Equity-Driven Workload Allocation for Crowdsourced Last-Mile Delivery

Abhay Sobhanan[1], Hadi Charkhgard[1], and Iman Dayarian[2]

[1]Department of Industrial and Management Systems Engineering, University of South Florida, Tampa, Florida, USA

[2]Culverhouse College of Business, University of Alabama, Tuscaloosa, Alabama, USA

Abstract

Crowdshipping, a rapidly growing approach in Last-Mile Delivery (LMD), relies on independent crowdworkers for delivery orders. Building a sustainable network of crowdshippers is essential for the survival and growth of such systems, while their participation is primarily motivated by fair pay. Additionally, the financial well-being of crowdworkers is sensitive to fair compensation, especially for those who depend on crowdwork as their main source of income. Therefore, equitable workload allocation and compensation mechanisms in crowdsourcing platforms will benefit both platforms and crowdworkers. We aim to answer several questions gig-economy platforms interested in fair pay may ask: How to measure equity, assess the cost benefits, and manage potential drawbacks? Our main contribution is the proposal of a practical equity-oriented framework tailored to crowdshipping within an LMD environment. This framework draws inspiration from the real-world operations of a group of crowdshipping platforms and operates in real-time. At its core is a bi-objective optimization process that balances equity and cost, aiming to address the study's main research questions. Built on a theoretical foundation, it enables the use of various equity measures and allows us to identify the equity measure that most reliably explores the trade-offs between cost and equity. We show that even a marginal sacrifice in cost efficiency (e.g., 2.5%) can significantly improve equity, potentially up to 39%. We provide actionable recommendations for practitioners, offering insights into selecting equity measures. We demonstrate that significant improvements in pay equity can be achieved with minimal increases in company's operational costs. Our experiments reveal that the best level of equity is achieved when the pool of employed crowdshippers is kept as small as possible. We quantify the loss of high and low-performing crowdshippers as the crowdshipper pool size increases, offering further insights for workforce management.

Key words: Crowdshipping, Last-mile delivery, Equitable workload allocation, Fair pay, Bi-objective optimization

1 Introduction

Last-mile delivery (LMD) consistently represents a significant expense within the logistics framework. According to CapGemini Research Institute (2019), LMD stands out as the primary cost driver in the supply chain, a fact accentuated by the exponential growth of e-commerce, which exposes the unsustainability of current delivery models. Vehicle routing in last-mile logistics is often considered non-value-added due to resource underutilization, high delivery costs, and loss of business time. In response to the growing demand for faster delivery services, several retailers are now adopting platforms allowing them to employ independent gig workers for order fulfillment. Such platforms facilitate the onboarding of gig workers based on demand dynamics, lowering operational costs compared to maintaining permanent employees, as noted by Fatehi and Wagner (2022). The employment of gig workers in the context of last-mile delivery is often referred to as crowdshipping. In a typical implementation of crowdshipping, a group of ad-hoc/occasional drivers takes over the delivery task of one or multiple online orders to online shopper locations (Alnaggar et al., 2019; Archetti et al., 2016; Arslan et al., 2019; Dayarian and Savelsbergh, 2020; Macrina et al., 2017; Soto Setzke et al., 2017). The most distinguishing characteristic of crowdshipping from conventional delivery methods in LMD is that crowdshippers are not employees of the company providing the delivery service; they become available to render service only occasionally and usually utilize spare space in their personal vehicles to make such deliveries on the way to their next destination. Thus, crowdshippers differ in terms of the time they become available, the capacity of their vehicles, and their next destinations as the determinant of their coverage area, among other factors.

Therefore, while crowdshipping revolutionizes LMD, its distinguishing factors introduce new operational challenges. One significant challenge, as discussed in Allon et al. (2023), is ensuring the uninterrupted and sustained arrival of participating crowdshippers. This is mostly determined by crowdshippers' satisfaction from their past participation. Satisfied crowdshippers are more likely to undertake subsequent jobs, maintaining operational flow. To satisfy crowdshippers, a fair and equitable workload allocation and compensation mechanism is required (Bai et al., 2019; Dayarian and Pazour, 2022). An equitable workload allocation creates equal profit-making opportunities for all participating crowdshippers at a time, given the system state and crowdshippers individual competencies. This not only promotes financial stability among crowdworkers, particularly those who rely on crowdwork as their primary income source, but also provides a sense of insurance, ensuring that even in the worst-case scenario, outcomes match or exceed those of an efficiency-focused workload allocation approach. However, achieving workload equity often comes with two types of costs: operational cost and opportunity cost. The operational cost is proportional to the deviation of the equity-oriented workload allocation from the allocation that would have been made if an efficiency-oriented objective had been adopted. The opportunity cost represents the potential reduction in the revenue of a subset of participating crowdshippers by altering the workload allocation to improve the payoff to the remaining crowdshippers to create a more equitable compensation allocation.

We focus on a setting in which a company, convinced of the benefits of an equitable workload allocation, is willing to take on the operational cost of equity. Passing on any part of the operational cost of equity to the crowdshippers would result in a reduction in the total payoff

(opportunity) offered, which goes against the spirit of equity used as an encouragement for a higher participation rate of the crowd. Opportunity cost, on the other hand, is an inherent aspect of equity considerations. By definition, it implies that some opportunities will be redistributed among crowdshippers, resulting in gains for some and losses for others compared to purely efficiency-oriented allocations. While increased participation may be expected among those with an improved payoff, those experiencing losses, despite recognizing the ethical principle of equity, may find it discouraging. This concern is particularly pertinent for high-performing crowdshippers, who make available higher resource levels (e.g., larger vehicles) and thus expect greater returns. Mitigating this drawback is a primary objective of our study, achieved through the implementation of an effective workforce size management strategy.

The main contribution of this paper lies in the proposal of a practical equity-oriented framework tailored to crowdshipping within an LMD environment. This framework seeks to mitigate the adverse effects of operational and opportunity costs while maximizing the benefits of equitable workload allocation. The considered framework is inspired by the real-world activities of a group of crowdshipping platforms. We refer to the identified problem as the Dispatch Zone-Wave Problem (DZWP). In conjunction with the problem formulation, we propose a solution method that provides effective strategies for addressing the DZWP. The method is tailored to optimize decision-making processes, thereby enhancing the overall effectiveness of the equity-aware framework in LMD contexts with a similar format. Numerical experiments are conducted to evaluate the efficiency and performance of the proposed method, providing empirical evidence of its capability to address practical-sized scenarios.

The proposed framework focuses on equity-aware workload allocation within a free-market environment, specifically aiming to benefit for-profit companies. The desired benefits materialize when: (1) the operational cost of equity remains sufficiently low to uphold the company's competitive edge, and (2) equity-driven outcomes favor all crowdshippers uniformly to the highest possible degree including the high-performing crowdshippers. To address these aspects, first, we design a mechanism to mathematically express crowdshippers' workloads (equity metric), which must be equalized. This would also necessitate the design of equity measures, to gauge the level of equality/inequality achieved. In this study, we propose a novel equity metric for comprehensively capturing workload allocation within a diverse courier pool. As for the equity measures, we adopt several established functions in the economics literature. Second, we acknowledge that at the heart of our proposed framework lies a bi-objective optimization process that determines the trade-offs between equity and cost, referred to as the 'nondominated frontier'. To preserve the company's competitive edge, we propose that the company sets an upper bound on the maximum operational cost they are willing to take on when balancing equity and cost. Utilizing our proposed framework, we aim to answer three primary research questions through extensive computational analysis:

1. *Which equity measure(s) prove most reliable within the scope of our study, given that there is no unique way of defining an equity objective function?* Using a multi-objective optimization analysis, we first prove that none of the equity measures in the literature of economics can be theoretically dismissed in the context of the DZWP. We then provide a novel solution approach to identify the most reliable equity measure(s) numerically. A measure is considered reliable

if its solutions, corresponding to its nondominated frontier, yield high-quality approximations of the nondominated frontier when moved to the criterion space of other equity measures. Our numerical findings highlight *Coefficient of Variation* as particularly promising within the context of our research.

2. *How much can equity be enhanced with varying equity budgets by the close of a working day?* We showcase that adhering to our framework can lead to an equity improvement of up to about 39% within a working day, with a cost increase of no more than 2.5%. Furthermore, this enhancement can extend to about 70% while incurring a cost increment of no more than 10%

3. *What potential loss might high-performing crowdshippers incur due to the company's equity considerations?* Our numerical evidence underscores our framework's aim to minimize losses for high-performing crowdshippers. To provide a more holistic perspective, we also study low-performing crowdshippers. A significant finding of this study is the alignment of interests between low-performing and high-performing drivers, as well as equity-focused for-profit companies. The advantage for all involved is in minimizing the number of crowdshippers in the system. This approach not only reduces the company's expenses but also minimizes opportunity costs for both high-performing and low-performing drivers.

The remainder of this paper is organized as follows: Section 2 provides a brief literature review. Section 3 discusses the theoretical foundation of the research. Section 4 details the DZWP. Our proposed optimization models and solution methods are in Section 5. Section 6 presents a solution to identify the most reliable equity measure(s). Section 7 presents a detailed computational study to tackle the research questions. Finally, Section 8 provides a discussion and concluding remarks.

2 Literature Review

In this section, we classify the literature relevant to our study into two categories: optimization approaches for the LMD and fairness in resource allocation. Among the optimization techniques, we investigate exact methods, machine learning models, and heuristics. We then review fairness principles and models in resource allocation literature, and explore equity considerations within the vehicle routing problem (VRP) literature.

2.1 Optimization Approaches for the Last-mile Delivery

Among the solution approaches, exact methods are devoted to achieving optimal solutions for optimization problems. To ensure their effectiveness, exact solution methods often necessitate extensive customization for each variant of the VRP. Notable contributions in the existing literature include the generic exact methods proposed by Baldacci and Mingozzi (2009) and Pessoa et al. (2020), which address multiple VRP variants. Additionally, Arslan et al. (2019) iteratively employs an exact method to solve the crowdsourced dynamic pickup and delivery problem as new information emerges. Furthermore, Fatehi and Wagner (2022) addresses a robust

counterpart of crowdshipping using queuing theory and robust optimization techniques. In a recent study, Raghavan and Zhang (2024) employed a branch-cut-and-price algorithm to study the impact of driver-aide on improving delivery speed. Despite these advancements, solving even moderately sized instances to optimality remains computationally demanding. For an extensive review of exact methods and heuristics applied to VRPs in freight transportation, readers can refer to Konstantakopoulos et al. (2022).

In recent years, machine learning (ML) techniques have gained popularity, either as standalone tools (Kwon et al., 2020) or in conjunction with exact or heuristic approaches (Morabit et al., 2021; Sobhanan et al., 2023) to solve VRPs. Behrendt et al. (2023) specifically focuses on the courier scheduling problem within the context of crowdsourced same-day deliveries. However, ML encounters challenges such as distribution dependency and extended training times, critical factors in the face of evolving data trends. Furthermore, the performance of many ML models is constrained to a specific problem size. Both exact and ML methods are tailored to specific problem contexts and may not be easily adaptable to problems with additional features.

As a result, heuristics emerged as the most widely employed and researched method for tackling large-scale optimization problems. Macrina et al. (2020) considers a variant of the crowdshipping problem involving intermediate depots and employs a variable neighborhood search heuristic. Among heuristics, the genetic algorithm stands out as a popular and widely accepted heuristic for solving the VRPs. Genetic algorithms operate on a population of solutions, providing increased flexibility in managing the population to guide the search toward optimal solutions. Vidal (2022) provides an open-source package to solve capacitated VRP with a genetic algorithm and finds near-optimal solutions with remarkably short computational times. Genetic algorithms have also been proven beneficial in effectively addressing multiple VRP variants. Some examples include routing problems with multiple depots and periodicity (Vidal et al., 2012), or truck-drone mixed fleet (Mahmoudinazlou and Kwon, 2024).

2.2 Fairness in Resource Allocation

Fairness in resource allocation is a critical concern that has been addressed in various fields such as healthcare, public policy, economics, and ethics. Readers may refer to Mandell (1991) for an earlier study that explores the trade-off between cost and equity in public service delivery, as opposed to LMD in for-profit companies. Bertsimas et al. (2012) addresses the trade-off between efficiency and fairness in resource allocation using the notion of α-fairness, where the parameter α provides flexibility in adjusting this trade-off. For a comprehensive examination of various mathematical models that attempt to balance efficiency and fairness, one can refer to Xinying Chen and Hooker (2023). The authors assert that no single approach is generally suitable to all problems, as the principles of fairness vary across contexts. Chen and Hooker (2020) presents a related approach by introducing social welfare functions (SWFs) that combine fairness and utilitarianism to propose a mixed integer linear programming model. Their sequential optimization prioritizes less advantaged recipients, but this lexicographic approach limits a balanced fairness perspective.

Readers are directed to Matl et al. (2018) for a comprehensive discussion of workload equity in VRPs. Their study covers desirable properties of equity functions and compares commonly used

ones. It also addresses the VRP with route balancing, which considers fairness in route length among couriers. Matl et al. (2019) conduct a bi-objective optimization analysis of classic VRPs. Other studies analyze fairness in transportation problems from a single-objective perspective. McCoy and Lee (2014) quantify the efficiency and equity of allocating health workers from clinics to outreach sites. Equity considerations are increasingly prominent in LMD decision-making. Ibarra-Rojas and Silva-Soto (2021) investigate equity in demand satisfaction when only a fraction of orders are fulfilled. Yu et al. (2024) explore workload balancing among couriers in a consistent VRP, focusing on route, courier, and time consistency.

3 Theoretical Foundation

In this section, we establish a theoretical framework applicable to various problems, including the one explored in this paper. Consider an entity G that needs to allocate N indivisible goods of varying sizes among M independent agents, each seeking to maximize their utility. Each good $i \in \mathcal{N} = 1, \ldots, N$, with size q_i, is assigned to one agent. The set of all feasible allocations is denoted by $\mathcal{S}$, with s_m representing the set of goods allocated to agent $m \in \mathcal{M} = \{1, \ldots, M\}$ under allocation $s \in \mathcal{S}$. Each allocation s incurs a cost, denoted by $\texttt{Cost}(s)$, for entity G and generates utility $u(s_m)$ for agent m. The utility function $u(\cdot)$ is identical for all agents. Entity G aims to minimize costs and ensure fairness in allocation, creating a multi-objective optimization problem with $M + 1$ objectives: minimizing G's costs and optimizing each agent's utility. To manage the complexity, we reduce the $M + 1$ objectives to two by replacing the utility functions of agents with an equity measure $\texttt{EquityMeasure}(s)$ that we seek to minimize. $\texttt{EquityMeasure}(s)$ is a function of $u(s_1), \ldots, u(s_M)$ and captures the disparity among agents' utilities; smaller values of $\texttt{EquityMeasure}(s)$ indicate lower disparities. Using this bi-objective conversion, entity G seeks to identify the trade-offs between cost and equity, determining what is known as the "nondominated frontier". Each point on this frontier, called a "Pareto-optimal" solution, represents a feasible solution within the objective function space, where improving one objective compromises the other. The final choice of a Pareto-optimal solution depends on the specific application. We propose a tailored selection approach, particularly suited for for-profit entities, which we elaborate on later in this study.

