivalently be framed as an allocation problem with hard inventory constraints, where the capacity for opportunity i is equal to m_i . This is a well-studied problem where the best-possible competitive ratio can be achieved via an exponential balancing algorithm when capacities are large (see, e.g., [Mehta et al. 2007b](#) and [Golrezaei et al. 2014](#)). More generally, for $\gamma > 0$, our problem can be viewed as an allocation problem with *soft* capacity constraints where we can exceed the capacities, but only at a “cost” of $1 - \gamma$. This model would apply in retail settings without lost sales; instead, when a customer chooses a product that is out of stock, the company can pay some cost (e.g., the cost of expedited shipping from a different location). Our goal in the subsequent section can therefore equivalently be thought of as investigating how the achievable guarantees of an online algorithm depend on this cost.

2.2. Algorithm Design and Analysis

Having introduced the model in Section 2.1, we now focus on understanding the achievable performance of online algorithms in our setting. We begin this section by establishing an upper bound on the performance of any online algorithm. Then we introduce a practical class of algorithms, called *balancing* algorithms, and establish a parametric lower bound on their competitive ratio. Finally, we show that a particular algorithm from this class achieves the best-possible competitive ratio in the limit where the minimum equity thresholds go to infinity.

Proposition 1 (Upper Bound on Competitive Ratio) *For any $m \in \mathbb{N}$ and any $\gamma \in [0, 1]$, no online algorithm (deterministic or randomized) can obtain a competitive ratio (see Definition 2) greater than $(e - 1 + \gamma)/e$.*

We prove Proposition 1 by making a connection to the online b -matching problem studied in Mehta et al. (2007b). The online b -matching problem corresponds to the following special case of our setting: (i) each volunteer only considers the top-ranked opportunity and signs up for it with probability one if they are compatible (i.e., there is an edge between them in the underlying bipartite graph); (ii) each opportunity has equity threshold b (corresponding to the capacity of a bin in the b -matching setting); and (iii) $\gamma = 0$. Using this connection, we analyze the class of hard instances introduced in Mehta et al. (2007b) under our bi-criteria objective function (i.e., where each additional sign-up above m_i generates a reward of $\gamma \in [0, 1]$). Roughly speaking, in the b -matching setting Mehta et al. (2007b) show that for this instance class an online algorithm cannot utilize a $1/e$ fraction of the arrivals, leading to an upper bound of $1 - 1/e$. However, under our objective function, an online algorithm can still generate a reward of γ/e from those “unutilized” arrivals in our setting, leading to the upper bound of Proposition 1. We present the full proof details in Appendix A.1.

Proposition 1 suggests that the efficiency weight γ plays a crucial role in the achievable performance guarantee: in the extreme where $\gamma = 0$, we cannot achieve a competitive ratio better than $(1 - 1/e)$ — which is the upper bound for the online b -matching problem (Mehta et al. 2007b). In the other extreme where $\gamma = 1$, we can achieve a competitive ratio of 1 by myopically maximizing the sign-up probability for each volunteer. In the following example, we show that the efficiency weight γ can have an even more profound impact on the performance of a *fixed* algorithm:

Example 1 *Suppose there are n opportunities with identical equity thresholds of m . Also, suppose there are $m \cdot n$ volunteers with an identical choice model that works as follows: each volunteer only considers the top-ranked opportunity; if it is opportunity 1, they sign up w.p. 1. Otherwise, they sign up for that opportunity w.p. $p \in (0, 1)$.*

Consider a myopic algorithm that maximizes the sign-up probability for each volunteer t by always displaying opportunity 1 as the top-ranked opportunity. If $\gamma = 1$, this algorithm achieves an objective value of $m \cdot n$ and a competitive ratio of one. Thus, this “appeal-based” ranking is the best-possible algorithm.

On the other hand, when $\gamma = 0$, such an algorithm would perform arbitrarily bad: by consistently ranking opportunity 1 on top, it only achieves an objective value of m . A different simple algorithm that alternates the top-ranked opportunity in a round-robin fashion achieves an objective value of at least $m \cdot n \cdot p$ (in expectation), which is much larger than m for $p >> 1/n$.

The above example suggests that maintaining *balance* in exposure can be an effective way to achieve equity. However, displaying opportunities in top positions regardless of their appeal may be inefficient: in the toy example above, ranking opportunity $i \geq 2$ at the top only generates a sign-up with probability p (compared to a sign-up probability of 1 if opportunity $i = 1$ is top-ranked). Thus, we also need to incorporate the sign-up probability of each opportunity into the ranking decision, and the degree to which we take this sign-up probability into account influences the potential trade-off between efficiency and equity. Building on this simple idea, we consider a general class of balancing algorithms that can be tailored based on the platform's objective function, defined as follows:

Definition 3 (Balancing Algorithms) *Recall that the number of sign-ups for opportunity i after the arrival of volunteer $t \in [T]$ is given by $Z_{i,t} \in \mathbb{Z}_+$. A balancing algorithm with penalty function ψ recommends the ranking \vec{S}_t that satisfies*

$$\vec{S}_t = \operatorname{argmax}_{\vec{S} \in \mathcal{S}} \sum_{i \in \vec{S}} \psi(Z_{i,t-1}/m_i) \cdot \phi_{i,t}(\vec{S}), \quad (2)$$

where $\psi : \mathbb{R}_+ \rightarrow [0, 1]$ is an arbitrary non-increasing function.⁶

This class of algorithms balances allocation across opportunities by applying a penalty function for each opportunity that is (weakly) increasing in their current “fill rate” (i.e., their number of sign ups normalized by their threshold). This class of algorithms has proven to be robust in theory and practice in similar online allocation problems with efficiency-oriented objectives such as online budgeted allocation and assortment (see, e.g., Mehta et al. 2007b and Golrezaei et al. 2014). Under our setting and for our bi-criteria objective function, we establish the following lower bound on the competitive ratio of this class of algorithms.

Theorem 1 (Lower Bound on Competitive Ratio for Balancing Algorithms) *Consider a balancing algorithm π^{BAL} with penalty function ψ . For any $\underline{m} \in \mathbb{Z}_+$ and any $\gamma \in [0, 1]$, we have:*

$$CR(\pi^{\text{BAL}}, \underline{m}, \gamma) \geq \alpha \cdot \beta \cdot \kappa, \quad (3)$$

where $\alpha = \min \left\{ 1, \frac{\gamma}{\psi(1)} \right\}$, $\beta = \min \left\{ 1, \lim_{x \rightarrow \infty} \frac{\psi(x)}{\gamma \cdot \psi(0)} \right\}$,⁷ and

$$\kappa = \min_{m, z \in \mathbb{Z}_+, m \geq \underline{m}} \psi(z/m) + (1/m) \sum_{k=1}^{\min\{z, m\}} \left(1 - \psi((k-1)/m) \right). \quad (4)$$

⁶ Based on Assumption 1, the platform can efficiently solve the above maximization problem.

⁷ If a term in either minimum has a denominator equal to 0, we follow the convention that the value of that minimum is 1.

Theorem 1 establishes a guarantee that holds for any balancing algorithm and any choice of parameters for the objective function. Note that the three constants α , β , and κ depend on the choice of penalty function as well as parameters γ and \underline{m} . The complete proof can be found in Appendix A.2.

To preview the proof, we adopt the LP-free framework developed in Goyal and Udwani (2019) and prove this result in three steps. (i) First, we define pseudo-rewards that depend on the particular choice of penalty function used by a balancing algorithm π^{BAL} (see Equations (12) and (13)). (ii) Then, we show that the expected value of π^{BAL} is at least a constant fraction of the expected sum of the pseudo-rewards (Lemma 1). (iii) We subsequently show that the expected sum of the pseudo-rewards is at least a constant fraction of the expected value of the clairvoyant benchmark (Lemma 2).

Leveraging Theorem 1, fixing a weight on efficiency γ we identify a penalty function that achieves the best possible competitive ratio in the limit $\underline{m} \rightarrow \infty$. In the following corollary, proven in Appendix A.3, we introduce this penalty function (parameterized by γ) and place a lower bound on its competitive ratio.

Corollary 1 (Asymptotically Optimal Algorithm) *Given an $\underline{m} \in \mathbb{Z}_+$ and a $\gamma \in [0, 1]$, for a balancing algorithm π^{BAL} with penalty function $\psi(x) = 1 - (1 - \gamma) \exp(-1 + \min\{x, 1\})$, we have*

$$CR(\pi, \underline{m}, \gamma) \geq e^{-1/\underline{m}} \left(\frac{e - 1 + \gamma}{e} \right),$$

which matches the upper bound of Proposition 1 in the limit as $\underline{m} \rightarrow \infty$.

It is instructive to examine the form of this asymptotically optimal penalty function: the penalty for each new sign-up exponentially increases, but only up to the equity threshold. After an opportunity i has reached its equity threshold (i.e., a fill rate of one), its penalty term remains constant and equal to γ . To better understand why, we make the following observations: (i) a less severe penalty may mistakenly prioritize efficiency, but (ii) continuing to penalize an opportunity may hurt efficiency without a sufficient gain in equity.⁸

Re-visiting the Connection to Retail Settings As discussed in Section 2.1, when $\gamma = 0$, our problem is equivalent to an inventory management problem with hard capacity constraints where the capacity for opportunity i is equal to m_i . In such a setting, the algorithm identified in Corollary 1 exactly corresponds to the algorithm presented in Theorem 2 of Golrezaei et al. 2014.

For $\gamma > 0$ (i.e., when capacities can be exceeded for a cost), the optimal penalty function that we identify imposes a less severe penalty for each sign-up, compared to the setting where $\gamma = 0$ (i.e.,

⁸ Roughly speaking, these two countervailing forces correspond to the factors α and β , respectively, in the competitive ratio given by Theorem 1.

hard capacity constraints). Intuitively, this is because the cost of exhausting capacity is diminished for larger γ .

Our setting can be directly compared to the setting in Section 6.2 of [Buchbinder et al. \(2007\)](#), which considers a matching problem with piecewise-linear concave rewards. Roughly speaking, their algorithm consists of the following approach (using terminology from a retail setting): consider two copies of each good, one with a capacity of m_i that generates a reward of 1, and another with unlimited capacity that generates a reward of γ . When determining which good to match, use the optimal penalty function with hard capacity constraints (i.e., $\psi(x) = 1 - e^{-1+x}$), but include both copies of each good as options. In the limit of large capacities, the lower bound on the competitive ratio of this algorithm approaches $(e - \gamma)/(e - 1)$. In contrast, the algorithm in Corollary 1 imposes a less severe penalty after each sign-up and achieves a strictly better competitive ratio for $\gamma \in (0, 1)$.

Taking inspiration from our modeling framework, next we turn our attention to the VM platform and design a new display ranking algorithm similar to the class of balancing algorithms while respecting various practical considerations.

3. Toward a More Equitable Sign-up Distribution on VM

In this section, we describe our applied work in collaboration with VM. We begin in Section 3.1 by revisiting VM's platform design, describing their current ranking algorithm, and specifying objectives that correspond to their desired outcomes. Translating VM's problem into the bi-criteria setting introduced in Section 2.1, we then turn our attention to algorithm design. In Section 3.2, we introduce our new ranking algorithm (SmartSort), which is inspired by our theoretical development while also incorporating a variety of practical considerations.

3.1. Additional Platform Background and Defining Objectives

As discussed in Section 1, VM operates similarly to many other two-sided online marketplaces. Nonprofit organizations post opportunities on VM's website, and prospective volunteers can search and sign-up for opportunities according to the process illustrated in the left and middle panels of Figure 1 in the Introduction. Specifically, when a volunteer visits the VM website, they are assigned a pre-populated search location based on information derived from their IP address or browser cookies. Upon selecting "Find Opportunities," a volunteer is recommended a *ranked list*—consisting of 25 opportunities per page—with a title and brief description for each opportunity. The volunteer can then browse these opportunities and may choose to sign-up by selecting "I Want to Help" for one or more of them.

It is well documented that when online platforms display many options, users exhibit "position bias", namely, they are more likely to view options in top positions (see, e.g., [Ursu \(2018\)](#) and references therein). Hence, the sign-up distribution on VM is crucially influenced by the display

rankings that VM presents to volunteers. Therefore, we next describe VM's default display ranking algorithm.⁹

VM's default display algorithm (CP). Prior to our collaboration, the default display ranking algorithm on VM (which is called “Near Me & New” and hereafter referred to as *Current Practice* (CP)) was appeal-based: roughly speaking, it displays opportunities in decreasing order of their perceived appeal to the searching volunteer (as defined by VM). As its name suggests, the “Near Me & New” algorithm displays closer and more recently-updated opportunities higher in the display ranking, as these are two of the only features that VM has data on that are correlated with the appeal of an opportunity toward a volunteer.

More precisely, let \mathcal{O} be the platform's current set of posted opportunities. Whenever a volunteer t searches from a specific location, the platform assigns every opportunity $i \in \mathcal{O}$ a tuple, $(g(d_{i,t}), u_{i,t})$, specific to that volunteer. The scalar $d_{i,t}$ represents the distance between the volunteer's search location and the opportunity's platform-assigned location,¹⁰ while the scalar $u_{i,t}$ represents the number of days between volunteer t 's search time and the time of the opportunity's most recent update. The following function $g : \mathbb{R}_+ \rightarrow \mathbb{Z}_+$ then maps every distance into a discrete classification that VM calls a *distance band*:

$$g(d) = \begin{cases} 0, & \text{if } d \leq 1 \\ \lceil d/5 \rceil, & \text{otherwise} \end{cases} \quad (5)$$

The display ranking limits itself to the set of opportunities that lie within 20 miles of the search location and that were posted or updated within the past 90 days, $\{i \in \mathcal{O} : d_{i,t} \leq 20, u_{i,t} \leq 90\}$. These opportunities are then presented to the volunteer sorted in descending order of $g(d_{i,t})$. Within the distance bands, ties are broken in favor of the more recently updated opportunities, that is, those with smaller $u_{i,t}$.¹¹

Under CP, certain opportunities were displayed in the top positions disproportionately often. For instance, during the same (short) window of time, all volunteers searching from the same zipcode will see the exact same ranking of opportunities. This may have contributed to imbalanced weekly sign-up distributions such as the one shown in the right panel of Figure 1.

⁹ Volunteers have the ability to access alternative display rankings, either by applying filters that specify the type of opportunities they are searching for (e.g., opportunities that support the elderly), or by changing the sorting criteria. However, the vast majority of volunteers (around 95%) do not exercise these alternative options.

¹⁰ Distance is measured in miles using the Haversine distance formula, based on the latitude and longitude corresponding to the opportunity and the search locations.

¹¹ Formally, the opportunities are displayed lexicographically in descending order, i.e., first in descending order of $g(d_{i,t})$, and then in descending order of $u_{i,t}$. As an example, the opportunity displayed first will be the most recently updated opportunity within 1 mile of the search location.

Defining objectives and outcome metrics. VM places a high priority on generating a large volume of sign-ups, as this is a key metric for demonstrating the success of their platform: their website prominently displays the total number of volunteer sign-ups that they have facilitated (18 million, as of 2023). However, conditional on maintaining the same sign-up volume, VM would prefer sign-ups to be more uniformly distributed across opportunities (compared to the current situation).

An important practical consideration is the time horizon over which we will be evaluating the shape of the sign-up distribution. To illustrate, suppose we have ten opportunities, o_w , $w \in [10]$, and let us consider two scenarios. In the first scenario, suppose in week $w \in [10]$, opportunity o_w gets 10 sign-ups, and the other opportunities get none. In the second scenario, each opportunity gets one sign-up per week for $w \in [10]$. In both scenarios, the total number of sign-ups per week is identical, regardless of the time horizon (e.g., ten weeks vs. one week). By contrast, the shape of the sign-up distribution can change dramatically depending on the time horizon. If we set the horizon to be ten weeks, both scenarios have a uniform sign-up distribution; however, if we set the horizon to be one week, then for each $w \in [10]$, the weekly sign-up distributions in the first scenario are highly imbalanced, but remain uniform in the second scenario.

