THE UNIVERSITY OF CHICAGO

ESSAYS ON MANAGING RESOURCES IN THE SHARING ECONOMY

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE UNIVERSITY OF CHICAGO
BOOTH SCHOOL OF BUSINESS
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
TAHSIN DENIZ AKTURK

CHICAGO, ILLINOIS
JUNE 2023

To my family

CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

The journey of completing a Ph.D. degree at the University of Chicago Booth School of Business has been long and enlightening. I am very grateful to those who have accompanied and supported me in this process.

First, I would like to express my deepest gratitude to my advisors, Prof. Ozan Candogan and Prof. Varun Gupta, who has been there since my Ph.D. journey. I am very fortunate to work with them, and this thesis is possible due to their extended support, guidance, and understanding. I am grateful to Prof. Candogan for always guiding me to be a better researcher and showing me what excellence is in research. I am also grateful to Prof. Gupta for the vast amount of time he spent sharing his knowledge and teaching me advanced concepts, which not only helped me form this thesis but will guide me in my future research. Finally, I want to thank both of them for helping my career and being patient while I was still learning to be rigorous in research.

I also want to thank my dissertation committee members Prof. John Birge, Prof. Amy Ward, Prof. Linwei Xin and Prof. Yuan Zhong for their continuous support. Prof. Birge and Prof. Xin were always generous in providing valuable insight, which was very helpful in shaping my papers. Prof. Ward and Prof. Zhong were very supportive of my academic applications during my job market, which I am grateful for. I also want to thank Prof. Zhong for all of my TA'ing in Booth, where I was lucky to learn from an excellent teacher.

I am grateful to the Booth Ph.D. staff Malaina Brown, Amity James, Cynthia Hillman, and Kimberly Mayer for all their assistance with administrative work, classes, reimbursements, and visa applications throughout the years.

Studying at Booth was only possible through the support of many in Turkey. First, I want to thank my professors at the Industrial Engineering department at Bilkent University. I am fortunate for the support of Prof. Cagin Ararat, Prof. Oya Karasan, and Prof. Hande Yaman, who motivated me during my applications and my Ph.D. I also want to thank my friends Haluk Akay, Okan Atalar, Mert Aydin, Prof. Suleyman Kerimov, and Ayaz Zamanov,

who have always been there for me and supported me to be better.

The best times at Chicago Booth were with my OM friends, who always made me feel at home. I want to thank Nasser Barjesteh, Kevin Chen and Cem Randa for their warm welcome and know-how, Charlie Hannigan and Gizem Yilmaz for being the best classmates, and Amir Alwan, Zuguang Gao, Emin Ozyoruk, Gulin Tuzcuoglu, Gorkem Unlu, and Naz Yetimoglu for making my years at Booth memorable.

I want to thank my closest friend Yang Wu, who had to support me during my most challenging times. However, meeting her made the challenges achievable.

Most importantly, I want to take my loving family. I want to thank my brother for always being there for me, my mother for always filling the family with love, and my father, for being my first and forever teacher and advisor. Starting from my grandmother, the sacrifices and opportunities provided by my family members made it possible for me to attend Booth and complete this dissertation.

ABSTRACT

The dissertation mainly focuses on the spatial imbalance of resources observed in the sharing economy. It consists of two main chapters, with the first chapter focusing on designing policies to solve this imbalance and the second chapter focusing on evaluating the performance of easy-to-implement policies on solving this imbalance of resources. In the third chapter, we discuss possible extensions and other applications for the methodology we provided.

In the first chapter, we consider the problem of managing resources in shared micro-mobility systems (bike-sharing and scooter-sharing). An important task in managing such systems is periodic repositioning/recharging/sourcing units to avoid stockouts or excess inventory at nodes with unbalanced flows. We consider a discrete-time model: each period begins with an initial inventory at each node in the network, and then customers (demand) materialize at the nodes. Each customer picks up a unit at the origin node and drops it off at a randomly sampled destination node with an origin-specific probability distribution. We model the above network inventory management problem as an infinite horizon discrete-time discounted Markov Decision Process and prove the asymptotic optimality of a novel mean-field approximation to the original MDP as the number of stations becomes large. To compute an approximately optimal policy for the mean-field dynamics, we provide an algorithm with a running time that is logarithmic in the desired optimality gap. Lastly, we compare the performance of our mean-field-based policy to state-of-the-art heuristics via numerical experiments, including experiments using Austin scooter-sharing data.

The second chapter considers the joint optimization of rebalancing/sourcing inventory on a graph. We focus on the lost-sales setting with customer-induced relocations. Through a coupling analysis, we provide worst-case performance bounds, with tight instances, for policies commonly used in practice. We provide further insights into the performance of these policies and discuss cost regimes where they are effective.

CHAPTER 1

MANAGING RESOURCES FOR SHARED MICROMOBILITY: APPROXIMATE OPTIMALITY IN LARGE-SCALE SYSTEMS

1.1 Introduction

In recent years, customer interest in shared micromobility systems (bike-sharing and scooter-sharing) has surged, prompting these systems to expand their service regions and the number of customers they serve. With the increase in customer demand and the number of service stations, the problem of imbalances in inventory accumulation has become even more prevalent. Stemming from the uncertainty of demand and correlation in customer preferences, rental systems which allow flexibility in where customers drop off the rented vehicles observe excessive unit accumulation in some stations and lack of available units in others (Gossett [2021]). Facing such a problem, the system operator has to regularly reposition units to avoid negative externalities arising from this imbalance. In specific systems such as scooter-sharing, the operator is also tasked with regularly recharging units (in conjunction with repositioning decisions) to maintain the availability of units to customers. Motivated by this issue, in the present paper we look at a discrete-time inventory management problem on a network where the system operator uses two levers to optimize the operations *(i)* repositioning/recharging units across nodes of the network, and *(ii)* adding/removing units in circulation to dynamically adjust the fleet size.

Making the right rebalancing decisions involves a trade-off as both rebalancing and recharging are costly, requiring physical labor and inducing other transportation costs. Furthermore, computing the optimal repositioning/recharging policy is a difficult problem as the size of the state space of the Markov Decision Process (MDP) grows exponentially in the number of nodes, and the curse of dimensionality limits our ability to handle large scale instances encountered in practice. To get around this curse of dimensionality, in this paper, we will work with an asymptotic mean-field setting to gain insights into the design of good inventory rebalancing/recharging policies. The mean-field model removes the stochasticity in the problem by replacing the detailed state space where we track the inventory of each station, which are random variables, with the empirical distribution of the inventory among station, which is deterministic in the many-station limit.

Existing work on rebalancing policies relies on the convexity of the value function for the MDP, which only holds under restrictive assumptions on cost parameters and other structural assumptions such as linear rebalancing costs. While helpful in the derivation and presentation of results, such assumptions present difficulties and limitations for implementation into real-life applications. Furthermore, computational methods proposed in similar problems are generally applicable for a few nodes. This paper specifically looks at large networks to obtain asymptotically optimal policies.

We make weak assumptions on the cost parameters and allow less restrictive system mechanics. We prove that the optimal policy for the mean-field dynamics corresponding to the original MDP provides an asymptotically optimal policy as the number of stations grows. We also provide an algorithm to compute a near-optimal policy for the mean-field system whose optimality gap decreases exponentially in the computational effort.

There are several important real-life applications of the problem setting discussed. While our results directly apply to any setting where customers rent products and control where to return them, this paper will primarily focus on applying our results to shared micromobility systems rapidly growing in popularity. While both scooter-sharing and bike-sharing have been introduced to urban settings to be a sustainable complement to public transportation systems, the differences in the characteristics of the two services provide individual challenges in managing the resources of each system. Most notably, a scooter-sharing system requires both regular recharging and repositioning. In contrast, a bike-sharing system requires regular repositioning and occasional maintenance of units[1]. To provide a unified framework to

1. Under this split of the two micromobility systems, e-bikes share similar characteristics to scooter-sharing systems.

address both challenges, we present a model that provides a unified view to study the two systems.

We organize the paper as follows: In Section 1.2, we present the model through which we aim to capture the first-order dynamics of the problem setting we introduced.

In Section 1.3, we introduce the asymptotic scaling leading to the mean-field dynamics. We prove that the optimal policy obtained through solving the deterministic mean-field model is asymptotically optimal for the original problem. Next, we move on to the mean-field model in Section 1.4, where we provide an algorithm to find a near optimal control policy for the mean-field dynamics. We do this by solving for the optimal steady state of the mean-field system while taking into account the minimum cost transient path to reach this state, given an initial inventory position and a transient period duration. This necessitates a further restatement of the mean-field problem in a compact space for technical reasons. We prove that the optimality gap of our algorithm (compared to the optimal infinite horizon discounted cost for the mean-field system) is exponentially decreasing with respect to the duration of the transient period. We also briefly discuss how our results are different from, and add to, the literature on deterministic dynamic programming. Then, in Section 1.5, we compare the numerical performance of our algorithm with some simple heuristics and some more sophisticated benchmarks. We evaluate the performance of our policy through both synthetic experiments and using the Austin scooter-sharing data-set.

Lastly, in Section 1.6, we discuss extensions such as heterogeneous costs (where different stations have different penalty and holding costs), settings with seasonal/cyclical demand patterns, heterogeneous depletion probabilities, separately modeling damaged/missing units, travel times, and fixed costs. Such extensions are important in the applications we consider, and we show that our results extend to these settings.

1.1.1 Literature Review

Inventory rebalancing has been a primary area of research with various problems in transshipment and fleet assignment. While there is a long line of important results in inventory rebalancing, most existing results do not cover features such as reusable resources. The context of our results is different from the following two papers which are the most similar in terms of the system dynamics they study (Benjaafar et al. [2022a], He et al. [2020]) as we work in a large-scale setting, integrate recharging and rebalancing decisions, work with a more general cost structure which includes the fleet-sizing aspect of the problem in addition to rebalancing, and our results readily extend to more general settings including non-stationarity and fixed costs.

Inventory Management and Transshipment

One of our related applications is container rentals, where transshipment companies rent containers in docks to carry their units. Van Mieghem and Rudi [2002] extend the newsvendor problem into a multi-period, multi station setting and provides properties on the structure of the optimal policy under certain conditions. Abouee-Mehrizi et al. [2015] prove that the optimal policy at a station for a two station system has a threshold structure with the values of the thresholds depending on the inventory position of the other station. An important difference between this line of work and ours is that in the aforementioned papers, units are consumed at stations and depart the network, whereas in our case customers move units between stations, making the network structure play an important role.

Fleet Assignment

Another closely related line of research is on fleet assignment problems, which look at the assignment problem of a set of units to a set of customers (based on different attributes of customers such as destination and price). Topaloglu and Powell [2006] provide an iterative value function approximation scheme for large scale problems. Adelman [2007] provides a pricing based approach for the fleet assignment problem. The distinction between this line of work of ours is the "assignment" of units, wherein the fleet assignment literature, the central planner "assigns" available units based on the destination. This is different from our case, where the customers pick up the available units in a first-come-first-serve manner, the system randomly moves to a new inventory position through customer movements, and we try to "fix" this inventory position through rebalancing.

Bike-sharing

Operations of bike-sharing systems have seen increased research activity of late, specifically relating to fleet-sizing and rebalancing problems. Benjaafar et al. [2022b] focus on the fleet-sizing problem where they optimize over the total number of units in circulation, given a service guarantee or revenue objective. Shu et al. [2013] look at the repositioning problem in order to maximize customer utilization without incorporating rebalancing costs. O'Mahony and Shmoys [2015] focus on developing repositioning heuristics for rebalancing (for a fixed number of units). We consider operational details associated with the penalty, rebalancing, and holding costs in our work. Our approach can be viewed as joint fleet sizing and rebalancing, which is in contrast to these papers, which focus one of the two.

Scooter-sharing

There has been a limited amount of work looking at problems faced in managing scooter-sharing and other mobility systems. Osorio et al. [2021] provide a discrete inventory routing model to streamline rebalancing/recharging operations and studies the performance of this approach through numerical examples. Greening and Erera [2021] provide simulation models for the vehicle distribution problem in scooter-sharing systems. Most works in this area focus on providing good heuristics, which is different from our approach of developing algorithms with theoretical guarantees.

Pricing

Another related line of research looks at the control of shared vehicle systems through pricing, using fluid models to understand the performance of different policies. Bimpikis et al. [2019] look at the effect of spatial price discrimination on platform profits. Banerjee et al. [2022] prove a worst-case performance bound for a fluid-based policy under a steady-state objective and prove asymptotic optimality of their policy when the number of units in circulation goes to infinity, while the number of stations grows sublinearly in the number of units. Lastly Balseiro et al. [2021] use a Lagrangian based methodology and focuses on hub-and-spoke networks. These works focus on solving unit availability through pricing alone, whereas we focus on unit availability through inventory repositioning. Furthermore, most asymptotic convergence results (such as in Balseiro et al. [2021]) focus on an average optimality criteria, where our result holds for a discounted problem where we optimize over both the transient period and the steady state.

Rebalancing

A feature common to most of the works cited in the previous paragraph is that the system operator decides to allocate units/set prices based on the origin *and* destination of the customer. Many recent papers study inventory systems where the central planner does not differentiate which customers receive units based on the destination, as in our problem. Zhao et al. [2020] look at the characterization of the optimal policy for a two-region setting with fixed costs. Hosseini et al. [2022] use a closed queuing network problem to construct a fluid approximation and uses the shadow prices to make rebalancing decisions. He et al. [2020] provide a distributionally robust optimization model under limited data and shows its solid numerical performance for a large-scale problem. Yang et al. [2022] differ from the other papers (including ours) where it looks at the repositioning problem through ex ante decisions, where the central-planner makes a decision for the target inventory position every period and rebalances after demand realization to that inventory period. The paper proves that a steady state policy is optimal for an average cost objective. Lastly, Benjaafar et al. [2022a] prove important structural properties of the value function of the proposed repositioning problem and the associated optimal policies, as well as proposes a cutting plane based algorithm to find policies with good performance. The work of He et al. [2020] and Benjaafar et al. [2022a] are perhaps the most relevant to the current paper. This being said, unlike these two papers, our work also considers the fleet sizing problem and the rebalancing problem (the number of units in circulation is a decision variable, whereas it is assumed to be constant in the other two papers). We believe this is an important issue to consider, especially considering real life examples such as Citi Bike in New York, which has a varying number of units in circulation from 7000 to 9000 (see Jay [2018]). Compared to these papers, which assume all units returned are re-usable, our paper also integrates possible recharging decisions. Furthermore, both papers provide results under the assumption of convexity of the value function with respect to the inventory position, which requires an assumption on the cost parameters. In contrast, we do not make any such assumption. Our results are, thus, applicable to broader settings where the cost of rebalancing is high and the assumption of convexity fails. Both papers also look at continuous approximations of the problem where they allow non-discrete inventory at stations while not considering the additional cost incurred through rounding. With a large number of stations, we believe the inventory at each station is not large enough to neglect rounding-based costs and prove a bound for the rounding based costs we observe through using the mean-field model. Lastly, we prove that our work can be extended to settings with fixed costs for repositioning, a setting very important for the applications we discussed.

1.1.2 Notation

We use $\mathbb{N}$ to denote the set of strictly positive integers, $\mathbb{N}_0$ to denote the set of nonnegative integers, and $\mathbb{R}_+$ to denote the set of nonnegative real numbers. We use $[n]$ to denote the set $\{1, \ldots, n\}$ for $n \in \mathbb{N}$. Pois(μ) denotes a Poisson random variable with mean μ. We use

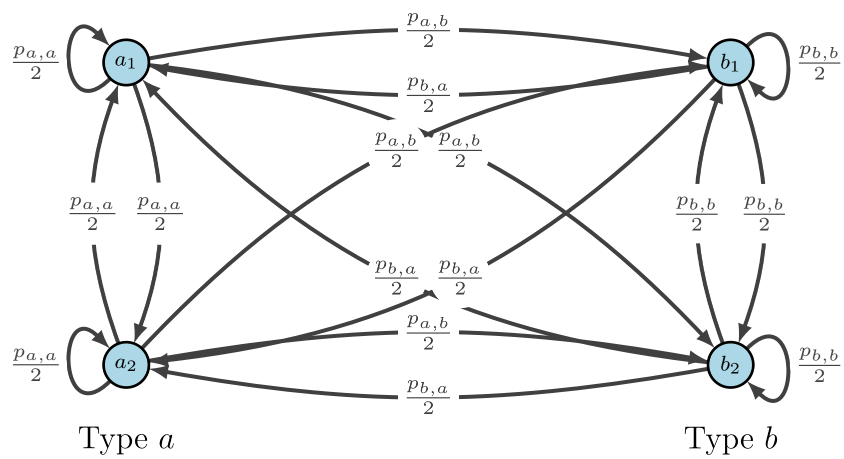


Figure 1.1: Types Network

$\mathcal{O}$ for the asymptotic big-O notation: $f(n) = \mathcal{O}(n)$ implies that $\exists M > 0, n_0 \in \mathbb{R}_+$ such that $\forall n \geq n_o$, $\left|f(n)\right| \leq Mn$. Similarly, we use $\Omega(n)$ for big-Omega notation: $f(n) = \Omega(n)$ implies that $\exists M > 0, n_0 \in \mathbb{R}_+$ such that $\forall n \geq n_o$, $f(n) \geq Mn$. The expression $(x)^+$ is shorthand for $\max(x, 0)$. The expression $\mathcal{L}(X)$ denotes the law of random variable X. The expression $\mathbb{P}\left[X = d\right]$ denotes the probability that X equals d. The notation $f \circ g(x)$ denotes function composition $f(g(x))$. Bold type characters denote vector quantities. Lastly, a summary table of the symbols used in the paper is provided in Section 1.8.