The process described above holds relevance across a wide spectrum of applications, and we apply it within the scope of this paper as well. Nevertheless, it is imperative to underscore a critical aspect when endeavoring to reduce the full-dimensional criterion space, i.e., the $M + 1$-dimensional criterion space, into a two-dimensional counterpart. This process of reduction essentially constitutes a mapping from a higher dimension to a lower dimension, and we define a 'proper mapping' as follows:

Definition 1. *Reducing the full-dimensional criterion space to a two-dimensional criterion space constitutes a 'proper mapping' if every Pareto-optimal solution within the bi-objective criterion space also remains Pareto-optimal within the full-dimensional criterion space.*

According to Definition 1, a proper mapping ensures that solving the bi-objective optimization problem guarantees a subset of Pareto-optimal solutions from the full-dimensional criterion space. Failure to ensure proper mapping could result in generating suboptimal outcomes, where

improving one agent's utility does not necessarily result in the deterioration of others' utilities. To ensure proper mapping, the structure of the utility function and the choice of equity measures are pivotal. For instance, consider using the 'range' of utility values as the equity measure to minimize. Without a specific structure in the utility function, achieving proper mapping using the range becomes improbable. This is because the range aims to minimize the gap between the best and worst utility values, neglecting those in between. Consequently, this approach could lead to arbitrarily poor outcomes for the agents whose utilities are not at the extremities of the range. Hence, such solutions may not qualify as Pareto-optimal for the full-dimensional criterion space.

The above discussion highlights that for an arbitrary utility function, not all equity measures are suitable. Specifically, those equity measures that do not result in a proper mapping can be ignored.

Theorem 1. *If $u(I)$ is an increasing function of $\sum_{i\in I} q_i$ for all $I \subseteq \mathcal{N}$, then the proper mapping is guaranteed for any chosen equity measure.*

Proof. We aim to prove the assertion through a contradiction. Consider an arbitrary equity measure, and suppose $u(I)$ is an increasing function of $\sum_{i\in I} q_i$, with $I \subseteq \mathcal{N}$, but this mapping is not a proper one. In such a scenario, there must exist a Pareto-optimal solution $s \in \mathcal{S}$ in the bi-objective optimization problem, which is not Pareto-optimal in the full-dimensional criterion space counterpart. This suggests the existence of a feasible solution $s' \in \mathcal{S}$ that dominates s in the full-dimensional criterion space. Formally, we require $\texttt{Cost}(s') \leq \texttt{Cost}(s)$ and $u(s'_m) \geq u(s_m)$ for all $m \in \mathcal{M}$, with at least one strict inequality.

Now, we consider two cases: First, suppose there exists $m \in \mathcal{M}$ with $u(s'_m) > u(s_m)$. In this case, it immediately follows that $\sum_{i\in s'_m} q_i > \sum_{i\in s_m} q_i$ by assumptions. However, since we know that $\sum_{i\in\mathcal{N}} q_i$ is a constant, there must exist $l \in \mathcal{M}$ such that $\sum_{i\in s'_l} q_i < \sum_{i\in s_l} q_i$. Based on our assumptions, this implies $u(s'_l) < u(s_l)$, leading to a contradiction as s' cannot dominate s. Alternatively, if $u(s'_m) = u(s_m)$ for all $m \in \mathcal{M}$, which implies

$$\texttt{EquityMeasure}(s') = \texttt{EquityMeasure}(s),$$

then for s' to dominate s, we must have $\texttt{Cost}(s') < \texttt{Cost}(s)$. However, this directly contradicts the Pareto-optimality of s in the bi-objective version. □

Theorem 1 suggests that as long as the utility function of agents is an increasing function of the summation of item sizes allocated to them, then none of the equity measures can be considered unsuitable in theory. Therefore, the task of selecting the most appropriate equity measure(s) becomes primarily a computational challenge. Later in Section 6, we outline a computational method for this selection process. The underlying concept of our proposed approach lies in solving a bi-objective optimization problem for each equity measure. If the Pareto-optimal solutions of one equity measure closely approximate those of the others, we designate it as the most reliable equity measure. This is because by selecting that equity measure, we obtain solutions that are good across a variety of equity measures.

4 Problem Description

Our study focuses on a company (i.e., Entity G) operating from a single depot, aiming to conduct last-mile deliveries (i.e., Indivisible Goods of varying sizes) using crowdshippers (i.e., Agents). Each day is divided into equal-length time intervals (e.g., every hour) that online shoppers can choose among for their delivery to take place. Each interval is available to the online shoppers by a fixed cutoff time, after which it cannot be selected for delivery. The group of orders received for delivery within a time interval is referred to as a "wave". Moreover, delivery orders within a wave are further broken down based on their delivery locations. Specifically, we assume that the geographical area covered by the depot is divided into predefined delivery zones to categorize orders based on their delivery locations. The compensation offered for deliveries made at different delivery zones can vary based on their characteristics (further discussed in Section 4.1).

To streamline our discussion, we label each delivery zone within a wave as a "zone-wave", presuming the company devises a plan for each zone-wave independently. Once the orders for a given zone-wave are identified, the decision maker faces the task of securing sufficient transportation capacity to deliver the orders of each interval. As crowdshippers become available, they choose the zone-wave they wish to serve. Subsequently, the company adopts a first-come, first-served approach for selecting crowdshippers for each zone-wave. Note that this assumption is both practical and fair. It is practical because from a crowdshipper's perspective, once they see a company's request for a given zone-wave in their platforms, they expect prompt accommodation rather than being put on hold for an extensive time to know if they are selected. It is fair because it prevents the company from discriminating against crowdshippers. We propose that the company selects the minimum number of crowdshippers necessary to fulfill orders for each zone-wave. Later in this paper, we explore the consequences of the company deviating from this assumption on both the low- and high-performing crowdshippers. Following crowdshipper selection, the company endeavors to allocate jobs (and their corresponding routes) to them by solving a variant of vehicle routing problem (VRP) incorporating two conflicting objectives: cost and equity.

4.1 Cost Objective Function

The foundation of our cost objective function lies in the compensation policy established by the company. Therefore, we first discuss this policy before presenting the cost function. The company's compensation policy encompasses two primary facets: per-delivery compensation and mileage compensation. Per-delivery compensation is contingent upon both the delivery zone and the quantity of deliveries. Specifically, a delivery in zone z is compensated proportional to the size/quantity of the delivery by a factor $\beta_z \in \mathbb{R}_+$, denoting the per-quantity compensation allocated by the company for Zone z. The value of β_z is customizable based on factors such as demand volume, traffic, safety, and proximity to the depot, among others. On the other hand, mileage compensation is proportional to the traveled distance by a crowdshipper to make assigned deliveries by a factor $\alpha \in [0, 1]$. In essence, α denotes the portion of routing costs covered by the company, and its determination in this study is grounded in two pivotal assumptions:

Assumption 1. *The company absorbs all operational costs of equity, ensuring that none are passed on to the crowdshippers.*

Assumption 2. *Crowdshippers should not be able to garner additional profits by deviating from the routes assigned to them by the company (the shortest route to make deliveries).*

Therefore, in our setting, the company compensates mileage costs only based on the optimal routes. Additionally, setting $\alpha = 1$ (full compensation of the mileage cost by the company) implies that within a designated zone-wave, the collective profits earned by all crowdshippers remain constant, equating to the total compensation per delivery. This collective profit resembles a fixed-size cake, with varying workload allocations akin to different cutting strategies implemented by the company. Striving for fairness, the company aims to distribute the cake more equitably while minimizing overall costs. Since the compensation per delivery remains fixed, the company's expenses are primarily influenced by routing costs. Therefore, the objective function for cost optimization is its total routing cost.

4.2 Equity Objective Function

This objective function relies on two critical concepts (Matl et al., 2018; Xinying Chen and Hooker, 2023):

- *Equity Metric* is the utility function of agents (e.g., crowdshippers), representing the aspect (e.g., profits, tour length) to be equalized. It quantifies resource allocation according to predefined fairness guidelines.
- *Equity Measure* is a mathematical function used to assess equality (e.g., range and standard deviation). It serves as the fairness criterion for equitable resource distribution.

Combining these two concepts yields the equity objective function, which requires optimization. However, there is no unique method to define this function. In the scope of our research, the challenge lies in delineating both the equity metric and the equity measure. This challenge emerges due to the varying characteristics among crowdshippers, such as vehicle capacities. While profits may initially appear as the most fitting equity metric since all crowdshippers strive to maximize their earnings, excluding crowdshippers' investments (in terms of resources they make available) from the equity metric may result in biased and inconclusive outcomes. This implies that crowdshippers can be viewed as rational investors who, from their perspective, anticipate higher returns with increased investment. Consequently, the equity metric should encompass crowdshippers' investments.

In our research context, crowdshippers offer the company two resources: their time and vehicle capacities. Since each zone-wave is tackled independently, crowdshippers chosen for a specific zone-wave are already largely standardized in terms of time, as they are allocated to the same zone-wave. Therefore, each zone-wave offers similar time commitments and compensation policy for crowdshippers interested in servicing it. Consequently, within a given zone-wave, the sole resource requiring adjustment to compute crowdshippers' profits is their vehicle capacities.

With this consideration, we introduce the following equity metric for each crowdshipper, termed "adjusted profit":

$$\text{Adjusted Profit} = \frac{\text{Total Profit Earned}}{\text{Vehicle Capacity}} \times \frac{\text{Total Capacity Used}}{\text{Vehicle Capacity}}. \tag{1}$$

Equation (1) reflects the preference of each crowdshipper to maximize profits by fully utilizing their vehicle capacity. The first term of the equation captures the latter preference, while the second term encapsulates the former. Based on our discussions in Section 4.1, "Total Profits Earned" can be calculated by taking the product of β_z and the "Total Capacity Used" representing the total quantity assigned to a crowdshipper. This implies that in EQ. (1), both the numerator and denominator are consistent in their quadratic form. Squaring inherently renders the equity objective function sensitive to outliers, similar to the concept of mean square errors in statistical methods.

Observation 1. *'Adjusted Profit' is an increasing function of 'Total Capacity Used'.*

Now that we have established the equity metric, the primary question arises: what form should the equity measure take? The challenge is related to the existence of multiple measures in the literature. Therefore, we approach the selection of an equity measure as a research question. Specifically, within the context of our research, we examine five well-known equity measures, i.e., range, mean absolute deviation, standard deviation, coefficient of variation, and Gini coefficient. While there are other measures, we do not include them in this study as they perform similarly to those we have already considered. For instance, Min-Max, a popular method, shares similarities with the range. We choose to use range since our research emphasizes distributional justice over the condition of the worst-paid-off crowdshipper.

The equity objective functions can be defined as follows. Let $\boldsymbol{x}$ represent any feasible solution to the problem at hand. In this study, we denote the adjusted profit for each crowdshipper $m \in \mathcal{M}$, corresponding to solution $\boldsymbol{x}$, as $\bar{p}_m(\boldsymbol{x})$. Essentially, $\bar{p}_m(\boldsymbol{x})$ quantifies the utility of crowdshipper m under solution $\boldsymbol{x}$. With these notations, the equity objective functions considered in this study are Range, Mean Absolute Deviation (MAD), Standard Deviation (SD), Coefficient of Variation (CV), and Gini Coefficient (GINI), represented as set $\Theta := \{\text{Range, MAD, SD, CV, GINI}\}$. The mathematical expressions of these functions are provided in Appendix A. In closing this subsection, we emphasize the following result:

Proposition 1. *The equity objective functions considered in this study offer proper mappings.*

Proof. This follows directly from Theorem 1 and Observation 1. □

In Section 6, we will present an approach to numerically compare and select the most reliable equity measure(s) for the problem under study.

4.3 Balancing Objectives: Equity and Cost

As previously noted, the equity and cost objective functions are two antithetical objectives, making it unlikely for a solution to simultaneously optimize both. Consequently, in tackling such bi-objective optimization problems, the primary aim for the company is to identify the

trade-offs between the objectives, determining what is known as the "nondominated frontier". Recall that each point on this frontier represents a feasible solution, known as a "Pareto-optimal" solution, within the objective function space. These solutions are labeled Pareto-optimal because improving one objective unavoidably results in a compromise in the other.

With the nondominated frontier calculated, the company faces the task of selecting a nondominated point (and its corresponding Pareto-optimal solution) that effectively balances the equity and cost objectives from its perspective. This selection process is closely tied to the specific problem at hand and is heavily influenced by the application's context and underlying research philosophy. In our research context, the guiding philosophy centers on favoring an equity-driven workload allocation, i.e., balancing adjusted profits, within a free-market environment, with a focus on benefiting for-profit companies. Benefits are attained when (1) the cost of equity remains sufficiently low to preserve the company's competitive edge, and (2) the outcomes driven by equity are uniformly favorable for all crowdshippers to the highest degree feasible. In essence, the cost of equity borne by the company should not yield undesirable consequences, resulting in uniformly unfavorable outcomes for all crowdshippers.

The latter condition is addressed by the proposed compensation policy outlined earlier in Section 4.1, which guarantees that crowdshippers will not bear the burden of equity costs. The former condition is pivotal in balancing the cost and equity objective functions. This means that the company can only afford to sacrifice a small fraction of its least-achievable costs to achieve a higher workload allocation for crowdshippers. Consequently, the company selects a point from the nondominated frontier with a capped percentage increase in costs, denoted by γ (e.g., 5%), and then chooses the closest feasible nondominated point to it—effectively selecting the nondominated point that maximizes the equity objective function while adhering to the cost cap. As an aside, we sometimes refer to γ as the equity budget of the company in this paper.

5 Optimizing a Zone-Wave

As mentioned previously, the company must independently explore each zone-wave. Therefore, this section provides additional details about the optimization problems that arise in each zone-wave, along with explanations of our proposed solution approaches. Our research context involves two optimization problems: crowdshipper selection and workload allocation. The former is a single-objective optimization problem to determine the minimum number of crowdshippers required to fulfill all orders within a targeted zone-wave. The latter, a bi-objective optimization process, focuses on generating the nondominated frontier for cost and equity objective functions.

5.1 The Crowdshipper Selection Problem

The problem is essentially a variation of the classical bin packing problem. It entails dynamically solving the classical bin packing problem as crowdshippers join the system and enroll for the targeted zone-wave. Specifically, crowdshippers are accepted on a first-come, first-served basis. A crowdshipper is admitted only if the current capacities of accepted crowdshippers are insufficient to meet the demand in the targeted zone-wave. Thus, admitting a crowdshipper involves solving a classical bin packing problem, which is not computationally expensive and can be quickly

addressed using a commercial integer linear programming solver (e.g., Gurobi or CPLEX). We assume that a sufficient number of crowdshippers will register in the system to fulfill orders in any zone-wave. This assumption is not restrictive because if there are insufficient crowdshippers, the company can redistribute some orders (in a first-come, first-served manner) to the immediate next wave within the same zone. Interested readers can find the mathematical formulation of the crowdshipper selection problem and further details in Appendix B.