After discussion with VM managers, we concluded that aiming for uniformity in the *weekly* sign-up distributions best reflects their goals, for two main reasons: (i) If an opportunity receives a large number of sign-ups within a week, this can overwhelm their volunteer management staff, leading to unnecessary costs and/or underutilized volunteers. (ii) As volunteers are the short side of the market on VM—especially when focusing on a single week—a uniform sign-up distribution corresponds to more opportunities receiving at least one sign-up. Providing a consistent supply of volunteers incentivizes opportunities to continue engaging with the platform; in contrast, if opportunities do not receive any sign-ups, this can lead to departure from the platform and potential damage to VM's reputation. Going beyond a week, VM is willing to tolerate larger imbalances in sign-ups because opportunities posted on VM can persist for varying amounts of time and may need to replenish their supply of volunteers at different rates.

With these considerations in mind, in collaboration with VM we agreed to consider the following two quantitative metrics:

- a) **Efficiency:** The total number of sign-ups in a week.
- b) **Equity:** The number of opportunities with at least one sign-up in a week.

Going forward, we will use the terms “efficiency” and “equity” to specifically refer to these two metrics. As discussed in detail in Section 4.1, these two objectives will be our primary outcome metrics when measuring the impact of our new algorithm in our field experiments.

Focusing on a single week, these two metrics correspond to the two objective criteria in our model (see Section 2.1) when $m_i = 1$ for each opportunity i . Inspired by our theoretical results—which show that constant-factor guarantees can be obtained in such a setting by balancing algorithms (see Theorem 1)—we aim to use a balancing algorithm to help VM improve its equity objective, ideally, with minimal harm to efficiency.

3.2. Algorithm Design: Introducing SmartSort

In this section we introduce *SmartSort* (SS), our new algorithm — inspired by the balancing algorithms introduced in Section 2.2 — that takes into account various practical considerations on VM. Our design of SS involved several steps which we describe in the rest of this section. First, in Step 1, we discuss various practical limitations which led us to limit our choice of a ranking algorithm to a score-based one. Next, in Step 2, we introduce the form of our scoring function inspired by balancing algorithms. Then in Steps 3 and 4, we describe how we decided on the key components of the scoring function based on our conversations with VM managers, simulations, and data analysis based on the Dallas-Fort Worth geographical region, where our first field experiment was launched.

Step 1: Practical Considerations. When designing our new algorithm we kept in mind the following constraints, which can be roughly divided into two categories: data limitations and implementation constraints.

Data limitations: In designing a ranking algorithm, ideally, we would want to estimate (or learn) the underlying choice model for different types of volunteers based on features beyond their location (e.g., demographic information). In order to do so, we would need to leverage fine-grained data on volunteers' browsing and viewing activities and as well as volunteer-specific characteristics. Unfortunately, VM does not record session-level data and does not have reliable access to other volunteer-specific features. As such, directly implementing a ranking algorithm that relies on a choice model (such as the balancing algorithm specified in Definition 3) is not feasible under the current circumstances.

Implementation constraints: Putting aside the data limitations, implementing an algorithm such as a balancing algorithm faces computational and communication challenges. Note that even this simple and myopic class of algorithms requires optimizing the ranking according to (2). Re-optimizing the ranking of hundreds of opportunities upon each volunteer arrival poses a substantial computational challenge for VM. In contrast, VM's current ranking algorithm—as described in 3.1—is a *score-based* ranking (meaning that it simply assigns a score to each opportunity and then ranks opportunities in a descending order of those scores). Such a ranking algorithm is practically appealing because it is intuitive, easy to communicate, and

computationally fast. To get buy-in from VM managers, we decided to limit ourselves to the class of score-based rankings. Thus, our algorithm design entails defining a scoring function which we describe next.

Step 2: Defining the Scoring Function. In light of the above limitations and inspired by the class of balancing algorithms, we proposed SmartSort, an algorithm that ranks opportunities in decreasing order of a scoring function that depends on the opportunity appeal and the number of sign-ups the opportunity has recently received. Concretely, whenever a volunteer t arrives, SmartSort will display opportunities in descending order of the following scoring function:

$$\text{score}_{i,t} = a_{i,t} \cdot \left(1 - \exp(-1 + \min\{c_1 \cdot Z_{i,t}, c_2\}) \right) / \left(1 - \exp(-1) \right). \quad (6)$$

The first factor, $a_{i,t}$, represents the pair-specific appeal of opportunity i to volunteer t , which we use as a proxy for volunteer's choice probability. The second factor represents the penalty term — much like the penalty function in an exponential balancing algorithm— where $Z_{i,t}$ denotes the total number of sign-ups for opportunity i within the past week. This rolling count of sign-ups is in line with VM's desire to impact sign-up distributions on a weekly basis (as discussed in Section 3.1). Due to technical limitations, VM cannot update the sign-up counts for opportunities in real time; only a nightly update is feasible. With this constraint in mind, and after discussion with VM's engineering team, we define the counter $Z_{i,t}$ as the total number of sign-ups that opportunity i received between 1 and 8 days before the arrival of volunteer t . The denominator is simply a normalization factor. Having defined the form of the scoring function, the next two steps involve specifying the two key components of the scoring function, i.e., the pair-specific appeals ($a_{i,t}$) and the parameters c_1 and c_2 in the penalty term.

Step 3: Specifying Opportunity Appeals. As mentioned earlier, given VM's data limitations, precisely estimating the appeal of each opportunity is infeasible: though we observe which opportunities receive the most sign-ups, we cannot observe what opportunities volunteers see before they sign up, nor do we have any information on volunteers who leave without signing up. Moreover, the lack of volunteer-specific data implies that we will be unable to personalize the offerings to volunteers beyond their location. Therefore, consistent with CP, we approximate the pair-specific appeal between an opportunity i and a volunteer t using distance between the two ($d_{i,t}$) and the number of days between the volunteer's search time and the opportunity's last update ($u_{i,t}$) as the only two relevant factors. However, we refine the formula such that the influence of each factor is continuous.

More specifically, upon arrival of a volunteer t , we likewise limit the display ranking to the set of opportunities that lie within 20 miles of the search location and that were posted or updated

within the past 90 days, $\{i \in \mathcal{O} : d_{i,t} \leq 20, u_{i,t} \leq 90\}$. Then we construct a proxy for the appeal of each opportunity i , denoted $a_{i,t}$, according to the following formula:

$$a_{i,t} = 10 - \frac{d_{i,t}}{10} - \frac{u_{i,t}}{200}. \quad (7)$$

This estimated appeal decreases the further away the opportunity is from the search location and the longer ago that it was updated. An opportunity that is exactly at the volunteer's search location and was updated on the day of search will have an appeal of 10. Each mile of distance will decrease the appeal by 0.1, while each day since the last update will decrease the appeal by 0.005. The respective weights on $d_{i,t}$ and $u_{i,t}$ were chosen to mirror VM's view with regard to the relative importance of distance vs. recency. For example, consider volunteer t and two opportunities, $o_1, o_2 \in \mathcal{O}$; suppose o_1 is 5 miles closer to t compared to o_2 . The definition of distance band, given in (5), implies that VM thinks o_1 should be ranked above o_2 regardless of the recency factor (i.e., how $u_{1,t}$ compares to $u_{2,t}$). Our definition of appeal also respects the same order: $a_{1,t} > a_{2,t}$, regardless of the values of $u_{1,t}$ and $u_{2,t}$. (Remark that because o_1 is in the subset of displayed opportunities, by definition, we have $u_{1,t} \leq 90$. Same holds for o_2 .)

Step 4: Calibrating the Penalty Term. Our final design step is concerned with fine tuning the parameters c_1 and c_2 . As will become clear later, making an appropriate choice for c_1 is more critical than c_2 . (Roughly speaking c_2 can be viewed as a safeguard to avoid overly penalizing an opportunity with a surge of sign-ups. See Remark 1 at the end of this section.) As such, we first assume $c_2 = 1$ and focus on how to set the parameter c_1 .¹²

Note that c_1 (combined with the exponential form of the penalty term) allows us to directly control the severity of the penalty for each additional sign-up: if $c_1 = 0$ (i.e., opportunities are not penalized at all after receiving a sign-up, as was the case under CP) then efficiency should be high, but equity may be low. In the other extreme, where opportunities are effectively banished from the display ranking after receiving a sign-up, equity may increase but at a potentially large cost to efficiency.

To better understand how our choice of c_1 impacts equity and efficiency, we conducted extensive simulations under various scenarios for market characteristics. These simulations involve constructing problem instances with different volunteer choice models, appeal distributions, and supply-to-demand ratios of volunteers to opportunities. A key guiding insight of our simulations is the following: *the impact of parameter c_1 depends on the number of high appeal opportunities relative to the number of volunteers.*

To illustrate this point, we present a simple yet instructive simulation result. We construct three instances where the only difference across instances is the distribution of appeals (i.e., the number

¹² Note that setting $c_1 > 1$ may result in negative scores; as such it is only meaningful to consider $c_1 \in [0, 1]$.

of volunteers and opportunities, volunteers' choice model, etc., remain the same across instances.)¹³ The appeal distribution of Instances 1, 2, and 3 are plotted in the right panel of Figure 3. For each instance, we vary the parameter c_1 between 0 and 1, and report the average efficiency and equity for each value in the left panel of Figure 3.

We make the following qualitative observations: (i) As expected, increasing c_1 improves equity and harms efficiency. (ii) More intriguingly, however, the impact of parameter c_1 on equity and efficiency (and the steepness of their trade-off) crucially depends on the relative number of high appeal opportunities: in Instance 1 where there are very few high-appeal opportunities, increasing c_1 substantially hurts efficiency with moderate gain on equity. This is because penalizing high appeal opportunities may lead to filling top positions with low-appeal opportunities. In contrast, in Instance 3 where there are many more high-appeal opportunities, increasing c_1 comes with a small loss in efficiency and substantial gain in equity. This is because there are enough unpenalized high appeal opportunities to fill the top positions. (iii) Another important observation is that increasing c_1 only impacts the two objective up to some instance-specific "convergence" threshold. This threshold is lower when there are more high-appeal opportunities (e.g., 0.8 in Instance 1 vs. 0.4 for Instance 3). This is a consequence of the underlying position bias which can be viewed as a "first page effect", i.e., volunteers only consider opportunities that are ranked on the first page.¹⁴ For a given c_1 , if opportunity o is moved out of the first page after a sign-up, then o is effectively banished from the ranking. As such, increasing c_1 does not meaningfully impact the ranking of o or the outcomes. When there are many high appeal opportunities, even a small penalty c_1 may move opportunity o out of the first page, explaining the early convergence.¹⁵

Equipped with the above observations, we now describe our thought process in setting parameter c_1 in SmartSort. First, we remind that as discussed in Section 3.1, VM's desirable outcome is to improve their metric of equity without harming efficiency. Further, due to technical limitations, we needed to set a universal parameter c_1 regardless of potential variations across underlying appeal distributions. However, as discussed above imposing a large penalty for each sign-up can pose a substantial risk to efficiency in situations where high appeal opportunities are scarce, and it may not even help in situations where high appeal opportunities are abundant. (For example compare

¹³ We refer the reader to Appendix B.1 for a detailed description of how the instances are constructed.

¹⁴ Indeed there is evidence in other contexts that users do not go beyond the first page; for example research shows that on Amazon 70% of users never go past the first page (<https://searchengineland.com/how-to-navigate-amazons-sponsored-brand-ads-updates>).

¹⁵ Similar observations can be made with respect to other marketplace characteristics: for example, the trade-off shown in the left panel of Figure 3 becomes steeper when the number of opportunities (relative to volunteers) is smaller or when the volunteers' choice model is more sensitive to the opportunity appeals (e.g., if the sign-up probability is convex in the appeal as opposed to linear). For the sake of brevity, we omit additional simulations.

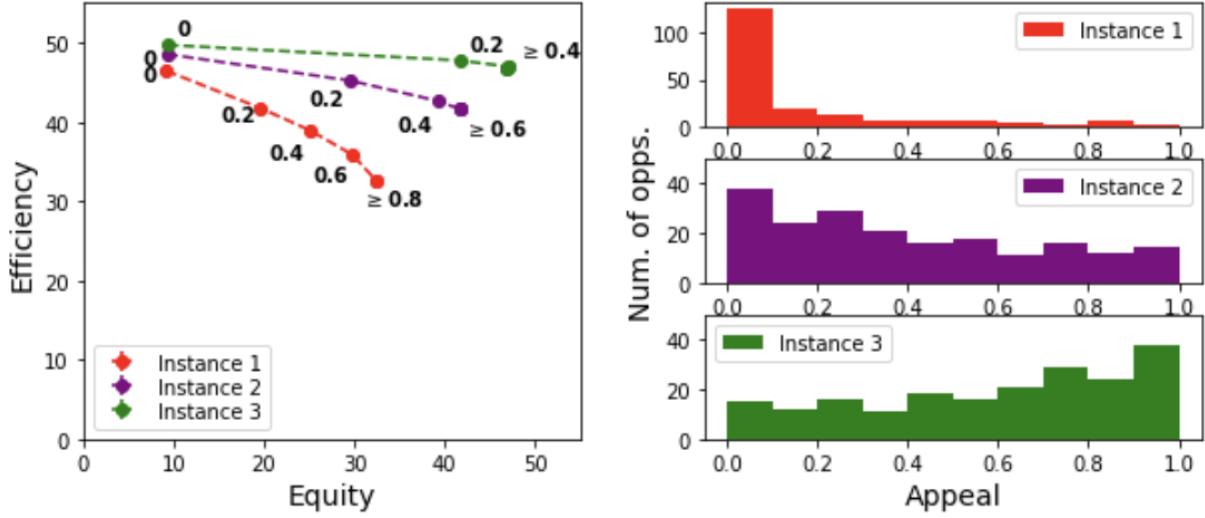


Figure 3 The right three panels show three instances of opportunity appeal distributions while the left panel shows the efficiency and equity outcomes for each of these instances under varying values for the penalty parameter c_1 .



Figure 4 The left (resp. right) panel shows the average daily distribution of opportunity appeals — as defined in Equation (7) — for the 100 most appealing opportunities for a representative volunteer from zip code 75039 (resp. 75201).

The solid black lines correspond to the average appeal score of the last opportunity to make the front page.

c_1 of 0.2 vs. 0.6 in Instances 1 and 3 in Figure 3). As such, we erred on the side of caution and set c_1 to be relatively small and equal to 1/12.

Our rationale behind the specific choice of 1/12 is attributed to the following data analysis and thought experiment: To understand the level of heterogeneity in appeal distribution, we considered the following two zip codes in the Dallas-Forth Worth region. Zip code 75039 is at the outskirts of Dallas with few nearby opportunities. In contrast, zip code 75201 is at the center of Dallas where many opportunities are located. Suppose every day between March 15th and May 23rd a representative volunteer searches from zip code 75039 (resp. 75201).

The left (resp. right) panel of Figure 4 shows the average daily distribution of opportunity appeals — as defined in Equation (7) — for the 100 most appealing opportunities for this representative volunteer.¹⁶ These empirical distributions suggest considerable differences in the appeal distribution. To highlight this, in the same histograms, we also plot a vertical line to indicate the average score of the 25th most appealing opportunity. We remind that each page on VM displays 25 opportunities. Thus in the absence of any penalty (i.e., if $c_1 = 0$), the black vertical line indicates the appeal “cutoff” for first page. We observed that this cutoff is considerably lower in zip code (75039) compared to 75201 (9.33 vs. 9.81).¹⁷

Based on this data analysis, the following thought experiment contributed to our choice of $c_1 = 1/12$. Suppose initially, no opportunity has a sign-up. In zip code 75039 (resp. 75201) consider an opportunity o_i (resp. o_j) with the maximal appeal of 10. Without a sign-up, opportunity o_i (resp. o_j) would be ranked on top. (Let us ignore the possibility of ties for the ease of argument.) When $c_1 = 1/12$, one sign-up reduces the score of opportunity o_i (resp. o_j) from 10 to approximately 9.5, according to SmartSort’s formula given by Equation (6). In zip code 75039, this would move opportunity o_i a few places down the display ranking, but it would likely remain on the first page. On the other hand, in zip code 75201, o_j would likely be moved off the front page; however due to the abundance of high appeal opportunities, it will be replaced by an opportunity with a similar appeal of approximately 9.9 (i.e., an opportunity within a mile of the volunteer’s search). Under the (conservative but reasonable) assumption that volunteers only pay attention to the first page of search results, we see that this level of penalization is small enough to keep the high appeal opportunity on the first page in situations where there is a scarcity of high appeal opportunities (i.e., zip code 75039) thus reducing the risk of efficiency loss. At the same time, it is sufficient to prioritize high appeal opportunities with no sign-ups. In Appendix B.2, we repeat the above thought experiment for a larger value of c_1 , i.e., $1/5$, and explain why such a larger penalty term could potentially harm efficiency.