1.2 Problem Definition

The paper aims to develop near-optimal rebalancing/sourcing/recharging policies for units distributed across a network, where customers pickup/drop-off units at nodes (stations) and arcs represent these units' movement through the stations. We model the problem as an infinite horizon, discrete-time, discounted cost MDP with a discrete state and action space. We let γ denote the discount rate, with $\gamma < 1$. We use $t \in \mathbb{N}$ to index the periods. Each period consists of two phases.

First, we have a rebalancing/sourcing/recharging phase where the central planner recharges units (if needed), moves units between stations, or moves units from/to a warehouse, with infinite capacity, to/from the network nodes.

In the second phase, demand materializes at the network nodes, customers pick up units subject to availability, and return them at (randomly sampled) destination nodes in either a re-usable or a depleted state[2] based on the outcome of a Bernoulli process.

The standard approach to modeling such problems includes aggregating multiple stations at a service region as a single station to reduce the state space of the problem. To expand on this structure and allow variations in the inventory of stations in the same region, we consider a network structure where stations are grouped into clusters. We also assume that all stations in the same cluster are homogeneous, meaning they have identical demand distributions, cost parameters, and thresholds (we will expand more on these terms later)[3]. Motivated by the homogeneity in these different dimensions, in the remainder of the paper we refer to these clusters as types.

While the technical motivation behind these homogeneity assumptions becomes clear from the mean-field formulation in Section 1.3, the managerial reasoning for this assumption is that it allows us to focus on the flow of units between neighborhoods[4]. Imposing stronger assumptions on properties of stations within a type allows us to model flow of units between types more accurately, with less restrictive assumptions (while also being able to prove our results). In addition, arguably, local-level inaccuracies are less costly to fix as local-level rebalancing is cheaper than rebalancing between distant locations, which motivates us to conduct our analysis under this assumption.

We let $\hat{e} < \infty$ be the total number of types. We use n_e to denote the number of stations of type e, and n as the total number of stations: $\sum_{e=1}^{\hat{e}} n_e = n$. We consider a heterogeneous network between types where the likelihood of customers departing a type e' station to drop

2. For ease of exposition, we assume that depletion covers all events (unit has no charge remaining, is broken) which result in the unit being unusable without intervention by the planner. In Section 1.6, we look at an extension of our model, which differentiates between such events.

3. In Section 1.5, we will use clustering algorithms to form these types.

4. For most practical applications, stations in the same cluster are closely located.

their unit at a type e station is captured by weight $w_{e',e}$, with $0 \leq w_{e',e} < \infty$. Through these weights, the probability $p_{e',e}$ that a customer departing a type e' station drops their unit at a type e station is calculated as:

$$p_{e',e} = \frac{n_e w_{e',e}}{\sum_{e_2=1}^{\hat{e}} w_{e',e_2} n_{e_2}}.$$

Furthermore, we assume that all stations of a type have identical inflow probabilities so the probability that the unit goes to a specific station i in type e is $\frac{p_{e',e}}{n_e}$.

The dynamics are as follows: At the start of each period t, the central planner observes the state, which corresponds to inventory position $\boldsymbol{x}^t = [x^t_{e,r}, x^t_{e,1}, \cdots, x^t_{e,n_e}]_{e=1}^{\hat{e}}$, where $x^t_{e,r}$ denotes the number of depleted units in type e[5] and $x^t_{e,i}$ denotes the number of units at station i of type e. We label the set of all inventory positions $\boldsymbol{x}^t$ as $\boldsymbol{X}$ where $\boldsymbol{X} = \mathbb{N}_0^{n+\hat{e}}$. Then, the planner takes an action, defined as a vector $\boldsymbol{a}^t = [a^t_{e,r}, a^t_{e,1}, \cdots, a^t_{e,n_e}]_{e=1}^{\hat{e}}$, whose elements define the inventory position post-rebalancing/recharging/sourcing. Specifically, $a^t_{e,r}$ denotes the number of depleted units in type e and $a_i^{t,e}, i \in [n_e]$, denotes the inventory position of station i of type e, after the planner's action.

Next, customers arrive, pick up units (one unit per customer subject to availability), and depart to their destinations. Three stochastic outcomes are realized, one arising from the arrival process of customers, one arising from the random choice of destinations of the customers departing stations with a unit, and lastly, one from a random Bernoulli process on whether the dropped unit is re-usable or requires recharging. We express the number of customers arriving to each station in period t through the vector $\boldsymbol{D}^t = [D^t_{e,1}, ..., D^t_{e,n_e}]_{e=1}^{\hat{e}}$. Referred to as (*demand*), we assume that the distribution of demand at each station of the same type is identical with $\mathcal{L}(D^e) = \mathcal{L}(D^t_{e,i}) \, \forall i \in [n_e]$. Furthermore, as units are returned, some become depleted (unusable) with probability $1 - q$ (we extend our model

5. The aggregation of the depleted units in each type does not imply a separate location where customers return depleted units but assumes that the exact location customers return depleted units in a type is redundant. We further discuss this, and other assumptions, in Section 1.2.2.

to location heterogeneous depletion probabilities in Section 1.6). To calculate the number of units returned to each station (and the number of units returned depleted to each type), we let $R^t_{e',e,r}$ denote the total inflow of depleted units to type e from type e' and let $R^t_{e',e,i}$ denote the number of units departing type e' and returned to station i of type e charged. Together, we call the vector of returns originating from each type e' as *trips* where $\boldsymbol{R}^t_{e'} = [\boldsymbol{R}^t_{e',1}, \boldsymbol{R}^t_{e',2}, \cdots, \boldsymbol{R}^t_{e',\hat{e}}]$, with $\boldsymbol{R}^t_{e',e} = [R^t_{e',e,r}, R^t_{e',e,1}, \cdots, R^t_{e',e,n_e}]$. $\boldsymbol{R}^t_{e'}$ follows a multinomial distribution with $\sum_{i=1}^{n_{e'}} \min(a^t_{e',i}, D^t_{e',i})$ trials and success probability $(1-q)p_{e',e}$ for $R^t_{e',e,r}$ and success probabilities $\frac{qp_{e',e}}{n_e}$ for the remaining terms $R^t_{e',e,1}, \cdots, R^t_{e',e,n_e}$. All travel times are assumed to be one unit so that all units are returned within the current period. Thus, the state at time $t+1$ is given by the following equations:

$$x^{t+1}_{e,i} = a^t_{e,i} - \min(a^t_{e,i}, D^t_{e,i}) + \sum_{e'=1}^{\hat{e}} R^t_{e',e,i} \qquad \forall i \in [n_e], \tag{1.2.1}$$

$$x^{t+1}_{e,r} = a^t_{e,r} + \sum_{e'=1}^{\hat{e}} R^t_{e',e,r} \quad \forall e. \tag{1.2.2}$$

We assume that $\{D^e, D^{t,e}_i\} \in \mathbb{N}_0 \, \forall i \in [n_e], e, t$, is a sequence of independent random variables with finite mean μ_e and finite variance. All demand distributions and transition probabilities are assumed to be known and stationary. Furthermore, we assume that the action for charged units is between a lower bound $\boldsymbol{L} = [L^1, L^2, \cdots, L^{\hat{e}}]$ and an upper bound $\boldsymbol{U} = [U^1, U^2, \cdots, U^{\hat{e}}]$, and the action for depleted units is between a lower bound $\boldsymbol{L}_r = [n_1 L^{r,1}, n_2 L^{r,2}, \cdots, n_{\hat{e}} L^{r,\hat{e}}]$ and an upper bound $\boldsymbol{U}_r = [n_1 U^{r,1}, n_2 U^{r,2}, \cdots, n_{\hat{e}} U^{r,\hat{e}}]$. That is,

$$\begin{aligned} &0 \le L^{r,e} \le \frac{a^t_{e,r}}{n_e} \le U^{r,e} < \infty && \forall e, \\ &0 \le L^e \le a^t_{e,i} \le U^e < \infty && \forall e, i \in [n_e]. \end{aligned}$$

We denote the set of all possible actions $\boldsymbol{a}^t$ as $\boldsymbol{A}$.

As previously discussed, the planner makes multiple decisions in the first phase, where they recharge depleted units, move units between stations, and move units between stations and the warehouse. To simplify notation, we will embed recharging and sourcing decisions into rebalancing decisions using dummy types. To do so, we model recharging as rebalancing units from depleted state to charged state and sourcing as rebalancing units from the warehouse to stations[6]. Specifically, we let $e = 0$ denote the warehouse (where $n_0 = 1, c_{0,0} = 0$) and introduce $e_2 = \hat{e} + e_1$ as an auxiliary type whose units will capture the number of depleted units in type e_1 (with $n_e = 1 \, \forall e > \hat{e}, c_{e,e} = 0 \, \forall e > \hat{e}$). Different c_{e_1,e_2} correspond to the cost of different type of interventions taken by the planner depending on whether either of e_1, e_2 is equal to 0, lies in $\{0, \cdots, \hat{e}\}$ or in $\{\hat{e}+1, \cdots, 2\hat{e}\}$. For example, $c_{e,0}$ for $e > \hat{e}$ corresponds to the cost of moving a depleted unit from type $e - \hat{e}$ to the warehouse. For the values which these parameters can take, we will only assume that

$$c_{e_1,e_2} \geq \frac{c_{e_1,e_1}}{2} + \frac{c_{e_2,e_2}}{2} \quad \forall e_1, e_2.$$

The assumption above ensures that it costs more to move units between two different types than within one of the two of those types. Expressly, we do not assume symmetry where c_{e_1,e_2} can have a different value than c_{e_2,e_1}. The rebalancing cost function $c(\boldsymbol{x}^t, \boldsymbol{a}^t)$, which gives us the minimum cost to move to action $\boldsymbol{a}^t$ from initial inventory position $\boldsymbol{x}^t$, is expressed as

6. In the remainder of the paper, all rebalancing/sourcing/recharging actions taken by the planner will be referred to as rebalancing for simplicity.

follows:

$$
\begin{aligned}
c(\boldsymbol{x}^t, \boldsymbol{a}^t) = \min_{y_{i,j}^{e_1,e_2}} & \sum_{e_1=0}^{2\hat{e}} \sum_{e_2=0}^{2\hat{e}} \sum_{i=1}^{n_{e_1}} \sum_{j=1}^{n_{e_2}} c_{e_1,e_2} y_{i,j}^{e_1,e_2} \\
\text{s.t.} \quad & x_{e,i}^t - a_{e,i}^t = \sum_{e_1=0}^{2\hat{e}} \sum_{j=1}^{n_{e_1}} (y_{i,j}^{e,e_1} - y_{j,i}^{e_1,e}) && \forall e \in [\hat{e}], i, \\
& x_{e-\hat{e},r}^t - a_{e-\hat{e},r}^t = \sum_{e_1=0}^{2\hat{e}} \sum_{j=1}^{n_{e_1}} (y_{i,j}^{e,e_1} - y_{j,i}^{e_1,e}) && \forall e > \hat{e}, i, \\
& y_{i,j}^{e_1,e_2} \geq 0 && \forall e_1, e_2, i, j.
\end{aligned}
$$

While the definition of this cost function is natural, it requires us to solve for the optimal inflow/outflow at each station. The following proposition establishes that the rebalancing cost function can be expressed through a simpler structure composed of the distance between action and initial inventory position at each station and a linear program solving for the aggregate minimum cost flow of units between types. This modification reduces the number of decision variables in the optimization program providing the rebalancing cost.

Proposition 1.2.1. *For all $\boldsymbol{x}^t, \boldsymbol{a}^t$, the rebalancing cost $c(\boldsymbol{x}^t, \boldsymbol{a}^t)$ can be alternatively expressed as:*

$$
\begin{aligned}
c(\boldsymbol{x}^t, \boldsymbol{a}^t) = & \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{i=1}^{n_e} \left| a_{e,i}^t - x_{e,i}^t \right| + \min_{y^{e_1,e_2}} \sum_{e_1=0}^{2\hat{e}} \sum_{e_2=0, e_2 \neq e_1}^{2\hat{e}} \left(c_{e_1,e_2} - \frac{c_{e_1,e_1}}{2} - \frac{c_{e_2,e_2}}{2} \right) y^{e_1,e_2} \\
s.t. \quad & \sum_{i=1}^{n_e} (x_{e,i}^t - a_{e,i}^t) = \sum_{e_1=0}^{2\hat{e}} (y^{e,e_1} - y^{e_1,e}) \qquad \forall e \in [\hat{e}], \\
& x_{e,r}^t - a_{e,r}^t = \sum_{e_1=0}^{2\hat{e}} (y^{e+\hat{e},e_1} - y^{e_1,e+\hat{e}}) \qquad \forall e \in [\hat{e}], \\
& y^{e_1,e_2} \geq 0 \qquad \forall e_1, e_2.
\end{aligned}
$$

Here, we label $\sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{i=1}^{n_e} \left| a_{e,i}^t - x_{e,i}^t \right|$ as the L_1 component, as it evaluates the ℓ_1 distance between the vectors $\boldsymbol{x}, \boldsymbol{a}$. We let z^* denote the optimal value of the optimization program in Proposition 1.2.1[7], and label this as the flow component, as it solves for the minimum cost flow between types which provides the post-rebalancing action $\boldsymbol{a}$.

Our model aims to capture both the rebalancing and fleet sizing aspects of the problem by including holding costs. Units in circulation can depreciate and lose some of their value. To capture this, after rebalancing, a cost of c_h is incurred for each unit in circulation (i.e., excluding the units at the warehouse), for an aggregate holding cost of $c_h \sum_{e=1}^{\hat{e}} a_{e,r}^t + c_h \sum_{e=1}^{\hat{e}} \sum_{i=1}^{n_e} a_{e,i}^t$. Lastly, c_p is the penalty cost incurred per customer who does not find a unit available. Letting $\boldsymbol{d}^t = [d_{e,1}^t, ..., d_{e,n_e}^t]_{e=1}^{\hat{e}}$ be demand realization at period t, a total penalty cost of $c_p \sum_{e=1}^{\hat{e}} \sum_{i=1}^{n_e} (d_{e,i}^t - a_{e,i}^t)^+$ is incurred for that period. We combine the holding and penalty costs into the *newsvendor cost* function:

$$N(\boldsymbol{a}^t) = c_h \sum_{e=1}^{\hat{e}} \sum_{i=1}^{n_e} a_{e,i}^t + c_h \sum_{e=1}^{\hat{e}} a_{e,r}^t + c_p \sum_{e=1}^{\hat{e}} \sum_{i=1}^{n_e} \mathbb{E}\left[(D^e - a_{e,i}^t)^+\right]. \tag{1.2.3}$$

The total cost at period t is the sum of the newsvendor and rebalancing cost functions.

1.2.1 Policies and the Value Function

A policy $\boldsymbol{\pi} = \{\pi^t\}_{t \in \mathbb{N}}$, where $\pi^t : \mathbb{N}_0^{n+\hat{e}} \to \boldsymbol{A}$, is a collection of mappings (indexed by period t) from $\boldsymbol{X}$ to $\boldsymbol{A}$. We label the set of all policies $\boldsymbol{\pi}$ as $\boldsymbol{\Pi}$. Given a policy $\boldsymbol{\pi}$, we denote by $V_{\boldsymbol{\pi}}(\boldsymbol{x}^t)$, the discounted cost-to-go under policy $\boldsymbol{\pi}$, starting with inventory position $\boldsymbol{x}^t$. Formally:

$$V_{\boldsymbol{\pi}}(\boldsymbol{x}^t) = \mathbb{E}_{\{\boldsymbol{D}^{t+k}, \boldsymbol{R}^{t+k}\}_{k=0}^{\infty}} \left[\sum_{s=t}^{\infty} \gamma^{s-t} \Big(c(\boldsymbol{x}^s, \pi^s(\boldsymbol{x}^s)) + N(\pi^s(\boldsymbol{x}^s)) \Big) \right].$$

The value function is defined as

$$V(\boldsymbol{x}^t) = \min_{\boldsymbol{\pi} \in \boldsymbol{\Pi}} V_{\boldsymbol{\pi}}(\boldsymbol{x}^t).$$

7. with $c(\boldsymbol{x}^t, \boldsymbol{a}^t) = \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{i=1}^{n_e} \left| a_{e,i}^t - x_{e,i}^t \right| + z^*$.

The value function can be alternatively expressed through the fixed point of the Bellman recursion:

$$V(\boldsymbol{x}^t) = \min_{\boldsymbol{a}^t \in \boldsymbol{A}} c(\boldsymbol{x}^t, \boldsymbol{a}^t) + N(\boldsymbol{a}^t) + \gamma \mathbb{E}_{\{\boldsymbol{D}^t, \boldsymbol{R}^t\}} \Big[V(\boldsymbol{x}^{t+1}) \Big].$$

For any inventory position $\boldsymbol{x}^t$ and period t, an optimal policy π^* satisfies $\pi^*(\boldsymbol{x}^t) = \boldsymbol{a}^{t^*}$ with $\boldsymbol{a}^{t^*}$ given by

$$\boldsymbol{a}^{t^*} \in \arg\min_{\boldsymbol{a}^t \in \boldsymbol{A}} c(\boldsymbol{x}^t, \boldsymbol{a}^t) + N(\boldsymbol{a}^t) + \gamma \mathbb{E}_{\{\boldsymbol{D}^t, \boldsymbol{R}^t\}} \Big[V(\boldsymbol{x}^{t+1}) \Big].$$

1.2.2 Relevance of Model to Micromobility Systems

In the formulation of the mathematical model, which aims to capture the first order properties of managing micromobility systems, we make several assumptions. Some of these assumptions, such as stationarity and the homogeneity of the holding and penalty cost parameters, are for ease of exposition and are relaxed in Section 1.6. In this section, we discuss the remaining assumptions and argue that while they were mainly motivated by tractability concerns, they are also well-aligned with the practice of resource management in micromobility systems[8].