Table 1: Notations Used in Workload Allocation Optimization

Notation	Description
Sets:	
$\mathcal{N}$	Set of all customers with a cardinality of N
$\overline{\mathcal{N}}$	Set of all vertices, $\mathcal{N} \cup \{0\}$ where 0 is the depot
$\mathcal{A}$	Set of all arcs
$\mathcal{M}$	Set of all selected crowdshippers with a cardinality of M
Parameters:	
c_{ij}	Cost of traversing the arc (i, j)
q_i	Order size of customer $i \in \mathcal{N}$
p_i	Per delivery compensation for fulfilling the order of customer $i \in \mathcal{N}$
Q_m	Vehicle capacity of crowdshipper $m \in \mathcal{M}$
Q	Maximum vehicle capacity among crowdshippers selected, i.e., $\max(Q_m)_{m \in \mathcal{M}}$
Variables:	
x_{ijm}	Binary decision variable which equals one if crowdshipper $m \in \mathcal{M}$ traverses the arc from i to j, 0 otherwise
u_i	Continuous decision variable designed for subtour elimination constraints, where $i \in \mathcal{N}$

5.2 The Workload Allocation Problem

After completing the crowdshipper selection, the next step is determining workload allocation, which involves two conflicting objectives: cost and equity. This problem can be seen as a variation of the capacitated VRP with three main distinguishing characteristics:

- *Heterogeneous Vehicles*: Unlike traditional VRP variants with homogeneous vehicles, our problem involves a limited fleet of vehicles with varying loading capacities.
- *Open Routes*: In our research context, each crowdshipper begins their route from the depot and serves assigned customers based on prescribed routing. Unlike traditional VRP, crowdshippers are not required to return to the depot after completing their tasks.
- *Bi-objective Optimization*: In contrast to traditional VRP with a single objective, our problem introduces an additional objective function alongside cost, namely equity. The equity objective function can be highly non-linear depending on the underlying equity measure.

Using the notation outlined in Table 1, the workload allocation problem for any zone-wave can be formulated as follows:

$$\min \quad \Big\{ \texttt{Cost}(\boldsymbol{x}),\ \texttt{EquityMeasure}(\boldsymbol{x}) \Big\} \tag{2a}$$

$$\text{s.t.} \quad \sum_{i \in \overline{\mathcal{N}}} \sum_{m \in \mathcal{M}} x_{ijm} = 1 \qquad \forall j \in \mathcal{N} \tag{2b}$$

$$\sum_{i \in \overline{\mathcal{N}}} x_{ijm} = \sum_{i \in \overline{\mathcal{N}}} x_{jim} \qquad \forall j \in \overline{\mathcal{N}}; m \in \mathcal{M} \tag{2c}$$

$$\sum_{j \in \mathcal{N}} x_{0jm} = 1 \qquad \forall m \in \mathcal{M} \tag{2d}$$

$$\sum_{m\in\mathcal{M}} x_{iim} = 0 \qquad \forall i \in \overline{\mathcal{N}} \tag{2e}$$

$$q_i \leq u_i \leq \sum_{m\in\mathcal{M}} Q_m \sum_{j\in\overline{\mathcal{N}}} x_{jim} \qquad \forall i \in \mathcal{N} \tag{2f}$$

$$u_i - u_j + Q \sum_{m\in\mathcal{M}} x_{ijm} \leq Q - q_j \qquad \forall i,j \in \mathcal{N}; i \neq j \tag{2g}$$

$$u_i \in \mathbb{R} \qquad \forall i \in \mathcal{N} \tag{2h}$$

$$x_{ijm} \in \{0,1\} \qquad \forall i,j \in \overline{\mathcal{N}}; m \in \mathcal{M} \tag{2i}$$

The first objective function aims to minimize the total routing cost incurred by all crowdshippers, and it can be stated as:

$$\texttt{Cost}(\boldsymbol{x}) := \sum_{i\in\overline{\mathcal{N}}} \sum_{j\in\mathcal{N}} \sum_{m\in\mathcal{M}} c_{ij} x_{ijm}.$$

The second objective function retains its abstraction, allowing for the integration of any equity measure to evaluate the disparity in adjusted profits. Specifically, define

$$\mathcal{N}_m(\boldsymbol{x}) := \left\{ j \in \mathcal{N} : \exists\ i \in \overline{\mathcal{N}} \text{ such that } x_{ijm} = 1 \right\}$$

as the set of customers to be serviced by crowdshipper $m \in \mathcal{M}$ under solution $\boldsymbol{x}$. Utilizing this notation, the adjusted profit for crowdshipper $m \in \mathcal{M}$ under solution $\boldsymbol{x}$, denoted by $\bar{p}_m(\boldsymbol{x})$, is expressed as:

$$\bar{p}_m(\boldsymbol{x}) = \left(\frac{\sum_{i\in\mathcal{N}_m(\boldsymbol{x})} p_i}{Q_m} \right) \left(\frac{\sum_{i\in\mathcal{N}_m(\boldsymbol{x})} q_i}{Q_m} \right). \tag{3}$$

With this groundwork, any of the $\texttt{Range}(\boldsymbol{x})$, $\texttt{MAD}(\boldsymbol{x})$, $\texttt{SD}(\boldsymbol{x})$, $\texttt{CV}(\boldsymbol{x})$, and $\texttt{GINI}(\boldsymbol{x})$ measures, as stated in Appendix A, can be employed as $\texttt{EquityMeasure}(\boldsymbol{x})$ in the optimization model.

Constraints (2b) ensures the fulfillment of all customer demands. It is important to note that while constraints (2c) enforce flow conservation at each node, the first objective function solely minimizes traversal costs for active routes. Hence, the cost of returning from the last served location to the depot, although considered in the constraint set, is omitted from the objective function. Crowdshippers depart from the depot according to constraints (2d), and self-visits are prohibited by constraints (2e). Subtour elimination constraints (2f)–(2g), also known as Miller-Tucker-Zemlin constraints (Miller et al., 1960), ensure the connectivity of crowdshipper routes based on heterogeneous load capacities. Variables u_i are assigned real values as per (2h), yet their range is bounded by constraints (2f). Finally, the integrality conditions of routing decision variables are enforced by constraints (2i). It is important to note that we have not explicitly introduced capacity constraints as they are inherently captured by constraints (2f)–(2g). This is because each customer must be satisfied by a crowdshipper, implying that customer $i \in \mathcal{N}$ must be in the sequence of customers to be serviced by a crowdshipper. Thus, variable u_i can be interpreted as the capacity occupied by customer i and its preceding customers in that crowdshipper's route.

5.2.1 Proposed Solution Approach

Solving the workload allocation problem (2) is challenging. This is evident as even classical VRPs with a single objective pose considerable difficulty. In our research, we confront a variant of the VRP that is notably more complex due to its bi-objective nature, particularly when one objective is nonlinear (partly due to the nonlinear nature of the adjusted profits). Consequently, employing exact multi-objective optimization methods for solving the workload allocation problem becomes impractical. Therefore, we propose a custom-built heuristic approach tailored to address the problem, irrespective of the equity measure employed. Our proposed method efficiently approximates the nondominated frontier for any instance of the workload allocation problem. It combines the core principles of the well-known Non-dominated Sorting Genetic Algorithm (NSGA-II; Deb et al. (2002)) with the concepts of Multi-Directional Local Search (MDLS; Tricoire (2012)), offering an effective solution approach for handling the complexities of the workload allocation problem.

We refer to our proposed solution method as HybridNsgaII. At the heart of our proposed method is NSGA-II, known for its computational efficiency in approximating optimal trade-offs between conflicting objectives by ranking and selecting solutions based on their dominance relationship (Jozefowiez et al., 2005; Srivastava et al., 2021; Zhou et al., 2018; Gullotta et al., 2021). Specifically, during each iteration, NSGA-II evaluates the rank of each candidate solution using a non-dominated sorting technique. This technique categorizes the solutions into different approximations of the Pareto-optimal frontiers based on their dominance relationship with others. This implies that the first approximation has rank one and is better than the second, and the second is better than the third one, and so forth. To generate a new population of offspring solutions, NSGA-II then performs selection, crossover, and mutation operations. Additionally, to ensure well-distributed objective function values or diversity in the population of solutions, NSGA-II uses a crowding distance metric.

To effectively solve the workload allocation problem, we either customize some of the main operations or add some new operations to NSGA-II. Customized operations include:

- *Solution Representation:* We employ a sequence-based solution representation for N customers and M crowdshippers by a two-part chromosome: a giant tour gene of length N specifying the order of customer visits, and a breakpoint gene of length $M-1$ indicating where the giant tour is split into crowdshipper routes. The breakpoint gene contains increasing positive integers less than N, ensuring each crowdshipper serves at least one customer.

- *Initialization:* The initial population is generated using both targeted and random methods. While targeted methods such as the Clarke-Wright savings heuristic and nearest neighbor algorithm provide good initial solutions that improve convergence speed, random permutations of customers and breakpoints increase population diversity.

- *Genetic Algorithm Operations:* Parent selection, a critical step in the genetic algorithm, is implemented using binary tournament selection based on fitness scores. To generate offspring, we employ effective crossover operations specifically designed for sequence-based chromosome representations. Additionally, mutations are introduced with a predetermined probability and tournament selection.

Newly integrated operations include:

- *Local Search:* We attempt to refine the set of best solutions by employing MDLS at each iteration. This process prioritizes cost and equity separately, leveraging Large Neighborhood Search and workload reassignment strategies to achieve targeted improvements.
- *Infeasible Population Management:* Due to vehicle capacity constraints, some solutions generated during the search may be infeasible. Instead of discarding these solutions, we either attempt to repair them or preserve them in a separate population. This diversifies the total population and aids in escaping local optima. To ensure proper integration into the selection process, the objective values of infeasible solutions are penalized based on the extent of capacity violations, weighted by an adaptive penalty factor.
- *Diversity Score:* Our method enhances population diversity using a diversity score and alternately focuses on cost, equity, and solution rank throughout the search process to ensure a balanced exploration of the solution space.

Additional information about HybridNsgaII is available in Appendix C for interested readers. HybridNsgaII serves as a key tool for addressing the research questions in this study. Therefore, in our computational analysis, we prioritize demonstrating its reliability before utilizing it to tackle the research questions. To show its reliability, we borrow two measures from the literature of multi-objective optimization:

- *Cardinality*: The Cardinality, sometimes referred to as 'Card.' in this study, shows the number of approximate nondominated points found. We would like to highlight that, unlike exact methods, in heuristic solution approaches, the number of approximate nondominated points does not necessarily increase over time, as it is possible for a point to be found in one iteration that dominates a subset of the points in the previous iteration. However, one can hope that the cardinality measure gets stabilized over time.
- *Hypervolume*: This is a widely used performance indicator in multi-objective optimization (Pal and Charkhgard, 2019), and we sometimes refer to it as 'HV' in this study. It is a single value reflecting the area of the criterion space that is dominated by the approximate nondominated frontier found and is trapped between the approximated frontier and a given reference point. Larger values for the Hypervolume indicate better approximations. To ensure this premise, it is critical for the reference point to be selected carefully and remain fixed when comparing multiple approximations. In this study, because we are minimizing both objective functions, the correct selection of the reference point means each component of the reference point should be larger than or equal to the same component among all points across all approximations.

6 Identifying the Most reliable Equity Measure(s)

In this section, we explore the process of identifying the most reliable equity measures for solving the workload allocation problem within a given zone-wave. As per Proposition 1, all equity measures considered in this study offer proper mappings. Therefore, selecting the most reliable

measure(s) is primarily an empirical question, best addressed through computational analysis. Thus, we propose a method for this purpose in this section. The approach proposed here conducts a comprehensive analysis over the entire set of approximate Pareto-optimal solutions found using each measure. Later in Section 7.2 and Appendix E, we highlight other alternative techniques that rely on analyzing a subset of Pareto-optimal solutions found using each measure. However, such approaches raise some major questions that are not trivial how they can be addressed. We posit that a measure identified as the most reliable using full-set analysis is likely to perform well under subset-based analysis but not necessarily vice versa. We provide numerical evidence supporting this claim in Section 7.2 and Appendix E. With this in mind, we next present our proposed selection approach.

Our approach starts with generating approximate frontiers for each equity measure using HybridNsgaII for each instance. Given that we are investigating five equity measures in this study, we produce five distinct approximate nondominated frontiers for each instance. The proposed approach involves mapping the approximate Pareto-optimal solutions identified under each equity measure into the bi-objective space of every other equity measure. For each pair of equity measures, this mapping can be expressed as follows:

$$\texttt{Map}\big((\text{Cost}, \theta^I) \rightarrow (\text{Cost}, \theta^{II})\big), \quad \forall \theta^I, \theta^{II} \in \Theta := \{\text{Range, MAD, SD, CV, GINI}\}.$$

For instance, let θ^I be Range and θ^{II} be MAD. This indicates a desire to map the approximate Pareto-optimal solutions obtained by HybridNsgaII for Cost and Range into the space of Cost and MAD. To achieve this, we simply compute the values of Cost and MAD for each solution. While the Cost values remain constant as in the Cost and Range space, this mapping generates several feasible points in the Cost and MAD space. However, some of these feasible points may dominate others in this new space, necessitating the removal of dominated points. Through a removal process, an approximate nondominated frontier is derived for the Cost and MAD space using the solutions originally found for the Cost and Range space.

Considering the exploration of five distinct equity measures in Θ, for every equity measure $\theta^{II} \in \Theta$, we obtain five different approximated nondominated frontiers through this mapping procedure, namely

$$\texttt{Approximation}(\theta^I, \theta^{II}) : \texttt{Map}\big((\text{Cost}, \theta^I) \rightarrow (\text{Cost}, \theta^{II})\big), \quad \forall \theta^I \in \Theta.$$

Note that among these approximations, $\texttt{Approximation}(\theta^{II}, \theta^{II})$, where the approximate Pareto-optimal solutions obtained by θ^{II} is re-mapped into the space of θ^{II}, is expected to be the best approximation of the nondominated frontier for the space of Cost and θ^{II}. This is because such an approximation implies that HybridNsgaII has directly generated an approximate nondominated frontier for the space of Cost and θ^{II}, rendering the mapping redundant. However, such certainty is not guaranteed given that HybridNsgaII is a heuristic approach. Hence, we regard $\texttt{Approximation}(\theta^{II}, \theta^{II})$ as a benchmark for assessing the quality of all $\texttt{Approximation}(\theta^I, \theta^{II})$, $\forall \theta^I \in \Theta \setminus \theta^{II}$. By comparing the Hypervolumes of all five approximations to that of $\texttt{Approximation}(\theta^{II}, \theta^{II})$, we can compute the Hypervolume gaps. Naturally, the Hypervolume gap will be 0% for $\texttt{Approximation}(\theta^{II}, \theta^{II})$. However, for each

$\theta^I \in \Theta \setminus \theta^{II}$, the Hypervolume gap corresponding to $\texttt{Approximation}(\theta^I, \theta^{II})$ may vary positively or negatively. Smaller Hypervolume gaps across all approximations denote better performance.