Overall, although we were keen to select parameters that would noticeably increase equity, we were also wary of a significant decrease in efficiency that could deter VM from further experimentation. Based on stakeholder feedback from VM, we determined that $c_1 = 1/12$ was an appropriately cautious choice, as it was unlikely to harm efficiency but could still lead to equity gains. Another advantage of having $c_1 = 1/12$ is that its impact is easy to communicate: with $c_1 = 1/12$, the first sign-up would de-prioritize an opportunity as if it were 5 miles farther (i.e., it would have the same

¹⁶ Note that Dallas-Forth Worth was the site of our first experiment and the time period March 15th to May 23rd coincides with its corresponding pre-experiment period. See Section 4.1 for more details.

¹⁷ Note that the dashed lines show one standard deviation of this cutoff value and the red bars indicate the standard error of the height of each bar.

score as another opportunity with no sign-ups and 5 miles further from the volunteer). Finally, we explain our choice of the second parameter c_2 in the remark below.

Remark 1 (Setting Parameter c_2) *As mentioned at the beginning of Step 4, our main consideration in calibrating c_2 was to avoid excessively penalizing an opportunity with a surge of sign-ups. First, we remind that setting $c_2 > 1$ may result in negative scores; as such, we only consider $c_2 \in [0, 1]$. If we set c_2 to be $z \cdot c_1 = z/12$, then if an opportunity receives weakly larger than 12 sign-ups then its score will be simply 0. Prior to implementing SmartSort, we observe that a few opportunities would occasionally get a large number of weekly sign-ups (as evidenced by the histogram on the right panel of Figure 1). Although we expect that under SmartSort such events rarely occur, to convince the VM's engineering team, we decided to safeguard against such extreme penalization by setting z to be 10 (i.e., $c_2 = 10/12$).*

VM implemented SmartSort in a limited roll-out designed for experimentation. In the following section, we describe the experimental setting and study SmartSort's causal impact on various key outcomes related to sign-ups. We conclude this section by highlighting that despite the limitations accounted for in our implementation of SmartSort, our design is quite flexible and could potentially be improved if given access to more granular data. In the future, VM is interested in expanding its data infrastructure for this purpose.

4. Experimental Results

In collaboration with VM, we designed and implemented two field experiments to measure the impact of SmartSort on efficiency and equity. We describe our experimental design, outcome metrics, and identification strategy in Section 4.1. In 4.2.1, we provide some descriptive evidence for the impact of SmartSort which also serves to motivate our subsequent analysis. We describe our empirical approach and present our findings in Section 4.2.2. In the main text, we focus on only one of our experiments. Our approach and findings are nearly identical in the second experiment. As such, we defer all the details of the second experiment to Appendices D and E, and comment briefly on the findings.

4.1. Experimental Setting

Experimental setup. Our experiments ran at the “city”-level, and use the difference-in-differences methodology to compare the efficiency and equity outcomes (defined below) in a treatment geographical area that uses SmartSort, with those same outcomes in a control geographical area. As we explain in Section 4.1.1, we opted for a city-level design in order to minimize interference.

Our first experiment started on Tuesday May 24, 2022, and ran for 10 consecutive weeks. During this period, SmartSort was the default display ranking algorithm for volunteers in the entire geographical region of Dallas-Fort Worth (DFW). Formally, the treatment was applied as the default

display ranking algorithm to any volunteer if the latitude and longitude of their search location was in a treated zip code. In DFW, a zip code was treated if and only if its geographic center is within a 40-mile radius of either the Dallas or Fort Worth city centers (measured using Haversine distance). We designed our experiment to run for 10 weeks and we measure outcomes on a weekly basis.¹⁸ Participants on the VM platform were at no point given any indication of SmartSort's implementation status.

Outcome metrics. As discussed in earlier sections, we design SmartSort with an eye toward creating a more equitable sign up distribution without significantly harming the number of sign-ups. Thus in our experimentation, our objective is to understand the effect of SmartSort on these two dimensions: the number of sign-ups (efficiency) and the distribution of these sign-ups (equity). To define these metrics, we track sign-ups aggregated at a regional level. Specifically, our metrics for the treatment region will consider the sign-ups that *all* opportunities whose zip code was within a 40-mile radius of either the Dallas or Fort Worth city centers (measured using Haversine distance); see Section 4.1.1 for a more detailed discussion. Each metric will be tracked on a weekly basis, aligned with VM's desire for opportunities to have frequent access to volunteers on a weekly basis (see Section 3.2 for more details on the design of SmartSort). Formally, the metrics for efficiency and equity¹⁹ are defined as follows:

1. *Efficiency.* Defined as the total weekly sign-ups for opportunities located within the set of zip codes of the region of interest.
2. *Equity.* Defined as the weekly number of opportunities (located within the set of zip codes of the region of interest) with *at least* one sign-up.

Difference-in-differences methodology. To establish that any observed changes in the outcome metrics were indeed due to the change in the display ranking algorithm and not due to seasonal trends or other shocks, we use the difference-in-differences (DID) methodology. On a high level, DID methods identify the average causal effects of treatments by comparing the outcomes of (i) units that receive and do not receive a treatment (ii) both before and after the treatment is administered. Intuitively, any change in the average outcomes of the treated units that occurs between the pre-experiment and experiment periods reflects the net sum of the treatment effect and the pre-existing time trend affecting outcomes. By effectively subtracting the pre-existing trend as observed through the control group between the pre-experiment and experiment periods, the method identifies the treatment effect alone.

¹⁸ Through simulations and discussions with VM, we determined a ten-week time-frame struck the appropriate balance between statistical power and experiment duration.

¹⁹ To understand the distributional impact of SmartSort and carry out some of the robustness analyses, we introduce auxiliary outcome metrics throughout but omit them from this preview for the sake of brevity.

We take Houston (HOU), the second-largest VM geographic region in Texas as the comparison city (concretely, any zip code within a 40-mile radius of HOU city center with distance measured using Haversine distance). DFW and HOU are similar across a spectrum of observable characteristics (see Table 4 of Appendix C.1) and exhibit similar seasonal trends in volunteering, making HOU an apt comparison. Additionally, as both are cities within Texas, they would be subject to similar state-level shocks. We define control units identically, i.e., city-level weekly outcomes for the weeks defined above.

DID methods rely critically on the parallel trend assumption that the outcomes' underlying time trends are shared between the treated and control units (Angrist and Pischke 2013). By taking the HOU geographic region as the control, we assume that trends in sign-ups are primarily driven by state-wide seasonal variability and occur in parallel across the two cities. We provide evidence in support of this parallel trends assumption in Appendix C.2.

4.1.1. Discussion on Experimental Design. Our experiment was run at the city-level. As discussed by the recent literature in experimental design (Holtz et al. 2020, Rios et al. 2022) running the experiment at the volunteer level would lead to potentially biased estimates due to interference. That is, suppose that we use SmartSort on a randomly-selected treatment group of volunteers, and use the current algorithm for the other (control) volunteers. In that setting, sign-ups generated by the control users will affect the ranking of the opportunities displayed to the treated users. Therefore, we opted to design our experiment at a city-level, a design that was also adopted by other papers in matching markets (Holtz et al. 2020, Rios et al. 2022).

One may also wonder whether we could have treated a smaller area. Observe that the nature of our intervention necessitates treatment of a large contiguous area due to network effects and contamination at the boundaries. To see why, note that because the treatment was assigned on the volunteer level, opportunities outside the treated area may nevertheless be impacted by the treatment whenever volunteers search within the treated area but close to the boundary (of treated and untreated areas). Similarly, opportunities within the treated area may still be displayed according to CP whenever volunteers search just outside the treated area. To limit such effects we selected the treated region such that there is little activity on the boundary: Only 3.9% of sign-ups and 4.5% of opportunities receiving sign-ups within DFW exist within 10 miles of the boundary. Additionally, only 7% of the DFW population reside within 10 miles of the boundary based on data from the US Census Decennial Survey (2020). HOU is similar, with only 3.8% of sign-ups and 4.1% of opportunities receiving sign-ups within 10 miles of the boundary and only 7% of the HOU population residing within 10 miles of the boundary.

Finally, note that SmartSort was only used as the default search algorithm. That is, if users were to change the sorting method from the default “Near me and new” to a different option, then the

opportunities will be displayed using one of VM's original algorithms. However, we do not expect this to have a significant impact on the results, as > 95% of volunteers used the default sorting.

4.2. Results

In this section we present the results of our field experiment. In Section 4.2.1 we present preliminary descriptive evidence of the impact of SmartSort on the efficiency and equity metrics. We formalize these findings in Section 4.2.2 using a difference-in-differences approach. Finally, in Section 4.3, we discuss possible mechanisms for these findings.

Recall that SmartSort was launched in DFW on Tuesday May 24, 2022, which is the 21st Tuesday of the year. We evaluate outcomes for ten weeks (Tuesday to Monday) following the implementation date. These ten weeks are weeks 21 through 30 in 2022, and in the subsequent sections we refer to them as the *experiment* period. We compare this period with the 10 prior weeks of 2022 (weeks 11 through 20), which we refer to as the *pre-experiment* period.

The data we will be analyzing is based on sign-ups in both the pre-experiment and experiment periods for opportunities located within the treated area of DFW and the control area of HOU. To avoid overly influential outliers, we drop the top 1% of opportunities in each geographic region by sign-ups received. The resulting dataset, which we will use throughout this section, consists of sign-ups for 1,360 unique opportunities in DFW and 898 unique opportunities in HOU.

4.2.1. Descriptive Evidence of the Impact of SmartSort Data in Table 1 presents descriptive statistics of the outcome metrics for DFW and HOU regions in the pre-experiment and experiment periods. As seen in the first column, the average total weekly sign-ups (i.e., efficiency) increased dramatically in DFW during the experiment period (from 261.6 to 308.8). However, SmartSort was not expected to have a positive impact on efficiency, and the comparison to HOU clarifies that this increase in efficiency is not specific to DFW. Both geographic regions saw a nearly identical percentage change in total sign-ups (around 18%), presumably due to similar seasonal trends, e.g., volunteering activity typically increases during the summer.

Given the increase in the total number of sign-ups during the experiment period, it is not surprising that the number of opportunities with at least one sign-up (i.e., equity) also increased in both geographic regions, as shown in the second column of Table 1. However, in contrast to the similar increase in efficiency, there are clear differences when comparing the change in equity across the two geographic regions. During the experiment period in DFW, there was a sizable expansion in the weekly number of opportunities with at least one sign-up (+21.5%) while in HOU, the percentage increase in this metric was remarkably smaller (+12.3%).

This breakdown of the changes in outcome metrics across the two geographic regions serves as a rudimentary form of a difference-in-differences (DID) analysis that does not account for fixed

Table 1 Platform's city-wide weekly average counts, DFW vs HOU

City	Period	Efficiency		Equity	
		Mean	Std.	Mean	Std.
Treated (DFW)	Pre-experiment	261.6	67.9	182.7	33.9
	Experiment	308.8	65.5	221.9	31.5
Control (HOU)	Pre-experiment	151.2	25.9	118.3	17.5
	Experiment	179.9	21.3	132.8	16.5
Change in the treated city (DFW)		+18.0%		+21.5%	
Change in the control city (HOU)		+18.9%		+12.3%	
Net change in DFW over HOU		-0.9%		+9.2%	

effects but is transparently suggestive. We note that while DFW and HOU are the two largest VM geographic regions in Texas, they are nevertheless operating at different scales. The seasonal expansion is proportional to the VM market size, hence we focus on a comparison of the percentage change in outcome metrics as displayed in the last row of Table 1. Based on these numbers, the introduction of SmartSort is associated with a negligible relative impact on efficiency (-0.9%) but a notable relative increase in equity ($+9.2\%$), which are both in line with our expectations. Roughly speaking, these results correspond to the log-transformed outcome metrics that we consider in our subsequent DID analyses in Section 4.2.2.

Before moving on to a more rigorous analysis of the causal impact of SmartSort in DFW, we highlight that a similar rudimentary analysis on the impact of SmartSort in SCA (using California's San Francisco/San Jose region as an appropriate point of comparison) suggests an almost identical effect to what we observe in the last row of Table 1. Specifically, we identify net percent changes of -0.9% in efficiency and $+8.9\%$ in equity. Table 7 in Appendix D.1 reports the full set of statistics for SCA corresponding to the columns of Table 1.

4.2.2. Difference-in-differences specification Under the assumption that the parallel trends hold, a DID analysis will identify SmartSort's effect on DFW opportunities in the experiment period, relative to any seasonal trends observed as a baseline in HOU during the same time period. We use the following DID regression as our base specification to estimate the causal effect of introducing SmartSort on the metric Y_{wc} , where w represents the week and c represents the geographic region. This generic outcome variable stands in for the metrics we previously introduced: the weekly total number of sign-ups (i.e., efficiency) and the weekly number of opportunities with at least one sign-up (i.e., equity). We consider the log-transformed version of these metrics to account

for seasonal trends that are proportional to market size, as discussed in Section 4.2.1; for the log-transformed outcome, the resulting coefficient on the treatment (denoted β_{treat}) can be interpreted as a percentage change on the linear outcome via the reverse transformation ($e^{\beta_{\text{treat}}} - 1$)%.

$$\log(Y_{wc}) = \beta_0 + \beta_{\text{treat}} \cdot \mathbf{1}_{\{w \in [21, 30]\}} \cdot \mathbf{1}_{\{c = \text{DFW}\}} + \beta_c \cdot \mathbf{1}_{\{c = \text{DFW}\}} + \beta_{\text{post}} \cdot \mathbf{1}_{\{w \in [21, 30]\}} + \varepsilon_{wc}, \quad (8)$$

In order to generate more precise estimates, we also include a richer specification that controls for weekly variability by including week fixed effects, δ_w .

$$\log(Y_{wc}) = \beta_0 + \beta_{\text{treat}} \cdot \mathbf{1}_{\{w \in [21, 30]\}} \cdot \mathbf{1}_{\{c = \text{DFW}\}} + \beta_c \cdot \mathbf{1}_{\{c = \text{DFW}\}} + \delta_w + \varepsilon_{wc}, \quad (9)$$

In the remainder of this section, we detail the causal economic impacts of SmartSort's introduction in DFW as evaluated through these DID regression analyses.

Causal effect on efficiency. We present the results of our DID analyses for the metric of total sign-ups in the first two columns of Table 2, where the first column corresponds to our base specification in Equation 8 and the second column adds week fixed effects as covariates according to Equation 9. In both specifications, we find no evidence that SmartSort impacted the efficiency on VM. Our aggregate estimate of the weekly city-level treatment effect is shown in the first row of Column 1 of Table 2. The estimates can be interpreted as a small, statistically insignificant 0.6% decrease in efficiency resulting from the introduction of SmartSort. Before discussing the DID results on equity, we note that the treatment effects in SCA are very similar, with a statistically insignificant 1% decrease in efficiency (see Table 8 in Appendix D.1).

Causal effect on equity. We next consider our main equity metric, the weekly number of opportunities with at least one sign-up. In Columns 3 through 5 of Table 2, we present the results from three different model specifications. Our results support the conclusion that SmartSort's implementation in DFW caused an increase in equity of around 8.5%, in stark contrast to the minimal observed change in efficiency.

In Column 3 (corresponding to our base specification in Equation 8), the estimated treatment effect is not statistically significant due to large week-to-week variability in our outcome metric. In order to obtain more precise estimates, we consider a new specification that controls for the (log transformed) efficiency, which has a strong positive correlation with the (log transformed) equity but is not highly correlated with the binary variable indicating whether the treatment was applied to a given combination of week and city once we control for the other covariates. Details are provided in Appendix C.3.