One important assumption pertains to depleted units: We aggregate the total number of depleted units at a type instead of the exact number of depleted units at each station. This is partially motivated by the fact that in practice, the cost of the recharging operation does not differ for different stations of a type. For instance, in recharging operations for scooter systems a contractor picks up a unit, takes the unit home and charges it, and brings the unit to an indicated location before a certain deadline (see, e.g., Helling [2022]). While the payment the contractor receives depends on multiple factors, there is no secondary differentiation on the exact station from which the unit is picked up as all stations of the same type are geographically close and recharging is not done on the spot but rather an another location. In the context of our model this renders it unnecessary to keep track of

8. We relegate discussion on various other aspects of the model to Section 1.10 to keep the exposition focused in this section.

the exact stations of the depleted units, and allows us to instead focus on depleted units in a type.

Second, we consider a network of stations. This is a natural representation for station-based systems but can be at first viewed as being incompatible with free-floating systems[9]. We believe that free-floating systems can also be represented through stations. First, the feasible regions for picking up/dropping units are restricted due to city ordinances (units can only be dropped in places that will not disturb movement). Second, as highlighted in Helling [2022], scooter-sharing companies such as Bird have specific spots in each neighborhood (called Bird Nests) where contractors must drop units. As a result, in these free-floating systems, many units can only be picked up in specific locations in the morning. Then, we can consider the close vicinity of these nests as stations and utilize our station-based analysis for both station-based and free-floating systems.

Finally, it is assumed that there are no spillovers of demand between stations: If a customers arrives at a station and cannot find a unit there, he/she does not move to other stations to pickup a unit. While no empirical analysis calculating the rate of spillovers (percentage of customers which move to other stations due to stockouts) has been done for scooter-sharing systems, Kabra et al. [2020] has shown that for the London bike-sharing system, "only 5.070% (±0.538%) of a stocked-out stations unserved users substitute to other stations."

1.3 Mean-Field Model

The model defined in Section 1.2, which we will refer to as the original model in the remainder of the paper, suffers from curse of dimensionality since the size of the state space is exponential in the number of stations. When n is large, we can more succinctly approximate the evolution of the state through mean-field dynamics where the state represents the

9. A free-floating system is one where customers can drop units in any part of a side-walk, which is in contrast to a station-based system where customers have to drop units at docks.

empirical distribution of inventory across the nodes, and state transitions are deterministic. We will first construct the mean-field model, then introduce the asymptotic scaling used, and lastly, prove the asymptotic optimality of the mean-field optimal policy for the original problem.

We proceed by defining the mean-field quantities analogous to the variables we introduced in Section 1.2 (state, action, policies, and value functions). The mean-field empirical representation for inventory of type e at period t is denoted by $\hat{\boldsymbol{x}}_e^t = [\hat{x}_{e,r}^t, \hat{x}_{e,d}^t]_{d=0}^{\infty}$, where $\hat{x}_{e,d}^t$ denotes the proportion of type-e stations that have inventory position d at period t, and $\hat{x}_{e,r}^t$ denotes the scaled number of depleted units in type e at period t. The scaling for $\hat{x}_{e,r}^t$ is done by dividing the total number of depleted units in type e by the total number of stations, n. The aggregate vector of inventory of all types is given by:

$$\hat{\boldsymbol{x}}^t = [\hat{\boldsymbol{x}}_e^t]_{e=1}^{\hat{e}},$$

where components of $\hat{\boldsymbol{x}}^t$ satisfy:

$$\begin{aligned}
\hat{x}_{e,d}^t &\geq 0 && \forall e, d, && \text{(non-negativity)} \\
\sum_{d=0}^{\infty} \hat{x}_{e,d}^t &= \frac{n_e}{n} && \forall e, && \text{(fraction of stations of type } e \text{ equals } \frac{n_e}{n}\text{)} \\
\hat{x}_{e,r}^t &\in \mathbb{R}_+ && \forall e. && \text{(number of depleted units is non-negative)}
\end{aligned}$$

We label the set of all feasible $\hat{\boldsymbol{x}}^t$ vectors as $\hat{\boldsymbol{X}}$.

Similarly, for actions, we define $\hat{\boldsymbol{a}}_e^t = [\hat{a}_{e,r}^t, \hat{a}_{e,d}^t]_{d=0}^{\infty}$, where $\hat{a}_{e,d}^t$ denotes the proportion of type-e stations that have d units after rebalancing at period t, and $\hat{a}_{e,r}^t$ denotes the scaled number of depleted units in type e, post-rebalancing at period t. The aggregate vector of action of all types is given by:

$$\hat{\boldsymbol{a}}^t = [\hat{\boldsymbol{a}}_e^t]_{e=1}^{\hat{e}},$$

where components of $\hat{\boldsymbol{a}}^t$ satisfy:

$$\hat{a}^t_{e,d} \geq 0 \qquad \forall e, d, \qquad \text{(non-negativity)}$$

$$\hat{a}^t_{e,d} = 0 \qquad \forall e, d \notin \{L^e, \dots, U^e\}, \qquad \text{(threshold constraint on stations)}$$

$$\sum_{d=L^e}^{U^e} \hat{a}^t_{e,d} = \frac{n_e}{n} \qquad \forall e, \qquad \text{(fraction of type } e \text{ stations equals } \frac{n_e}{n}\text{)}$$

$$\hat{a}^t_{e,r} \in [L^{r,e}, U^{r,e}]. \qquad \text{(threshold constraint on depleted units)}$$

We label the set of all feasible $\hat{\boldsymbol{a}}^t$ values as $\hat{\boldsymbol{A}}$.

Under the mean-field model, we construct deterministic approximations for the demand and trip processes. Given action $\hat{\boldsymbol{a}}^t$ for period t, the state at period $t+1$ under the mean-field model is obtained through the following equations:

$$\hat{x}^{t+1}_{e,d} = \sum_{b=L^e}^{U^e} \hat{a}^t_{e,b} \mathbb{P}\left[b - \min(b, D^e) + \hat{R}^e = d\right] \qquad \forall e, d, \tag{1.3.1}$$

$$\hat{R}^e \stackrel{d}{=} \text{Pois}\left(q \frac{n}{n_e} \sum_{e'=1}^{\hat{e}} p_{e',e} \sum_{b=L^{e'}}^{U^{e'}} \hat{a}^t_{e',b} \mathbb{E}\left[\min(b, D^{e'})\right]\right) \qquad \forall e, \tag{1.3.2}$$

$$\hat{x}^{t+1}_{e,r} = \hat{a}^t_{e,r} + (1-q) \sum_{e'=1}^{\hat{e}} p_{e',e} \sum_{b=L^{e'}}^{U^{e'}} \bar{a}^t_{e',b} \mathbb{E}\left[\min(b, D^{e'})\right] \qquad \forall e. \tag{1.3.3}$$

Remark 1.3.1. *For the special case of $q = 0$, the above equations reduce to:*

$$\hat{x}^{t+1}_{e,d} = \sum_{b=L^e}^{U^e} \hat{a}^t_{e,b} \mathbb{P}\left[b - \min(b, D^e) = d\right] \qquad \forall e, d,$$

$$\hat{x}^{t+1}_{e,r} = \hat{a}^t_{e,r} + \sum_{e'=1}^{\hat{e}} p_{e',e} \sum_{b=L^{e'}}^{U^{e'}} \bar{a}^t_{e',b} \mathbb{E}\left[\min(b, D^{e'})\right] \qquad \forall e.$$

An important property of mean-field approximation compared to other fluid limits is that we can use the exact demand distribution D^e at each type e compared to using just its mean, which gives higher fidelity to the mean-field model. While we exactly capture the demand distributions, we approximate the trip processes to simplify the dynamics and model state transitions through independent processes.

First, as trips are multinomially distributed, the total number of arrivals to each station is binomial, which can be approximated with a Poisson distribution. While the sum of the total number of arrivals to all stations is a constant in the original system (equal to the total number of departures), our approximation treats the total arrivals at each station as an independent Poisson random variable. As a result, the reason $\hat{R}^e$ is distributed Poisson is independent of the demand distribution but instead because Poisson is a suitable approximation for the multinomial distribution. Then, to form the state transition function of this deterministic system, given $\hat{a}^t_{e,b}$ values, we first calculate the probability that stations of type e with b units will end up with d units. As we are working with a mean-field model, these probabilities are equal to the proportion of stations which move to d units from b units (with type e), and multiplied with $\hat{a}^t_{e,b}$, gives us the proportion of stations which had b units previously but will end up with d units. Repeating this process for all possible e, b, d combinations provides Equation 1.3.1. For depleted units, because we can aggregate the total number of depleted units in a type, we approximate their inflows in the mean-field setting through their expected inflows in the stochastic setting, giving us Equation 1.3.3.

Under the mean-field model, for rebalancing, we will evaluate the least amount of rebalancing to move from the current inventory distribution to the target inventory distribution. This metric is the earth mover's distance or the 1st Wasserstein distance. For any two distributions y, z with a non-negative and discrete domain, we express the (1st) Wasserstein distance as:

$$W(y, z) = \sum_{x=0}^{\infty} |F_y(x) - F_z(x)|,$$

where F is the cumulative mass function.

We next introduce the rebalancing cost function for the mean-field model:

$$\begin{aligned}
\hat{c}(\hat{\boldsymbol{x}}^t, \hat{\boldsymbol{a}}^t) =& n \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{b=0}^{\infty} \left| \sum_{d=0}^{b} (\hat{x}_{e,d}^t - \hat{a}_{e,d}^t) \right| \\
& + n \min_{y^{e_1,e_2}} \sum_{e_1=0}^{2\hat{e}} \sum_{e_2=0, e_2 \neq e_1}^{2\hat{e}} \left(c_{e_1,e_2} - \frac{c_{e_1,e_1}}{2} - \frac{c_{e_2,e_2}}{2} \right) y^{e_1,e_2} \\
\text{s.t.} \quad & \sum_{b=1}^{\infty} b(\hat{x}_{e,b}^t - \hat{a}_{e,b}^t) = \sum_{e_1=0}^{2\hat{e}} (y^{e,e_1} - y^{e_1,e}) && \forall e \in [\hat{e}], \\
& \hat{x}_{e,r}^t - \hat{a}_{e,r}^t = \sum_{e_1=0}^{2\hat{e}} (y^{e+\hat{e},e_1} - y^{e_1,e+\hat{e}}) && \forall e \in [\hat{e}], \\
& y^{e_1,e_2} \geq 0 && \forall e_1, e_2.
\end{aligned}$$

Similar to the original model formulation, we label $n \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{b=0}^{\infty} \left| \sum_{d=0}^{b} (\hat{x}_{e,d}^t - \hat{a}_{e,d}^t) \right|$ as the *Wasserstein component*, as it evaluates the Wasserstein distance between the vectors $\hat{\boldsymbol{x}}, \hat{\boldsymbol{a}}$. We let $\hat{z}^*$ denote the optimal value of second component of the optimization program above[10], and label this as the *flow component*, as it solves for the minimum cost flow between types which provides the post-rebalancing action $\boldsymbol{a}$. The flow component is very similar to the flow component of the original model, with the difference of using mean-field model variables instead. As for the Wasserstein component, we prove in Section 1.11 that the L_1 component in Proposition 1.2.1 can be represented through the Wasserstein component of the mean-field model.

Next, we introduce the newsvendor cost function for the mean-field model

$$\hat{N}(\hat{\boldsymbol{a}}^t) = nc_h \sum_{e=1}^{\hat{e}} \sum_{b=L^e}^{U^e} b\hat{a}_{e,b}^t + nc_h \sum_{e=1}^{\hat{e}} \hat{a}_{e,r}^t + nc_p \sum_{e=1}^{\hat{e}} \sum_{b=L^e}^{U^e} \hat{a}_{e,b}^t \mathbb{E}\left[(D^e - b)^+\right]. \tag{1.3.4}$$

This is very similar to the newsvendor cost function of the original model, with the only difference being that action variables of the original model are replaced with their mean-field

10. with $\hat{c}(\hat{\boldsymbol{x}}^t, \hat{\boldsymbol{a}}^t) = n \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{b=0}^{\infty} \left| \sum_{d=0}^{b} (\hat{x}_{e,d}^t - \hat{a}_{e,d}^t) \right| + \hat{z}^*$.

counterparts.

Lastly, we define a new class of policies for the mean-field model. Similar to the previous section, we introduce $\hat{\boldsymbol{\pi}} = \{\hat{\pi}^t\}_{t\in\mathbb{N}}$, where $\hat{\pi}^t : \hat{\boldsymbol{X}} \to \hat{\boldsymbol{A}}$. We label the set of all policies $\hat{\boldsymbol{\pi}}$ as $\hat{\boldsymbol{\Pi}}$. Given a policy $\hat{\boldsymbol{\pi}}$, we denote by $\hat{V}_{\hat{\boldsymbol{\pi}}}(\hat{\boldsymbol{x}}^t)$, the discounted cost-to-go under policy $\hat{\boldsymbol{\pi}}$, starting with inventory position $\hat{\boldsymbol{x}}^t$, under the mean-field model. Formally:

$$\hat{V}_{\hat{\boldsymbol{\pi}}}(\hat{\boldsymbol{x}}^t) = \sum_{s=t}^{\infty} \gamma^{s-t}\Big(\hat{c}(\hat{\boldsymbol{x}}^s, \hat{\pi}^s(\hat{\boldsymbol{x}}^s)) + \hat{N}(\hat{\pi}^s(\hat{\boldsymbol{x}}^s))\Big).$$

As in Section 1.2, the optimal value function can be defined as the minimizer of the expression above:

$$\hat{V}(\hat{\boldsymbol{x}}^t) = \min_{\hat{\boldsymbol{\pi}}\in\hat{\boldsymbol{\Pi}}} \hat{V}_{\hat{\boldsymbol{\pi}}}(\hat{\boldsymbol{x}}^t),$$

or as the fixed point of the Bellman recursion:

$$\hat{V}(\hat{\boldsymbol{x}}^t) = \min_{\hat{\boldsymbol{a}}^t\in\hat{\boldsymbol{A}}} \hat{c}^n(\hat{\boldsymbol{x}}^t, \hat{\boldsymbol{a}}^t) + \hat{N}(\hat{\boldsymbol{a}}^t) + \gamma\hat{V}(\hat{\boldsymbol{x}}^{t+1}).$$

Using the fact that the mean-field model is deterministic, an equivalent characterization can be given in terms of the following deterministic infinite-dimensional minimization problem:

$$\hat{V}(\hat{\boldsymbol{x}}^t) = \min_{\{\hat{\boldsymbol{a}}^s\in\hat{\boldsymbol{A}}\}_{s=t}^{\infty}} \sum_{s=t}^{\infty} \gamma^{s-t}\Big(\hat{c}(\hat{\boldsymbol{x}}^s, \hat{\boldsymbol{a}}^s) + \hat{N}(\hat{\boldsymbol{a}}^s)\Big).$$

For any inventory position $\hat{\boldsymbol{x}}^t$ and period t, the optimal policy of the mean-field model, $\hat{\pi}^*$, satisfies

$$\hat{\boldsymbol{a}}^{t^*} \in \arg\min_{\hat{\boldsymbol{a}}^t\in\hat{\boldsymbol{A}}} \hat{c}(\hat{\boldsymbol{x}}^t, \hat{\boldsymbol{a}}^t) + \hat{N}(\hat{\boldsymbol{a}}^t) + \gamma\hat{V}(\hat{\boldsymbol{x}}^{t+1}).$$

1.3.1 Lifting the Mean-field Policy to the Stochastic Setting

Until now, we defined a separate model, referred to as the mean-field model, without connecting it to the original model. Our goal is to solve the mean-field model, obtain a policy, and use it to define a policy for the original model. To do that, we need to first introduce mappings between the two models. We first introduce $g(\boldsymbol{x}^t)$, where g projects inventory positions in $\boldsymbol{X}$ (the original stochastic model) to inventory positions in $\hat{\boldsymbol{X}}$ (deterministic mean-field model) with

$$
\begin{aligned}
g_{e,d}^{t}(\boldsymbol{x}^{t}) &= \frac{\sum_{i=1}^{n_e} \mathbb{I}\left\{x_{e,i}^{t} = d\right\}}{n} && \forall e, d, \\
g_{e,r}^{t}(\boldsymbol{x}^{t}) &= \frac{x_{e,r}^{t}}{n} && \forall e.
\end{aligned}
$$

Second, per our goal of using the mean-field system to obtain feasible policies for the original stochastic model, we also need to project the optimal actions of the mean-field model to feasible actions for the original model. However due to the structure of $\hat{\boldsymbol{A}}$, $n\hat{a}_{e,d}^{t}$, which denotes the number of type-e stations with inventory d units after rebalancing at period t, need not be an integer and consequently may not be a feasible action for the original model. To address this issue we introduce $\bar{\boldsymbol{a}}_e^t = [\bar{a}_{e,r}^t, \bar{a}_{e,d}^t]_{d=0}^{\infty}$ and define $\bar{\boldsymbol{A}} \subset \hat{\boldsymbol{A}}$ as the subset of mean-field action space that is both feasible for the original model and dependent on n, with:

$$
\bar{\boldsymbol{a}}^t = [\bar{\boldsymbol{a}}_e^t]_{e=1}^{\hat{e}} \in \bar{\boldsymbol{A}} \iff \bar{\boldsymbol{a}}^t \in \hat{\boldsymbol{A}};\; n\bar{a}_{e,d}^t \in \mathbb{N}_0 \;\forall e, d;\; n\bar{a}_{e,r}^t \in \mathbb{N}_0 \,\forall e.
$$

We define function f as a mapping from $\hat{\boldsymbol{A}}$ to $\bar{\boldsymbol{A}}$, where, for depleted units, the function f simply rounds $n\hat{a}_{e,r}^t$ to the nearest integer value. For charged units, f maps units through the Largest Remainder Method. Under this method, applied separately for each type e, given a set of fractional $n\hat{a}_{e,d}^t$ values, we first allocate the whole parts to each corresponding $n\bar{a}_{e,d}^t$ variable. Then, we compute the amount of units remaining to sum up to n_e, and allocate these units one by one, in the order of decreasing fractional parts[11].