After computing all five different approximations for each equity measure, we can now readily identify the most reliable measure(s). An equity measure earns this distinction if it not only produces the best-known approximation for itself but also comes close to achieving the same level of accuracy for other measures through mapping. We gauge this performance using the Hypervolume gap. Clearly, determining the best measure through this method requires a comprehensive computational study across various instances to assess the average performance of each equity measure and its corresponding mappings.

7 Computational Study

In this section, we conduct an extensive computational study using large-sized instances to not only showcase the performance of HybridNsgaII but also to address the main research questions delineated in Section 1. All simulation and optimization tools necessary for conducting the experiments in this paper are implemented and executed in Julia 1.9.4. The experiments are carried out on a MacBook Pro featuring an Apple M1 Chip with 16GB RAM, running on macOS Sonoma 14.1.2. In this paper, we conduct two types of experiments, each necessitating a unique instance generation approach. The first type focuses on operations at a single zone-wave level, exclusively tackling the workload allocation problem while omitting the crowdshipper selection aspect. Conversely, the second type explores the comprehensive operations of the company throughout a typical working day. These experiments encompass multiple zones and waves, addressing both the crowdshipper selection and workload allocation problems on a frequent basis.

For the first set of experiments, we create numerous instances of varying sizes, with the specific quantity and dimensions outlined individually for each experiment of this type. Here, we concentrate on delineating the instance generation process. Each instance comprises depots and customers randomly positioned across a grid with coordinates ranging from $[1, 100] \times [1, 100]$. Customer demands are integer values within the interval $[1, 100]$, while vehicle capacities are generated as τq_m where τ is randomly drawn from a uniform distribution in the range $[0.7, 3.3]$ and q_m represents the mean demand per vehicle. Additionally, the parameter β_z is randomly selected from the interval of $[0.7, 1.0]$. In the second set of experiments, we consider a working day as ten waves of customer orders, each wave containing between 150 and 200 orders. Orders are divided into four geographical zones and each zone is assigned a compensation factor β_z from the set $\{0.6, 0.7, 0.8, 0.9\}$ based on demand volume. Within each zone, two types of crowdshippers are available to fulfill the orders, differentiated by their vehicle capacity. Our simulation assumes a scenario where approximately 80% of crowdshippers are low-performing, with vehicle capacities $L \in \{150, 200, 250\}$, and the remaining 20% are high-performing, with capacities $H \in \{450, 500\}$. We denote the set of low-performing crowdshippers and the set of high-performing crowdshippers by $\mathcal{M}_L$ and $\mathcal{M}_H$, respectively.

7.1 Validating the Proposed Heuristic Approach

In this subsection, our goal is to present numerical evidence demonstrating the high-quality performance of HybridNsgaII in solving the workload allocation problem. This demonstration inherently asserts that HybridNsgaII is indeed a fitting tool for tackling our research questions. We provide two types of analysis to demonstrate the high performance of HybridNsgaII. First, we compare its performance with an off-the-shelf exact solver. Secondly, we provide some results about its workings over time, reflecting that it quickly generates high-quality approximations and can improve its approximations over time.

Table 2: Comparing HybridNsgaII to an Exact Solver using Range as the Equity Measure

N	M	Exact Solver			HybridNsgaII			HV Gap
		Card.	HV	Time (sec.)	Card.	HV	Time (sec.)	(%)
8	2	4.8	475.32	2.05	4.8	475.32	3.74	0.00
8	3	8.7	468.12	4.26	8.7	468.12	3.92	0.00
10	2	7.7	462.42	23.79	7.7	462.42	4.18	0.00
10	3	10.8	452.03	84.29	10.4	451.80	5.21	0.05
12	2	6.3	439.78	969.98	6.2	439.46	4.77	0.07
12	3	14.6	423.71	1,625.90	13.4	423.39	6.19	0.07

In light of the complex nature of the VRP variation we are examining, exact multi-objective optimization solvers face significant challenges in solving instances unless the equity measure chosen is Range from our list of options. This preference stems from the inherently linear nature of Range, coupled with the fact that existing off-the-shelf multi-objective optimization tools are primarily designed for integer linear programs. Consequently, our comparison with an exact off-the-shelf solver is focused solely on Range. However, we note that despite the Range's linear nature, the range of adjusted profits is non-linear due to the underlying non-linearity of adjusted profits (i.e., Equation (3)). To address this, we have decided to employ the range of vehicle utilization values (specifically, $\frac{\sum_{i\in\mathcal{N}_m(\boldsymbol{x})} q_i}{Q_m}$) for our comparison. For this purpose, we have employed the Julia package *MultiObjectiveAlgorithms*, leveraging the ϵ-constraint method and *Gurobi Optimizer*.

We did not impose any time limit for the Exact Solver. Consequently, six instances took up to 75 minutes to be solved using the exact solver, while the rest were solved within 30 minutes. The results of our comparison across various instance sizes are presented in Table 2, with each row displaying averages from 10 randomly generated instances (60 instances in total). In the final column of the table, we showcase the gap between the HV reported by our proposed approach and the exact method for the fixed reference point of $(700, 1)$.

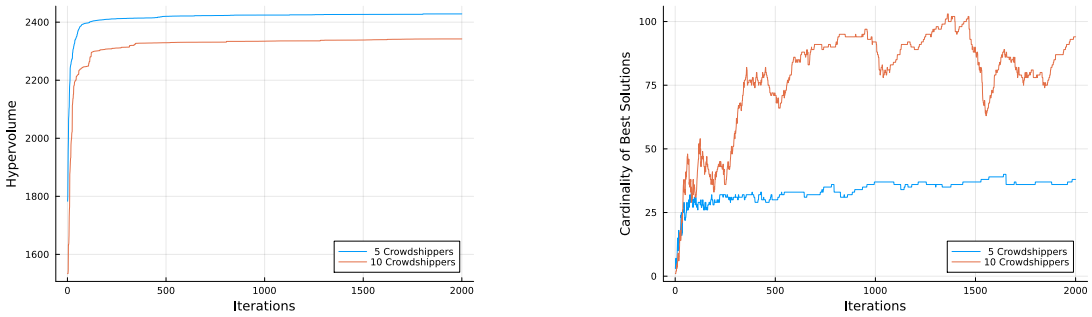
Overall, our findings underscore the efficacy of our proposed approach, which significantly outpaces the exact solver in terms of speed. Notably, for networks of size 8, our approach successfully generated all nondominated points. This trend continues for networks of size 10 with 2 crowdshippers. Furthermore, our observation from the table reveals that even for larger-sized instances, the proposed approach consistently finds almost all nondominated points. For additional details, readers may refer to Appendix D.

Now we focus on assessing the overall performance of the HybridNsgaII across various equity measures and its evolution over time. To accomplish this, we generated 80 random

Table 3: Average Run Time of HybridNsgaII across different Equity Measures

N	M	Average CPU Time (sec.)				
		Range	MAD	SD	CV	GINI
25	3	2.10	2.49	2.17	1.94	1.96
25	5	4.06	3.09	4.25	3.59	2.80
50	5	3.80	4.65	5.02	4.53	5.76
50	10	7.24	7.32	8.32	9.17	9.49
75	8	9.72	13.63	13.36	15.61	15.59
75	15	11.85	19.94	19.90	20.75	20.51
100	10	18.62	22.16	27.14	24.57	29.94
100	20	16.17	27.81	27.90	25.55	29.39
Average		9.20	12.64	13.51	13.21	14.43

instances using our instance generator across 8 classes of different sizes, each containing up to 100 nodes and 20 crowdshippers. The solution times of HybridNsgaII under different equity measures are detailed in Table 3, with the values in each row representing averages across 10 instances. Observe that the proposed approach can generate an approximate frontier within a matter of seconds even for large-sized instances. Unsurprisingly, we observe a relationship between the solution time of HybridNsgaII and the complexity of the chosen equity measure. The equity measure with the least complexity, Range, exhibits an average solution time of 9.2 seconds, while the most complex measure, GINI, necessitates 14.43 seconds on average to solve an instance.



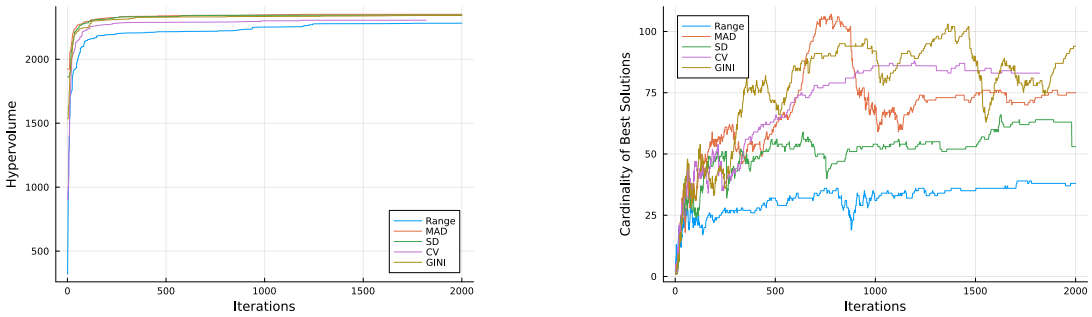
(a) Hypervolume

(b) Cardinality

Figure 1: Progression of HybridNsgaII results until termination under GINI for two instances with $N = 50$

The progression of HybridNsgaII in terms of both Hypervolume and Cardinality at each iteration is illustrated in Figures 1 and 2. Specifically, Figure 1 focuses solely on the GINI measure, depicting the progress of two instances simultaneously. Both instances have a population size of $N = 50$ deliveries, with one having $M = 5$ crowdshippers and the other $M = 10$. The latter instance is further depicted in Figure 2 across all equity measures considered. From these figures, it is evident that almost after 200 iterations, Hypervolume has nearly converged, indicating that high-quality approximations have been generated within only 10% of the total runtime of the HybridNsgaII. This implies that early termination can still result in high-quality approximations by HybridNsgaII, which can be useful for solving instances significantly larger than those used in this study.

Regarding Cardinality, natural fluctuations are observed, indicating continuous improvement



(a) Hypervolume

(b) Cardinality

Figure 2: Progression of HybridNsgaII results under various equity measures for an instance with $N = 50$ and $M = 10$

of its best-known approximate nondominated frontier over time, facilitated by various enhancement techniques such as dynamic weight adjustments and periodic repopulation of the solution set. It is noteworthy that a decrease in the nondominated frontier can occur when a newly found feasible point dominates a subset of points in the best-known approximate nondominated frontier, resulting in a reduction in Cardinality as dominated points are removed. Furthermore, an interesting but expected observation is that more complex equity measures lead to larger Cardinality values. This implies that in practice, companies may struggle to enhance equity when employing simpler measures like Range. The challenge lies in selecting a point from the approximate nondominated frontier that balances equity without significantly impacting costs. Due to the limited options available-attributed to the small cardinality under these simpler equity measures-companies may find it difficult to achieve substantial improvements.

7.2 The Most Reliable Equity Measures

In this subsection, we embark on exploring our first research question: which equity measure(s) prove most reliable within the scope of our study? To achieve this goal, we utilize the selection method introduced in Section 6. This method identifies a measure as most reliable when it can generate a set of Pareto-optimal solutions using HybridNsgaII that remain robust across all measures. As discussed in Section 6, alternatives to our proposed approach may involve analyzing a subset of Pareto-optimal solutions rather than the entire set holistically. However, such approaches encounter challenges as determining the appropriate subset selection method is not straightforward. Additionally, we posit that a method identified as most reliable through holistic analysis is likely to maintain its efficacy relative to subset-based approaches, but the reverse may not hold true. This implies that selections made using our proposed approach are expected to possess an additional layer of robustness against subset-based approaches. For those interested, Appendix E provides numerical evidence and further discussions supporting this claim, focusing on a special case of subset-based approaches when the subset size is one.

Given these considerations, we revisit the 80 instances outlined in Table 3 from the preceding subsection. These instances are evenly distributed across four classes, categorized by the number of customers in their networks, denoted as $N \in \{25, 50, 75, 100\}$. Utilizing the mapping techniques detailed in Section 6, we attempt to identify the most reliable measure(s). Table 4

provides a comprehensive summary of the HV gap for each observed instance class in the mapping experiments. Each row reports approximate Pareto-optimal solutions obtained using a specific equity function, reevaluated against the remaining functions investigated in this study. To maximize the reliability of our findings, we first solve each instance 10 times using HybridNsgaII for every equity measure. The best approximation, based on HV, is then selected from these 10 runs for each equity and will be used for subsequent mapping processes across various criterion spaces for that instance. Furthermore, to ensure consistent comparison of the Hypervolume gap across diverse problem sizes, each experiment class uses a fixed reference point. This point is determined by the maximum observed cost within the respective class.

Table 4: Mapping Results: Hypervolume Gaps

(a) Class $N = 25$

Map	HV Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	0.40	-0.03	0.39	-0.04	0.14
MAD	-0.25	0.00	-0.21	0.02	-0.36	-0.16
SD	0.44	0.93	0.00	0.92	0.31	0.52
CV	-0.38	-0.02	-0.39	0.00	-0.42	-0.24
GINI	0.10	0.57	-0.29	0.55	0.00	0.19

(b) Class $N = 50$

Map	HV Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	1.47	0.95	1.65	1.02	1.02
MAD	-0.86	0.00	-0.05	0.29	-0.34	-0.19
SD	-0.43	0.93	0.00	1.16	0.25	0.38
CV	-1.42	-0.26	-0.45	0.00	-0.62	-0.55
GINI	-0.61	0.50	0.00	0.75	0.00	0.13

(c) Class $N = 75$

Map	HV Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	1.17	0.93	1.30	1.31	0.94
MAD	-0.47	0.00	-0.34	0.25	-0.01	-0.11
SD	0.49	2.12	0.00	2.21	1.41	1.25
CV	-1.16	-0.20	-0.93	0.00	-0.29	-0.52
GINI	-0.69	0.48	-0.96	0.65	0.00	-0.10

(d) Class $N = 100$

Map	HV Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	1.37	0.53	1.81	2.30	1.20
MAD	0.71	0.00	-0.04	0.60	0.82	0.42
SD	1.16	3.04	0.00	3.42	2.95	2.11
CV	-1.44	-0.48	-2.20	0.00	-0.09	-0.84
GINI	-1.18	-0.46	-1.64	0.08	0.00	-0.64

Note that a negative HV gap in Table 4 indicates that the approximation found through mapping is better than the original approximation for that criterion space. This can partly be attributed to the fact that HybridNsgaII is a heuristic approach; however, by running each

instance 10 times, we aim to minimize its impact. Overall, the table shows that the average HV gap through mapping is up to about 2% better or worse than the original approximation. The mapping results demonstrate that CV is the most reliable equity measure, consistently outperforming others across all instance classes. Notably, GINI also exhibits improved performance as the problem size increases. However, while the Range equity function is computationally efficient, it underperforms due to its inability to capture variations in adjusted profits effectively.