We would expect this new specification to generate a more precise estimate for the increase in equity caused by SmartSort, and this estimate will be unbiased *assuming* that efficiency is

Table 2 Log Transformed City Level Analyses, DFW vs HOU

	Efficiency		Equity		
	(1)	(2)	(3)	(4)	(5)
Treated	-0.006 (0.123)	-0.006 (0.064)	0.082 (0.097)	0.086** (0.027)	0.085** (0.025)
Post	0.181* (0.087)		0.119 (0.068)	-0.018 (0.020)	
City	-0.533*** (0.087)	-0.533*** (0.045)	-0.430*** (0.068)	-0.028 (0.027)	-0.121* (0.052)
Log Sign-ups				0.755*** (0.036)	0.580*** (0.091)
Intercept	5.538*** (0.062)	5.575*** (0.075)	5.193*** (0.048)	1.013*** (0.201)	1.928** (0.510)
Week FE	No	Yes	No	No	Yes
Obs.	40	40	40	40	40
R ²	0.696	0.959	0.747	0.981	0.992
Adj. R ²	0.670	0.912	0.726	0.979	0.982
RSE	0.195	0.101	0.153	0.042	0.039
F Stat.	27.45***	20.23***	35.459***	455.423***	98.638***

Note:

*p<0.05; **p<0.01; ***p<0.001

unaffected by the treatment (Angrist and Pischke 2013); we further discuss this specification in Appendix C.3. The results reported in Column 4 match our expectations: the coefficient for this new covariate, denoted by “Log Sign-ups,” is large and significant. Further, the standard error on the coefficient for the treatment effect decreases sharply (from 0.097 in Column 3 to 0.027 in Column 4), while the point estimate barely changes (from 0.082 to 0.086). Additionally including week-level fixed effects (Column 5) produces similar results. Likewise, the treatment effects in SCA are very similar (reported in Appendix D.1); the corresponding treatment effect estimates for columns 3 and 4 of Table 8 show that the point estimate stays roughly the same (0.075 and 0.082), while the standard errors decrease dramatically (from 0.053 to 0.018). Including week fixed effects in column 5 leads to similar results.

4.3. Distributional Impacts of SmartSort

We now conduct a deeper investigation into the mechanism by which SmartSort achieves gains in equity (i.e., the weekly number of opportunities with at least one sign-up). Our analysis in Section 4.2.2 suggests that deployment of SmartSort did not meaningfully impact efficiency (i.e., total weekly sign-ups) in DFW. This implies that any increase in equity must be the result of a redistribution of sign-ups from opportunities that previously received multiple sign-ups in a week. To understand the impact of SmartSort on distributional outcomes beyond our primary outcome metric of equity (defined in Section 4.1), here we consider the following auxiliary outcome metrics:

1. The weekly number of opportunities (located within the set of zip codes of the region of interest) with *exactly* one sign-up.
2. The weekly number of opportunities (located within the set of zip codes of the region of interest) with *at least* two sign-ups.
3. The weekly number of opportunities (located within the set of zip codes of the region of interest) with *at least* three sign-ups.
4. The percentage of weekly sign-ups that are in *excess* of one sign-up. Said differently, the percentage of weekly sign-ups that go to opportunities receiving more than one sign-up. Hereafter we refer to this metric as *excess*.

Table 3 Platform's city-wide weekly distributional counts, DFW vs HOU;

		Opportunities receiving			% of Sign-ups in Excess of 1 sign-up		
		= 1 sign-up	≥ 2 sign-ups	≥ 3 sign-ups	Mean	Std.	
City	Period	Mean	Std.	Mean	Std.	Mean	Std.
Treated (DFW)	Pre-experiment	137.0	19.3	45.7	16.4	18.0	9.1
	Experiment	165.3	17.6	56.6	15.9	17.5	7.4
Control (HOU)	Pre-experiment	95.6	11.8	22.7	6.4	6.5	2.8
	Experiment	102.4	12.7	30.4	6.9	8.7	2.7
Change in the treated city (DFW)		+20.7%		+23.9%	-2.7%	-1.8%	
Change in the control city (HOU)		+7.1%		+33.9%	+33.8%	+4.8%	
Net change in DFW over HOU		+13.6%		-10.0%	-36.5%	-6.6%	

Following the same steps as described in Section 4.2.1, in Table 3, we provide descriptive evidence on the impact of SmartSort on these outcome measures. (The order of columns matches the order in which we define these metrics above.)

These aggregate-level analysis suggests that the equity gain is primarily driven by an increase in the number of opportunities that receive exactly one sign-up: During the experiment period, the weekly number of opportunities with exactly one sign-up increases in both DFW and HOU (resp. 20.7% and 7.1%), however the increase is substantially higher in DFW (a net change of 13.6%). On the other hand, the net change in the weekly number of opportunities receiving 2 or more (resp. 3) is negative -10% (resp. -36.5%). The greater effect on the weekly number of opportunities receiving 3 or more sign-ups is particularly interesting, as it suggests that SmartSort does not just redistribute sign-ups away from opportunities at the margin.

Both of these are aligned with our expectation: because VM is under-supplied, every week there are many opportunities with no sign-ups. Among opportunities with similar levels of appeal,

SmartSort will prioritize those with no sign-ups, thus increasing the number of opportunities getting a sign-up at the expense of those that would have gotten multiple sign-ups under CP.

Another way to quantify the re-distributive impact of SmartSort is the last outcome metric defined above: the excess. Reducing this metric implies that sign-ups are more uniformly distributed. (In the extreme case where no opportunity gets more than one sign-up, this metric is zero). In the last column of Table 3, we observe that during the experiment period, excess slightly decreased in DFW (-1.8%) whereas it increased in HOU (4.8%), resulting in a net change of -6.6% . In Appendix C.4, we provide a formal difference-in-differences analysis that supports our observations based on the above rudimentary analysis. Additional results and robustness checks are available in an electronic companion and omitted here for brevity.

5. Conclusion

In both the public and private sectors, two-sided platforms play an increasingly essential role in facilitating social, economic, and civic connections. While maximizing total number of connections (i.e., efficiency) remains a prominent goal for these platforms, catering to both sides of the platform gives rise to the importance of other objectives such as equitable distribution of these connections. One of the platform's key levers for achieving these goals is their display ranking algorithms (also referred to as search or recommendation algorithms). Our work illustrates how a platform such as VM can move toward increasing equity through re-designing its display ranking algorithm.

In collaboration with VM, we designed and implemented a new ranking algorithm, SmartSort, that aims to produce more equitable weekly sign-up (i.e., connection) distribution without harming efficiency. Our experimental results — based on field experiments in two geographical regions — provided convincing evidences on the positive impact of SmartSort to the extent that VM has recently deployed it nation-wide. Further experimentation to improve SmartSort design as well as understanding its impact on other outcomes such as volunteer experience are interesting directions to pursue once VM's data infrastructure is expanded to collect finer grained data.

Motivated by our collaboration with VM, we also developed a theoretical framework to design and evaluate robust (i.e., under the adversarial approach) online ranking algorithms for platforms with concerns about distributional outcomes. While VM's desire for equity partly stems from its egalitarian nature — to help all opportunities — e-commerce and social media platforms may have similar desire for reasons such as long-term engagement of suppliers and content creators (Chen et al. 2022). The equity notion we consider in this paper is aligned with VM's vision for equity but it also has the practical advantages of being easy to communicate and easy to measure. Defining other practically motivated notions of fairness, and designing online ranking algorithms for them is a fruitful direction for future research on the theoretical side.

References

- Olufunke Adebola, Priyank Arora, and Can Zhang. Farm equipment sharing in emerging economies. *Available at SSRN 4190725*, 2022.
- Narges Ahani, Tommy Andersson, Alessandro Martinello, Alexander Teytelboym, and Andrew C Trapp. Placement optimization in refugee resettlement. *Operations Research*, 69(5):1468–1486, 2021a.
- Narges Ahani, Paul Götz, Ariel D Procaccia, Alexander Teytelboym, and Andrew C Trapp. Dynamic placement in refugee resettlement. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 5–5, 2021b.
- Maxwell Allman, Itai Ashlagi, Irene Lo, Juliette Love, Katherine Mentzer, Lulabel Ruiz-Setz, and Henry O’Connell. Designing school choice for diversity in the san francisco unified school district. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 290–291, 2022.
- Mohammad Reza Aminian, Vahideh Manshadi, and Rad Niazadeh. Fair markovian search. *Available at SSRN 4347447*, 2023.
- Joshua D. Angrist and Joörn-Steffen Pischke. *Mostly harmless econometrics: An empiricists companion*. Cram101 Publishing, 2013.
- Jerry Anunrojwong, Krishnamurthy Iyer, and Vahideh Manshadi. Information design for congested social services: Optimal need-based persuasion. *Management Science*, 2022.
- Ali Aouad and Daniela Saban. Online assortment optimization for two-sided matching platforms. *Management Science*, 69(4):2069–2087, 2023.
- Makis Arsenis and Robert Kleinberg. Individual fairness in prophet inequalities. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 245–245, 2022.
- Susan Athey, Dean Karlan, Emil Palikot, and Yuan Yuan. Smiles in profiles: Improving fairness and efficiency using estimates of user preferences in online marketplaces. Technical report, National Bureau of Economic Research, 2022.
- Jackie Baek and Vivek Farias. Fair exploration via axiomatic bargaining. *Advances in Neural Information Processing Systems*, 34:22034–22045, 2021.
- Kirk Bansak and Elisabeth Paulson. Outcome-driven dynamic refugee assignment with allocation balancing. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 1182–1183, 2022.
- Surender Baswana, Partha Pratim Chakrabarti, Sharat Chandran, Yashodhan Kanoria, and Utkarsh Patange. Centralized admissions for engineering colleges in india. *INFORMS Journal on Applied Analytics*, 49(5):338–354, 2019.
- Asia J Biega, Krishna P Gummadi, and Gerhard Weikum. Equity of attention: Amortizing individual fairness in rankings. In *The 41st international acm sigir conference on research & development in information retrieval*, pages 405–414, 2018.

- Niv Buchbinder, Kamal Jain, and Joseph Seffi Naor. Online primal-dual algorithms for maximizing ad-auctions revenue. In *European Symposium on Algorithms*, pages 253–264. Springer, 2007.
- Qinyi Chen, Negin Golrezaei, Francisca Susan, and Edy Baskoro. Fair assortment planning. *arXiv preprint arXiv:2208.07341*, 2022.
- Maxime C Cohen, Sentao Miao, and Yining Wang. Dynamic pricing with fairness constraints. *Available at SSRN 3930622*, 2021.
- Maxime C Cohen, Adam N Elmachtoub, and Xiao Lei. Price discrimination with fairness constraints. *Management Science*, 68(12):8536–8552, 2022.
- Jose Correa, Andres Cristi, Paul Duetting, and Ashkan Norouzi-Fard. Fairness and bias in online selection. In *International Conference on Machine Learning*, pages 2112–2121. PMLR, 2021.
- José Correa, Natalie Epstein, Rafael Epstein, Juan Escobar, Ignacio Rios, Nicolás Aramayo, Bastián Bahamondes, Carlos Bonet, Martin Castillo, Andres Cristi, et al. School choice in chile. *Operations Research*, 70(2):1066–1087, 2022.
- Joann F de Zegher and Irene Lo. Crowdsourcing market information from competitors. *Available at SSRN 3537625*, 2020.
- Nikhil R Devanur and Kamal Jain. Online matching with concave returns. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 137–144, 2012.
- Yiding Feng, Rad Niazadeh, and Amin Saberi. Near-optimal bayesian online assortment of reusable resources. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 964–965, 2022.
- Daniel Freund, Thodoris Lykouris, Elisabeth Paulson, Bradley Sturt, and Wentao Weng. Group fairness in dynamic refugee assignment. *arXiv preprint arXiv:2301.10642*, 2023.
- Ragnar Frisch and Frederick V. Waugh. Partial time regressions as compared with individual trends. *Econometrica*, 1(4):387–401, 1933. ISSN 00129682, 14680262. URL <http://www.jstor.org/stable/1907330>.
- Negin Golrezaei, Hamid Nazerzadeh, and Paat Rusmevichientong. Real-time optimization of personalized assortments. *Management Science*, 60(6):1532–1551, 2014.
- Xiao-Yue Gong, Vineet Goyal, Garud N Iyengar, David Simchi-Levi, Rajan Udwani, and Shuangyu Wang. Online assortment optimization with reusable resources. *Management Science*, 2021.
- Vineet Goyal and Rajan Udwani. Online matching with stochastic rewards: Optimal competitive ratio via path based formulation. *arXiv preprint arXiv:1905.12778*, 2019.
- Vineet Goyal, Garud Iyengar, and Rajan Udwani. Asymptotically optimal competitive ratio for online allocation of reusable resources. *arXiv preprint arXiv:2002.02430*, 2020.
- Swati Gupta and Vijay Kamble. Individual fairness in hindsight. *The Journal of Machine Learning Research*, 22(1):6386–6420, 2021.

- David Holtz, Ruben Lobel, Inessa Liskovich, and Sinan Aral. Reducing interference bias in online marketplace pricing experiments. *arXiv preprint arXiv:2004.12489*, 2020.
- David Kempe and Mohammad Mahdian. A cascade model for externalities in sponsored search. *Internet and Network Economics*, pages 585–596, 2008.
- Varun Krishnan, Ramon Iglesias, Sebastien Martin, Su Wang, Varun Pattaahiraman, and Garrett Van Ryzin. Solving the ride-sharing productivity paradox: Priority dispatch and optimal priority sets. *INFORMS Journal on Applied Analytics*, 52(5):433–445, 2022.
- Max Kuhn and Kjell Johnson. *Applied Predictive Modeling*. Springer, 2013. ISBN 978-1-4614-6848-6.
- Retsef Levi, Manoj Rajan, Somya Singhvi, and Yanchong Zheng. The impact of unifying agricultural wholesale markets on prices and farmers' profitability. *Proceedings of the National Academy of Sciences*, 117(5):2366–2371, 2020.
- Irene Lo, Vahideh Manshadi, Scott Rodilitz, and Ali Shamel. Commitment on volunteer crowdsourcing platforms: Implications for growth and engagement. *Available at SSRN 3802628*, 2021.
- Will Ma and David Simchi-Levi. Algorithms for online matching, assortment, and pricing with tight weight-dependent competitive ratios. *Operations Research*, 68(6):1787–1803, 2020.
- Vahideh Manshadi and Scott Rodilitz. Online policies for efficient volunteer crowdsourcing. *Management Science*, 2022.
- Vahideh Manshadi, Rad Niazadeh, and Scott Rodilitz. Fair dynamic rationing. *Available at SSRN 3775895*, 2022a.
- Vahideh Manshadi, Scott Rodilitz, Daniela Saban, and Akshaya Suresh. Online algorithms for matching platforms with multi-channel traffic. *arXiv preprint arXiv:2203.15037*, 2022b.
- Duncan C McElfresh, Christian Kroer, Sergey Pupyrev, Eric Sodomka, Karthik Abinav Sankararaman, Zack Chauvin, Neil Dexter, and John P Dickerson. Matching algorithms for blood donation. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 463–464, 2020.
- Aranyak Mehta, Amin Saberi, Umesh Vazirani, and Vijay Vazirani. Adwords and generalized online matching. *J. ACM*, 54(5):22–es, oct 2007a. ISSN 0004-5411. doi: 10.1145/1284320.1284321. URL <https://doi.org/10.1145/1284320.1284321>.
- Aranyak Mehta, Amin Saberi, Umesh Vazirani, and Vijay Vazirani. Adwords and generalized online matching. *Journal of the ACM (JACM)*, 54(5):22–es, 2007b.
- Hao Yi Ong, Daniel Freund, and Davide Crapis. Driver positioning and incentive budgeting with an escrow mechanism for ride-sharing platforms. *INFORMS Journal on Applied Analytics*, 51(5):373–390, 2021.
- Jörn-Steffen Pischke. Empirical methods in applied economics: Lecture notes. *Downloaded July*, 24:2015, 2005.
- Canice Prendergast. The allocation of food to food banks. *Journal of Political Economy*, 130(8):000–000, 2022.