Consequently, the composite policy $f \circ \hat{\pi}^*$ gives actions in $\bar{\boldsymbol{A}}$. Finally, we need a function which maps these projected actions to feasible actions for the original model. Since $\bar{\boldsymbol{a}}^t$ only specifies the empirical distribution of inventory among stations, there are possibly multiple actions $\boldsymbol{a}^t$ for the original stochastic system consistent with the empirical distribution. To

11. Under the Largest Remainder Method, Birkhoff [1976] proves that $f(\hat{\boldsymbol{a}}) \in \arg\min_{\bar{\boldsymbol{a}} \in \bar{\boldsymbol{A}}} \sum_{e=1}^{\hat{e}} \sum_{b=L^e}^{U^e} \sum_{d=0}^{b} \left| n\hat{a}_{e,d} - n\bar{a}_{e,d} \right| + \sum_{e=1}^{\hat{e}} \left| n\hat{a}_{e,r} - n\bar{a}_{e,r} \right|$.

select among these actions, we adopt the natural approach of choosing the action which minimizes the ℓ_1 distance. Specifically, we use the function $h(\bar{\boldsymbol{a}})$ defined below to project actions in $\bar{\boldsymbol{A}}$ to actions in $\boldsymbol{A}$ using the ℓ_1 norm (we suppress the dependence on $\boldsymbol{x}$ for ease of notation):

$$
\begin{aligned}
h(\bar{\boldsymbol{a}}) \; & \in \arg\min_{\boldsymbol{a}\in\boldsymbol{A}} \sum_{e=1}^{\hat{e}} \sum_{i=1}^{n_e} \left| a_{e,i}^t - x_{e,i}^t \right| + \sum_{e=1}^{\hat{e}} \left| a_{e,r}^t - x_{e,r}^t \right| && (1.3.5) \\
\text{s.t.} \quad & \frac{\sum_{i=1}^{n_e} \mathbb{I}\left\{ a_{e,i}^t = d \right\}}{n} = \bar{a}_{e,d}^t && \forall e, d, \\
& a_{e,r}^t = n\bar{a}_{e,r}^t && \forall e.
\end{aligned}
$$

To solve the matching/optimization problem in (1.3.5), we introduce a new ordering for stations, where stations are ordered in their post-rebalancing inventory. We assign the station with the highest pre-rebalancing inventory of a type e, the highest post-rebalancing inventory of the same type e, assign the station with the second highest pre-rebalancing inventory of a type e, the second highest post-rebalancing inventory of the same type e and so on. We repeat this assignment for all types and, through this assignment, solve for the minimizing action for (1.3.5) (Bobkov and Ledoux [2019], Lemma 4.1).

In summary the g, h and f functions are used to obtain a feasible policy through the mean-field model as follows: Given an initial inventory position $\boldsymbol{x}^t$ for the stochastic system, the mean-field model observes the projected inventory position $g(\boldsymbol{x}^t)$ and gives the optimal mean-field action $\hat{\pi}^* \circ g(\boldsymbol{x}^t)$. This is projected to the original model first by discretizing via f and then lifting to $\boldsymbol{A}$ via h as $h \circ f \circ \hat{\pi}^* \circ g(\boldsymbol{x}^t)$.

1.3.2 Constructing the Asymptotic Scaling

In order to construct the asymptotic scaling, we look at a sequence of systems indexed by the number of stations n. The $(n+1)$'st system is constructed by adding to the n'th system one additional station with type e' (out of the possible $\hat{e}$ distinct types), an initial inventory

(inventory at $t = 1$) of $m \in \mathbb{Z}_0$ units; as well as adding $k \in \mathbb{Z}_0$ depleted units to type e'. Other than the assumption that the number of types $\hat{e}$ is finite, we do not make any assumptions on how these values are generated. The parameters of the $(n+1)$'st system can therefore be expressed in terms of the parameters of the n'th system as follows[12]:

$$\begin{aligned}
x^{1,n+1}_{e',n_e+1} &= m, \\
x^{1,n+1}_{e',r} &= x^{1,n}_{e',r} + k, \\
n^{n+1}_{e'} &= n^{n}_{e'} + 1, \\
n^{n+1}_{e} &= n^{n}_{e} && \forall e \neq e', \\
p^{n+1}_{e,e'} &= \frac{(n^{n}_{e'} + 1) w_{e,e'}}{\sum_{e_2=1}^{\hat{e}} w_{e,e_2} n^{n+1}_{e_2}} && \forall e.
\end{aligned}$$

Under this scaling, we will see that the optimality gap of using the mean-field optimal policy on the original model grows sub-linearly with each added station, in contrast to the total cost under an optimal policy, which grows linearly.

Remark 1.3.2. *Readers familiar with mean-field analysis will realize that the asymptotic scaling we defined above differs from the usual convention where the initial conditions of the sequence of systems are assumed to converge uniformly to a limiting distribution, and this limiting distribution is used to define the mean-field dynamics. The difference is that in our work, we will study the optimality gap in the pre-limit for the n'th system by defining the mean-field model using the initial conditions of the n'th system itself. For example, the fraction of stations of type e in the mean-field model used to approximate the n'th system will be $\frac{n^n_{e'}}{n}$, rather than a limiting distribution P^e, where P^e denotes the proportion of the number of stations of each type. In addition to proving an optimality gap which is practically more useful, this methodology ensures that the action $g(\boldsymbol{a}^t)$ is feasible for the mean-field model, allowing us to establish the $\mathcal{O}(\sqrt{n})$ sub-optimality of the mean-field optimal policy*

12. To make the asymptotic scaling clearer, we add n as a superscript to indicate the dependence of parameters to n.

(see Proposition 1.3.3). The assumption that our actions are bounded ensures that any irregularities in the initial inventory distribution disappear after $t = 1$.

1.3.3 Asymptotic Optimality of the Mean-Field Policy

This section develops the asymptotic optimality result for the mean-field based policy and some supporting lemmas. The first step to establish asymptotic optimality is to show that our mean-field state transition functions are good approximations of the original model's state transition functions. To this end, we will show that given the same post-rebalancing inventory distributions ($\hat{\boldsymbol{a}}^{\boldsymbol{t}} = g(\boldsymbol{a}^t)$) in period t, the expected mean-field rebalancing cost of moving from $g(\boldsymbol{x}^{t+1})$ (inventory distribution the system will be in period $t+1$ according to the original model) to $\hat{\boldsymbol{x}}^{t+1}$ (inventory distribution the system will be in period $t+1$ according to the mean-field model), grows with at most rate $\mathcal{O}(\sqrt{n})$. Essentially, as the number of stations increases, mean-field state transitions accurately predict the expected inventory distribution of the next period, and the rebalancing cost the planner can pay to follow the mean-field trajectory is of order $\mathcal{O}(\sqrt{n})$.

Proposition 1.3.3. *For an arbitrary action* $\boldsymbol{a}^t$ *taken in the original model, let* $\boldsymbol{x}^{t+1}$ *be the state evolution defined in* (1.2.1)-(1.2.2)*, and let* $\hat{\boldsymbol{x}}^{t+1}$ *be the state evolution of the mean-field dynamics defined in* (1.3.1)-(1.3.3) *under the projected mean-field action* $g(\boldsymbol{a}^t)$*. Then,*

$$\mathbb{E}_{\{\boldsymbol{D}^t,\boldsymbol{R}^t\}}\left[\hat{c}(g(\boldsymbol{x}^{t+1}),\hat{\boldsymbol{x}}^{t+1})\big|\hat{\boldsymbol{a}}^t = g(\boldsymbol{a}^t)\right] \leq \mathcal{O}(\sqrt{n}) \quad \forall \boldsymbol{a}^t \in \boldsymbol{A}.$$

Consequently, given the same initial action, the expected rebalancing cost between next period's inventory distribution under the original and mean-field models is of order $\mathcal{O}(\sqrt{n})$*.*

The proof for Proposition 1.3.3 is technical and invokes the Central Limit Theorem (CLT) for Wasserstein distances. The main challenge is that the arrival process of stations are correlated with each other, as all station inflows depend on the multinomial distribution based on the departures of all stations. To overcome this challenge, we prove that this distribution can be approximated by n independent Poisson random variables, within $\mathcal{O}(\sqrt{n})$ error. After this transformation, an application of CLT gives the desired result.

While Proposition 1.3.3 proves that the mean-field state transitions are good approximations of the original stochastic state transitions for the many station setting, this result itself is not sufficient to establish asymptotic optimality. The primary remaining challenge is that as we want to show that the projected mean-field policy to the original model is asymptotically optimal, we need to show that the modifications on the policy through the functions h, f have only a limited impact on the resultant cost. The following Theorem establishes this and proves that the mean-field optimal policy is $\mathcal{O}(\sqrt{n})$ sub-optimal in the original stochastic setting.

Theorem 1.3.4. *Let $\hat{\pi}^*$ be an optimal policy for the mean-field model. Then, the lifted mean-field policy $h \circ f \circ \hat{\pi}^* \circ g$ satisfies*

$$V_{h \circ f \circ \hat{\pi}^* \circ g}(\boldsymbol{x}^t) - V(\boldsymbol{x}^t) \leq \mathcal{O}(\sqrt{n}) \quad \forall \boldsymbol{x}^t \in \boldsymbol{X}.$$

Consequently, the optimality gap of the composite policy $h \circ f \circ \hat{\pi}^* \circ g$ is at most $\mathcal{O}(\sqrt{n})$.

The importance of Theorem 1.3.4 is better understood through Corollary 1.3.5, which shows that the mean-field policy is asymptotically optimal. This is proven by using the fact that $N(\boldsymbol{a}^t)$ is of order $\Omega(n)$, as both the holding and the expected penalty costs grow linearly with n.

Corollary 1.3.5. *For any inventory position $\boldsymbol{x}^t \in \boldsymbol{X}$,*

$$\lim_{n \to \infty} \frac{V_{h \circ f \circ \hat{\pi}^* \circ g}(\boldsymbol{x}^t)}{V(\boldsymbol{x}^t)} = 1.$$

Thus, the mean-field model provides an asymptotically optimal policy for the original stochastic model.

The upshot of Corollary 1.3.5 is that if we can solve for the optimal policy for the infinite horizon discounted mean-field model, then we obtain an asymptotically optimal policy for the original stochastic model (we present the numerical results in Section 1.5). However, solving for the optimal mean-field policy is an infinite dimensional problem and hence intractable, and does not lend itself to iterative methods such as value or policy iteration. It turns out that we can cast the problem of obtaining a policy with arbitrarily small optimality gap as an optimization problem with a restricted number of decision variables, and we do so in Section 1.4. However, before doing that, to reduce the state-space and establish an improved performance bound for the case of $q = 0$, we take a further step to transform the current mean-field model so that the state space is compact (as we have defined it so far, only the action space is compact). We perform this transformation in the next subsection.

Remark 1.3.6. *Our objective in this paper has been to obtain near optimal policies rather than to characterize the structure of optimal policies as is done in Benjaafar et al. [2022a] and He et al. [2020], who show that the optimal policy has a threshold-like structure, albeit under cost assumptions which ensure convexity of the value function. We do not make such assumptions on costs, and as a result we can handle more general cost functions, where the resultant mean-field policy may differ from a threshold policy.*

1.3.4 A Compact State Space Reformulation of the Mean-Field Model

Under the mean-field model, $\hat{\boldsymbol{A}}$ is compact, whereas $\hat{\boldsymbol{X}}$ is not. This is reflected in the infinite-dimensional state-space required for $\hat{\boldsymbol{X}}$, where due to Poisson inflow, for any type e, $\mathbb{E}\left[\hat{R}^e\right] > 0$ implies $\hat{x}_{e,d}^{t+1} > 0 \; \forall d$. This section shows that we can aggregate inventory variables corresponding to stations with a high number of units, reducing the state-space significantly. This aggregation will also allow us to express the mean-field in a compact space, which will be necessary to develop our performance bound in Section 1.4. We provide the necessary definitions in this section and provide the technical details on how these functions/variables are obtained, as well as the mathematical intuition, in Section 1.13.

We let $\breve{\boldsymbol{x}}^t = [\breve{x}_e^t]_{e=1}^{\hat{e}}$ and denote the new state space by $\breve{\boldsymbol{X}}$ where

$\breve{\boldsymbol{x}}_e^t = [[\breve{x}_{e,r}^t, \breve{x}_{e,s}^t, \breve{x}_{e,u}^t, \breve{\boldsymbol{x}}_{e,d}^t]_{d=0}^{\infty}]_{e=1}^{\hat{e}} \in \breve{\boldsymbol{X}}$. The state transitions under this setting are:

$$\breve{x}_{e,d}^{t+1} = \sum_{b=L^e}^{U^e} \hat{a}_{e,b}^t \mathbb{P}\left[b - \min(b, D^e) + \hat{R}^e = d\right] \qquad \forall e, t, d \leq U^e, \tag{1.3.6}$$

$$\breve{x}_{e,d}^{t+1} = 0 \qquad \forall e, t, d > U^e, \tag{1.3.7}$$

$$\breve{x}_{e,r}^{t+1} = \hat{a}_{e,r}^t + (1-q) \sum_{e'=1}^{\hat{e}} p_{e',e} \sum_{b=L^{e'}}^{U^{e'}} \hat{a}_{e',b}^t \mathbb{E}\left[\min(b, D^{e'})\right] \qquad \forall e, t, \tag{1.3.8}$$

$$\breve{x}_{e,s}^{t+1} = \sum_{d=U^e+1}^{\infty} d \sum_{b=L^e}^{U^e} \hat{a}_{e,b}^t \mathbb{P}\left[b - \min(b, D^e) + \hat{R}^e = d\right] \qquad \forall e, t, \tag{1.3.9}$$

$$\breve{x}_{e,u}^{t+1} = \sum_{d=U^e+1}^{\infty} (d - U^e) \sum_{b=L^e}^{U^e} \hat{a}_{e,b}^t \mathbb{P}\left[b - \min(b, D^e) + \hat{R}^e = d\right] \qquad \forall e, t. \tag{1.3.10}$$

Here, Equation (1.3.6) preserves the state transitions for stations with less then or equal to U^e units, and stations with more then U^e units are represented in Equations (1.3.9),(1.3.10), where the total number of units at these stations are aggregated through two different functions. Lastly, the state transition function for depleted units is identical to the mean-field formulation.

As we are working with a new state-space, we have to introduce a new rebalancing cost function. The new rebalancing cost function $\hat{c}_2(\breve{\boldsymbol{x}}^t, \hat{\boldsymbol{a}}^t)$ is expressed as:

$$\begin{aligned}
\hat{c}_2(\breve{\boldsymbol{x}}^t, \hat{\boldsymbol{a}}^t) = {} & n \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \breve{x}_{e,u}^t + n \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{b=0}^{U^e-1} \left| \sum_{d=0}^{b} (\breve{x}_{e,d}^t - \hat{a}_{e,d}^t) \right| \\
& + n \min_{y^{e_1,e_2}} \sum_{e_1=0}^{2\hat{e}} \sum_{e_2=0, e_2 \neq e_1}^{2\hat{e}} \left(c_{e_1,e_2} - \frac{c_{e_1,e_1}}{2} - \frac{c_{e_2,e_2}}{2} \right) y^{e_1,e_2} \qquad (1.3.11) \\
\text{s.t.} \quad & \breve{x}_{e,s}^t + \sum_{b=1}^{U^e} b(\breve{x}_{e,b}^t - \hat{a}_{e,b}^t) = \sum_{e_1=0}^{2\hat{e}} (y^{e,e_1} - y^{e_1,e}) \qquad \forall e \in [\hat{e}], \\
& \breve{x}_{e,r}^t - \hat{a}_{e,r}^t = \sum_{e_1=0}^{2\hat{e}} (y^{e+\hat{e},e_1} - y^{e_1,e+\hat{e}}) \qquad \forall e \in [\hat{e}], \\
& y^{e_1,e_2} \geq 0 \qquad \forall e_1, e_2.
\end{aligned}$$

In Proposition 1.13.3, we prove that working with these updated state-transition dynamics and the updated state space does not alter the optimal action, and an equivalent representation of the mean-field model, where the state is in $\breve{\boldsymbol{X}}$, exists. We do this by showing that for each type e, we can aggregate stations with more than U^e (upper threshold for type e) units. This is because these stations have "extra" units which need to be removed, and knowing the total number of these "extra" units at each type is sufficient to make an optimal decision.

Consequently, in the remainder of the paper, we will use $\breve{\boldsymbol{X}}$ as the state space of the mean-field model. Given an initial inventory position $\breve{\boldsymbol{x}}^t \in \breve{\boldsymbol{X}}$, we express the value function through the fixed point of the Bellman recursion as

$$\hat{V}(\breve{\boldsymbol{x}}^t) = \min_{\hat{\boldsymbol{a}}^t \in \hat{\boldsymbol{A}}} \hat{c}_2(\breve{\boldsymbol{x}}^t, \hat{\boldsymbol{a}}^t) + \hat{N}(\hat{\boldsymbol{a}}^t) + \gamma \hat{V}(\breve{\boldsymbol{x}}^{t+1}), \tag{1.3.12}$$

where $\breve{\boldsymbol{x}}^{t+1}$ is given by Equations (1.3.6)-(1.3.10).