7.3 Quantifying Potential Equity Improvement

In this subsection, we explore our second research question: How much can equity improve with varying equity budgets by the end of a working day? The goal is to demonstrate that equity can be significantly enhanced across various zones even with only a small increase in the operational cost for the company in different zones. To explore the question, we consider different equity budgets for the company, ranging from $\gamma \in \{2.5\%, 5\%, 7.5\%, 10\%\}$, and conduct a full simulation of the entire working day. We report the average improvement for each zone in Table 5, with each row showing averages across all 10 waves. This section focuses on CV, identified previously as one of the most reliable equity measures. Within each zone-wave, we select the minimum necessary crowdshippers based on the first-come, first-served principle. Our method for selecting a point from the approximate nondominated frontier follows the approach in Section 5.2.1, choosing the point that maximizes equity for the given equity budget.

Table 5: Summary of Results Obtained by Simulating a Working Day

Zone	β_z	N	$\lvert\mathcal{M}_L\rvert$	$\lvert\mathcal{M}_H\rvert$	Card.	Best Cost	Improvement in CV (%)			
							$\gamma = 2.5\%$	$\gamma = 5\%$	$\gamma = 7.5\%$	$\gamma = 10\%$
1	0.6	85.6	68.2	17.4	24.3	1,771.2	31.69	41.80	45.73	49.44
2	0.7	37.4	30.1	7.3	19.9	499.6	32.41	51.53	62.60	69.75
3	0.8	36.2	29.1	7.1	19.8	465.1	38.58	52.54	61.64	72.17
4	0.9	14.1	9.9	4.2	8.1	108.9	23.45	47.33	59.47	65.41

Table 5 demonstrates that even a minor deviation of 2.5% from the least-cost solution can significantly enhance equity in workload allocation. Across the four zones, distinguished by order volume and compensation policies, we observe an average equity improvement of 31.53% when allowing for this slight cost increase. For example, in Zone 1, a 2.5% deviation permits the selection of a solution with a route cost averaging 1,815.48, compared to the average least cost of 1,771.2. Similarly, with an equity budget of 5%, the average equity improvement is observed as 41.8% compared to the least-cost solution. This highlights that even with a modest equity budget, the company can foster a more equitable workload distribution, incentivizing participation from crowdshippers and aiding a sustainable business model.

7.4 Quantifying the Loss of Crowdshippers

In this subsection, we explore our third research question: How much potential loss might a high-performing crowdshipper face due to the company's equity considerations? Gains (or losses) refer to the difference between adjusted profits under an equity-aware solution and those under the least-cost solution. We provide insights on both low-performing and high-performing

crowdshippers in this section. The response to the research question depends on the number of crowdshippers in the system. Our proposed approach, by construction, targets the minimum number of crowdshippers (denoted as M) employed in a first-come-first-served framework, when solving the crowdshipper selection problem. It is anticipated that deviating from this minimum count and admitting more crowdshippers will increase the loss of crowdshippers. Thus, we examine the research question across different scenarios, contingent upon the crowdshipper count. Specifically, we increase M by 50% and 100%, i.e., $1.5M$ and $2M$, as shown in Table 6. Echoing the preceding subsection, we also explore various equity budgets for the company, spanning $\gamma \in \{2.5\%, 5\%, 7.5\%, 10\%\}$, and conduct a comprehensive simulation covering the entire operational day.

Table 6: Comparing the Impacts of Different Numbers of Crowdshippers by Simulating a Working Day

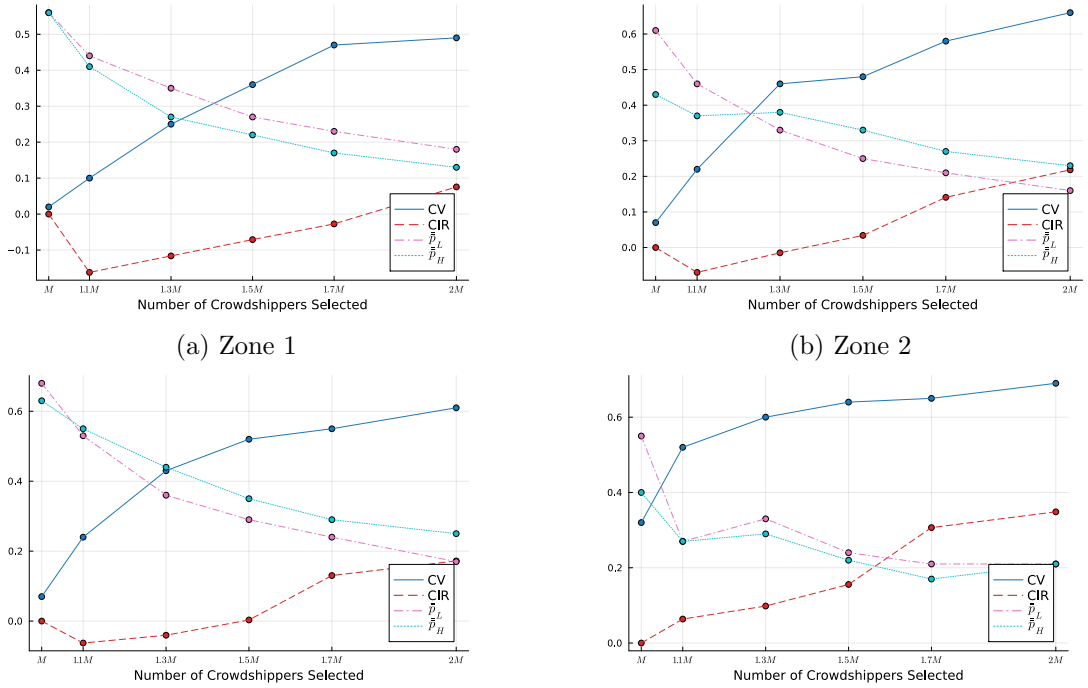
Zone	Equity Budget (%)	Number of Crowdshippers Selected											
		M				$1.5M$				$2M$			
		Cost	CV	$\bar{\bar{p}}_L$	$\bar{\bar{p}}_H$	Cost	CV	$\bar{\bar{p}}_L$	$\bar{\bar{p}}_H$	Cost	CV	$\bar{\bar{p}}_L$	$\bar{\bar{p}}_H$
1	0.0	1,771.2	0.04	0.56	0.56	1,631.1	0.49	0.28	0.24	1,888.4	0.65	0.18	0.15
	2.5	1,797.3	0.02	0.56	0.56	1,669.7	0.36	0.27	0.22	1,933.1	0.49	0.18	0.13
	5.0	1,831.1	0.02	0.56	0.56	1,709.4	0.31	0.27	0.21	1,980.8	0.42	0.17	0.12
	7.5	1,848.9	0.01	0.56	0.56	1,750.3	0.27	0.27	0.21	2,027.6	0.38	0.17	0.12
	10.0	1,891.9	0.01	0.56	0.56	1,791.2	0.24	0.27	0.21	2,069.8	0.35	0.16	0.12
2	0.0	499.6	0.10	0.61	0.44	514.3	0.64	0.26	0.37	607.7	0.82	0.18	0.24
	2.5	508.2	0.07	0.61	0.43	525.5	0.48	0.25	0.33	619.2	0.66	0.16	0.23
	5.0	520.7	0.04	0.61	0.44	538.5	0.39	0.24	0.33	635.0	0.57	0.16	0.21
	7.5	531.6	0.03	0.62	0.43	550.7	0.32	0.25	0.31	651.7	0.46	0.17	0.17
	10.0	546.0	0.03	0.61	0.43	563.8	0.27	0.24	0.30	664.7	0.40	0.16	0.17
3	0.0	465.1	0.12	0.66	0.66	466.7	0.65	0.28	0.40	546.0	0.80	0.17	0.28
	2.5	474.6	0.07	0.68	0.63	476.1	0.52	0.29	0.35	556.2	0.61	0.17	0.25
	5.0	483.7	0.05	0.68	0.63	488.6	0.40	0.29	0.33	571.5	0.49	0.17	0.22
	7.5	491.4	0.04	0.69	0.62	500.1	0.35	0.29	0.31	585.7	0.41	0.18	0.20
	10.0	505.5	0.03	0.68	0.62	511.1	0.29	0.29	0.30	597.3	0.35	0.17	0.20
4	0.0	108.9	0.40	0.54	0.42	125.7	0.77	0.25	0.26	145.8	0.86	0.22	0.23
	2.5	109.9	0.32	0.55	0.40	127.0	0.64	0.24	0.22	148.2	0.69	0.21	0.21
	5.0	112.3	0.20	0.49	0.40	130.4	0.51	0.26	0.22	151.1	0.55	0.23	0.16
	7.5	114.0	0.14	0.51	0.39	133.7	0.38	0.25	0.21	155.1	0.43	0.20	0.15
	10.0	115.6	0.13	0.51	0.39	136.8	0.27	0.25	0.21	157.8	0.34	0.17	0.15

The zone-wave experiment results for different numbers of crowdshippers are summarized in Table 6, with each row representing averages across all 10 waves. Note that $\bar{\bar{p}}_L$ and $\bar{\bar{p}}_H$ are the average adjusted profits for low-performing and high-performing crowdshippers, respectively. Results within each zone are categorized by the equity budgets discussed earlier. We observe the following five key managerial insights:

- *Insight I:* Expanding the number of crowdshippers generally drives up operational costs for the company. Yet, exceptions arise due to the requirement of utilizing all selected crowdshippers on a first-come-first-served basis. This variance is notable in Zone 1, where up to 70% increase in crowdshipper numbers has reduced the company's operational costs, while a 100% increase has increased the cost.

- *Insight II:* Even if the company hires more crowdshippers than necessary, it can still notably improve equity across all zones with a small equity budget. This improvement is evident in the shifts in the CV values for each crowdshipper size.

- *Insight III:* While equity gains remain notable even with a limited equity budget across different workforce sizes, the disparities between the average adjusted profits of high-performing and low-performing crowdshippers exhibit minimal variation. In essence, $\bar{\bar{p}}_L$ and $\bar{\bar{p}}_H$ exhibit comparable magnitudes across different budget allocations within each zone and the same workforce size. Thus, according to the definition of CV, equity gains predominantly arise from reductions in standard deviations.
- *Insight IV:* The equity deteriorates with an increase in the number of crowdshippers in general. This is apparent from the rising trend in the CV values across each row. According to the definition of CV, this suggests that as the number of crowdshippers increases, the average adjusted profits decrease significantly, to an extent where even a decrease in standard deviation cannot compensate for it.
- *Insight V:* Interestingly, selecting more crowdshippers than needed increases opportunity costs for both low- and high-performing crowdshippers. This is evident in the reduced adjusted profits for both types as the number of crowdshippers rises.



(a) Zone 1 (b) Zone 2

(c) Zone 3 (d) Zone 4

Figure 3: Average trends in Equity, Cost Increase Ratio, and Adjusted Profits when $\gamma = 2.5\%$

To demonstrate the broader applicability of our findings beyond the $1.5M$ and $2M$ thresholds, Figure 3 showcases the average values of CV, $\bar{\bar{p}}_L$, $\bar{\bar{p}}_H$, and the Cost Increase Ratio (CIR) for all four zones across various workforce sizes, with an equity budget set at 2.5%. The CIR quantifies the percentage increase or decrease in costs compared to hiring the minimum number of crowdshippers. For instance, a CIR of 0.1 indicates a 10% increase, while -0.1 implies a 10% decrease in costs. Overall, Figure 3 reaffirms that our key findings extend beyond the $1.5M$ and $2M$ thresholds. Consequently, the managerial imperative highlighted in this section is to prioritize selecting the minimum number of crowdshippers.

8 Discussion

This study addresses three fundamental questions that crowdshipping-based LMD platforms, with a focus on fair pay, may pose to foster a sustainable crowd for the longevity and expansion of their systems: How can equity be effectively measured? How can the cost benefits be assessed? And what are the potential downsides of using more couriers? The first question tackles the challenge of selecting an appropriate equity measure from the long list of available measures in statistics and economics literature. The second question emphasizes the imperative for LMD platforms, operating under for-profit entities, to consider costs thoughtfully while striving for equity. This necessitates the development of tailored tools capable of evaluating both costs and benefits. Finally, the last question focuses on minimizing a significant negative outcome of equity consideration, which is linked to the notion of opportunity cost. This concept is an inherent part of equity considerations, indicating that the redistribution of opportunities results in gains for some and losses for others. Consequently, a primary negative consequence of equity consideration is the discouragement of individuals, especially high-performing crowdshippers, who experience reduced opportunities. Hence, it is crucial to mitigate this negative impact.

To address these research questions, we developed a practical equity-oriented framework inspired by the real-time operations of a group of crowdshipping platforms. The framework comprises two key elements: (1) a fair crowdshipper selection process based on the principle of first-come, first-served, and (2) an innovative bi-objective optimization process for workload allocation and optimal route selection aimed at balancing equity and cost. This optimization process utilizes an effective heuristic approach known as HybridNsgaII, and its effectiveness was demonstrated through a comprehensive computational analysis.

To tackle the first question, we laid down theoretical groundwork rooted in the concept of Pareto-optimality in multi-objective optimization, identifying which equity measures are not desirable. Subsequently, we outlined a systematic selection process to determine the most reliable equity measures. Our approach is straightforward: if solutions generated under a given equity measure perform well across all measures, then that measure is deemed the most reliable. Implementing this process alongside HybridNsgaII, we found that among established methods such as Range, MAD, SD, CV, and GINI, CV stands out as the most reliable equity measure within our study's scope. Furthermore, GINI emerged as a viable alternative for larger problem instances. However, Range might appeal more to those prioritizing computational speed, boasting an average execution time of 9.2 seconds, compared to 13.21 seconds for CV and 14.43 seconds for GINI.

We explored the tradeoffs between cost and equity from the company's perspective using HybridNsgaII to address our second research question. By exploring various equity budget scenarios that might interest a for-profit enterprise, we unveiled a significant finding: even a slight increase in operational expenses can yield a substantial enhancement in equity. Notably, our study reveals a key managerial insight, showcasing a logarithmic-like relationship between equity improvement and budget allocation. This suggests that modest equity budgets drive the most significant improvements, while larger allocations result in only marginal gains. Through our computational analysis, we underscore that a mere 2.5% increase in the equity budget could potentially boost equity by approximately 40%, whereas a subsequent 7.5% investment by the

company may only yield an additional 30% improvement in equity.

To address the final question, we examined the number of crowdshippers as a key factor in mitigating the negative effects of equity considerations. Our analysis revealed a significant insight: all stakeholders, including low- and high-performing crowdshippers and equity-focused for-profit companies, benefit from minimizing crowdshippers in a first-come, first-served environment. Excess crowdshippers increase company costs and inflate opportunity costs for all drivers. Additionally, regardless of the equity budget, the level of equity decreases with an excessive number of crowdshippers.