- Ignacio Rios, Daniela Saban, and Fanyin Zheng. Improving match rates in dating markets through assortment optimization. *Manufacturing & Service Operations Management*, 2022.
- Alvin E Roth, Tayfun Sönmez, and M Utku Ünver. A kidney exchange clearinghouse in new england. *American Economic Review*, 95(2):376–380, 2005.
- Paat Rusmevichientong, Mika Sumida, and Huseyin Topaloglu. Dynamic assortment optimization for reusable products with random usage durations. *Management Science*, 66(7):2820–2844, 2020.
- Ashudeep Singh and Thorsten Joachims. Fairness of exposure in rankings. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2219–2228, 2018.
- Mika Sumida, Guillermo Gallego, Paat Rusmevichientong, Huseyin Topaloglu, and James Davis. Revenue-utility tradeoff in assortment optimization under the multinomial logit model with totally unimodular constraints. *Management Science*, 67(5):2845–2869, 2021.
- Raluca M Ursu. The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science*, 37(4):530–552, 2018.
- Vijay V Vazirani. Online bipartite matching and adwords. In *47th International Symposium on Mathematical Foundations of Computer Science*, 2022.
- Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data, Second Edition*. MIT Press, Cambridge, 2010. ISBN 9780262296793.
- Andrew Chi-Chin Yao. Probabilistic computations: Toward a unified measure of complexity. In *18th Annual Symposium on Foundations of Computer Science (sfcs 1977)*, pages 222–227. IEEE Computer Society, 1977.

Appendix A: Omitted Details from Section 2.2

A.1. Proof of Proposition 1

This proof adapts the proof of Theorem 7.1 in Mehta et al. (2007b) to our setting, which involves presenting a ranking to arriving volunteers as well as a more general concave objective function. Capacities in their setting can be viewed of as an opportunity's equity threshold in our setting, and the only difference lies in the reward of γ for a sign-up (i.e., a match) beyond an opportunity's equity threshold (i.e., capacity). To show that no online algorithm obtains a competitive ratio greater than $\frac{e-1+\gamma}{e}$, it is sufficient to show that there exists a distribution over a class of instances for which no deterministic online algorithm obtains more than a $\frac{e-1+\gamma}{e}$ fraction of the optimal clairvoyant benchmark (Yao 1977).

For a fixed efficiency weight γ , consider an instance with n opportunities indexed $i \in \{1, \dots, n\}$ with homogeneous equity thresholds of m . In this instance there are exactly $n \cdot m$ arriving volunteers, each with the same choice model: an arriving volunteer only considers the top-ranked opportunity and will sign up if and only if it is *compatible* with that opportunity. The compatibility structure is defined as follows: the first m volunteers are compatible with all opportunities, the next m volunteers are only compatible with opportunities $i > 1$, and so on. In other words, volunteers $(j-1) \cdot m + 1$ through $j \cdot m$ are only compatible with opportunities $i \geq j$.

We now consider the class of instances that we can obtain by permuting the indices of the opportunities in the instance described above. Specifically, we apply a permutation ρ to the indices of opportunities, such that they are now indexed $\{\rho(1), \dots, \rho(n)\}$. Given this indexing, volunteers $(j-1) \cdot m + 1$ through $j \cdot m$ are only compatible with opportunities i where $\rho^{-1}(i) \geq j$.

In any such permuted instance, the optimal clairvoyant benchmark obtains a reward of $n \cdot m$. To see this, consider a clairvoyant solution that for each volunteer $t \in \{(j-1) \cdot m + 1, \dots, j \cdot m\}$, it presents a ranking with opportunity $\rho^{-1}(j)$ ranked first and the remaining opportunities ranked arbitrarily. These m volunteers are all compatible with the top-ranked opportunity, and thus they deterministically sign up. Each opportunity then gets exactly m sign-ups, leading to an objective value of $n \cdot m$.

We now bound the performance of any deterministic online algorithm on the uniform distribution over this class of instances.

Claim 1 Suppose that the permutation ρ is drawn uniformly at random from the set of all permutations of n opportunities. For any deterministic online algorithm π ,

$$\mathbb{E}[\pi] \leq \left(\frac{e-1+\gamma}{e}\right) n \cdot m, \quad (10)$$

where the expectation is taken over permutations ρ .

Proof of Claim 1: To aid in this proof, let us define $v_{i,j}$ as the number of volunteers from the set $\{(j-1) \cdot m + 1, \dots, j \cdot m\}$ that sign up for opportunity $\rho^{-1}(i)$. Based on the compatibility structure, no online algorithm can distinguish between opportunities that are still compatible with arriving volunteers. Therefore, we have:

$$E_\rho[v_{i,j}] \leq \begin{cases} \frac{m}{n-j+1}, & \text{if } i \geq j \\ 0 & \text{if } i < j \end{cases}$$

To see this, note that for each volunteer t between $t \in \{(j-1) \cdot m + 1, j \cdot m\}$, there are a total of $n - j + 1$ compatible opportunities, i.e., any opportunity i where $\rho^{-1}(i) \geq j$. Since the permutation is chosen uniformly at random, all remaining compatible opportunities appear identical to the online algorithm. This means that the number of volunteer sign-ups for each opportunity cannot exceed the average of $\frac{m}{n-j+1}$, when taking an expectation over permutations.

After all volunteers arrive, opportunity $\rho^{-1}(i)$ has received at most $\sum_{j=1}^i \frac{m}{n-j+1}$ sign-ups. As the objective function is concave in the number of sign-ups for each opportunity, the expected objective value for an online algorithm π is at most:

$$E_\rho[\pi] \leq \sum_{i=1}^n \min \left\{ m, \sum_{j=1}^i \frac{m}{n-j+1} \right\} + \gamma \max \left\{ 0, \sum_{j=1}^i \frac{m}{n-j+1} - m \right\} := \text{RHS} \quad (11)$$

Taking the limit as n and m approach infinity, we have

$$\lim_{n,m \rightarrow \infty} \frac{\text{RHS}}{n \cdot m} = \left(\frac{e-1+\gamma}{e} \right).$$

This completes the proof of Claim 1. \square

In combination with Yao's lemma, this proves that no online algorithm (deterministic or randomized) can obtain a competitive ratio greater than $\frac{e-1+\gamma}{e}$. To see that this holds for any finite minimum equity threshold \underline{m} , note that we can simply add an additional opportunity with an arbitrary equity threshold and no compatible volunteers to the set of instances described. \square

A.2. Proof of Theorem 1

We will prove this result in three steps. (i) First, we will define pseudo-rewards that depend on the particular choice of penalty function used by a balancing algorithm π^{BAL} (see Definition 3). (ii) Then, in Lemma 1 we will show that the expected value of the algorithm is at least a constant fraction of the expected sum of the pseudo-rewards. (iii) We subsequently show in Lemma 2 that the expected sum of the pseudo-rewards is at least a constant fraction of the expected value of the clairvoyant benchmark (see Definition 1). Combining the three steps establishes a lower bound on the competitive ratio of π^{BAL} (see Definition 2). The bound on the competitive ratio is equal to the product of the constant-factors from Steps (ii) and (iii), which are both parameterized by γ and \underline{m} .

Step (i): Defining Pseudo-Rewards. Consider a balancing algorithm π^{BAL} with penalty function ψ , which is a non-increasing function of the “fill rate” of an opportunity, i.e., the ratio between its number of sign-ups and its equity threshold. For such an algorithm, we will define pseudo-rewards which we emphasize are purely for accounting purposes. In the subsequent steps of the proof, we will use the pseudo-rewards as an intermediary to compare the performance of π^{BAL} with that of the clairvoyant benchmark.

Recall that $X_{i,t}(\vec{S})$ is an indicator random variable that equals 1 if and only if volunteer t signs up for opportunity i when presented with ranking \vec{S} . For each arriving volunteer $t \in \{1, \dots, T\}$, let \vec{S}_t^{BAL} denote the ranking provided by π^{BAL} , and let \vec{S}_t^{OPT} denote the ranking provided by our clairvoyant benchmark. Furthermore, recall that $Z_{i,t}$ denotes the number of sign-ups for opportunity i after the arrival of volunteer

t , and let $f_{i,t} = Z_{i,t}/m_i$ denote the fill rate of the opportunity after the arrival of volunteer t . With this notation in mind, we define

$$\lambda_t = \sum_{i=1}^n \psi(f_{i,t-1}) X_{i,t}(\vec{S}_t^{\text{OPT}}) \quad (12)$$

$$\theta_i = \sum_{t=1}^T \left(1 - \psi(f_{i,t-1})\right) \mathbb{1}\{Z_{i,t} \leq m_i\} X_{i,t}(\vec{S}_t^{\text{BAL}}) \quad (13)$$

We note that these pseudo-rewards are random and depend on the volunteers' choices.

Step (ii): Lower Bounding $\mathbb{E}[\pi^{\text{BAL}}(\mathcal{I})]$ with Pseudo-Rewards. We use the pseudo-rewards defined in Step (i) to establish a lower bound on the expected value of π^{BAL} , as formalized in the following lemma.

Lemma 1 *Consider a balancing algorithm π^{BAL} with penalty function ψ . For any instance \mathcal{I} with an efficiency weight γ , we have*

$$\mathbb{E}[\pi^{\text{BAL}}(\mathcal{I})] \geq \alpha \cdot \mathbb{E} \left[\sum_{i=1}^n \theta_i + \sum_{t=1}^T \lambda_t \right],$$

where $\alpha = \min\{1, \gamma/\psi(1)\}$ and where the expectation is taken over the volunteers' choices.²⁰

Proof of Lemma 1 Recall that (OBJ) can be re-written as the objective function in Equation (1). Taking the expectation over the volunteers' choices, we have the following series of inequalities:

$$\mathbb{E}[\pi^{\text{BAL}}(\mathcal{I})] = \mathbb{E} \left[\sum_{i=1}^n \sum_{t=1}^T \mathbb{1}\{Z_{i,t} \leq m_i\} X_{i,t}(\vec{S}_t^{\text{BAL}}) + \gamma \mathbb{1}\{Z_{i,t} > m_i\} X_{i,t}(\vec{S}_t^{\text{BAL}}) \right] \quad (14)$$

$$\begin{aligned} &= \mathbb{E} \left[\sum_{i=1}^n \sum_{t=1}^T \left(1 - \psi(f_{i,t-1})\right) \mathbb{1}\{Z_{i,t} \leq m_i\} X_{i,t}(\vec{S}_t^{\text{BAL}}) + \psi(f_{i,t-1}) \mathbb{1}\{Z_{i,t} \leq m_i\} X_{i,t}(\vec{S}_t^{\text{BAL}}) \right. \\ &\quad \left. + \gamma \mathbb{1}\{Z_{i,t} > m_i\} X_{i,t}(\vec{S}_t^{\text{BAL}}) \right] \end{aligned} \quad (15)$$

$$= \mathbb{E} \left[\sum_{i=1}^n \theta_i + \sum_{i=1}^n \sum_{t=1}^T \psi(f_{i,t-1}) \mathbb{1}\{Z_{i,t} \leq m_i\} X_{i,t}(\vec{S}_t^{\text{BAL}}) + \gamma \mathbb{1}\{Z_{i,t} > m_i\} X_{i,t}(\vec{S}_t^{\text{BAL}}) \right] \quad (16)$$

$$\begin{aligned} &\geq \mathbb{E} \left[\sum_{i=1}^n \theta_i + \sum_{i=1}^n \sum_{t=1}^T \psi(f_{i,t-1}) \mathbb{1}\{Z_{i,t} \leq m_i\} X_{i,t}(\vec{S}_t^{\text{BAL}}) + \alpha \cdot \psi(f_{i,t-1}) \mathbb{1}\{Z_{i,t} > m_i\} X_{i,t}(\vec{S}_t^{\text{BAL}}) \right] \\ &\quad (17) \end{aligned}$$

$$\begin{aligned} &\geq \alpha \cdot \mathbb{E} \left[\sum_{i=1}^n \theta_i + \sum_{i=1}^n \sum_{t=1}^T \psi(f_{i,t-1}) \mathbb{1}\{Z_{i,t} \leq m_i\} X_{i,t}(\vec{S}_t^{\text{BAL}}) + \psi(f_{i,t-1}) \mathbb{1}\{Z_{i,t} > m_i\} X_{i,t}(\vec{S}_t^{\text{BAL}}) \right] \\ &\quad (18) \end{aligned}$$

$$= \alpha \cdot \mathbb{E} \left[\sum_{i=1}^n \theta_i + \sum_{i=1}^n \sum_{t=1}^T \psi(f_{i,t-1}) X_{i,t}(\vec{S}_t^{\text{BAL}}) \right] \quad (19)$$

$$\geq \alpha \cdot \mathbb{E} \left[\sum_{i=1}^n \theta_i + \sum_{i=1}^n \sum_{t=1}^T \psi(f_{i,t-1}) X_{i,t}(\vec{S}_t^{\text{OPT}}) \right] \quad (20)$$

$$= \alpha \cdot \mathbb{E} \left[\sum_{i=1}^n \theta_i + \sum_{t=1}^T \lambda_t \right] \quad (21)$$

²⁰If the denominator $\psi(1) = 0$, we set $\alpha = 1$.

There are two key steps in this series of inequalities. First, in Line (18), we rely on the fact that whenever the indicator $\mathbb{1}\{Z_{i,t} > m_i\}$ is 1, we must have $f_{i,t-1} \geq 1$. Furthermore,

$$f_{i,t-1} \geq 1 \implies \alpha \cdot \psi(f_{i,t-1}) \leq \alpha \cdot \psi(1) = \min\{\psi(1), \gamma\} \leq \gamma,$$

where the first inequality holds because ψ is weakly decreasing.

Then, in Line (20), we use the optimality criteria of the exponential balancing function π^{BAL} , namely, that π^{BAL} presents a ranking that maximizes $\sum_{i \in \mathcal{S}} \psi(Z_{i,t}) \cdot \phi_{i,t}(\vec{S})$. This implies that

$$\mathbb{E}\left[\sum_{i=1}^n \sum_{t=1}^T \psi(f_{i,t-1}) X_{i,t}(\vec{S}_t^{\text{BAL}})\right] \geq \mathbb{E}\left[\sum_{i=1}^n \sum_{t=1}^T \psi(f_{i,t-1}) X_{i,t}(\vec{S}_t^{\text{OPT}})\right],$$

due to the fact that $\mathbb{E}[X_{i,t}(\vec{S})] = \phi_{i,t}(\vec{S})$ and the tower property of expectations. (These steps are identical to the rigorous application of the tower property of expectations in a similar proof in [Manshadi et al. \(2022b\)](#).)

□

Upper Bounding $\mathbb{E}[\text{OPT}(\mathcal{I})]$ with Pseudo-Rewards. We now establish an upper bound on the clairvoyant solution for any instance using the expected sum of the pseudo-rewards under π^{BAL} .

Lemma 2 *Consider a balancing algorithm π^{BAL} with penalty function ψ . For any instance \mathcal{I} with an efficiency weight γ and a minimum equity threshold \underline{m} , we have*

$$\mathbb{E}\left[\sum_{i=1}^n \theta_i + \sum_{t=1}^T \lambda_t\right] \geq \beta \cdot \kappa \cdot \mathbb{E}[\text{OPT}(\mathcal{I})], \quad (22)$$

where the expectation is taken over the volunteers' choices, $\beta = \min\left\{1, \lim_{x \rightarrow \infty} \frac{\psi(x)}{\gamma \cdot \psi(0)}\right\}$, ²¹ and where

$$\kappa = \min_{m, Z \in \mathbb{Z}_+, m \geq \underline{m}} \psi(Z/m) + (1/m) \sum_{k=1}^{\min\{Z, m\}} \left(1 - \psi((k-1)/m)\right). \quad (23)$$

Proof of Lemma 2 We will show that a tighter version of Equation (22) holds for each opportunity and for any realization of volunteer choices (similar to the approach taken in, e.g., [Manshadi et al. 2022b](#)). Specifically, letting \mathcal{O}_i denote the set of volunteers who sign up for opportunity i under the optimal clairvoyant policy, we will show

$$\theta_i + \sum_{t \in \mathcal{O}_i} \lambda_t \geq \beta \cdot \kappa \cdot (\gamma |\mathcal{O}_i| + (1 - \gamma) \min\{m_i, |\mathcal{O}_i|\}) \quad (24)$$

Summing over all opportunities and taking expectation over the volunteers' choices completes the proof.