1.4 An Approximation Scheme for the Optimal Mean-Field Policy

Having introduced the mean-field model, we present an approximation scheme to obtain control policies for the mean-field model in this section. These policies provide an arbitrarily small optimality gap for the mean-field model. At a high level, our algorithm solves for a "good" recurrent steady-state action and a transient finite horizon policy terminating in this steady-state. We prove that the optimality gap decays exponentially in the length of the transient period.

1.4.1 Formulation of the Algorithm

Given an initial inventory position $\breve{\boldsymbol{x}}^t$ and a parameter $T \geq 0$, our policy is specified by a sequence of $T+1$ actions where the $(T+1)$'st action is repeated for all future periods. To arrive at policy $\tilde{\boldsymbol{\pi}} = \{\tilde{\pi}^{t+k}\}_{k \in \mathbb{N}_0}$ (dependencies on $\breve{\boldsymbol{x}}^t, T$ are suppressed for ease of notation), we first solve for:

$$\{\tilde{\boldsymbol{a}}^{t+k}\}_{k=0}^{T} \in \arg \min_{\{\hat{\boldsymbol{a}}^{t+k} \in \hat{\boldsymbol{A}}\}_{k=0}^{T}} \sum_{k=0}^{T-1} \gamma^k \Big(\hat{c}_2(\breve{\boldsymbol{x}}^{t+k}, \hat{\boldsymbol{a}}^{t+k}) + \hat{N}(\hat{\boldsymbol{a}}^{t+k}) \Big)$$
$$+ \gamma^T \hat{c}_2(\breve{\boldsymbol{x}}^{t+T}, \hat{\boldsymbol{a}}^{t+T}) + \frac{\gamma^T}{1-\gamma} \Big(\gamma \hat{c}_2(\breve{\boldsymbol{x}}^{t+T+1}, \hat{\boldsymbol{a}}^{t+T}) + \hat{N}(\hat{\boldsymbol{a}}^{t+T}) \Big). \quad (1.4.1)$$

In (1.4.1), the first summation denotes the per-period costs of the first T periods, term $\gamma^T \hat{c}_2(\breve{\boldsymbol{x}}^{t+T}, \hat{\boldsymbol{a}}^{t+T})$ denotes the cost of moving from the initial inventory distribution of $(T+1)$'st period to the repeated action, and $\frac{\gamma^T}{1-\gamma} \Big(\gamma \hat{c}_2(\breve{\boldsymbol{x}}^{t+T+1}, \hat{\boldsymbol{a}}^{t+T}) + \hat{N}(\hat{\boldsymbol{a}}^{t+T}) \Big)$ denotes the costs paid in the remaining periods, with the recursion obtained through the condition that the $(T+1)$'st action is repeated for all future periods (the same costs are observed each period due to the model being deterministic).

Given the solution of the above program, $\tilde{\boldsymbol{\pi}}$ is constructed as follows:

$$\tilde{\pi}^{t+k}(\breve{\boldsymbol{x}}^{t+k}) = \tilde{\boldsymbol{a}}^{t+k} \quad \forall k < T, \breve{\boldsymbol{x}}^{t+k} \in \breve{\boldsymbol{X}}, \quad (1.4.2)$$

$$\tilde{\pi}^{t+k}(\breve{\boldsymbol{x}}^{t+k}) = \tilde{\boldsymbol{a}}^{t+T} \quad \forall k \geq T, \breve{\boldsymbol{x}}^{t+k} \in \breve{\boldsymbol{X}}. \quad (1.4.3)$$

Furthermore, the cost of policy $\tilde{\boldsymbol{\pi}}$, $\hat{V}_{\tilde{\boldsymbol{\pi}}}(\breve{\boldsymbol{x}}^t)$, is equal to the optimal value of the optimization problem in (1.4.1).

Remark 1.4.1. *The equivalence of $\hat{V}_{\tilde{\boldsymbol{\pi}}}(\breve{\boldsymbol{x}}^t)$ to the optimal value of the optimization problem in* (1.4.1) *is due to the mean-field model being deterministic where given a sequence of actions, the inventory evolution is exactly known. When we consider the stochastic setting, the realized inventory positions will be different than the mean-field forecast, and therefore the realized cost of $\tilde{\boldsymbol{\pi}}$ will be different from $\hat{V}_{\tilde{\boldsymbol{\pi}}}(\breve{\boldsymbol{x}}^t)$. In Section 1.5, we look at a variant where we recompute the policy at every period to solve the issue of mismatch between the true stochastic and mean-field trajectories.*

For $T = \infty$, this policy is optimal because the model is deterministic. For finite T, such a policy is computationally advantageous compared to other possible methods such as model predictive control since it is a single-shot problem unlike model predictive control, which requires re-solving at each period (we will compare the performance of model predictive control with our control policy in Section 1.5). It is also computationally advantageous compared to methods such as value iteration, as it also avoids the need to approximate the value function at all states.

Another advantage of our proposed policy is that it provides a useful way practitioners should make rebalancing/recharging decisions. In most micromobility systems, platforms have a general idea of the ideal initial distribution of inventory amongst different stations. Furthermore, for a variety of (possibly unmodeled) reasons (e.g., limitations in the rebalancing/recharging capacities), such systems often observe different distributions of inventory. Many responses to this problem are either too limited, where the planner allows the system to drift to extreme inventory positions, or too excessive, where the planner implements excessive control to move the system to its ideal position. The algorithm we propose dilutes this excessive control over multiple periods, incorporating the repositioning done by customers in order to avoid paying excessive rebalancing costs.

While analyzing algorithm's performance, we separately derive a bound for a special case of the model where $q = 0$. The reason why $q = 0$ motivates a separate analysis is that under this assumption, the state transitions provided in Equations (1.3.6)-(1.3.10) are affine[13]. Having affine state transitions provides important properties for the algorithm, such as allowing the algorithm to be reformulated exactly as a linear program (we provide the resultant linear program under this reformulation in Section 1.14). Furthermore, we show that under affine state transitions, the mean-field model admits a fixed point (an inventory position such that taking an optimal action for that inventory position results in moving to the same position next period) and develop a better worst-case bound (for systems with $\gamma > 0.5$). We provide the technical details on this fixed point analysis, together with the proof of Theorem 1.4.2, in Section 1.15.

13. For $q > 0$, arrival rate $\hat{R}^e$ has a non-linear dependence on $\hat{\boldsymbol{a}}$ causing non-affine state transitions.

1.4.2 Worst-Case Performance of the Control Algorithm

The following theorem is one of our main results, establishing that the optimality gap decreases exponentially with respect to the transient period T used to find the approximately optimal policy for the mean-field model.

Theorem 1.4.2. *The optimality gap of the policy* $\tilde{\boldsymbol{\pi}}$ *obtained via* (1.4.1)-(1.4.3) *decreases exponentially with respect to length* T *of the transient horizon:*

$$\hat{V}_{\tilde{\boldsymbol{\pi}}}(\breve{\boldsymbol{x}}^t) - \hat{V}(\breve{\boldsymbol{x}}^t) \leq \frac{\gamma}{1-\gamma} C \gamma^T \quad \forall \breve{\boldsymbol{x}}^t \in \breve{\boldsymbol{X}},$$

where

$$C = 2\sum_{e=1}^{\hat{e}} \Big(n_e \max(c_{0,e}, c_{e,0})(U^e - L^e) + n \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0})(U^{r,e} - L^{r,e}) \Big).$$

Furthermore, we can derive an alternative bound for $q = 0$ *with:*

$$\hat{V}_{\tilde{\boldsymbol{\pi}}}(\breve{\boldsymbol{x}}^t) - \hat{V}(\breve{\boldsymbol{x}}^t) \leq C\gamma^T \quad \forall \breve{\boldsymbol{x}}^t \in \breve{\boldsymbol{X}}.$$

We provide two separate analyses for the two structurally different models to construct these bounds. For the general case (for all q values including $q = 0$), we observe that any optimal policy's first $T+1$ actions are a feasible sequence of action for the algorithm. For the special case $q = 0$, we add the observation that the system admits a fixed point that the control algorithm can reach in period $T+1$. The constant C observed in both bounds is an upper bound to two times the maximum rebalancing cost paid to move between two feasible actions, with

$$\begin{aligned} C \geq\, & 2 \max_{\breve{\boldsymbol{x}} \in \breve{\boldsymbol{X}}, \hat{\boldsymbol{a}} \in \hat{\boldsymbol{A}}} \hat{c}_2(\breve{\boldsymbol{x}}, \hat{\boldsymbol{a}}) & \\ \text{s.t.} \quad & \breve{x}^t_{e,s}, \breve{x}^t_{e,u} = 0 & \forall e, \\ & \breve{x}^t_{e,r} \in [L^{r,e}, U^{r,e}] & \forall e. \end{aligned}$$

The theorem states that extending the transient horizon of the algorithm results in an exponential decrease in the upper-bound of the optimality gap of the algorithm (in the mean-field setting). While it may not be possible to consider very long forecast horizons in real-life settings due to limitations on forecasting capabilities for very distant periods and the added computational cost for the algorithm (computational cost of the algorithm is linear in the transient horizon), the result suggests that any feasible extension of this horizon is beneficial. This benefit is even more valuable for systems with limited discounting (γ close to 1), where the result indicates that investment in extending forecasting and computational capabilities can positively impact profits.

Remark 1.4.3. *Both bounds in Theorem 1.4.2 may seem weak for problems with γ close to one (compared to problems with small γ), as in that case γ^T can be non-negligible even for large T. However, the optimal cost also grows approximately with rate $\frac{1}{1-\gamma}$ so while the optimality gap may remain significant, the ratio of the gap with the optimal cost remains small for all discount rates.*

1.5 Numerical Experiments

In this section, we look at the performance of our control policy in a stochastic environment, and compare it with some simple heuristics as well as sophisticated approaches.

1.5.1 Description of Simulated Policies

Our numerical studies will explore seven different policies which include standard benchmark policies, variations of our control policy, and policies based on other fluid approximations. We provide the basic analysis of how these policies are constructed in this section. Additional discussion on how we solve these policies for our model can be found in Section 1.17. Specifically, the policies which we consider are:

1. Resultant policy of the control algorithm we presented in Section 1.4 (which we label as

static control). This policy is a static policy where we first use initial inventory position $\boldsymbol{x}^0$ (which the mean-field based policy observes as $g(\boldsymbol{x}^0)$) to obtain $\tilde{\boldsymbol{\pi}}$. Then, at each period t, we take action $h(f(\tilde{\pi}^t(\breve{\boldsymbol{x}}^t)))$, regardless of how much mean-field and stochastic inventory transitions deviate.

2. Control policy with re-solving, which was briefly mentioned in Remark 1.4.1. Under this policy, at every period, we use $\boldsymbol{x}^t$ to obtain $\tilde{\boldsymbol{\pi}}$, implement the resultant optimal action for the first period, but then re-solve the algorithm next period for the newly induced inventory position $\boldsymbol{x}^{t+1}$. Through this policy, we address the issue of stochasticity, allowing the algorithm to adapt to volatility in demand realization. We label this policy as *re-solving control*.

3. Model predictive control (labeled *MPC*). Under MPC, at every period, we solve for the optimal initial action of the corresponding *mean-field* finite horizon problem with $T+1$ periods ($T+1$ is selected as the period length to ensure consistency with the control algorithm), with the optimal initial action given by:

$$\hat{\boldsymbol{a}}^{t*} \in \arg \min_{\{\hat{\boldsymbol{a}}^{t+k} \in \hat{\boldsymbol{A}}\}_{k=0}^{T}} \sum_{k=0}^{T} \gamma^k \bigg(\hat{c}_2(\breve{\boldsymbol{x}}^{t+k}, \hat{\boldsymbol{a}}^{t+k}) + \hat{N}(\hat{\boldsymbol{a}}^{t+k}) \bigg).$$

Then, at every period, we take the designated action (which is again mapped to the stochastic environment through the h and f functions), move to the newly induced inventory position next period, and re-solve for the optimal initial action. The difference between MPC and re-solving control is in the objective function giving the initial action. Re-solving control assumes the $(T+1)$'st action is repeated indefinitely and uses the value function of the corresponding state as the terminal cost in period $T+1$.

Remark 1.5.1. *The first three policies are constructed with the assistance of the mean-field model, where we solve optimization programs based on the mean-field model, and then use the h, f functions to map their optimal solutions to the stochastic environment. The remaining* *policies are independent of the mean-field model and will be directly based on the original model formulation.*

4. *Newsvendor* action policy, which solves the problem assuming that rebalancing is free. Under this assumption, the optimal solution is decomposable over the stations and is based on finding the optimal trade-off between holding and penalty costs. Formally, the newsvendor action policy solves for an $\boldsymbol{a}_N$ satisfying:

$$\boldsymbol{a}_N \in \arg\min_{\{\boldsymbol{a}^t \in \boldsymbol{A}\}} N(\boldsymbol{a}^t),$$

and rebalances to this action every period, regardless of the initial inventory position.

5. *Hybrid no rebalancing* policy, which does not rebalance units between types (unless threshold constraints are violated) and only recharges depleted units while keeping them in the same type. This policy is an enhancement of the standard no-rebalancing policy (where no units are recharged/sourced/rebalanced) as the standard no-rebalancing policy: (i) does not recharge units resulting in all units being depleted (for $q < 1$) as time progresses, (ii) does not source units (given that the initial inventory position satisfies thresholds) so the fleet size is equal to the initial inventory position. We enhance the performance of this policy by allowing it to recharge depleted units each period (the recharged units are distributed among the stations of the same type) and rebalance once in the initial period so that the policy can select a favorable fleet-size and then rebalance if action thresholds (upper/lower limits on the number of units a station can have post-rebalancing) are violated. To solve for the fleet-size, we use a line-search algorithm where we compare different initial numbers of units (assumed to be equal at every station) and use Monte-Carlo simulation to solve for the best starting position.

6. *Myopic* policy, which solves for the single-period variation of the original stochastic model.

Formally, at each period t, the myopic policy solves for an $\boldsymbol{a}_M^t$ satisfying:

$$\boldsymbol{a}_M^t \in \arg\min_{\boldsymbol{a}^t \in \boldsymbol{A}} c(\boldsymbol{x}^t, \boldsymbol{a}^t) + N(\boldsymbol{a}^t). \tag{1.5.1}$$

While the myopic policy does not require the evaluation of costs of future periods, it requires solving a non-linear integer program (due to the structure of $N(\boldsymbol{a}^t)$) where the number of variables increase with n. Nevertheless, we are able to reformulate (1.5.1) as a mixed-integer program through the assistance of an interim model. We provide the details on this reformulation in Section 1.17.

7. *Large Market* policy, which solves for a policy through the fluid approximation based on the construction provided in Braverman et al. [2019]. Under this fluid approximation, we look at a setting where both the fleet size and demand at stations tend to infinity, which allows considering demand to be deterministic and depletion, unit drop-offs to be exactly proportional to the respective $q, p_{i,j}$ values (as the resultant inventory positions are fractional, we extend the action space $\boldsymbol{A}$ to allow for fractional action's which we then round when implementing the policy). Thus, under the large market approximation, the state at period $t+1$ is given by the following equations:

$$x_{e,i}^{t+1} = a_{e,i}^t - \min(a_{e,i}^t, \mu_e) + \sum_{e'=1}^{\hat{e}} \sum_{j=1}^{n_{e'}} \frac{q p_{e',e}}{n_e} \min(a_{e',j}^t, \mu_{e'}) \qquad \forall i \in [n], \tag{1.5.2}$$

$$x_{e,r}^{t+1} = a_{e,r}^t + (1-q) \sum_{e'=1}^{\hat{e}} p_{e',e} \sum_{j=1}^{n_{e'}} \min(a_{e',j}^t, \mu_{e'}) \quad \forall e. \tag{1.5.3}$$

The resultant value function can then be expressed as

$$V^M(\boldsymbol{x}^t) = \min_{\boldsymbol{a}^t \in \boldsymbol{A}} c(\boldsymbol{x}^t, \boldsymbol{a}^t) + c_h \sum_{e=1}^{\hat{e}} \sum_{i=1}^{n_e} a_{e,i}^t + c_h \sum_{e=1}^{\hat{e}} a_{e,r}^t + c_p \sum_{e=1}^{\hat{e}} \sum_{i=1}^{n_e} (\mu_e - a_{e,i}^t)^+ + \gamma V^M(\boldsymbol{x}^{t+1}). \tag{1.5.4}$$

Both the large market and the mean-field models are deterministic approximations which help obtain policies for the original stochastic problem. They also suffer from the same problem where solving for the optimal policy is an infinite dimensional problem (as in the case of the mean-field model)[14]. As a result, we adapt the static control algorithm to the large market model, where we again solve a $T+1$ period finite horizon problem with the $T+1$'th period repeated for all periods. The reason we adapt the static control policy rather than the resolving control policy is that the adapting the resolving control policy requires solving a non-linear optimization problem every period (even after our reformulation), and the number of variables linearly increase with n.

One difficulty of using a static policy for this fluid approximation is that only inventory distributions (rather than the exact inventory position of each station) of types for future periods can be accurately predicted as the number of stations increases. As a result, specifying an exact action to a station without observing the demand realization and the initial inventory position of that station may result in additional rebalancing costs. To minimize rebalancing costs, we use the g and h functions where given any period $s > t$, and action $\boldsymbol{a}^s$ (solved for at period t), we implement action $h(g(\boldsymbol{a}^s))$ at period s. The g and h functions take $\boldsymbol{x}^s$ as an input and modify action $\boldsymbol{a}^s$ to match the order statistics of pre-rebalancing inventory, in each type, to avoid paying additional rebalancing costs.