References

Allon, G., M. C. Cohen, W. P. Sinchaisri. 2023. The impact of behavioral and economic drivers on gig economy workers. *Manufacturing & Service Operations Management* **25**(4) 1376–1393.

Alnaggar, A., F. Gzara, J. H. Bookbinder. 2019. Crowdsourced delivery: A review of platforms and academic literature. *Omega* 102139.

Archetti, C., M. Savelsbergh, M. G. Speranza. 2016. The vehicle routing problem with occasional drivers. *European Journal of Operational Research* **254**(2) 472–480.

Arslan, A. M., N. Agatz, L. Kroon, R. Zuidwijk. 2019. Crowdsourced delivery—a dynamic pickup and delivery problem with ad hoc drivers. *Transportation Science* **53**(1) 222–235.

Bai, J., K. C. So, C. S. Tang, X. Chen, H. Wang. 2019. Coordinating supply and demand on an on-demand service platform with impatient customers. *Manufacturing & Service Operations Management* **21**(3) 556–570.

Baldacci, R., A. Mingozzi. 2009. A unified exact method for solving different classes of vehicle routing problems. *Mathematical Programming* **120** 347–380.

Behrendt, A., M. Savelsbergh, H. Wang. 2023. A prescriptive machine learning method for courier scheduling on crowdsourced delivery platforms. *Transportation Science* **57**(4) 889–907.

Bertsimas, D., V. F. Farias, N. Trichakis. 2012. On the efficiency-fairness trade-off. *Management Science* **58**(12) 2234–2250.

CapGemini Research Institute. 2019. The last-mile delivery challenge. Tech. rep., CapGemini Research Institute.

Chen, V. X., J. N. Hooker. 2020. Balancing fairness and efficiency in an optimization model. arXiv preprint arXiv:2006.05963.

Clarke, G., J. W. Wright. 1964. Scheduling of vehicles from a central depot to a number of delivery points. *Operations Research* **12**(4) 568–581.

Dayarian, I., J. Pazour. 2022. Crowdsourced order-fulfillment policies using in-store customers. *Production and Operations Management* **31**(11) 4075–4094.

Dayarian, I., M. Savelsbergh. 2020. Crowdshipping and same-day delivery: Employing in-store customers to deliver online orders. *Production and Operations Management* **29**(9) 2153–2174.

Deb, K., A. Pratap, S. Agarwal, T. Meyarivan. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* **6**(2) 182–197.

Dixon, P. M., J. Weiner, T. Mitchell-Olds, R. Woodley. 1987. Bootstrapping the gini coefficient of inequality. *Ecology* **68**(5) 1548–1551.

Fatehi, S., M. R. Wagner. 2022. Crowdsourcing last-mile deliveries. *Manufacturing & Service Operations Management* **24**(2) 791–809.

Gullotta, A., A. Campisano, E. Creaco, C. Modica. 2021. A simplified methodology for optimal location and setting of valves to improve equity in intermittent water distribution systems. *Water Resources Management* **35** 4477–4494.

Ibarra-Rojas, O., Y. Silva-Soto. 2021. Vehicle routing problem considering equity of demand satisfaction. *Optimization Letters* **15**(6) 2275–2297.

Jozefowiez, N., F. Semet, E.-G. Talbi. 2005. Enhancements of NSGA II and its application to the vehicle routing problem with route balancing. *International Conference on Artificial Evolution (Evolution Artificielle)*. Springer, 131–142.

Konstantakopoulos, G. D., S. P. Gayialis, E. P. Kechagias. 2022. Vehicle routing problem and related algorithms for logistics distribution: A literature review and classification. *Operational Research* **22**(3) 2033–2062.

Kwon, Y.-D., J. Choo, B. Kim, I. Yoon, Y. Gwon, S. Min. 2020. Pomo: Policy optimization with multiple optima for reinforcement learning. *Advances in Neural Information Processing Systems* **33** 21188–21198.

Larranaga, P., C. M. H. Kuijpers, R. H. Murga, I. Inza, S. Dizdarevic. 1999. Genetic algorithms for the travelling salesman problem: A review of representations and operators. *Artificial Intelligence Review* **13** 129–170.

Macrina, G., L. D. P. Pugliese, F. Guerriero, D. Laganà. 2017. The vehicle routing problem with occasional drivers and time windows. *International Conference on Optimization and Decision Science*. Springer, 577–587.

Macrina, G., L. D. P. Pugliese, F. Guerriero, G. Laporte. 2020. Crowd-shipping with time windows and transshipment nodes. *Computers & Operations Research* **113** 104806.

Mahmoudinazlou, S., C. Kwon. 2024. A hybrid genetic algorithm with type-aware chromosomes for traveling salesman problems with drone. *European Journal of Operational Research* doi: https://doi.org/10.1016/j.ejor.2024.05.009. URL `https://www.sciencedirect.com/science/article/pii/S0377221724003461`.

Mandell, M. B. 1991. Modelling effectiveness-equity trade-offs in public service delivery systems. *Management Science* **37**(4) 467–482.

Matl, P., R. F. Hartl, T. Vidal. 2018. Workload equity in vehicle routing problems: A survey and analysis. *Transportation Science* **52**(2) 239–260.

Matl, P., R. F. Hartl, T. Vidal. 2019. Workload equity in vehicle routing: The impact of alternative workload resources. *Computers & Operations Research* **110** 116–129.

McCoy, J. H., H. L. Lee. 2014. Using fairness models to improve equity in health delivery fleet management. *Production and Operations Management* **23**(6) 965–977.

Miller, C. E., A. W. Tucker, R. A. Zemlin. 1960. Integer programming formulation of traveling salesman problems. *Journal of the ACM (JACM)* **7**(4) 326–329.

Morabit, M., G. Desaulniers, A. Lodi. 2021. Machine-learning–based column selection for column generation. *Transportation Science* **55**(4) 815–831.

Pal, A., H. Charkhgard. 2019. A feasibility pump and local search based heuristic for bi-objective pure integer linear programming. *INFORMS Journal on Computing* **31**(1) 115–133.

Pessoa, A., R. Sadykov, E. Uchoa, F. Vanderbeck. 2020. A generic exact solver for vehicle routing and related problems. *Mathematical Programming* **183** 483–523.

Raghavan, S., R. Zhang. 2024. The driver-aide problem: Coordinated logistics for last-mile delivery. *Manufacturing & Service Operations Management* **26**(1) 291–311.

Shaw, P. 1998. Using constraint programming and local search methods to solve vehicle routing problems. *International conference on principles and practice of constraint programming*. Springer, 417–431.

Sobhanan, A., J. Park, J. Park, C. Kwon. 2023. Genetic algorithms with neural cost predictor for solving hierarchical vehicle routing problems. arXiv preprint arXiv:2310.14157.

Soto Setzke, D., C. Pflügler, M. Schreieck, S. Fröhlich, M. Wiesche, H. Krcmar. 2017. Matching drivers and transportation requests in crowdsourced delivery systems .

Srivastava, G., A. Singh, R. Mallipeddi. 2021. NSGA-II with objective-specific variation operators for multiobjective vehicle routing problem with time windows. *Expert Systems with Applications* **176** 114779.

Tricoire, F. 2012. Multi-directional local search. *Computers & Operations Research* **39**(12) 3089–3101.

Vidal, T. 2022. Hybrid genetic search for the CVRP: Open-source implementation and swap* neighborhood. *Computers & Operations Research* **140** 105643.

Vidal, T., T. G. Crainic, M. Gendreau, N. Lahrichi, W. Rei. 2012. A hybrid genetic algorithm for multidepot and periodic vehicle routing problems. *Operations Research* **60**(3) 611–624.

Xinying Chen, V., J. Hooker. 2023. A guide to formulating fairness in an optimization model. *Annals of Operations Research* 1–39.

Yu, X.-P., Y.-S. Hu, P. Wu. 2024. The consistent vehicle routing problem considering driver equity and flexible route consistency. *Computers & Industrial Engineering* **187** 109803.

Zhou, L., N. Geng, Z. Jiang, X. Wang. 2018. Multi-objective capacity allocation of hospital wards combining revenue and equity. *Omega* **81** 220–233.

A Mathematical Expressions of the Considered Equity Objective Functions

The equity objective functions considered in this study include:

1. *Range*: This simple measure of inequality is calculated as the difference between the maximum and minimum adjusted profits, reflecting the maximum magnitude by which the workload allocation varies among the crowdshippers. Equitable optimization occurs when the range is 0. Mathematically, Range can be stated as:

$$\texttt{Range}(\boldsymbol{x}) = \max_{m \in \mathcal{M}} \bar{p}_m(\boldsymbol{x}) - \min_{m \in \mathcal{M}} \bar{p}_m(\boldsymbol{x}). \tag{4}$$

2. *Mean Absolute Deviation (MAD)*: This measure evaluates the average absolute difference between each adjusted profit and the mean of adjusted profits. Unlike the range, MAD considers every adjusted profit, offering a more comprehensive measure of equity. Mathematically, MAD can be stated as:

$$\texttt{MAD}(\boldsymbol{x}) = \frac{\sum_{m \in \mathcal{M}} |\bar{p}_m(\boldsymbol{x}) - \bar{\bar{p}}(\boldsymbol{x})|}{M} \tag{5}$$

   where

$$\bar{\bar{p}}(\boldsymbol{x}) := \frac{\sum_{m \in \mathcal{M}} \bar{p}_m(\boldsymbol{x})}{M}$$

   is the mean of adjusted profits.

3. *Standard Deviation (SD)*: This measure stands as one of the widely adopted statistical measures of dispersion, assessing the deviation of individual adjusted profits from their mean. While it offers a more comprehensive gauge of equality compared to the aforementioned measures, decision-makers might refrain from its use due to the additional computational demands it entails or its comparatively greater complexity relative to simpler methods. Mathematically, SD can be stated as:

$$\texttt{SD}(\boldsymbol{x}) = \sqrt{\frac{\sum_{m \in \mathcal{M}} \left(\bar{p}_m(\boldsymbol{x}) - \bar{\bar{p}}(\boldsymbol{x})\right)^2}{M}}. \tag{6}$$

4. *Coefficient of Variation (CV)*: Also known as the relative standard deviation, this measure compares the standard deviation to the mean. Its relative nature makes it more interpretable and intuitive for assessing inequality compared to the SD, which presents outcomes in the original unit. However, unlike the SD, the CV lacks translation invariance. This means that while shifting the entire dataset by a constant does not affect the SD, it does alter the mean, consequently impacting the CV value as well. Mathematically, CV can be stated as:

$$\texttt{CV}(\boldsymbol{x}) = \frac{\texttt{SD}(\boldsymbol{x})}{\bar{\bar{p}}(\boldsymbol{x})}. \tag{7}$$

5. *Gini Coefficient (GINI)*: This measure resembles the CV, albeit with a change in the numerator. The Gini coefficient stands as one of the most renowned measures of inequality,

frequently applied in economics to gauge wealth disparities. Its range spans from zero to a maximum value of one. The Gini index directly correlates with the area between the Lorenz curve and a diagonal line representing perfect equality. In cases of perfect equality, this area diminishes, resulting in an index value of 0. Mathematically, GINI can be stated as:

$$\texttt{GINI}(\boldsymbol{x}) = \frac{\frac{1}{2M^2}\sum_{m\in\mathcal{M}}\sum_{l\in\mathcal{M}}|\bar{p}_m(\boldsymbol{x}) - \bar{p}_l(\boldsymbol{x})|}{\bar{\bar{p}}(\boldsymbol{x})}. \tag{8}$$

When adjusted profits are arranged in ascending order, the expression in (8) simplifies to (9), resulting in quicker computation (Dixon et al., 1987).

$$\texttt{GINI}(\boldsymbol{x}) = \frac{\sum_{l\in\mathcal{M}}(2l - M - 1)\bar{p}_l(\boldsymbol{x})}{M\sum_{m\in\mathcal{M}}\bar{p}_m(\boldsymbol{x})}. \tag{9}$$

B Mathematical Formulation of the Crowdshipper Selection Problem

Let $\overline{\mathcal{M}} := \{1, 2, \ldots, \overline{M}\}$ represent the index set of crowdshippers dynamically joining the system and enrolling for a target zone-wave. This set evolves over time, with each element indicating the position number of the crowdshippers, where, for instance, '1' denotes the first crowdshipper to enroll. We assume that a sufficient number of crowdshippers will register in the system to fulfill orders in any zone-wave. This assumption is not limiting since if there is an insufficient number of crowdshippers, the company can redistribute some orders (in a first-come, first-served manner) to the immediate next wave within the same zone. With this premise, the objective is to determine the minimum number of crowdshippers, denoted by M, needed to fulfill all orders, considering crowdshippers are employed on a first-come, first-served basis. The load capacity of crowdshipper $m \in \overline{\mathcal{M}}$ is denoted by Q_m, and the index set of customers is represented by $\mathcal{N} = \{1, \ldots, N\}$, with q_i indicating the demand size of customer $i \in \mathcal{N}$. The crowdshipper selection problem is formulated as a bin packing problem as follows:

$$M = \min \quad \sum_{m\in\overline{\mathcal{M}}} z_m \tag{10a}$$

$$\text{s.t.} \quad \sum_{m\in\overline{\mathcal{M}}} y_{im} = 1 \qquad \forall i \in \mathcal{N}, \tag{10b}$$

$$\sum_{i\in\mathcal{N}} q_i y_{im} \le Q_m z_m \qquad \forall m \in \overline{\mathcal{M}}, \tag{10c}$$

$$z_{m+1} \le z_m \qquad \forall m \in \overline{\mathcal{M}}\backslash\{\overline{M}\}, \tag{10d}$$

$$y_{im} \in \{0, 1\} \qquad \forall i \in \mathcal{N}, \forall m \in \overline{\mathcal{M}}, \tag{10e}$$

$$z_m \in \{0, 1\} \qquad \forall m \in \overline{\mathcal{M}}, \tag{10f}$$

where y_{im} is a binary variable indicating whether customer i's order is assigned to crowdshipper m or not. Additionally, let z_m be a binary variable indicating whether or not crowdshipper m is selected. The objective function aims to minimize the number of selected crowdshippers.

Constraints (10b) ensure that each order is assigned to a crowdshipper. Constraints (10c) ensure that the capacities of selected crowdshippers are not exceeded. Finally, Constraints (10d) ensure that crowdshippers are selected in a first-come, first-served manner. It is evident that M must lie within the range $[M', n]$, where

$$M' = \min \{m' \in \overline{\mathcal{M}} : \sum_{j \in \mathcal{N}} q_j \leq \sum_{m=1}^{m'} Q_m\}.$$

This implies that in terms of implementation, the company will select all crowdshippers if the number of crowdshippers who have already signed up for the zone-wave is less than or equal to M'. Beyond this threshold, the optimization problem mentioned earlier needs to be solved whenever a new crowdshipper arrives. In cases of infeasibility, the company accepts the current crowdshipper and awaits further crowdshipper sign-ups. The last crowdshipper to accept is the one that renders the problem feasible for the first time.