To establish Equation (24), we will separately bound the value of θ_i and the value of the λ_t pseudo-rewards, for all $t \in \mathcal{O}_i$ (i.e., for all volunteers that sign up for opportunity i under the clairvoyant solution). Starting with the definition of λ_t in Equation (12),

$$\sum_{t \in \mathcal{O}_i} \lambda_t = \sum_{t \in \mathcal{O}_i} \psi(f_{i,t-1}) X_{i,t}(\vec{S}_t^{\text{OPT}}) \quad (25)$$

$$= \sum_{t \in \mathcal{O}_i} \psi(f_{i,t-1}) \quad (26)$$

$$\geq \sum_{t \in \mathcal{O}_i} \psi(f_{i,T}) \quad (27)$$

$$= |\mathcal{O}_i| \psi(f_{i,T}) \quad (28)$$

²¹ If the denominator $\gamma \cdot \psi(0) = 0$, we set $\beta = 1$.

Equality in Line (26) follows from the definition that $t \in \mathcal{O}_i$ only if the volunteer signs up for opportunity i when presented with the optimal clairvoyant benchmark's ranking, i.e., $X_{i,t}(\vec{S}_t^{\text{OPT}}) = 1$. The following inequality in Line (27) relies on the fact that ψ is a weakly decreasing function and $f_{i,T} \geq f_{i,t-1}$ for all $t \in \{1, \dots, T\}$.

Next, based on the definition of θ_i in Equation (13),

$$\theta_i = \sum_{t=1}^T \left(1 - \psi(f_{i,t-1})\right) \mathbb{1}\{Z_{i,t} \leq m_i\} X_{i,t}(\vec{S}_t^{\text{BAL}}) \quad (29)$$

$$= \sum_{k=1}^{Z_{i,T}} \left(1 - \psi((k-1)/m_i)\right) \mathbb{1}\{k \leq m_i\} \quad (30)$$

$$= \sum_{k=1}^{\min\{Z_{i,T}, m_i\}} \left(1 - \psi((k-1)/m_i)\right) \quad (31)$$

$$= \sum_{k=1}^{\min\{f_{i,T}, 1\}} \left(1 - \psi((k-1)/m_i)\right) \quad (32)$$

Based on our separate bounds for the pseudo-rewards, to complete the proof of Lemma 2, it is sufficient to establish that for all $|\mathcal{O}_i| \in \mathbb{Z}_+$ and all $f_{i,T} \in \mathbb{R}_+$, we have

$$|\mathcal{O}_i| \psi(f_{i,T}) + \sum_{k=1}^{\min\{f_{i,T}, 1\}} \left(1 - \psi((k-1)/m_i)\right) \geq \beta \cdot \kappa \cdot \left(\gamma |\mathcal{O}_i| + (1-\gamma) \min\{m_i, |\mathcal{O}_i|\}\right) \quad (33)$$

To prove Inequality (33), we define two functions, $L_i(y)$ and $R_i(y)$ for $y \in \mathbb{R}_+$, which represent the left and right hand sides of the inequality as a continuous function of $|\mathcal{O}_i|$. Specifically,

$$L_i(y) = y \cdot \psi(f_{i,T}) + \sum_{k=1}^{\min\{f_{i,T}, 1\}} \left(1 - \psi((k-1)/m_i)\right) \quad (34)$$

$$R_i(y) = \beta \cdot \kappa \cdot \left(\gamma \cdot y + (1-\gamma) \min\{m_i, y\}\right) \quad (35)$$

We make the following structural observations about the two functions: $L_i(y)$ is a linear function with $L_i(0) \geq 0$ and slope $\psi(f_{i,T})$. Furthermore, $R_i(y)$ is a piecewise-linear function with $R_i(0) = 0$, with slope $\beta \cdot \kappa$ for $y \in [0, m_i]$ and slope $\beta \cdot \kappa \cdot \gamma$ for $y > m_i$. Given these observations, to prove Inequality (33), it is sufficient to show that (i) $L_i(m_i) \geq R_i(m_i)$, and (ii) for $y > m_i$, the slope of $R_i(y)$ does not exceed that of $L_i(y)$.

To establish (i) $L_i(m_i) \geq R_i(m_i)$, we have

$$\begin{aligned} m_i \psi(f_{i,T}) + \sum_{k=1}^{\min\{f_{i,T}, 1\}} \left(1 - \psi((k-1)/m_i)\right) \\ = m_i \left(\psi(f_{i,T}) + (1/m_i) \sum_{k=1}^{\min\{f_{i,T}, 1\}} \left(1 - \psi((k-1)/m_i)\right) \right) \end{aligned} \quad (36)$$

$$\geq m_i \left(\min_{m, z \in \mathbb{Z}_+, m \geq m_i} \psi(z/m) + (1/m) \sum_{k=1}^{\min\{z, m\}} \left(1 - \psi((k-1)/m)\right) \right) \quad (37)$$

$$= m_i \cdot \kappa \quad (38)$$

$$\geq m_i \cdot \beta \cdot \kappa \quad (39)$$

To establish (ii) for $y > m_i$, the slope of $R_i(y)$ does not exceed that of $L_i(y)$, we have

$$\beta \cdot \kappa \cdot \gamma \leq \beta \cdot \psi(0) \cdot \gamma = \min\{\psi(0) \cdot \gamma, \lim_{x \rightarrow \infty} \psi(x)\} \leq \psi(f_{i,T}),$$

where the first inequality holds by plugging in $y = 0$ to the definition of κ in Equation 23 and the last inequality holds because ψ is weakly decreasing.

Therefore, we have shown that Inequality (33) holds for all $|\mathcal{O}_i| \in \mathbb{Z}_+$ and all $f_{i,T} \in \mathbb{R}_+$, which is sufficient to complete the proof of Lemma 2. \square

Based on the pseudo-rewards defined in Step (i) and using Lemmas 1 and 2, we can establish the following series of inequalities to bound the competitive ratio of a balancing algorithm π^{BAL} with penalty function ψ :

$$\mathbb{E}[\pi^{\text{BAL}}(\mathcal{I})] \geq \alpha \cdot \mathbb{E}\left[\sum_{i=1}^n \theta_i + \sum_{t=1}^T \lambda_t\right] \geq \alpha \cdot \beta \cdot \kappa \cdot \mathbb{E}[\text{OPT}(\mathcal{I})], \quad (40)$$

where $\alpha = \min\left\{1, \frac{\gamma}{\psi(1)}\right\}$, $\beta = \min\left\{1, \lim_{x \rightarrow \infty} \frac{\psi(x)}{\gamma \cdot \psi(0)}\right\}$, and κ is defined in Equation (23). The first inequality comes from Lemma 1, and the second comes from Lemma 2. This completes the proof of Theorem 1. \square

A.3. Proof of Corollary 1

By Theorem 1, the competitive ratio of a balancing algorithm consists of three terms. For the specific penalty function $\psi(x) = 1 - (1 - \gamma) \exp(-1 + \min\{x, 1\})$, we have $\psi(0) = 1 - (1 - \gamma)e^{-1}$ and $\psi(1) = \lim_{x \rightarrow \infty} \psi(x) = \gamma$. Therefore $\alpha = \min\left\{1, \frac{\gamma}{\psi(1)}\right\} = 1$ and $\beta = \min\left\{1, \lim_{x \rightarrow \infty} \frac{\psi(x)}{\gamma \cdot \psi(0)}\right\} = 1$.

For the final term,

$$\kappa = \min_{m,z \in \mathbb{Z}_+, m \geq \underline{m}} \psi(z/m) + (1/m) \sum_{k=1}^{\min\{z,m\}} (1 - \psi((k-1)/m)) \quad (41)$$

$$\geq \min_{m,z \in \mathbb{Z}_+, m \geq \underline{m}} \psi(z/m) + (1/m)e^{-1/m} \sum_{k=1}^{\min\{z,m\}} (1 - \psi(k/m)) \quad (42)$$

$$\geq e^{-1/\underline{m}} \min_{m,z \in \mathbb{Z}_+, m \geq \underline{m}} \psi(z/m) + (1/m) \sum_{k=1}^{\min\{z,m\}} (1 - \psi(k/m)) \quad (43)$$

$$\geq e^{-1/\underline{m}} \min_{m,z \in \mathbb{Z}_+, m \geq \underline{m}} \psi(z/m) + \int_0^{\min\{z/m, 1\}} 1 - \psi(x) \, dx \quad (44)$$

$$= e^{-1/\underline{m}} \min_{m,z \in \mathbb{Z}_+, m \geq \underline{m}} 1 - (1 - \gamma) e^{-1+\min\{z/m, 1\}} + (1 - \gamma) e^{-1+\min\{z/m, 1\}} - (1 - \gamma) e^{-1} \quad (45)$$

$$= e^{-1/\underline{m}} \min_{m,z \in \mathbb{Z}_+, m \geq \underline{m}} 1 - (1 - \gamma) e^{-1} \quad (46)$$

$$= e^{-1/\underline{m}} \left(\frac{e - 1 + \gamma}{e} \right) \quad (47)$$

The inequality in Line (42) relies on the algebraic fact that for this penalty function, $1 - \psi((k-1)/m_i) \geq e^{-1/\underline{m}}(1 - \psi(k/m_i))$ for all $k \in \mathbb{N}$. The inequality in Line (44) comes from the fact that $\psi(x)$ is weakly decreasing in x , and hence the prior summation is equivalent to a right Riemann sum of a weakly increasing function, which is then lower-bounded by the integral.

Due to the fact that $\lim_{m \rightarrow \infty} e^{-1/m} \left(\frac{e - 1 + \gamma}{e} \right) = \frac{e - 1 + \gamma}{e}$, this algorithm asymptotically obtains the best-possible competitive ratio, as it matches the upper bound for any online algorithm that was established in Proposition 1. \square

Appendix B: Missing Details for Section 3

B.1. Additional Details about Simulations in Figure 3

Market Size: All three instances consist of 200 opportunities and 100 homogeneous volunteers. Similar insights apply when varying this supply-to-demand ratio, for example in a setting with 50 opportunities.

Opportunity Appeals: Since volunteers are homogeneous, each opportunity i has a universal appeal of a_i . We assume that a_i is drawn independently from a beta distribution which takes two parameters α and β . We vary these parameters across the three instances:

- Instance 1: $\alpha = 0.25$, and $\beta = 1$
- Instance 2: $\alpha = 0.75$, and $\beta = 1$
- Instance 3: $\alpha = 1$, and $\beta = 0.75$

In the right three panels of Figure 3, we present the distribution of opportunity appeals corresponding to each instance. We note that in the first example (shown in red), there are very few high-appeal opportunities; in the second example (shown in purple), there are still relatively few high-appeal opportunities while in the third example (shown in green), there are many high-appeal opportunities.

Volunteer Choice: The probability that a volunteer signs-up for an opportunity depends on its position in the display ranking as well as its appeal. Specifically, suppose $\sigma_{i,k}(\vec{S})$ is an indicator which equals 1 if and only if opportunity i is displayed in position k under ranking \vec{S} .

We assume that the probability that volunteer t selects opportunity i when shown display ranking \vec{S} is given by

$$\phi_{i,t}(\vec{S}) = \frac{1}{20} \sum_{k=1}^{10} \sigma_{i,k}(\vec{S}) \left(2 - \frac{2k}{11}\right)^+ \cdot a_i. \quad (48)$$

For this volunteer choice model, only opportunities ranked in the top ten positions have positive probability of a sign-up (capturing, e.g., the impact of being on the first page of search results), and the total probability of receiving any sign-up is at most $1/2$ (which could be achieved if all opportunities had the maximum appeal of 1). This simple choice model has the following desirable property: under any balancing function, the optimal ranking (as specified in Definition 3, in particular Equation 2) coincides with the score-based ranking, i.e., ranking opportunities in descending order of $\psi(Z_{i,t}) \cdot a_i$. As discussed in Step 1, due to various practical considerations, we settled for a score-based ranking algorithm. Under this simple choice model, such a score-based ranking is equivalent to implementing the corresponding balancing algorithm. The optimality of this simple score-based ranking is our main motivation for focusing on this choice model in these illustrative examples. Also note this choice model can be viewed as a simplified version of a multinomial logit (MNL) model with position bias but without substitution effects.

Finally, we show the optimality of this score-based ranking by a contradiction argument: suppose opportunity i is ranked in position k while opportunity j is ranked in position $k+1$; however, opportunity i has the worse score, i.e., $\psi(Z_{i,t}) \cdot a_i \leq \psi(Z_{j,t}) \cdot a_j$. If we swap the ranks of opportunities i and j , the objective in Equation 2 will change by $(\psi(Z_{j,t}) \cdot a_j - \psi(Z_{i,t}) \cdot a_i)/110 \geq 0$, if $k \leq 10$; otherwise it will remain unchanged.

Number of Simulation Iterations: For each instance we run 2,000 iterations.

B.2. Additional Thought Experiment for Step 4

Let us repeat the thought experiment from Step 4 for a more aggressive penalty parameter, e.g., $c_1 = 1/5$.

In Zip code 75039 (resp. 75201) consider again an opportunity o_i (resp. o_j) with the maximal appeal of 10. Without a sign-up, opportunity o_i (resp. o_j) would be ranked on top. When $c_1 = 1/5$, one sign-up reduces the score of opportunity o_i (resp. o_j) to approximately 8.7, according to SmartSort's formula given by Equation 6. In Zip code 75039, opportunity o_i would now likely move off the front page and would be replaced by an opportunity with a score of approximately 9.4, which could be an opportunity without a sign-up but 6 miles away (thus potentially much less appealing). This could have a substantial negative impact on efficiency, especially if volunteers preferences for close opportunities are non-linear. In contrast, this change would not have an impact in Zip code 75201, relative to the previous discussion: opportunity o_j would still move off the front page of search results and would still be replaced by an opportunity with a similar score of approximately 9.9 (i.e., an opportunity without a sign-up within a mile of the volunteer's search).

Appendix C: Supplemental Material for the Dallas Experiment

C.1. Balance Tables

As mentioned in Section 4.1, DFW and HOU are the first and second largest regions in Texas. This makes HOU a natural comparison city. Drawing on zip-level data from the US Census American Community Survey 2020²², in Table 4 we show that in terms of population, demographic composition, and income level the regions included in our analysis are fairly comparable. Additionally, using aggregate city-level data on volunteer activity on VM provided by Google Analytics we also confirm that in our pre-experiment period (defined in Section 4.1) the number of visitors and sessions (i.e., the periods of time during which visitors interact with the VM site) are on the same order. However, we note that this data is not available on a more granular level (i.e., on a zip code level). Further it only comprises a subset of the full activity in the DFW and HOU regions. However, this is the only data available on users and sessions.

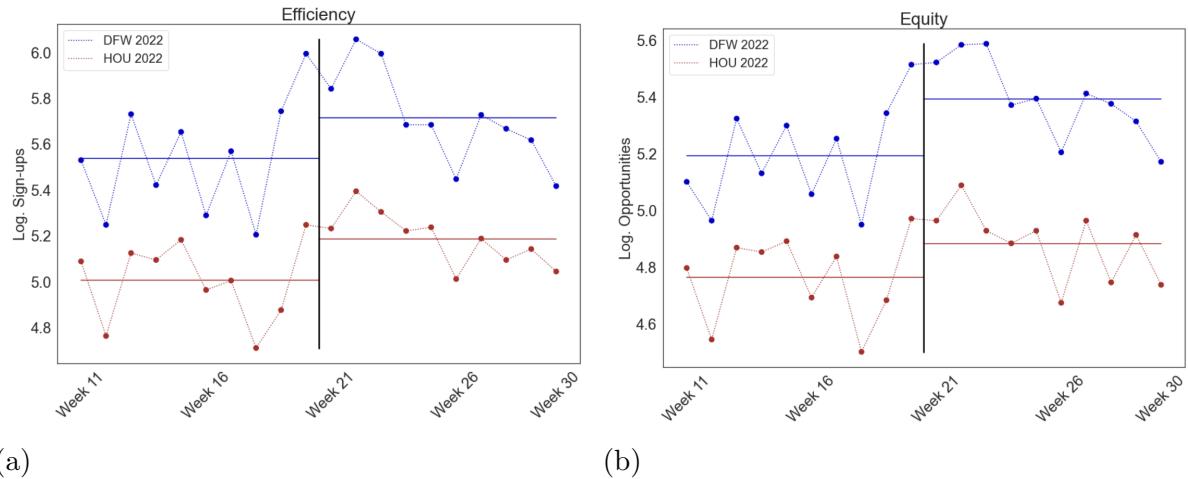
C.2. Parallel Trends Assumption

The key identifying assumption for the DID analyses is that of parallel trends between the treatment and control groups. In this appendix we provide evidence in support of the parallel trend assumption between DFW and HOU. First, in Figure 5 we simply plot the trends for our two outcome metrics across the two regions. Recall that the pre-experiment ran for 10 consecutive weeks from Tuesday March 15, 2022 (Week 11) - May 23, 2022 (Week 20) and the experiment period ran for 10 consecutive weeks starting from May 24, 2022 (Week 21) - August 1, 2022 (Week 30). For both metrics, we observe strong and reasonably consistent seasonal trends in the pre-experiment period, with a noticeable summertime boost in volunteer activity in both Texas cities.