When evaluating the performances of these policies, as an additional benchmark, we also compute the lower bound of the optimal cost obtained from the mean-field model. We solve for this value by computing the cost of the static control algorithm in the mean-field model for a large T value and then using Theorem 1.4.2.

14. The large market policy also includes piece-wise components in the state transition functions (due to the lost-sales setting), which cause non-linearity. In Section 1.17, we show an exact reformulation to overcome this issue.

1.5.2 Selection of the Cost Parameters

Our goal in constructing our model was to approach various micromobility systems, differing in terms of their type of units (scooters/bikes) and the setting these systems are used in (major cities/smaller locations). Consequently, the selection of cost parameters to analyze each of these systems differs. To this end, we run several experiments where we vary values of cost parameters to understand the robustness of our policy. Nevertheless, we will calibrate values for each cost parameter in our experiments as a starting point. We base these values on the scooter-sharing system in Austin to match our data-driven experiment. For space considerations, we list the values below while describing their calibration in detail in Section 1.17.3.

First, we assign the cost of rebalancing within a type to be \$1 and the cost of rebalancing units between types to be \$2. We also assign the cost of recharging to be \$3 (indicating that the cost is \$4 if a depleted unit is charged and dropped to the same type and \$5 if the unit is dropped to a different type). Finally, we assign a sourcing cost of \$6 as sourcing is expected to be the most expensive operation.

For the remaining values, we assign $\gamma = 0.95$, $c_h = 2.2$, $c_p = 11.3$, and $q = 0$. For space considerations, we describe their calibration in detail in Section 1.17.3.

1.5.3 Synthetic Experiments

We first construct a series of synthetic experiments to understand better the performance of our policy[15]. We let $\hat{e} = 4$, with $n_e = 20\,\forall e$. We assume that demand is distributed Poisson in each of the four types, albeit with different mean values. We also assume an imbalanced network between types with different p_{e_1,e_2} values. We assume that the upper threshold for charged units is 17 units and the upper threshold for depleted units is five units per station of the same type. We assume that lower threshold for both charged units and depleted

15. For space considerations, some supportive experiments can be found in the Section 1.17.

units is 0. We let $T = 3$ for static control, resolving control, MPC, and the large market policy. We simulate the performance of the policies through Monte-Carlo simulation, with 100 repetitions.

Impact of Penalty Cost

In our first experiment, we vary the penalty cost c_p (assumed to be \$11.2 previously). The results are shown in Figure 1.2.

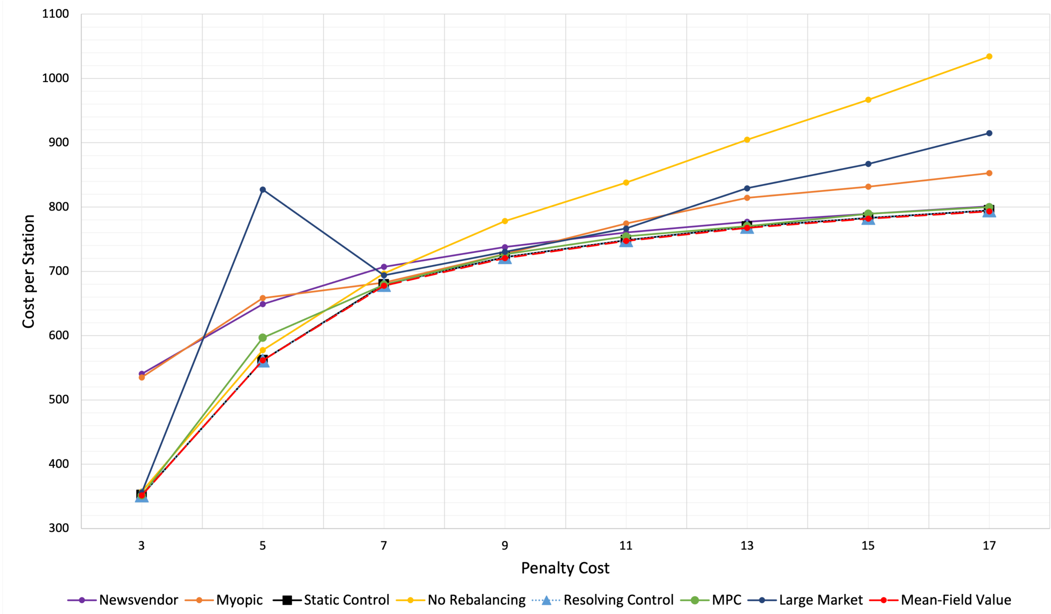


Figure 1.2: Impact of Penalty Cost on the Performance of Policies

Several points are worth highlighting: First, the performance of both static and resolving control policies are approximately equal to the mean-field lower bound. In addition to covering the fundamental trade-offs that the heuristic policies optimize, the two policies make multi-period decisions considering the spatial distribution of inventory. For systems with very low penalty costs, maintaining any service level is sub-optimal; hence these policies decrease the number of units in circulation until no units remain. In contrast, for systems with high penalty costs, the two policies observe that the initial number of units in circulation is insufficient; hence these policies source units (instead of rebalancing units already in circulation) to reach a good fleet size. As different micromobility systems require different cost values (which are also difficult to estimate), showing a strong performance for a wide range of settings is essential for implemented policies.

Another critical point is the varying performance of the large market policy, which is near-optimal for some penalty cost values and performs poorly for others. This is due to the assumption of deterministic demand that this policy uses. For example, given that the penalty cost is low, we expect the optimal policy only to recharge a unit if the probability that a customer will pick this unit up is close to 1. As a result, for effective policies, the number of charged units post-rebalancing should be much less than the mean demand. However, due to the deterministic demand assumption, this cutoff for the large market policy is equal to the mean demand.

Similarly, for systems with high penalty costs, the optimal policy will charge a unit even if the probability that the unit will be picked up is small. Again, this cutoff for the large market policy equals mean demand due to the deterministic demand assumption. Consequently, information on the variance of demand distributions at stations is necessary to make accurate decisions, which the large market policy fails to do.

Impact of Number of Stations

In our second experiment, we vary n (assumed to be 80 previously). The results are shown in Figure 1.3.

First, as the number of stations increases, static and resolving control policies approach the cost of the mean-field model (while performing better than all other policies for all network sizes). This is expected because the mean-field model better approximates the original model as the number of stations increases. For MPC, which is based on the mean-field formulation but uses smaller cost coefficients for the last period "steady-state" costs, we observe that the performance is slightly worse than the static/resolving policies. This decrease in performance of MPC is due to the selection of T, where for large T, the impact of

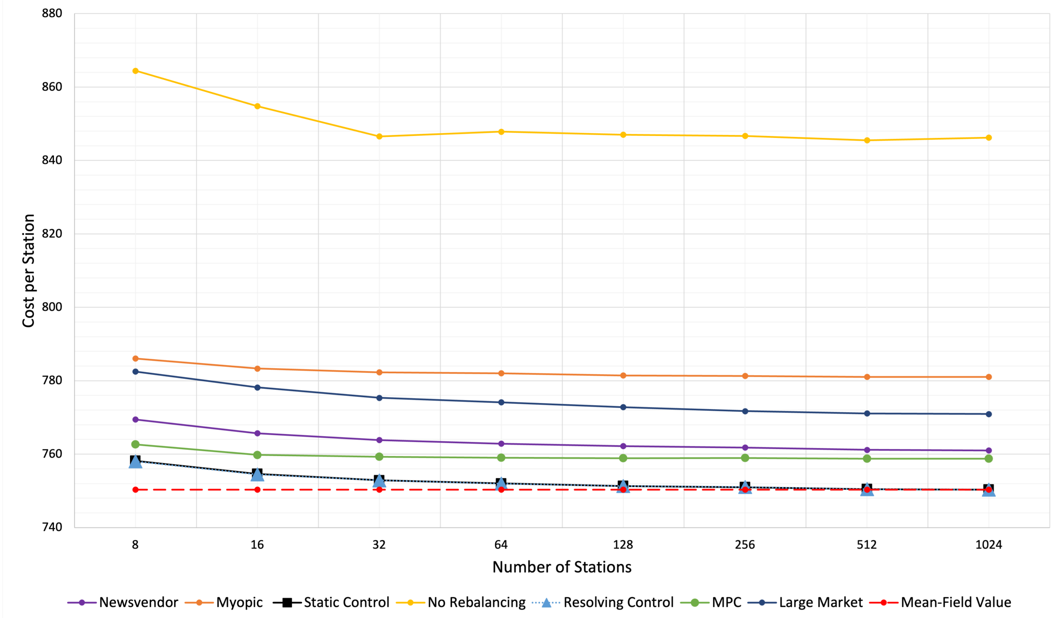


Figure 1.3: Impact of Number of Stations on the Performance of Policies

the steady-state modification on the initial action is lower, and MPC takes the same actions as resolving control. Nevertheless, the control algorithm allows us to better approximate the infinite-horizon problem through a much smaller T value.

Remark 1.5.2. *In both experiments, we observe that the static and resolving control policies provide similar results, which is counter-intuitive as we expect the resolving policy to perform better. In Section 1.17.5, we show that this is due to the selection of cost parameters, where moving units to different types costs one more dollar than moving the unit within the same type. The advantage of resolving is the ability to adapt to extreme demand realizations, which, for the $q = 0$ case, corresponds to re-solving based on the distribution of depleted units among types (possibly taking a lower action at a type if there are less depleted units to supply that action). As the additional cost of moving units to different types is low (compared to main cost drivers such as recharging cost and penalty cost), the resolving control policy takes almost identical actions to the static control policy, even when the realized distribution of depleted units differs from the mean-field expected distribution of depleted units. For systems where the cost of moving units between types is high, we observe a more significant gap between* *static and resolving control for small n (this gap vanishes for large n as extreme demand realizations are less common).*

Remark 1.5.3. *When comparing static and resolving policies, we omit an important fact that moving to a policy with re-solving (applicable for both MPC and re-solving control) is computationally costly (especially for systems with $q > 0$ as they require solving a non-linear model) while modifying the objective function to account for steady-state costs has negligible one-off computational cost. The former is an important decision which might not be feasible in a real-life setting where there are many types of stations and less time to make (or modify) decisions, while the latter is a simple modification of the policy which the experiments show improves performance, irrespective of whether re-solving is feasible or not.*

1.5.4 Austin Scooter-Sharing Data

Forming the Numerical Example

In order to better analyze the performance of our policy in a real-life setting, we conduct a numerical experiment using Austin scooter-sharing data (City of Austin Transportation Department [2022]). We use the Austin Shared Micromobility Vehicle Trips data-set, focusing on scooter trips between September 1, 2021, and September 15, 2021. Before providing the results, we briefly explain how we estimate model parameters such as demand, and probability transition matrix, using the available data.

The primary step in estimating our parameters is to form the types/clusters. In the provided dataset, the location granularity is census tracts, where we are given the starting and ending census tract (but not the exact coordinates) of each trip. In order to form the types, we use a 2-step approach, where in the first step, we use the weighted k-means clustering algorithm to construct five clusters based on the longitude/latitude values of the census tracts. After obtaining an initial solution for types/clusters, we compute the state transition probabilities of the census tracts to each type and the state transition probabilities between types. Finally, in the second step, we can re-assign census tracts to types that better match their trip patterns through these values.

We estimate the demand distributions and the number of stations for each type using the identification data on the census tracts belonging to that type. We use the average number of pickups per day[16] and the weighted density at each type (we use the population density information of each census tract, weighted by the proportion of pickups corresponding to that census tract). We let types with higher population densities have higher mean demands (and a lower number of stations indirectly), based on Kabra et al. [2020], which establishes the unwillingness of customers to use stations farther away.

Results

Running the experiment; we obtain that static control, re-solving control, and the mean-field lower bound provides a cost of \$2.017M (million dollars/rounded to nearest thousand dollars), newsvendor policy provides a cost of \$2.03M, MPC provides a cost of \$2.044M, no rebalancing provides a cost of \$2.079M, myopic policy provides a cost of \$2.178M, and the large market policy provides a cost of \$2.29M.

One major difference between the synthetic experiments and the data-driven experiment is the disparity in the properties of types. For example, in the data-driven experiment, the type with the highest mean demand (per station) has a mean demand of almost six times more then the type with the lowest mean demand. Under such disparities, policies need to account for state transitions. We see this through the poor performance of the myopic policy, which does not include state-transition probabilities while making decisions. In contrast, both static and re-solving control approximate these state transitions to make accurate decisions and thus accurately model systems with very heterogeneous clusters.

Lastly, we highlight the managerial implications of the experiment we conducted. In this

16. Due to the lack of exact pickup-dropoff coordinates, we rely on the censored trip data to estimate demand.

experiment, we differed from the previous sections by first estimating the model's parameters (we took them as given while proving our theoretical results) and then running our algorithm, highlighting its strong performance. While we introduced our method to collect census tracts into clusters, there are many papers (such as Dai et al. [2018]) that provide algorithms for clustering stations and predicting customer behavior. This aligns with the popularity of exploring patterns in shared mobility data. At its core, our algorithm utilizes these data-driven approaches to obtain effective rebalancing/recharging/sourcing policies. Our approach is universal for all shared micromobility systems (and other systems that allow customers the freedom to decide where to drop their units). It uses the existing patterns in customer behavior to construct the types and then applies the algorithm to obtain the rebalancing/recharging/sourcing policies. While our approach is primarily intended for large-scale systems (for the accuracy of the mean-field model), our algorithm also performs well for small network sizes, as seen in Section 1.5.3. While the realization of demand may differ from mean-field expectations in such networks, our algorithm still considers the fundamental trade-offs that the heuristic policies optimize and optimize the movement of inventory across time, taking the spatial distribution of inventory into account. Such an approach is essential for systems with balanced cost parameters, as such systems penalize both excessive and no rebalancing/recharging.

1.6 Extensions

This section relaxes some of the assumptions made and discusses possible extensions. While some of the assumptions we made were necessary to obtain our results, others were for clarity of writing. Some real-life applications of our model do require specific extensions such as heterogeneous costs across types, heterogeneous depletion probabilities, including damaged/missing units, travel times, and/or fixed costs. This section shows that we can extend our results to incorporate these extensions.

1.6.1 Heterogeneous Cost Parameters

While developing the model, we assumed that the holding cost c_h and penalty cost c_p are homogeneous across types. We made this assumption for clarity of exposition. However, in many relevant applications, these costs are heterogeneous, such as in systems where demand at certain types may have a higher priority. Different holding costs can also be assigned depending on whether a unit is charged or depleted. We generalize our model by allowing the cost parameters $c_{h,e}, c_{h,r,e}, c_{p,e}$ to take different values for different types to model such applications. We still assume that there are finitely many types, and the analysis for asymptotic optimality of the mean-field model and the methodology used for proving the worst-case bound for the control algorithm remain identical.

1.6.2 Extension to Non-Stationary Settings

Our paper assumes that demand distributions, state transition probabilities, and thresholds are stationary. We did this for clarity of presentation, and in many applications, cyclic or seasonal demand and trip patterns are observed. One example is from modeling rush hour traffic, where network flows in the morning rush hour differs from that in the evening rush hour. This subsection will extend our work to a class of non-stationary problems, known as periodic problems. To extend our approach to settings with cyclic demand patterns and seasonality effects, we assume a season length $H < \infty$, where for any period t and types e_1, e_2, the primitives satisfy: (i) $\mathcal{L}(D^{t+H,e}) = \mathcal{L}(D^{t,e})$, (ii) $p^{t+H}_{e_1,e_2} = p^{t}_{e_1,e_2}$, (iii) $\boldsymbol{L}^{t+H} = \boldsymbol{L}^{t}$, (iv) $\boldsymbol{U}^{t+H} = \boldsymbol{U}^{t}$, (v) $\boldsymbol{L}^{t+H,r} = \boldsymbol{L}^{t,r}$, (vi) $\boldsymbol{U}^{t+H,r} = \boldsymbol{U}^{t,r}$. Here, we make the dependence of $p^{t}_{e_1,e_2}, \boldsymbol{D}^t, \boldsymbol{L}^t, \boldsymbol{U}^t, \boldsymbol{L}^{t,r}, \boldsymbol{U}^{t,r}$ on t explicit to highlight the non-stationarity in the problem.

To extend our results, we first have to show that the mean-field model is asymptotically optimal. This is straightforward as the stationarity property is not used anywhere while proving asymptotic optimality. Moreover, as we keep the type structure intact, the assumption that $\hat{e} < \infty$ will be sufficient to obtain asymptotic optimality.

To extend the worst-case performance bound for our control algorithm, we follow the standard approach of studying such periodic MDP's, which is to define *composite periods*, with each composite period consisting of H periods of one cycle. This, in turn, yields an MDP where each composite period is *i.i.d.*, allowing us to establish a worst-case bound for the control algorithm. For completeness, we formalize these steps in Section 1.16 and obtain a similar bound to the one obtained in Theorem 1.4.2 in Theorem (1.16.1).

1.6.3 Extending the Depletion Probability

One of the main goals of our paper is to integrate recharging decisions into rebalancing decisions effectively. We considered a simple setting of depletion for clarity of writing, where a unit is depleted after a customer induced trip with a probability of $1 - q$. This section will discuss extensions to this setting, including heterogeneous depletion probabilities and different depletion events, and discuss the implications of not directly including battery levels in our model.