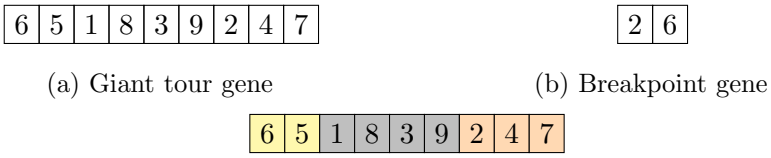
C Detailed Description of the Proposed Method

In this section, we explain the details of our proposed method, HybridNsgaII, designed to efficiently approximate the nondominated frontier, denoted by $\mathcal{Y}_N$, for any instance of the workload allocation problem. This algorithm merges the core principles of the well-known Non-dominated Sorting Genetic Algorithm (NSGA-II) with the concepts of Multi-Directional Local Search (MDLS), providing an effective approach to address the complexities of the workload allocation problem.

C.1 Non-dominated Sorting Genetic Algorithm

Non-dominated Sorting: Let $\mathcal{P}$ represent the set of points corresponding to a given population of initial solutions. Each iteration of the NSGA-II begins with the process of non-dominated sorting to the population. A notion of different fronts, denoted as $\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_k$, based on the dominance relation over a population is used in this method. Here, the points in $\mathcal{Y}_i$ are non-dominating to each other, and each front $\mathcal{Y}_i$ contains at least one point that dominates a point in $\mathcal{Y}_{i+1}$ if $|\mathcal{Y}_{i+1}| \geq 1$. In non-dominated sorting, the set of non-dominated points in $\mathcal{P}$ is allocated to the first front $\mathcal{Y}_1$. Subsequently, the set of non-dominated points in the remaining population is assigned to $\mathcal{Y}_2$, and this process continues until every point is allocated to a particular front.

Crowding Distance: Crowding distance serves as a metric to guide the solution selection process in NSGA-II when the number of required solutions is less than the cardinality of a front. It provides an estimate of the density of solutions surrounding a particular solution in the front. In the selection process using crowding distance, solutions are sorted based on each of their objective values. The crowding distance of a solution is subsequently calculated as the normalized sum of the difference between the objective values of neighboring points along each dimension. During selection, solutions with higher crowding distance values are prioritized. This preference promotes diversification in the next generation, leading to a set of well-distributed



(a) Giant tour gene (b) Breakpoint gene

(c) Customer visit sequence for each crowdshipper in distinct colors

Figure 4: Chromosome solution representation of an instance with $N = 9$ and $M = 3$

solutions in the solution space.

Outline: NSGA-II starts with an initial population $\mathcal{P}_0$ of size $2n_{\text{P}}$, where n_{P} represents the population size for the genetic algorithm. Let $\mathcal{P}_t$ denote the population at iteration t of the algorithm. At each iteration t, $\mathcal{P}_t$ undergoes non-dominated sorting where the population is sorted and rank or fitness levels are assigned to each individual. The rank of a solution is equal to its non-domination or front level. Clearly, $\mathcal{Y}_1$ comprises the best solutions within $\mathcal{P}_t$. Next, an elitist selection procedure forms a subset of the population $\mathcal{P}_t$, denoted as $\mathcal{Q}_t$, with a size of n_{P}. We start with an empty set and the solutions from each front are sequentially added to $\mathcal{Q}_t$ when applicable. For a front k, if the size of $\mathcal{Y}_k$ is less than or equal to $n_{\text{P}} - |\mathcal{Q}_t|$, it is entirely added to $\mathcal{Q}_t$. Otherwise, solutions in $\mathcal{Y}_k$ are sorted based on their crowding distances in descending order and the top $n_{\text{P}} - |\mathcal{Q}_t|$ individuals are selected. The resultant $\mathcal{Q}_t$ of size n_{P} serves as the parent population to generate a set of offspring solutions $\mathcal{R}_t$ of the same size n_{P} using the genetic algorithm. $\mathcal{P}_{t+1} = \mathcal{Q}_t \cup \mathcal{R}_t$ gives the new population for the next iteration of the algorithm. Refer to Deb et al. (2002) for further details on NSGA-II and the algorithm pseudocode. This iterative process is continued until convergence is observed or a desired stopping criteria is reached. More specifically, we stop the algorithm if the set of best solutions remains unchanged after t_{break} consecutive iterations or if the maximum number of iterations, t_{MaxIter}, is reached. Moreover, to escape local optima, if no improvements are observed for $t_{\text{repopulate}}$ consecutive iterations, we repopulate the solution space while retaining the best solutions identified up to that iteration.

C.2 Genetic Algorithm

Our tailored genetic evolution in the HybridNsgaII drives the improvement of solutions by employing iterative processes of parent selection, crossover, and mutations over successive generations based on a fitness score to evaluate the solution quality.

C.2.1 Chromosome Representation

For an instance with N customers and M crowdshippers, a solution consists of M routes that represent the order in which the crowdshippers fulfill the assigned customer demands in a given sequence. To represent such a solution, we use a two-part chromosome encoding: a gene of length N for the giant tour and a gene of length $M - 1$ that provides the breakpoint indices for crowdshippersin the giant tour. Since each crowdshipper needs to be utilized, it is safe to assume that the last crowdshipper visits at least one customer from the end of the giant tour. An example is illustrated in Figure 4. Note that the breakpoint gene contains positive integers in

a strictly increasing order and the last element of the gene must be less than N, which ensures that each crowdshipper serves at least one location.

C.2.2 Initial Solutions

The initialization of the population is a crucial step, where the solutions are either randomly generated or strategically use targeted rules to generate good initial solutions that can expedite the convergence of the methodology. First, we use the Clarke-Wright savings heuristic (Clarke and Wright, 1964) adapted for the heterogeneous capacitated VRP to find a good initial solution. Next, we employ the nearest neighbor algorithm to generate a giant tour, which is a greedy approach with an arbitrary origin and visiting the nearest unvisited location in each step. Here, the breakpoint gene is generated by a heuristic approach based on the average vehicle utilization level $\rho = (\sum_{i \in \mathcal{N}} q_i)/(\sum_{m \in \mathcal{M}} Q_m)$. Customers are assigned sequentially for each crowdshipper m until the expected load size of ρQ_m is attained. In addition to targeted initial solution generation, the inclusion of random solutions plays a crucial role in diversifying the population. The remaining individuals of the initial population are generated through random permutations of the customers with random and increasing breakpoint indices for the giant tour.

C.2.3 Fitness Scores

Each individual in the population is ranked based on a fitness score, which serves as a metric for solution quality. In order to promote diversity in the population and enhance the solution search space, a diversity factor is introduced in the fitness score, which measures the dissimilarity of an individual from its neighbors. The population is sorted based on both objective values, employing lexicographic ascending order to determine the neighbors of each individual. Specifically, the population can be first sorted based on the cost, and in cases where the costs of two solutions are equal, they are further sorted based on equity. The diversity factor $\delta(I)$ for an individual $I \in \mathcal{P}$, defined as (11), is the average normalized Hamming distance of I with its k nearest neighbors.

$$\delta(I) = \frac{1}{k(N+M-1)} \sum_{l=1}^{k} \sum_{i=1}^{N+M-1} \mathbf{1}(I[i] \neq I_l[i]) \tag{11}$$

In (11), $I_l[i]$ is the i-th index element of the chromosome I_l of length $N + M - 1$. Here, $\mathbf{1}(\cdot)$ is an indicator function that is equal to 1 if the condition in $(\cdot)$ is true, and 0 otherwise. For the first and last sorted individual, we set $k = 1$, while for the remaining individuals, $k = 2$.

We estimate three distinct fitness scores for an individual, considering objectives related to cost and equity, alongside the rank of an individual determined through non-dominated sorting.

$$\text{fitness}_{\text{cost}}(I) = C(I) \times \left(1 - \frac{n_{\text{E}}}{n_{\text{P}}}\right)^{\delta(I)}, \tag{12}$$

where $C(I)$ is the total routing cost of individual I and n_{E} is the number of elite individuals desired in the population. Similarly, fitness scores corresponding to equity objective and rank of a solution can be quantified based on the diversity factor.

C.2.4 Infeasible Solutions

Due to the capacity constraints of the problem, a generated solution may be infeasible. In the event of an infeasible solution generation, the candidate undergoes repair with a probability p_r. Otherwise, the infeasible solution is added to a dedicated population for the infeasible solutions. The maintenance of an infeasible population along with the feasible population diversifies the solution space and guides the algorithm towards better solutions by overcoming local optima. The objective values of infeasible solutions are penalized to appropriately integrate infeasible solutions into the selection process. The penalty is determined based on the margin by which the vehicle capacities are violated, weighted by a penalty factor w_p. During the NSGA-II iterations, w_p is adaptively adjusted to limit the maximum ratio of infeasible solutions in the population to a desired level.

C.2.5 Parent Selection, Crossover, and Mutation

Parent selection plays a pivotal element in guiding the current solutions toward optimality by favoring individuals with higher fitness scores. During each step of the genetic algorithm, a crossover is performed with a crossover probability p_c. Two distinct parents are selected using K tournament selection based on the fitness scores to execute the crossover operation. Larger tournaments favor better solutions but to mitigate the risk of premature convergence, we use a binary tournament selection procedure. In addition, an adaptive strategy is incorporated to dynamically determine which fitness score governs the parent selection process. Both the parents are selected using $\text{fitness}_{\text{rank}}(I)$ based on a probability $(1 - \frac{t-t_{\text{PF}}}{t})p_c$, where t_{PF} is the last iteration of NSGA-II when the set of efficient solutions was updated. Otherwise, one parent is chosen based on $\text{fitness}_{\text{cost}}(I)$, and the other parent is determined using $\text{fitness}_{\text{equity}}(I)$ using binary tournament selection.

Many sequence-based crossover operations are documented in the literature. Through numerical experiments, we have adopted three highly effective crossovers, namely partially mapped crossover (PMX), order crossover (OX1), and order-based crossover (OX2) for the giant tour genes. Additional details on the crossover operations are presented in Larranaga et al. (1999). The crossover applies to the giant tour gene of selected parents and transmits the ordering information from parents to their offspring. Then, one of our two crossover strategies is randomly selected to generate the breakpoint genes for the offspring. The first strategy involves directly copying a substring of the breakpoint gene from a parent to an offspring, followed by obtaining the remaining components based on the other parent. In cases where the insertion of substrings from the second parent to the offspring is invalid, i.e., the breakpoint gene components are not strictly increasing, a random integer is generated within the permissible range of values. Similarly, the breakpoint gene of the second offspring is completed by copying a substring from the second parent and utilizing the first parent gene to complete the incomplete information. The second strategy entails sequentially generating the breakpoint gene of an offspring by assigning a random value in the range of the minimum permissible value and the maximum of corresponding parent gene indices.

With a mutation probability p_m, we select a solution I based on its $\text{fitness}_{\text{rank}}(I)$ using binary tournament selection to apply mutation. One of the following five mutations is randomly

applied to I. Displacement mutation selects a random location, followed by its reinsertion to a different random position, and the exchange mutation picks two customers randomly and swaps them. Shuffle mutation (SM) scrambles the ordering of a subtour string of a random crowdshipper visit sequence. We also perform a cyclic shift mutation by a left shift of the giant tour gene from a randomly chosen starting point. The final mutation type involves a random change in a component of the breakpoint gene, ensuring that the chromosome representation remains valid after the alteration.

C.3 Multi-Dimensional Local Search (MDLS)

After each iteration of the algorithm, we perform the multi-dimensional local search (Tricoire, 2012) on the existing set of best solutions for each objective distinctly. Utilizing crowding distance as a metric, we focus on exploring solutions in less crowded regions of the frontier. A solution is chosen for education based on higher crowding distance value. To select individuals for education, one can opt for the top-ordered solutions or apply binary tournament selection based on the crowding distance for diversification. Upon discovering a new best solution, we update the existing best solutions in $\mathcal{Y}_1$. This process can be repeated for all the solutions in $\mathcal{Y}_1$ or a user-defined number of solutions, n_{MDLS}. We execute MDLS using Large Neighborhood Search (Shaw, 1998) and workload reassignment to attempt to improve the selected solution for cost and equity objectives, respectively. If improvement is achieved for either objective, the new solutions obtained are stored in the offspring population $\mathcal{R}_t$, where t is the current iteration. After the education of all selected solutions in an iteration, the set of current best solutions is updated by merging it with $\mathcal{R}_t$ and removing all dominated solutions.

C.4 Parameter Settings

We perform our experiments using commonly recommended parameter values from the recent genetic algorithm literature (Mahmoudinazlou and Kwon, 2024), with minor modifications tailored to suit the problem requirements. The parameters used in our experiments are $n_{\text{P}} = 200$, $n_{\text{E}} = 0.8n_{\text{P}}$, $p_c = 0.9$, $p_m = 0.1$, $p_r = 0.9$, $w_p = 2.0$, $n_{\text{MDLS}} = 0.2|\mathcal{Y}_1|$, $t_{\text{repopulate}} = 50$, $t_{\text{break}} = 100$, and $t_{\text{MaxIter}} = 2000$. Moreover, we dynamically adjust the penalty for infeasible solutions w_p by monitoring the average ratio a of infeasible solutions within the population over the last 10 iterations. When $a > 0.2$, we update w_p as $\max\{1000, 1.2w_p\}$, and if $a < 0.1$, w_p is reduced to $\min\{0.1, 0.8w_p\}$.

D Comparison With Exact Method

We elaborate on the findings presented in Table 2 from Section 7.1, illustrating the performance of both HybridNsgaII and the exact solution across each instance. The results, detailed in Table 7, highlight their speed and capability of the proposed method in generating nearly all nondominated points in each instance.