²² <https://www.census.gov/newsroom/press-kits/2021/acs-5-year.html>

Table 4 Balance Table - DFW and HOU

	DFW	HOU
Total Population	6,282,445	5,538,4552
Pct Population Female	50.41%	50.20%
Pct Population White	57.75%	51.54%
Median Age (years)	32.9	33.3
Median Income	\$54,747	\$53,600
Total Users (pre-experiment)	22,531	16,608
Total Sessions (pre-experiment)	28,421	21,814

**Figure 5** The weekly seasonal trends of efficiency and equity in DFW and HOU.

To provide quantitative evidence for the parallel trend assumption, first we follow the approach of comparing the slope of time trends in the pre-experiment period across two regions (see Rios et al. (2022) for an example of this method). To that end, we consider the following specification formula.

$$\log(Y_{wc}) = \beta_0 + \beta_c \cdot \mathbf{1}_{\{c=\text{DFW}\}} + \delta_w + \rho_c w + \varepsilon_{wc}$$

In the above, $w \in [11, 20]$ is the week, $c \in \{\text{DFW}, \text{HOU}\}$ is the city, β_c is the city fixed effect, δ_w are week fixed effects, and ρ_c are the time slope in each of the control and treated regions.

We estimate the ρ_c for both DFW and HOU and use a t-test to determine if they are statistically significantly different from each other. If they are not, we can conclude that parallel trends hold. The first two rows of Table 5 shows the estimated slopes for each of the outcome metrics in DFW and HOU. We see that there are significant seasonal effects pre-experiment, as both ρ_{DFW} and ρ_{HOU} are positive and statistically significant for both efficiency and equity. The final row shows the result of a t-test that considers the null hypothesis that $\rho_{\text{DFW}} = \rho_{\text{HOU}}$. When the t-test test statistic is small and statistically insignificant, we cannot reject the null hypothesis. The final row of Table 5 indicates that the difference between ρ_{DFW} and ρ_{HOU} is

small for both efficiency and equity (0.024 and 0.022 respectively) and that both differences are statistically insignificant. Thus, we cannot reject the null hypothesis that the slopes ρ_{DFW} and ρ_{HOU} are equal to each other for both efficiency and equity, suggesting that parallel trends hold.

Table 5 Comparison of Pre-trend Slopes in Log Outcomes in DFW vs HOU

	Efficiency	Equity
ρ_{DFW}	0.203*** (0.016)	0.183*** (0.008)
ρ_{HOU}	0.178*** (0.010)	0.161*** (0.005)
T-Test	0.024 (0.018)	0.022 (0.010)

Finally, to provide further evidence for parallel trends, we conduct an event study (Pischke 2005). The purpose of this analysis is to consider whether there is a significant difference in trend between the treated and the control group at any time pre-experiment. Concretely we consider the following specification:

$$\log(Y_{wc}) = \beta_c \cdot \mathbf{1}_{\{c=DFW\}} + \delta_w + \sum_{o=-9, o \neq 0}^{10} \beta_{w,c} \cdot \mathbf{1}_{\{w=20+o\}} \cdot \mathbf{1}_{\{c=DFW\}} + \varepsilon_{wc}, \quad (49)$$

In the above, $w \in [11, 30]$ is the week, $c \in \{DFW, HOU\}$ is the city, β_c is the city fixed effect, δ_w are week fixed effects, and $\beta_{w,c}$ are the leads and lags of the treatment. $\beta_{w,c}$ can be interpreted as the treatment effect if the treatment had occurred in that week, using the final pre-experiment week as a baseline (Pischke 2005).

We plot the coefficients $\beta_{w,c}$ in Figure 6, where the bars represent a 95% confidence interval with bootstrapped standard errors calculated through resampling. We use the standard bootstrap method of sampling with replacements Kuhn and Johnson (2013), which works in the following way to generate new instances of the treated and control cities:

1. We first define the total population of units as the sign-ups occurring in DFW over the 20 weeks of our analysis. Each of these units is an opportunity-week pair corresponding to the opportunity and the week in which the sign-up occurred. We then sample from this population with replacement to form a new set of opportunity-week pairs such that this set is of the same length as the original population. We then do the same for the total population of sign-ups occurring in HOU over the 20 weeks of our analysis.
2. We aggregate these resampled sets on the week level to generate a new instance of the treated and control cities. Then we estimate our coefficients $\beta_{w,c}$ according to Equation 49.
3. We repeat steps 1-2 for 1,000 iterations, creating a set of 1,000 estimates for $\beta_{w,c}$.
4. We calculate the standard error of $\beta_{w,c}$ over these iterations, and report these as the standard errors in Figure 6.

Our pre-trend analysis generally supports the parallel trends assumption for our data: we do not see a pattern of steady increase or decrease in the $\beta_{w,c}$ coefficients over the pre-experiment weeks, suggesting that the trends are neither diverging nor converging over time. In addition, across all metrics, very few of the coefficients for the leads of the treatment are statistically indistinguishable from 0, suggesting that parallel trends hold over the pre-experiment weeks.

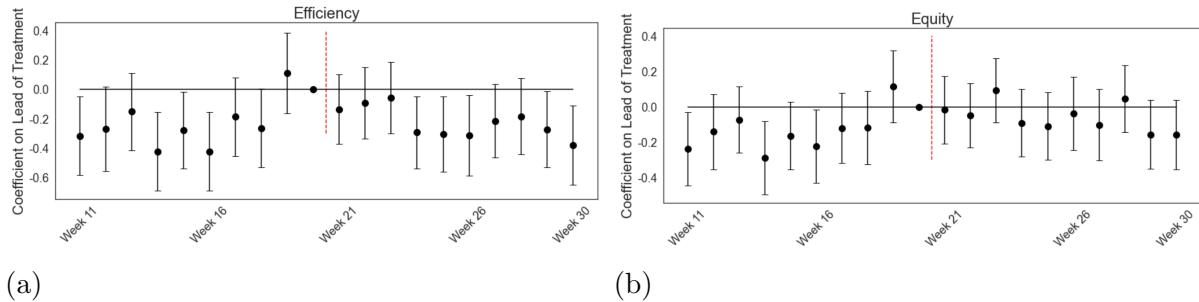


Figure 6 An event study pre-trend analysis on the seasonal trends of our outcome metrics.

C.3. Justification for Including Time-Varying Covariates in DID Analysis

Including a time-varying covariate, such as efficiency, to a difference-in-differences regression specification requires proper justification. Intuitively, if the weekly efficiency outcome explains some of the variance in the weekly equity outcome then not controlling for it could result in omitted variable bias, leading to imprecise measurement of the variance on the treatment coefficient. However, if efficiency is itself strongly correlated with the treatment, then including this variable would introduce co-linearity issues that could bias the interpretation of the treatment coefficient. Therefore, it would only be appropriate to control for the time-varying covariate of efficiency if efficiency and the equity outcome are highly correlated but efficiency and the treatment indicator are not correlated once we control for the other covariates. As mentioned in the main text, this is in fact the case in our data. We show the correlation values throughout this section.

To rigorously make the above point, let us first introduce new notations. First consider the base specification from Equation 8; let us refer to this as specification A (SA).

$$\log(Y_{wc}) = \beta_0 + \beta_{\text{treat}} \cdot T_{wc} + \beta_c \cdot \mathbf{1}_{\{c=\text{DFW}\}} + \beta_{\text{post}} \cdot \mathbf{1}_{\{w \in [21, 30]\}} + \varepsilon_{wc}, \quad (\text{SA})$$

Here T_{wc} is the indicator variable for whether the treatment is turned on or not in week $w \in [11, 30]$ and city $c \in \{\text{DFW}, \text{HOU}\}$ and Y_{wc} is the equity outcome in week w in city c . Note that ε_{wc} is the residual error term and thus assumed to be uncorrelated with all covariates. By definition, we have:

$$\mathbb{E}[\varepsilon_{wc}^2] := 1 - R_{(\text{SA})}^2$$

Now by the Frisch-Waugh-Lovell Theorem (Frisch and Waugh (1933), Angrist and Pischke (2013)), we can write the treatment coefficient in (SA) as follows (note that this will be an unbiased estimator).

$$\beta_{treat} = \frac{Cov(Y_{wc}, \tilde{T}_{wc}^A)}{Var(\tilde{T}_{wc}^A)},$$

where \tilde{T}_{wc}^A is defined as follows:

$$T_{wc} = \alpha_c \cdot \mathbf{1}_{\{c=\text{DFW}\}} + \alpha_w \cdot \mathbf{1}_{\{w \in [21, 30]\}} + \tilde{T}_{wc}^A \quad (50)$$

Next consider the specification that additionally controls for the log-transformed weekly number of sign-ups as in Column 4 in Table 2. Let us refer to this as specification B (SB), expressed as follows:

$$\log(Y_{wc}) = \gamma_0 + \gamma_{treat} \cdot T_{wc} + \gamma_c \cdot \mathbf{1}_{\{c=\text{DFW}\}} + \gamma_{post} \cdot \mathbf{1}_{\{w \in [21, 30]\}} + \gamma_L \cdot L_{wc} + \mu_{wc}, \quad (\text{SB})$$

In the above, L_{wc} represents the log-transformed number of sign-ups in week $w \in [11, 30]$ and city $c \in \{\text{DFW}, \text{HOU}\}$. Note that again μ_{wc} is the residual error and assumed to be uncorrelated with all covariates. By definition, we have:

$$\mathbb{E}[\mu_{wc}^2] := 1 - R_{(\text{SB})}^2$$

Applying the Frisch-Waugh-Lovell Theorem again, we can express the treatment coefficient in (SB) as follows ²³:

$$\gamma_{treat} = \frac{Cov(Y_{wc}, \tilde{T}_{wc}^B)}{Var(\tilde{T}_{wc}^B)},$$

where \tilde{T}_{wc}^B is similarly defined:

$$T_{wc} = \eta_c \cdot \mathbf{1}_{\{c=\text{DFW}\}} + \eta_w \cdot \mathbf{1}_{\{w \in [21, 30]\}} + \eta_L \cdot L_{wc} + \tilde{T}_{wc}^B \quad (51)$$

Estimate of γ_{treat} :

We want to know how controlling for L_{wc} will impact the estimate of the treatment effect; specifically, we want to show that in our setting $\gamma_{treat} \approx \beta_{treat}$. To do this, it suffices to show that $Var(\tilde{T}_{wc}^B) \approx Var(\tilde{T}_{wc}^A)$ and that $Cov(Y_{wc}, \tilde{T}_{wc}^B) \approx Cov(Y_{wc}, \tilde{T}_{wc}^A)$. And to do this, it suffices to show that $\tilde{T}_{wc}^B \approx \tilde{T}_{wc}^A$.

Towards this goal, we first use the Frisch-Waugh-Lovell Theorem to express the coefficient on L_{wc} in (51):

$$\eta_L = \frac{Cov(T_{wc}, \tilde{L}_{wc})}{Var(\tilde{L}_{wc})}$$

Here \tilde{L}_{wc} is defined as follows:

$$L_{wc} = \zeta_c \cdot \mathbf{1}_{\{c=\text{DFW}\}} + \zeta \cdot \mathbf{1}_{\{w \in [21, 30]\}} + \tilde{L}_{wc} \quad (52)$$

Notice that when the correlation between T_{wc} and \tilde{L}_{wc} ($\rho_{T, \tilde{L}}$) is low and all covariates are finitely bounded, we can conclude $\eta_L \approx 0$. In our setting $\rho_{T, \tilde{L}} = -0.01$, so $\eta_L \approx 0$. Next we make the crucial observation that \tilde{T}_{wc}^A and \tilde{T}_{wc}^B are residual error terms and thus assumed to be uncorrelated with the covariates in their

²³ Again, by the same theorem, this is an unbiased estimator.

specifications. Thus, having $\eta_L \approx 0$ implies $\tilde{T}_{wc}^B \approx \tilde{T}_{wc}^A$. Having $\tilde{T}_{wc}^B \approx \tilde{T}_{wc}^A$ implies both that $Var(\tilde{T}_{wc}^B) \approx Var(\tilde{T}_{wc}^A)$ and that $Cov(Y_{wc}, \tilde{T}_{wc}^B) \approx Cov(Y_{wc}, \tilde{T}_{wc}^A)$.

Thus, $\rho_{T,\tilde{L}} \approx 0$ implies that $\beta_{treat} \approx \gamma_{treat}$; since both β_{treat} and γ_{treat} are unbiased estimators, we expect to observe that $\hat{\beta}_{treat} \approx \hat{\gamma}_{treat}$. Table 2 shows exactly this pattern — the coefficient in column 3 (0.082) is approximately equal to the coefficient in column 4 (0.086) — suggesting that controlling for L_{wc} is appropriate. The same argument holds for the specification in column 5. For this specification, $\rho_{T,\tilde{L}} = -0.013$ and the coefficient of 0.085 is approximately equal to that in column 3, suggesting that controlling for L_{wc} is appropriate here as well.²⁴

Variance of $\hat{\gamma}_{treat}$:

Let us now consider what will happen to the variance on the treatment coefficient. Asymptotic normality of the OLS regression gives us the following expression for the estimated coefficient in (SA) (Wooldridge (2010)):

$$\sqrt{n}(\hat{\beta}_{treat} - \beta_{treat}) \xrightarrow{d} \mathcal{N}\left(0, \frac{\mathbb{E}[\varepsilon_{wc}^2 | T_{wc}, \mathbf{1}_{\{c=DFW\}}, \mathbf{1}_{\{w \in [21,30]\}}]}{Var(\tilde{T}_{wc}^A)}\right)$$

Given the assumptions on the residual term, we can simplify this to:

$$\sqrt{n}(\hat{\beta}_{treat} - \beta_{treat}) \xrightarrow{d} \mathcal{N}\left(0, \frac{(1 - R_{(SA)}^2)}{Var(\tilde{T}_{wc}^A)}\right)$$

Next the asymptotic distribution for the treatment coefficient in (SB) is given by:

$$\begin{aligned} \sqrt{n}(\hat{\gamma}_{treat} - \gamma_{treat}) &\xrightarrow{d} \mathcal{N}\left(0, \frac{\mathbb{E}[\mu_{wc}^2 | T_{wc}, \mathbf{1}_{\{c=DFW\}}, \mathbf{1}_{\{w \in [21,30]\}}, L_{wc}]}{Var(\tilde{T}_{wc}^B)}\right) \\ &\xrightarrow{d} \mathcal{N}\left(0, \frac{(1 - R_{(SB)}^2)}{Var(\tilde{T}_{wc}^B)}\right) \end{aligned}$$

Taking the ratio of the variances of $\hat{\gamma}_{treat}$ and $\hat{\beta}_{treat}$, we get the following expression:

$$\frac{(1 - R_{(SB)}^2)}{(1 - R_{(SA)}^2)} \cdot \frac{Var(\tilde{T}_{wc}^A)}{Var(\tilde{T}_{wc}^B)}$$

Notice, when the correlation between $\log(Y_{wc})$ and L_{wc} ($\rho_{\log(Y_{wc}), L}$) is high, the first term of the above expression will be considerably less than 1. When the correlation between T_{wc} and \tilde{L}_{wc} ($\rho_{T,\tilde{L}}$) is low, we have already shown that the second term of the above expression will be approximately 1. Thus, in this setting the variance of $\hat{\gamma}_{treat}$ is lower than that of $\hat{\beta}_{treat}$ asymptotically.