While we assumed that depletion probability is identical for all trips, in real-life, we would expect trips of longer distances to use more battery charge and thus have a higher probability of depletion. To incorporate such trips in our model, we can extend our results to an origin-destination dependent setting where an unit picked up from type e_1 and dropped to e_2 is still available with probability q_{e_1,e_2}. Under this setting, the success probabilities of the trip distributions will change, with $\boldsymbol{R}^t_{e',e} = [R^t_{e',e,r}, R^t_{e',e,1}, \cdots, R^t_{e',e,n_e}]$ being a multinomial distribution with $\sum_{i=1}^{n_{e'}} \min(a^t_{e',i}, D^t_{e',i})$ trials and success probability $(1-q_{e',e})p_{e',e}$ for $R^t_{e',e,r}$ and success probabilities $\frac{q_{e',e}p_{e',e}}{n_e}$ for remaining terms. As the original trip distribution changes, $\hat{R}^e$, which is the mean-field approximation of this process, will also change with

$$\hat{R}^e \stackrel{d}{=} \mathrm{Pois}\left(\frac{n}{n_e} \sum_{e'=1}^{\hat{e}} p_{e',e} q_{e',e} \sum_{b=L^{e'}}^{U^{e'}} \hat{a}^t_{e',b} \mathbb{E}\left[\min(b, D^{e'})\right]\right) \quad \forall e.$$

While we have changed the state transition equations, our methodology for obtaining asymptotic optimality as well as obtaining the worst-case bounds will remain identical. This is because our analysis already considers origin-destination heterogeneity of trips through the $p_{e',e}$ probabilities, and extending q to the same level of detail does not effect any of the steps.

Second, in our model, we assume no differentiation between depleted units other than the type they are physically at. As briefly mentioned in Section 1.2.2, the amount a platform pays for recharging a depleted unit depends on multiple different factors, including the difficulty of locating the unit, the neighborhood the unit is located in, and other factors. While our current cost structure only captures the neighborhood component, we can extend our model to allow for different recharging costs within the same type. To do this, we will extend the q probability such that the depletion probability is partitioned to map to different groups of depleted units. Specifically, in order to consider M different recharging costs within a type, we will introduce new inventory and action variables $x^t_{e,r,1}, ..., x^t_{e,r,M}$, $a^t_{e,r,1}, \ldots, a^t_{e,r,M}$ where the probability that a unit moving to a type e ends up in the k'th depleted group is q_k, with $\sum_{k=1}^M q_k = q$ (or $\sum_{k=1}^M q_{e_1,e_2,k} = q_{e_1,e_2}$ in the case of heterogeneous depletion probabilities). We can then assign a different recharging cost to each of these groups. Mathematically, given $M < \infty$, our results extend. The resultant mean-field state transitions for the depleted unit groups becomes

$$\hat{x}^{t+1}_{e,r,k} = \hat{a}^t_{e,r} + (1 - q_k) \sum_{e'=1}^{\hat{e}} p_{e',e} \sum_{b=L^{e'}}^{U^{e'}} \bar{a}^t_{e',b} \mathbb{E}\left[\min(b, D^{e'})\right] \quad \forall k \in \{1, \cdots, M\},$$

where we can show that splitting depleted units arriving to a type to a finite number of subsets preserves asymptotic optimality. The methodology used for the remaining results on the worst-case bound of the mean-field model is identical.

Lastly, we do not explicitly define battery charge level as an attribute in our model. Instead, we consider binary charge levels, where a unit is either charged or depleted. Unlike the previous paragraphs, our results do not readily extend to the battery charge setting. Such a setting requires working with the multi-product (different charge levels can be thought as different products) version of our problem, where there are different products available at a station and arriving customers can substitute between products based on availability[17]. Nevertheless, our approach covers the important properties of battery charge for large-scale problems. To demonstrate this, we can consider a simple setting with three charge levels (fully charged, half charged, and zero charged), and each trip takes half of the battery life. As a result, all fully charged units picked-up will be available next period, where all half charge units will be depleted with usage. Letting $q = \frac{1}{2}$ (2 being the number of trips a fully charged unit can take), in the large-scale setting, under our model, approximately half of the units in ongoing trips will be depleted. In the simple alternative setting, approximately half of the units will also be depleted (assuming customers have no preference between half charged and fully charged units). So, while our model can produce instances where specific units may be used several times; it is more accurate in capturing the total number of units depleted. The crucial step is to calculate the average battery consumption in trips between types, and assign the correct q_{e_1,e_2} values (where q_{e_1,e_2} is the inverse of the expected number of trips a scooter can take between types e_1, e_2 given full initial charge).

1.6.4 Damaged/Missing Units

In addition to rebalancing/sourcing/recharging units, a common task provided by micromobility systems is maintenance. Some of the returned units are damaged, and these units have to be attended by maintenance crews (as most repair operations cannot be crowd-sourced). In our model, we do not differentiate between depleted units and damaged units for ease of exposition. However, in real life, these are two different problems with different costs. Nevertheless, we can extend our model and our results to include damaged units. To do that, we will extend the state space to define a new type for damaged units (or multiple types if the region in which the unit is damaged has a large impact on cost). Labeling this type as type e_d, we let the probability that units at a type get damaged give the inflow

17. While we can also think of different charge levels as different stations/nodes, our results would require assuming that demand for each of the charge levels is independent, which is unrealistic as customers generally choose the closest unit with enough charge remaining.

probabilities. We assume that demand at this type is equal to zero to model the inability of customers to use these units. We can then assign rebalancing costs to this type to account for the average maintenance costs paid per unit.

A similar treatment can also be used for missing units. While we assume that every unit picked up will be returned (which follows the real-life practice of platforms placing high penalties to ensure customers return their units), units do end up missing due to theft and other forms of vandalism. Similar to the above discussion, we can introduce a new type e_m with the probability that units at a type are lost give the inflow probabilities for this type. Unlike damaged units, we will let the upper threshold for this type equal 0, so no holding costs are paid for these units and the fleet size decreases accordingly. We can then assign rebalancing costs to this type to account for the replacement cost of each unit.

In both cases, the extension of our results is straightforward as we are only adding dummy types to account for these particular realizations. Furthermore, in both cases, cost parameters should be assigned such that the costs reflect the condition of the units, and under these cost parameters an optimal action should be to remove these units[18]. Through forcing this optimal action (by the selection of problem parameters), we ensure that no decision variables are added for the control algorithm, and the computational effort required does not increase.

1.6.5 Travel Times

One of the assumptions we made for tractability was that customers return units by the end of the period. While this is a reasonable assumption for micromobility systems due to platform restrictions on usage times, travel times are an integral part of the problem for other relevant applications such as logistics and transshipment, where our results can be useful. Through a similar formulation as in Section 1.6.4, we model travel times by adding

18. This also reflects real-life operations as indicated in City of Chicago [2019], where it is stated that "Vendors must relocate any scooter parked outside of the pilot area, and any non-functional scooter, within 2 hours of notification by the City or a resident."

dummy types to the system. To highlight how dummy types are utilized, we provide the following example setting where we assume that there are two types e_1, e_2 such that it takes two periods for a unit departing from e_1 to move to e_2. We introduce a new dummy type e_3 such that all units going to e_2 from e_1 first end up in e_3. We assume that e_1 has no direct outflow to e_2 and the probability that the unit goes to e_2 is the inflow probability of e_3. Furthermore, we assume that all units leaving e_3 end up in e_2. We also need to ensure that all units entering e_3 depart e_3 and the platform does not rebalance any units into/out of transit, which is possible by assigning a large upper threshold, a lower threshold equal to zero, a large number of stations, deterministic demand with mean equal to the upper threshold, penalty cost equal to holding cost, and high rebalancing costs to e_3. Lastly, we assume that $q_{e_1,e_2} = 1$, and integrate the depletion event in this route through $q_{e_3,e_2} = q_{e_1,e_2}$.

Using this formulation, we incorporate travel times into our problem through this formulation without adding extra decision variables (as there is no decision making at the dummy types).

1.6.6 *Fixed Costs*

Another assumption we made was that the total rebalancing cost incurred was linear in the number of units repositioned/charged. In applications such as bike-sharing, where the bulk of the repositioning cost involves visiting a station for rebalancing rather than rebalancing of each unit, a more complex cost structure with fixed costs can be useful to better understand system dynamics[19]. To this end, we add the cost term $\sum_{e=1}^{\hat{e}} \sum_{i=1}^{n_e} c_{f,e} \mathbb{I} \left\{ x_{e,i}^t \neq a_{e,i}^t \right\}$ to the rebalancing cost function of the original model. This implies that in addition to a linear rebalancing/recharging costs, a type dependent fixed cost of $c_{f,e}$ is incurred if any unit is moved in or out of a station. To extend our analysis to this setting, we need to find an equivalent representation of the rebalancing cost for the mean-field model. The requisite

19. We discuss the linearity assumption in more detail in Section 1.10.

function for the rebalancing cost is given by:

$$
\begin{aligned}
\hat{c}(\boldsymbol{x}^t, \boldsymbol{a}^t) = n \sum_{e=1}^{\hat{e}} c_{f,e} \sum_{b=0}^{\infty} \frac{1}{2} \left| \hat{x}^t_{e,b} - \hat{a}^t_{e,b} \right| + n \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{b=0}^{\infty} \left| \sum_{d=0}^{b} (\hat{x}^t_{e,d} - \hat{a}^t_{e,d}) \right| \\
+ n \min_{y^{e_1,e_2}} \sum_{e_1=0}^{2\hat{e}} \sum_{e_2=0, e_2 \neq e_1}^{2\hat{e}} \left(c_{e_1,e_2} - \frac{c_{e_1,e_1}}{2} - \frac{c_{e_2,e_2}}{2} \right) y^{e_1,e_2} \qquad (1.6.1) \\
\text{s.t.} \quad \sum_{b=1}^{\infty} b(\hat{x}^t_{e,b} - \hat{a}^t_{e,b}) = \sum_{e_1=0}^{2\hat{e}} (y^{e,e_1} - y^{e_1,e}) \qquad \forall e \in [\hat{e}], \\
\hat{x}^t_{e,r} - \hat{a}^t_{e,r} = \sum_{e_1=0}^{2\hat{e}} (y^{e,e_1} - y^{e_1,e}) \qquad \forall \hat{e} < e, \\
y^{e_1,e_2} \geq 0 \qquad \forall e_1, e_2.
\end{aligned}
$$

To see why the above is true, we first need to modify the function h (previously defined in (1.3.5)) which, given the detailed inventory vector $\boldsymbol{x}^t$ and the empirical action $\bar{\boldsymbol{a}}^t$, gives the detailed action $\boldsymbol{a}^t$ with marginal $\bar{\boldsymbol{a}}^t$ that minimizes the rebalancing cost. Under fixed costs, the modified h function is given by:

$$
\begin{aligned}
h(\bar{\boldsymbol{a}}) \in \arg\min_{\boldsymbol{a} \in \boldsymbol{A}} \sum_{e=1}^{\hat{e}} \sum_{i=1}^{n_e} c_{f,e} \mathbb{I} \left\{ x^t_{e,i} \neq a^t_{e,i} \right\} + \sum_{e=1}^{\hat{e}} \sum_{i=1}^{n_e} \left| a^t_{e,i} - x^t_{e,i} \right| + \sum_{e=1}^{\hat{e}} \left| a^t_{e,r} - x^t_{e,r} \right| \qquad (1.6.2) \\
\text{s.t.} \quad \frac{\sum_{i=1}^{n_e} \mathbb{I} \left\{ a^t_{e,i} = d \right\}}{n} = \bar{a}^t_{e,d} \quad \forall e, d, \\
a^t_{e,r} = n \bar{a}^t_{e,r} \quad \forall e.
\end{aligned}
$$

It is straightforward to observe that the total variation distance component (first term) of (1.6.1) constitutes a lower bound on the fixed cost component (first term) of (1.6.2). To see why the minimum fixed cost paid under (1.6.2) equals the first term of (1.6.1) (while not affecting the linear rebalancing costs), we need to look back to the optimal coupling, which solves the optimal transport problem in (1.3.5). Without fixed costs, we showed that given an initial inventory vector and a target action vector for stations of a given type, sorting them in increasing order and then matching up the indices minimized total rebalancing cost. This cost was given by the Wasserstein distance between the two vectors. However, the optimal solution for minimizing the Wasserstein distance is not unique. In the presence of fixed costs, we need to ensure that we move from the current to target inventory position by changing inventory at as few stations as possible. To achieve this, we consider the following coupling: Focus on stations of type e. Let k_b^e be the smaller of $n\bar{a}_b^e$ and the number of type e stations with $x_{e,i} = b$. Then k_b^e type e stations will be assigned $x_{e,i} = a_{e,i} = b$ under $\boldsymbol{x}^t$ and $\boldsymbol{a}^t$ so that we do not pay any fixed cost for these stations. The remaining stations of type e will be ordered in increasing order of inventory and matched with the residual action vector as previously. Stations with identical pre- and post-rebalancing inventories do not contribute to the linear component of the rebalancing cost, and this ordering provides the minimum amount of fixed costs paid, giving the first term on the right-hand side of (1.6.1), while not changing the remaining cost terms.

One last issue is that in the proof of Proposition 1.3.3, we upper bound the sub-optimality of the mean field action by the Wasserstein distance between the stochastic inventory evolution and its mean-field evolution, and then utilize the CLT for Wasserstein distance between empirical distribution and the true distribution. However, the new rebalancing cost we provided is not equivalent to a Wasserstein distance so we cannot directly utilize the CLT for Wasserstein distances[20]. To resolve this issue, we first prove the following lemma:

Lemma 1.6.1. *Fixed cost component of the mean-field rebalancing cost satisfies*

$$\sum_{e=1}^{\hat{e}} c_{f,e} \sum_{b=0}^{\infty} \frac{1}{2}\left|\hat{x}_{e,b}^t - \hat{a}_{e,b}^t\right| \leq \sum_{e=1}^{\hat{e}} c_{f,e} \sum_{b=0}^{\infty}\left|\sum_{d=0}^{b}(\hat{x}_{e,d}^t - \hat{a}_{e,d}^t)\right| \quad \forall \hat{\boldsymbol{x}}^t \in \hat{X}, \hat{\boldsymbol{a}}^t \in \hat{A}.$$

Using this lemma, we can show that the fixed cost component can also be upper bounded

20. While not equivalent to the Wasserstein distance, the rebalancing cost under fixed costs is an earth mover distance with $d(u, v) = c_r|u - v| + c_f\mathbb{I}\{u \neq v\}$.

by a Wasserstein distance and follow our previous analysis of applying the CLT for Wasserstein distances on the upper bound, proving asymptotic optimality.

Once we prove the asymptotic optimality of the corresponding mean-field model, the remaining results follow as the methodology used for proving the worst-case bound for the cost of our control algorithm remains identical (including the special case of $q = 0$).

Impact of Adding Fixed Costs

Due to the great importance of fixed costs for some of our applications, we conduct a separate numerical experiment where we incorporate fixed costs. We use the same setup as the other synthetic experiments and vary the fixed cost c_f (assumed to be homogeneous among all types). We adapt all our policies and the worst-case bound, except for the large market policy, to this setting[21]. The results are shown in Figure 1.4.

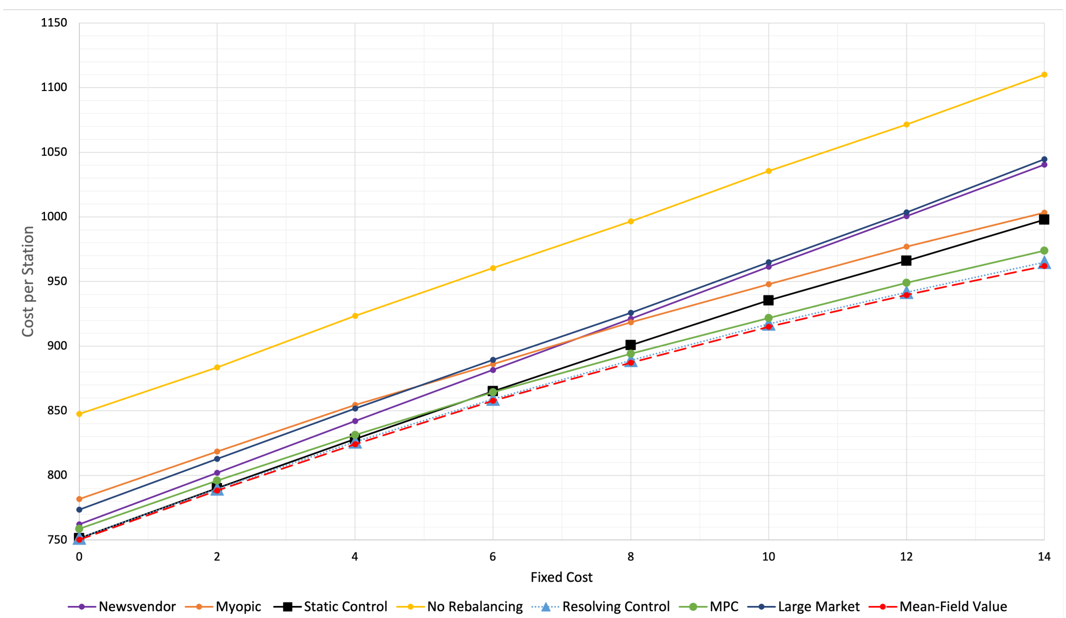


Figure 1.4: Impact of Adding Fixed Costs on the Performance of Policies

Through Figure 1.4, we observe that as fixed costs increase, the performance of the

21. Due to the large market policy structure, incorporating the indicator function providing the fixed costs requires adding an excessive number of binary variables, making the mixed-integer non-linear optimization program computationally difficult to solve.

static control policy decreases compared to resolving-based policies. As fixed costs increase, demand realizations have a more substantial impact on the action, as near-optimal policies aim to change inventory in as few stations as possible (and possibly move more than necessary units at a station for which the fixed cost is already paid). The resolving control policy can adapt to the demand realization and thus is near-optimal for all fixed costs, whereas the static control fails and thus performs worse for higher fixed costs. While not covered in this figure, as the number of stations increases, extreme demand realizations become less common and the static control policy performs well, even for systems with high fixed costs.