Table 7: Comparing HybridNsgaII to an Exact Solver on Each Instance Using Range as the Equity Measure

Instance	Exact Solver			HybridNsgaII			HV Gap
	Card.	HV	Time (sec.)	Card.	HV	Time (sec.)	(%)
T1-N8-D2	6	466.05	1.75	6	466.05	4.80	0.00
T2-N8-D2	5	521.74	2.10	5	521.74	4.60	0.00
T3-N8-D2	3	428.14	1.00	3	428.14	3.13	0.00
T4-N8-D2	5	467.30	1.77	5	467.30	3.13	0.00
T5-N8-D2	5	503.41	1.28	5	503.41	3.22	0.00
T6-N8-D2	5	411.27	4.71	5	411.27	3.30	0.00
T7-N8-D2	4	448.04	3.03	4	448.04	3.10	0.00
T8-N8-D2	4	469.12	1.56	4	469.12	5.82	0.00
T9-N8-D2	5	562.91	1.99	5	562.91	3.08	0.00
T10-N8-D2	6	475.20	1.27	6	475.20	3.24	0.00
T1-N8-D3	4	513.74	2.57	4	513.74	3.22	0.00
T2-N8-D3	10	457.67	3.27	10	457.67	4.21	0.00
T3-N8-D3	9	468.94	5.58	9	468.94	6.20	0.00
T4-N8-D3	9	440.71	2.97	9	440.71	3.37	0.00
T5-N8-D3	5	463.33	3.18	5	463.33	4.22	0.00
T6-N8-D3	10	491.44	5.40	10	491.44	3.26	0.00
T7-N8-D3	3	473.56	0.82	3	473.56	3.26	0.00
T8-N8-D3	9	455.39	5.06	9	455.39	3.19	0.00
T9-N8-D3	15	480.26	7.83	15	480.26	4.55	0.00
T10-N8-D3	13	436.17	5.93	13	436.17	3.74	0.00
T1-N10-D2	7	417.18	4.34	7	417.18	3.34	0.00
T2-N10-D2	13	485.47	8.17	13	485.47	3.90	0.00
T3-N10-D2	6	463.49	34.18	6	463.49	3.74	0.00
T4-N10-D2	3	472.88	12.82	3	472.88	3.25	0.00
T5-N10-D2	6	418.22	29.46	6	418.22	3.89	0.00
T6-N10-D2	8	437.10	18.90	8	437.10	6.04	0.00
T7-N10-D2	7	514.13	41.04	7	514.13	3.48	0.00
T8-N10-D2	5	456.26	23.95	5	456.26	3.49	0.00
T9-N10-D2	14	489.71	37.84	14	489.71	5.95	0.00
T10-N10-D2	8	469.74	27.22	8	469.74	4.72	0.00
T1-N10-D3	7	464.94	40.82	7	464.94	5.41	0.00
T2-N10-D3	14	424.11	127.37	14	424.11	3.49	0.00
T3-N10-D3	7	486.07	56.36	7	484.27	5.52	0.37
T4-N10-D3	17	384.88	166.43	17	384.88	6.09	0.00
T5-N10-D3	8	462.35	48.26	9	461.98	4.38	0.08
T6-N10-D3	18	511.57	140.08	13	511.52	3.90	0.01
T7-N10-D3	8	492.40	106.00	8	492.40	5.69	0.00
T8-N10-D3	7	407.16	79.26	7	407.16	5.26	0.00
T9-N10-D3	13	469.39	61.68	13	469.34	6.05	0.01
T10-N10-D3	9	417.40	16.67	9	417.40	6.30	0.00
T1-N12-D2	3	376.76	822.34	3	376.76	3.59	0.00
T2-N12-D2	4	444.06	497.20	3	441.52	3.02	0.58
T3-N12-D2	9	469.23	531.41	10	469.18	4.95	0.01
T4-N12-D2	8	448.29	635.93	8	447.64	5.98	0.15
T5-N12-D2	8	416.02	2,042.81	7	416.02	6.08	0.00
T6-N12-D2	6	452.45	78.54	6	452.45	4.48	0.00
T7-N12-D2	8	430.34	2,557.05	8	430.34	3.99	0.00
T8-N12-D2	4	492.51	508.81	4	492.51	3.94	0.00
T9-N12-D2	5	463.19	1,171.47	5	463.16	5.56	0.01
T10-N12-D2	8	404.97	854.20	8	404.97	6.11	0.00
T1-N12-D3	18	399.64	1,394.93	19	398.97	6.44	0.17
T2-N12-D3	12	429.81	1,270.76	12	429.55	5.11	0.06
T3-N12-D3	17	430.84	1,079.78	16	430.63	6.52	0.05
T4-N12-D3	12	435.09	311.29	10	434.98	6.32	0.03
T5-N12-D3	14	429.85	4,266.00	11	428.41	6.54	0.34
T6-N12-D3	17	378.79	2,275.71	14	378.63	6.45	0.04
T7-N12-D3	14	423.02	357.34	13	422.97	5.51	0.01
T8-N12-D3	14	418.27	2,018.39	12	418.11	6.30	0.04
T9-N12-D3	8	463.07	538.12	8	463.07	6.21	0.00
T10-N12-D3	20	428.69	2,746.69	19	428.61	6.48	0.02

E Single-Solution Mapping Results

Our systematic approach, detailed in Section 6, selects the most reliable equity measure by assessing the complete set of approximate Pareto-optimal solutions generated by each measure via a mapping procedure, and by comparing hypervolume values. This method provides a comprehensive view of the approximate nondominated frontier across various measures, aiding in the identification of the most reliable one. Alternatively, a different approach involves selecting a subset of the approximate Pareto-optimal solutions obtained by each measure and applying the same mapping process described in Section 6 for selection, rather than considering the entire approximate frontier. However, such a subset-based approach raises several questions, such as determining the subset size and the subset selection process, that are not trivial to answer.

These challenges are precisely why in this paper we advocate for a holistic analysis of the complete set of Pareto-optimal solutions found by each measure, rather than relying on a subset. However, to showcase that the effectiveness of CV extends beyond the holistic analysis of the approximate nondominated frontier, we examine a specific case: subset-based approaches with the subset size of one. In this scenario, we select a single Pareto-optimal solution from each measure and compare them using the same mapping procedure detailed in Section 6. An advantage of a subset size of one is the elimination of the need for Hypervolume, as we are not comparing sets. During the mapping procedure, we calculate the gap relative to the measure corresponding to the hosting space. For instance, if we are mapping into the Cost-and-Range criterion space, we evaluate the points based on their Range gaps instead of Hypervolume gaps. These gaps, termed *target-measure gaps*, are computed similar to Hypervolume Gaps, as described in Section 6.

The average target-measure gaps for different instance classes, including $N = 25$, $N = 50$, $N = 75$, and $N = 100$, are provided in Tables 8-11, respectively. To select a solution, we consider five different equity budget scenarios, with $\lambda \in \{0\%, 2.5\%, 5\%, 7.5\%, 10\%\}$. For each scenario, we choose the solution that maximizes equity from the set of approximate Pareto-optimal solutions reported by each measure and then perform the mapping process. Our analysis reveals that CV consistently performs well, ranking as the second-best choice after MAD. As discussed in Section 7.2, MAD performs poorly under full-set mapping assessment. However, we observe here that under single-solution mapping assessment, it excels. Thus, while MAD may lack robustness under full-set mapping, CV maintains robustness under single-solution mapping. Moreover, as mentioned in Section 7.2, GINI's performance improves with larger instance sizes under full-set mapping, a trend observed similarly in single-solution mapping. This further underscores the robustness of results derived from full-set mapping compared to single-solution mapping. In summary, CV demonstrates strong performance across both single-solution and full-set mapping analyses, suggesting its effectiveness across the entire spectrum of mapping assessments.

Table 8: Single-Point Mapping Results for Class $N = 25$: Target-Measure Gaps

(a) $\gamma = 0\%$

Map	Equity Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	0.28	-0.94	0.23	-0.17	-0.12
MAD	-1.17	0.00	-3.8	-0.21	-1.31	-1.30
SD	0.17	0.26	0.00	0.24	0.16	0.17
CV	-0.81	0.17	-3.46	0.00	-1.15	-1.05
GINI	0.00	0.12	-0.26	0.10	0.00	-0.01

(b) $\gamma = 2.5\%$

Map	Equity Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	0.47	-1.64	0.38	-0.29	-0.22
MAD	-1.77	0.00	-6.3	-0.34	-2.29	-2.14
SD	0.42	0.61	0.00	0.57	0.38	0.40
CV	-0.97	0.21	-4.49	0.00	-1.54	-1.36
GINI	0.13	0.50	-0.88	0.43	0.00	0.04

(c) $\gamma = 5\%$

Map	Equity Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	0.60	-1.96	0.49	-0.33	-0.24
MAD	-1.91	0.00	-6.82	-0.35	-2.55	-2.33
SD	0.57	0.79	0.00	0.75	0.50	0.52
CV	-1.09	0.23	-5.27	0.00	-1.81	-1.59
GINI	0.17	0.68	-1.20	0.59	0.00	0.05

(d) $\gamma = 7.5\%$

Map	Equity Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	0.64	-2.23	0.52	-0.43	-0.30
MAD	-1.96	0.00	-7.95	-0.35	-3.05	-2.66
SD	0.61	0.84	0.00	0.80	0.52	0.55
CV	-1.16	0.24	-5.83	0.00	-2.07	-1.76
GINI	0.18	0.73	-1.24	0.63	0.00	0.06

(e) $\gamma = 10\%$

Map	Equity Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	0.62	-2.40	0.51	-0.49	-0.35
MAD	-2.10	0.00	-8.31	-0.38	-3.15	-2.79
SD	0.60	0.84	0.00	0.80	0.53	0.55
CV	-1.14	0.25	-5.94	0.00	-2.06	-1.78
GINI	0.20	0.71	-1.27	0.62	0.00	0.05

Table 9: Single-Point Mapping Results for Class $N = 50$: Target-Measure Gaps

(a) $\gamma = 0\%$

Map	Equity Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	0.14	-0.25	0.12	-0.01	0.00
MAD	-0.65	0.00	-1.86	-0.08	-0.74	-0.67
SD	0.09	0.17	0.00	0.16	0.10	0.10
CV	-0.48	0.06	-1.22	0.00	-0.44	-0.42
GINI	-0.03	0.10	-0.16	0.09	0.00	0.00

(b) $\gamma = 2.5\%$

Map	Equity Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	0.41	-0.83	0.36	-0.09	-0.03
MAD	-2.75	0.00	-6.31	-0.36	-2.50	-2.38
SD	0.29	0.55	0.00	0.51	0.33	0.34
CV	-1.66	0.23	-4.14	0.00	-1.52	-1.42
GINI	-0.06	0.58	-0.90	0.51	0.00	0.03

(c) $\gamma = 5\%$

Map	Equity Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	0.53	-1.16	0.47	-0.16	-0.06
MAD	-3.36	0.00	-7.67	-0.46	-3.12	-2.92
SD	0.39	0.73	0.00	0.69	0.43	0.45
CV	-1.96	0.28	-4.99	0.00	-1.89	-1.71
GINI	-0.11	0.61	-0.96	0.51	0.00	0.01

(d) $\gamma = 7.5\%$

Map	Equity Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	0.64	-1.51	0.57	-0.27	-0.11
MAD	-3.70	0.00	-8.14	-0.49	-3.38	-3.14
SD	0.42	0.77	0.00	0.73	0.46	0.48
CV	-1.96	0.27	-5.26	0.00	-2.07	-1.80
GINI	-0.04	0.69	-0.98	0.60	0.00	0.05

(e) $\gamma = 10\%$

Map	Equity Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	0.69	-1.56	0.61	-0.27	-0.11
MAD	-3.62	0.00	-8.40	-0.47	-3.49	-3.20
SD	0.47	0.83	0.00	0.79	0.48	0.51
CV	-1.89	0.26	-5.45	0.00	-2.13	-1.84
GINI	-0.10	0.74	-1.05	0.63	0.00	0.04

Table 10: Single-Point Mapping Results for Class $N = 75$: Target-Measure Gaps

(a) $\gamma = 0\%$

Map	Equity Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	0.14	-0.18	0.13	0.00	0.02
MAD	-0.56	0.00	-1.12	-0.06	-0.46	-0.44
SD	0.00	0.00	0.00	0.00	0.00	0.00
CV	-0.63	0.06	-1.23	0.00	-0.49	-0.46
GINI	-0.07	0.22	-0.32	0.19	0.00	0.00

(b) $\gamma = 2.5\%$

Map	Equity Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	0.47	-0.61	0.42	0.00	0.06
MAD	-2.28	0.00	-4.45	-0.26	-1.88	-1.77
SD	0.10	0.22	0.00	0.20	0.12	0.13
CV	-1.76	0.20	-3.58	0.00	-1.39	-1.31
GINI	-0.15	0.42	-0.60	0.36	0.00	0.01

(c) $\gamma = 5\%$

Map	Equity Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	0.61	-1.09	0.55	-0.14	-0.01
MAD	-3.51	0.00	-6.75	-0.39	-2.91	-2.71
SD	0.28	0.56	0.00	0.53	0.32	0.34
CV	-2.17	0.23	-4.74	0.00	-1.99	-1.73
GINI	-0.14	0.60	-0.84	0.52	0.00	0.03

(d) $\gamma = 7.5\%$

Map	Equity Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	0.60	-1.19	0.55	-0.19	-0.05
MAD	-4.20	0.00	-7.41	-0.47	-3.17	-3.05
SD	0.34	0.74	0.00	0.70	0.42	0.44
CV	-2.28	0.25	-4.92	0.00	-2.03	-1.80
GINI	-0.23	0.70	-0.97	0.60	0.00	0.02

(e) $\gamma = 10\%$

Map	Equity Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	0.62	-1.07	0.57	-0.14	0.00
MAD	-4.73	0.00	-7.91	-0.54	-3.31	-3.30
SD	0.38	0.78	0.00	0.74	0.44	0.47
CV	-2.39	0.25	-4.92	0.00	-2.06	-1.82
GINI	-0.27	0.74	-1.11	0.63	0.00	0.00

Table 11: Single-Point Mapping Results for Class $N = 100$: Target-Measure Gaps

(a) $\gamma = 0\%$

Map	Equity Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	0.14	-0.18	0.13	0.00	0.02
MAD	-0.71	0.00	-1.00	-0.07	-0.44	-0.44
SD	0.01	0.04	0.00	0.04	0.02	0.02
CV	-0.53	0.05	-0.96	0.00	-0.38	-0.36
GINI	-0.10	0.23	-0.31	0.20	0.00	0.00

(b) $\gamma = 2.5\%$

Map	Equity Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	0.36	-0.36	0.33	0.04	0.07
MAD	-2.69	0.00	-4.00	-0.27	-1.73	-1.74
SD	0.01	0.12	0.00	0.11	0.07	0.06
CV	-1.60	0.16	-2.57	0.00	-1.01	-1.00
GINI	-0.20	0.42	-0.55	0.36	0.00	0.01

(c) $\gamma = 5\%$

Map	Equity Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	0.38	-0.49	0.35	-0.01	0.05
MAD	-3.99	0.00	-5.47	-0.41	-2.29	-2.43
SD	0.12	0.33	0.00	0.31	0.19	0.19
CV	-2.07	0.21	-3.50	0.00	-1.40	-1.35
GINI	-0.36	0.56	-0.75	0.47	0.00	-0.02

(d) $\gamma = 7.5\%$

Map	Equity Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	0.65	-0.73	0.59	0.03	0.11
MAD	-4.58	0.00	-6.27	-0.48	-2.62	-2.79
SD	0.18	0.55	0.00	0.51	0.31	0.31
CV	-2.53	0.23	-4.34	0.00	-1.82	-1.69
GINI	-0.44	0.66	-0.94	0.55	0.00	-0.03

(e) $\gamma = 10\%$

Map	Equity Gap (%) using					Average
From\To	Range	MAD	SD	CV	GINI	
Range	0.00	0.63	-0.85	0.58	-0.04	0.06
MAD	-5.34	0.00	-6.95	-0.58	-2.81	-3.14
SD	0.22	0.59	0.00	0.56	0.34	0.34
CV	-2.68	0.25	-4.39	0.00	-1.81	-1.73
GINI	-0.55	0.70	-1.05	0.58	0.00	-0.06