In our setting, the correlation between $\log(Y_{wc})$ and L_{wc} is $\rho_{\log(Y_{wc}), L} = 0.91$ while the correlation between T_{wc} and \tilde{L}_{wc} is $\rho_{T,\tilde{L}} = -0.01$. This suggests that after controlling for L_{wc} (i.e., the log-transformed efficiency), we should see a decrease in the variance of our treatment coefficient. Table 2 shows exactly this pattern — the error on the coefficient in column 4 (0.027) is smaller than the error on the coefficient in column 3 (0.097) — suggesting that controlling for L_{wc} is appropriate. Again, the same argument holds for the specification in column 5. For this specification, $\rho_{\log(Y_{wc}), L} = 0.91$ and $\rho_{T,\tilde{L}} = -0.013$. The error on the coefficient in column 5 (0.025) is lower than that in column 3, suggesting that controlling for L_{wc} is appropriate here as well.

²⁴ Note, here \tilde{L}_{wc} is the residualized version of L_{wc} based on the covariates in the specification of column 5.

C.4. DID Analysis on Distributional Impacts

In this section, we complement the aggregate-level analysis of Section 4.3 using a DID analysis with the following specification:

$$\log(Y_{wc}) = \beta_0 + \beta_{\text{treat}} \cdot \mathbf{1}_{\{w \in [21, 30]\}} \cdot \mathbf{1}_{\{c = \text{DFW}\}} + \beta_c \cdot \mathbf{1}_{\{c = \text{DFW}\}} + \delta_w + \beta_S \cdot \log(S_{wc}) + \varepsilon_{wc}, \quad (53)$$

This specification is identical to the one used in column 5 of Table 2; it controls for city and week fixed effects as well as the log-transformed efficiency, denoted by $\log(S_{wc})$.

Table 6 Distribution of Opps, DFW vs HOU

	Number of Opportunities Receiving			% of Sign-ups in
	= 1 sign-up	≥ 2 sign-ups	≥ 3 sign-ups	Excess of 1 sign-up
Treated	0.125** (0.043)	-0.060 (0.058)	-0.288 (0.158)	-0.063** (0.018)
City	-0.132 (0.090)	-0.147 (0.122)	0.683 (0.331)	0.085* (0.037)
Log Sign-ups	0.425* (0.160)	0.999*** (0.215)	3.116*** (0.584)	0.302*** (0.066)
Intercept	2.491* (0.891)	-1.772 (1.202)	-14.337*** (3.264)	-1.344** (0.369)
Week FE	Yes	Yes	Yes	Yes
Obs.	40	40	40	40
R ²	0.969	0.982	0.925	0.872
Adj. R ²	0.929	0.959	0.828	0.707
RSE	0.068	0.092	0.250	0.028
F Stat.	24.050***	42.466***	9.525***	4.983**

Note:

*p<0.05; **p<0.01; ***p<0.001

The results are presented in Table 6. We observe that the estimated treatment effects are directionally aligned with our aggregate-level analysis in Section 4.3 even though coefficients for the number of opportunities receiving at least 2 or 3 sign-ups are not statistically significant.

Appendix D: Empirical Analysis for the Southern California Experiment

D.1. Results in Southern California

Upon completion of the first wave of experimentation in DFW, we designed an additional wave of experimentation in three California regions. This wave of experimentation followed a staggered roll-out, such that each region adopted SmartSort at a different time. The order of roll-out was chosen at random.

The timeline for the roll-out in California is as follows:

- Phase I began on Saturday, November 5, 2022 in West LA, defined as all zip codes with a geographic center within the California state boundaries with a latitude between 34.5° N and 33.6° N and longitude west of 118.1° W.

2. Phase II began on Tuesday, December 6, 2022 in East LA, defined as all zip codes with a geographic center within the California state boundaries with a latitude between 34.5° N and 33.6° N and longitude east of 118.1° W.
3. Phase III began on Wednesday, January 4, 2023 in San Diego, defined as all zip codes with a geographic center within the California state boundaries with a latitude south of 33.6° N.

We emphasize that once the treatment was turned on in each of these regions, it was not turned off.

Given the variability in the time of treatment adoption for these geographic regions, we would ideally like to run a staggered treatment adoption analysis to obtain our difference-in-differences estimates. However, these three regions have contiguous borders (see Figure 2). Opportunities near these borders could be viewed by both volunteers that are experiencing SmartSort search rankings and those that are experiencing search rankings based on CP. This phenomenon could create contamination that would violate the no anticipation assumption required for a difference-in-differences analysis. That is, the introduction of SmartSort in West LA might create an effect in East LA *prior* to the introduction of SmartSort in East LA. To avoid this concern, we first consider an identical analysis to that in DFW in which we combine all California experimental regions into a single Southern California region (SCA) and we compare the ten weeks of the pre-experiment period when no part of the region was treated to ten weeks of the experiment period when the entire region was treated. Specifically we define the pre-experiment period to be Tuesday August 23, 2022 to Monday October 31, 2022 (corresponding to weeks 34-43 of 2022) and the experiment period to be Wednesday January 4, 2023 to Tuesday March 14, 2023 (corresponding to weeks 1-10 of 2023).²⁵

To control for seasonal trends, we choose a control region, the San Francisco/San Jose (denoted by SFJ), which consists of all zip codes that are within a 20-mile radius of either the San Francisco or San Jose city centers (measured using Haversine distance). This control region is not contiguous with the treated region. SCA and SFJ each contain the two largest cities in their area, are comparable over a spectrum of observable characteristics (see Table 11 in Appendix E.1), and experience similar seasonal trends (see Figure 7 in Appendix E.1). To be consistent with the terminology of our first experiment, we refer to SCA and SFJ as “city,” even though they are comprised of more than one city. The data we will be analyzing in the second experiment is based on sign-ups in both the pre-experiment and experiment periods for opportunities located within the treated area of SCA and the control area of SFJ. To avoid overly influential outliers, we drop the top 1% of opportunities in each of the treated and control regions by sign-ups received. The resulting dataset, which we will use throughout this section, consists of sign-ups for 5,886 unique opportunities in SCA and 2,298 unique opportunities in SFJ.

Aggregate statistics from this second experiment show a similar effect as we saw in our first experiment. Table 7 — which is the analog of Table 1 from Section 4.2.1 — shows that both SCA and SFJ experience an increase in efficiency (presumably due to an increase in volunteering activity at the beginning of the year);

²⁵ In keeping with the implementation schedule in our first experiment, we had intended the final California roll-out to be on Tuesday January 3, 2023. However, due to an administrative issue at VolunteerMatch the actual implementation date was Wednesday January 4, 2023, resulting in experiment period weeks that run from Wednesday to Tuesday of each week. An identical analysis that defines the pre-experiment period as Wednesday August 24, 2022 to Tuesday November 1, 2022 produces nearly identical results.

however the net change in SCA over SFJ is quite small (-0.9%). On the other hand, while both SCA and SFJ also experience an increase in equity, the net change in SCA over SFJ is large and positive ($+8.9\%$).

Table 7 Platform's city-wide weekly average counts, SCA vs SFJ

City	Period	Efficiency		Equity	
		Mean	Std.	Mean	Std.
Treated (SCA)	Pre-experiment	1258.9	91.9	851.3	57.8
	Experiment	1503.9	127.7	1035.0	59.4
Control (SFJ)	Pre-experiment	544.7	69.8	358.6	36.9
	Experiment	655.8	82.9	404.2	38.2
Change in the treated region (SCA)		$+19.5\%$		$+21.6\%$	
Change in the control region (SFJ)		$+20.4\%$		$+12.7\%$	
Net change in SCA over SFJ		-0.9%		$+8.9\%$	

The full difference-in-differences analysis supports these aggregate results. Table 8 (the analog of Table 2 in Section 4.2.2) summarizes this analysis. The coefficient of -0.010 in columns 1 and 2 can be interpreted as a small, statistically insignificant -1.0% decrease in the total number of sign-ups in a week after SmartSort was introduced in 2023 in SCA.

Meanwhile, we do see significant changes in the equity metrics; the treatment coefficients (0.075 , 0.082 , and 0.083 respectively) in columns 4, 5, and 6 correspond to an increase of (7.8% , 8.5% , 8.7% respectively) in equity after SmartSort was introduced in 2023 in SCA. The effect is statistically significant in columns 4 and 5, i.e. after controlling for log-transformed efficiency.²⁶

Finally an investigation into the distributional impacts of SmartSort in SCA shows that here too the equity gain was primarily driven by an increase in the number of opportunities receiving exactly one sign-up. Table 9 (the analog of Table 3 in section 4.3) presents the descriptive statistics for the outcome metrics described in section 4.3. We observe that during the experiment period, the weekly number of opportunities with exactly one sign-up increases in both SCA and SFJ (resp. 21.4% and 4.7%), however the increase is substantially higher in SCA (a net change of $+16.7\%$). On the other hand, the net change in the weekly number of opportunities receiving at least 2 (resp. 3) sign-ups is negative -9.5% (resp. -35.8%). This pattern of redistribution is again aligned with our expectation.

Finally looking at the measure of excess, in the last column of Table 9, we observe that during the experiment period excess decreased in SCA (-5%) whereas it increased in SFJ ($+9\%$), resulting in a net change of -14% . A difference-in-differences analysis identical to the one in Appendix C.4 and summarized in Table 10 corroborates these observational findings.

²⁶The conditions stated in Appendix C.3 that justified controlling for (log-transformed) efficiency in columns 4 and 5 also hold in our second setting. Details are omitted for the sake of brevity.

Table 8 Log Transformed City Level Analyses, SCA vs SFJ

	Efficiency		Equity		
	(1)	(2)	(3)	(4)	(5)
Treated	-0.010 (0.068)	-0.010 (0.044)	0.075 (0.053)	0.082*** (0.018)	0.083*** (0.015)
Post	0.187*** (0.048)		0.121*** (0.037)	-0.017 (0.015)	
City	0.843*** (0.048)	0.843*** (0.031)	0.867*** (0.037)	0.248*** (0.039)	0.232*** (0.034)
Log Sign-ups				0.734*** (0.043)	0.754*** (0.083)
Intercept	6.479*** (0.034)	6.303*** (0.051)	5.998*** (0.026)	1.241*** (0.281)	1.160** (0.523)
Week FE	No	Yes	No	No	Yes
Obs.	40	40	40	40	40
R ²	0.947	0.989	0.971	0.997	0.999
Adj. R ²	0.942	0.976	0.969	0.997	0.997
RSE	0.107	0.069	0.083	0.028	0.024
F Stat.	213.848***	76.349***	406.034***	2,785.415***	666.594***

Note:

*p<0.05; **p<0.01; ***p<0.001

Table 9 Platform's city-wide weekly distributional counts, SCA vs SFJ

City	Period	Opportunities receiving				% of Sign-ups	
		= 1 sign-ups	≥ 2 sign-ups	≥ 3 sign-ups	Excess of 1 sign-up	Mean	Std.
Treated (SCA)	Pre-experiment	613.1	40.1	238.2	27.8	85.5	13.4
	Experiment	744.4	42.9	290.6	33.0	91.6	24.2
Control (SFJ)	Pre-experiment	251.6	21.7	107.0	16.7	41.0	9.3
	Experiment	263.5	24.0	140.7	19.5	58.6	12.6
Change in the treated city (SCA)		+21.4%		+22.0%		+7.1%	-5.0%
Change in the control city (SFJ)		+4.7%		+31.5%		+42.9%	+9.0%
Net change in SCA over SFJ		+16.7%		-9.5%		-35.8%	-14.0%

Appendix E: Supplemental Evidence for the Southern California Experiment

E.1. San Francisco/San Jose as Control for Southern California

SCA and SFJ are two of the largest regions in California. This makes SFJ a natural comparison city. Drawing on zip-level data from the US Census American Community Survey 2020²⁷, in Table 11 we show that in terms of population, demographic composition, and income level the regions included in our analysis are fairly

²⁷ <https://www.census.gov/newsroom/press-kits/2021/acs-5-year.html>

Table 10 Distribution of Opps, SCA vs SFJ

	= 1 sign-up	Number of Opportunities Receiving ≥ 2 sign-up	≥ 3 sign-ups	% of Sign-ups in Excess of 1 sign-up
Treated	0.154** (0.027)	-0.067* (0.032)	-0.304*** (0.061)	-0.054*** (0.011)
City	0.388*** (0.090)	-0.101 (0.148)	-0.011 (0.281)	-0.150*** (0.049)
Log Sign-ups	0.598*** (0.147)	0.902*** (0.173)	1.276* (0.330)	0.159** (0.057)
Intercept	1.823* (0.928)	-2.164* (1.092)	-2.028 (2.079)	-0.678* (0.358)
Week FE	Yes	Yes	Yes	Yes
Obs.	40	40	40	40
R ²	0.997	0.994	0.973	0.911
Adj. R ²	0.993	0.986	0.939	0.797
RSE	0.043	0.051	0.097	0.017
F Stat.	236.329***	124.614***	28.054***	7.956***

Note:

*p<0.05; **p<0.01; ***p<0.001

comparable though SCA is much larger. Additionally, using aggregate city-level data on volunteer activity provided by Google Analytics we also confirm that in our pre-experiment period (defined in Appendix D.1) the number of visitors and sessions (i.e., the periods of time during which visitors interact with the VM site) in the two largest cities in each region are on similar orders; we note that this data is not available on a more granular level, and comprises a subset of the full activity in the SCA and SFJ regions.

Table 11 Balance Table - SCA and SFJ

	SCA	SFJ
Total Population	20,362,390	3,380,356
Pct Population Female	50.48%	50.17%
Pct Population White	53.44%	45.52%
Median Age (years)	36.6	38.2
Median Income	\$78,000	\$120,000
Total Users (pre-experiment)	40,903	18,109
Total Sessions (pre-experiment)	58,218	26,049

The key identifying assumption for the DID analyses is that of parallel trends between the treatment and control groups. Here we provide evidence in support of the parallel trend assumption between SCA and SFJ. In Figure 7 we simply plot the trends for our two outcome metrics across the two regions. Recall that the pre-experiment ran for 10 consecutive weeks from Tuesday August 23, 2022 - Monday October 31, 2022 (week 34

- week 43) and the experiment period ran for 10 consecutive weeks starting from Wednesday January 4, 2023 - Tuesday March 14, 2023 (week 1 - week 10). For both metrics, we observe strong and reasonably consistent seasonal trends in the pre-experiment period, with a noticeable start-of-the-year boost in volunteer activity in both California regions. Further evidence of the parallel trends assumption is available but omitted for brevity.

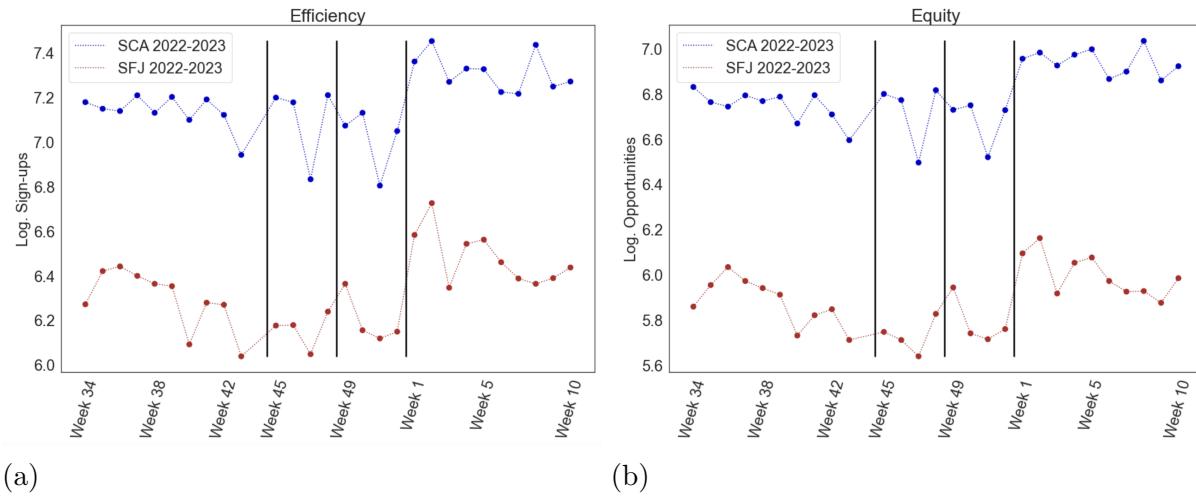


Figure 7 The weekly seasonal trends of efficiency and equity in SCA and SFJ.