Remark 1.6.2. *The main managerial implication is that resolving policies is more effective for systems with high fixed costs and fewer stations. Static policies, in contrast, perform well if the cost structure is more linear (as the dependence of action to initial inventory decreases) or if there are many stations (as state transitions become closer to what static policies expect).*

1.7 Concluding Remarks

We considered the problem of managing resources in shared micromobility systems and developed an MDP capturing the first-order trade-offs relevant to these systems. We proved that a policy based on a deterministic approximation to the stochastic model is asymptotically optimal. We provided an algorithm to solve the associated mean-field model and proved that the optimality gap obtained by using the algorithm decreases exponentially with respect to the transient period. Finally, we compared the performance of our algorithm with other policies numerically.

We believe our setup can be used as a basis to direct future research into expanding these results to a more general framework. Our work made some assumptions, such as restricting the class of networks to cluster-based networks and assuming that demand across stations is independent. This allowed us to use the empirical representation and consequently provided an important step for the asymptotic optimality result. Extending this analysis to more general networks and networks with spillover effects across nodes would provide an important step in expanding the applicability of results.

1.8 A Summary of Notation Used

γ : Discount rate (Section 1.2, Page 8).

n_e : Number of stations at type e (Section 1.2, Page 9).

$\hat{e}$: Number of types (Section 1.2, Page 9).

n : Number of stations (Section 1.2, Page 9).

$p_{e',e}$: Probability that a unit moves from type e' to type e (Section 1.2, Page 10).

$\boldsymbol{x}^t$: Number of units at each station (before rebalancing) (Section 1.2, Page 10).

$\boldsymbol{a}^t$: Number of units at each station (after rebalancing) (Section 1.2, Page 10).

$\boldsymbol{D}^t$: Demand vector (Section 1.2, Page 10).

$\boldsymbol{R}^t$: Trips vector (Section 1.2, Page 11).

$\boldsymbol{L}$: Vector of lower bounds for action (Section 1.2, Page 11).

$\boldsymbol{U}$: Vector of upper bounds for action (Section 1.2, Page 11).

c : Rebalancing cost function for original model (Section 1.2,, Page 12).

c_h : Holding cost (Section 1.2, Page 14).

c_p : Penalty cost (Section 1.2, Page 14).

N : Newsvendor cost function for original model (Section 1.2, Page 14).

V : Value function for original model (Section 1.2, Page 14).

$\hat{\boldsymbol{x}}^t$: Inventory vector for mean-field model (Section 1.3, Page 17).

$\hat{\boldsymbol{a}}^t$: Action vector for mean-field model (Section 1.3, Page 17).

$\hat{R}^e$: Random variable denoting inflow to a station of type e (Section 1.3, Page 18).

$\hat{c}$: Rebalancing cost function for mean-field model (Section 1.3, Page 20).

$\hat{N}$: Newsvendor cost function for mean-field model (Section 1.3, Page 20).

$\hat{V}$: Value function for mean-field model (Section 1.3, Page 21).

g : Function mapping $\boldsymbol{X} \to \hat{\boldsymbol{X}}$ (Section 1.3, Page 21).

$\bar{\boldsymbol{a}}^t$: Rounded action vector of mean-field model (Section 1.3, Page 22).

f : Function mapping $\hat{\boldsymbol{A}} \to \bar{\boldsymbol{A}}$ (Section 1.3, Page 22).

h : Function mapping $\bar{\boldsymbol{A}} \to \boldsymbol{A}$ (Section 1.3, Page 23).

$\breve{\boldsymbol{x}}^t$: Inventory vector for mean-field model restricted within thresholds (Section 1.3, Page 28).

$\hat{c}_2$: Modified rebalancing cost function for mean-field model (Section 1.3, Page 28).

T : Transient horizon for the control algorithm (Section 1.4, Page 29).

$\tilde{\boldsymbol{\pi}}$: Policy outputted through the control algorithm (Section 1.4, Page 30).

1.9 Proofs of Section 1.2 Results

In this section, we will prove an alternative formulation for the rebalancing function.

Proposition 1.2.1. *For all $\boldsymbol{x}^t, \boldsymbol{a}^t$, the rebalancing cost $c(\boldsymbol{x}^t, \boldsymbol{a}^t)$ can be alternatively expressed as:*

$$
\begin{aligned}
c(\boldsymbol{x}^t, \boldsymbol{a}^t) = \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{i=1}^{n_e} \left| a_{e,i}^t - x_{e,i}^t \right| + \min_{y^{e_1,e_2}} \sum_{e_1=0}^{2\hat{e}} \sum_{e_2=0, e_2 \neq e_1}^{2\hat{e}} \left(c_{e_1,e_2} - \frac{c_{e_1,e_1}}{2} - \frac{c_{e_2,e_2}}{2} \right) y^{e_1,e_2}
\end{aligned}
$$

$$
\begin{aligned}
s.t. \quad & \sum_{i=1}^{n_e} (x_{e,i}^t - a_{e,i}^t) = \sum_{e_1=0}^{2\hat{e}} (y^{e,e_1} - y^{e_1,e}) \qquad \forall e \in [\hat{e}], \\
& x_{e,r}^t - a_{e,r}^t = \sum_{e_1=0}^{2\hat{e}} (y^{e,e_1} - y^{e_1,e}) \qquad \forall e > \hat{e}, \\
& y^{e_1,e_2} \geq 0 \qquad \forall e_1, e_2.
\end{aligned}
$$

Proof. We can express the objective function of the original rebalancing cost function as:

$$
\begin{aligned}
\sum_{e_1=0}^{2\hat{e}} \sum_{e_2=0}^{2\hat{e}} \sum_{i=1}^{n_{e_1}} \sum_{j=1}^{n_{e_2}} c_{e_1,e_2} y_{i,j}^{e_1,e_2} &= \sum_{e_1=0}^{2\hat{e}} \sum_{e_2=0, e_2 \neq e_1}^{2\hat{e}} \sum_{i=1}^{n_{e_1}} \sum_{j=1}^{n_{e_2}} \left(c_{e_1,e_2} - \frac{c_{e_1,e_1}}{2} - \frac{c_{e_2,e_2}}{2} \right) y_{i,j}^{e_1,e_2} \\
&\quad + \sum_{e_1=0}^{2\hat{e}} \sum_{e_2=0, e_2 \neq e_1}^{2\hat{e}} \sum_{i=1}^{n_{e_1}} \sum_{j=1}^{n_{e_2}} \frac{c_{e_1,e_1}}{2} y_{i,j}^{e_1,e_2} \\
&\quad + \sum_{e_1=0}^{2\hat{e}} \sum_{e_2=0, e_2 \neq e_1}^{2\hat{e}} \sum_{i=1}^{n_{e_1}} \sum_{j=1}^{n_{e_2}} \frac{c_{e_2,e_2}}{2} y_{i,j}^{e_1,e_2} + \sum_{e_1=0}^{2\hat{e}} \sum_{i=1}^{n_{e_1}} \sum_{j=1}^{n_{e_2}} c_{e_1,e_1} y_{i,j}^{e_1,e_2} \\
&= \sum_{e_1=0}^{2\hat{e}} \sum_{e_2=0, e_2 \neq e_1}^{2\hat{e}} \left(c_{e_1,e_2} - \frac{c_{e_1,e_1}}{2} - \frac{c_{e_2,e_2}}{2} \right) \sum_{i=1}^{n_{e_1}} \sum_{j=1}^{n_{e_2}} y_{i,j}^{e_1,e_2} \\
&\quad + \sum_{e_1=0}^{2\hat{e}} \sum_{i=1}^{n_{e_1}} \frac{c_{e_1,e_1}}{2} \sum_{e_2=0}^{2\hat{e}} \sum_{j=1}^{n_{e_2}} y_{i,j}^{e_1,e_2} + \sum_{e_2=0}^{2\hat{e}} \sum_{j=1}^{n_{e_2}} \frac{c_{e_2,e_2}}{2} \sum_{e_1=0}^{2\hat{e}} \sum_{i=1}^{n_{e_1}} y_{i,j}^{e_1,e_2}. \qquad (1.9.1)
\end{aligned}
$$

Given this formulation, we aggregate the inflows and outflows to/from each station, with

$$
\begin{aligned}
y^{e_1,e_2} &= \sum_{i=1}^{n_{e_1}} \sum_{j=1}^{n_{e_2}} y_{i,j}^{e_1,e_2} && \forall e_1, e_2, \\
y_{i,\cdot}^{e_1,\cdot} &= \sum_{e_2=0}^{2\hat{e}} \sum_{j=1}^{n_{e_2}} y_{i,j}^{e_1,e_2} && \forall e_1, i, \\
y_{\cdot,j}^{\cdot,e_2} &= \sum_{e_1=0}^{2\hat{e}} \sum_{i=1}^{n_{e_1}} y_{i,j}^{e_1,e_2} && \forall e_2, j.
\end{aligned}
$$

Inputting these variables to Equation (1.9.1), we obtain

$$\begin{aligned}
\sum_{e_1=0}^{2\hat{e}}\sum_{e_2=0}^{2\hat{e}}\sum_{i=1}^{n_{e_1}}\sum_{j=1}^{n_{e_2}} c_{e_1,e_2} y_{i,j}^{e_1,e_2} &= \sum_{e_1=0}^{2\hat{e}}\sum_{e_2=0,e_2\neq e_1}^{2\hat{e}} \left(c_{e_1,e_2} - \frac{c_{e_1,e_1}}{2} - \frac{c_{e_2,e_2}}{2}\right) y^{e_1,e_2} \\
&\quad + \sum_{e_1=0}^{2\hat{e}}\sum_{i=1}^{n_{e_1}} \frac{c_{e_1,e_1}}{2} y_{i,\cdot}^{e_1,\cdot} + \sum_{e_2=0}^{2\hat{e}}\sum_{j=1}^{n_{e_2}} \frac{c_{e_2,e_2}}{2} y_{\cdot,j}^{\cdot,e_2} \\
&= \sum_{e_1=0}^{2\hat{e}}\sum_{e_2=0,e_2\neq e_1}^{2\hat{e}} \left(c_{e_1,e_2} - \frac{c_{e_1,e_1}}{2} - \frac{c_{e_2,e_2}}{2}\right) y^{e_1,e_2} \\
&\quad + \sum_{e_1=0}^{2\hat{e}}\sum_{i=1}^{n_{e_1}} \frac{c_{e_1,e_1}}{2} \left(y_{i,\cdot}^{e_1,\cdot} + y_{\cdot,i}^{\cdot,e_1}\right).
\end{aligned}$$

Also expressing the constraints in terms of the aggregated variables, we can rewrite the rebalancing cost $c(\boldsymbol{x}^t, \boldsymbol{a}^t)$ as

$$\begin{aligned}
c(\boldsymbol{x}^t, \boldsymbol{a}^t) = \min & \sum_{e_1=0}^{2\hat{e}}\sum_{e_2=0,e_2\neq e_1}^{2\hat{e}} \left(c_{e_1,e_2} - \frac{c_{e_1,e_1}}{2} - \frac{c_{e_2,e_2}}{2}\right) y^{e_1,e_2} \\
& + \sum_{e_1=0}^{2\hat{e}}\sum_{i=1}^{n_{e_1}} \frac{c_{e_1,e_1}}{2} \left(y_{i,\cdot}^{e_1,\cdot} + y_{\cdot,i}^{\cdot,e_1}\right) \\
\text{s.t.} \quad & a_{e,i}^t = x_{e,i}^t + y_{\cdot,i}^{\cdot,e} - y_{i,\cdot}^{e,\cdot} \qquad \forall e, i, \\
& \sum_{i=1}^{n_e} (x_{e,i}^t - a_{e,i}^t) = \sum_{e_1=0}^{2\hat{e}} (y^{e,e_1} - y^{e_1,e}) \qquad \forall e \in [\hat{e}], \\
& x_{e,r}^t - a_{e,r}^t = \sum_{e_1=0}^{2\hat{e}} (y^{e+\hat{e},e_1} - y^{e_1,e+\hat{e}}) \qquad \forall e \in [\hat{e}], \\
& y^{e_1,e_2}, y_{i,\cdot}^{e_1,\cdot}, y_{\cdot,i}^{\cdot,e_1} \geq 0 \qquad \forall e_1, e_2, i, j.
\end{aligned}$$

Now, we solve the problem sequentially where we fix some feasible $y^{e_1,e_2}\ \forall e_1, e_2$ and solve for the optimal $y_{i,\cdot}^{e_1,\cdot}, y_{\cdot,i}^{\cdot,e_1}\ \forall e_1, i$. As the cost coefficients of both $y_{i,\cdot}^{e_1,\cdot}, y_{\cdot,i}^{\cdot,e_1}$ are identical and non-negative, and the constraint provides $y_{\cdot,i}^{\cdot,e} - y_{i,\cdot}^{e,\cdot} = k$ for some constant k (possibly negative) while both $y^{\cdot,e}_{\cdot,i}, y^{e,\cdot}_{i,\cdot}$ are non-negative, we have that:

$$y^{\cdot,e}_{\cdot,i} * y^{e,\cdot}_{i,\cdot} = 0 \qquad \forall e, i,$$

with at most one of the two terms having a strictly positive value. As a result, the optimal solution of $y^{\cdot,e}_{\cdot,i}, y^{e,\cdot}_{i,\cdot}$ satisfies

$$y^{\cdot,e}_{\cdot,i} + y^{e,\cdot}_{i,\cdot} = |x^t_{e,i} - a^t_{e,i}| \qquad \forall e, i.$$

Inputting the above equation to the objective function, combined with the assumption that $c_{e_1,e_1} = 0 \ \forall e_1 > \hat{e}$, we prove the Proposition. □

1.10 Additional Model Discussions

In this section, we discuss some additional aspects of how our model assumptions relate to the practice of managing resources in micromobility systems.

One assumption which we make is on the linearity of the rebalancing cost function, where in contrast to other cost structures such as including fixed costs, where a fixed cost is paid if any units are moved in or out of a station (independent of the number of units moved in/out), and considering trucking routes, we assume that the cost of rebalancing is linear on the number of units rebalanced[22]. While a linear cost function is a standard modeling assumption in rebalancing papers such as in He et al. [2020] or Benjaafar et al. [2022a], we also believe it is consistent with practice in scooter-sharing systems and is becoming much more standard for practice in bike-sharing systems. For scooter-sharing, as highlighted in Helling [2022], large firms, such as Bird, crowd-source rebalancing/recharging operations by paying freelance contractors for moving or recharging units (these contractors find and complete these rebalancing/recharging operations through the app). The contractors are compensated per unit with no batching, providing a justification for our linearity assumption. For bike-

22. While we do consider fixed costs in Section 1.6, there are many different ways fixed costs can be incorporated into the rebalancing cost function which we do not discuss so we believe that the justification of our current rebalancing cost function is important.

sharing, while trucking based rebalancing still provides of a large portion of rebalancing operations for some firms, crowd-sourcing based alternatives rapidly gain popularity. One major example is the Bike Angels programs currently implemented in cities such as New York and Chicago, where individuals are awarded points for rebalancing individual units across stations, which can be redeemed for awards. Furthermore, trucking based rebalancing spends significant time on the handling of individual units (withdrawing/inputting units to docks or finding and loading/dropping units for dockless systems), which support our linearity assumption.

Another assumption we make is that the action at each station is restricted through an upper threshold. We believe that this assumption follows the practice for both docked and free-floating systems. For a docked system, the total number of docks at a station provides a natural upper threshold[23]. For flee-floating systems, while there is no limit on the total number of units returned to a specific location (customers, barring some urban feasibility constraints, are free to drop the units in locations of their choice), most platforms have a contractual obligation to remove these "excess" units. This obligation is best indicated in the permit provided by Los Angeles to free-floating micromobility systems, which states that "Operators shall remove electric scooters from the public right-of-way on a daily basis" (City of Los Angeles Department of Transportation [2018]).

1.11 An Alternative State-Space Representation

While proving the results of Section 1.3, we need the assistance of an interim model, which looks at inventory and action as an empirical distribution, as in the mean-field model, but also is stochastic and provides identical costs to the original model. This section introduces this model, which we label as the distributional model. We will first define quantities analogous to the variables we introduced in Section 1.3 (state, cost functions, policies). Then, we will

23. While docked systems also place a limit on the total number of units which can be returned to a station, such platforms have regularly implemented services such as valet service, where an employee is placed in high inflow stations to take excess units in order to eliminate this limit.

prove that this model is, in fact, equivalent to the original model and use this equivalence to prove the results of Section 1.3.

In accordance, we introduce an alternative empirical representation for inventory, where $\bar{\boldsymbol{x}}^{t,e} = [\bar{x}_d^{t,e}]_{d=0}^{\infty}$ with $\bar{\boldsymbol{X}} \subset \hat{\boldsymbol{X}}$ as a subset of mean-field inventory space satisfying:

$$\bar{\boldsymbol{x}}^t = [\bar{x}_r^{t,e}, \bar{\boldsymbol{x}}^{t,e}]_{e=1}^{\hat{e}} \in \bar{\boldsymbol{X}} \iff \bar{\boldsymbol{x}}^t \in \hat{\boldsymbol{x}};\ n\bar{x}_d^{t,e} \in \mathbb{N}_0\ \forall e, d;\ n\bar{x}_r^{t,e} \in \mathbb{N}_0\ \forall e.$$

For action, we use $\bar{\boldsymbol{a}}^t$, which we already defined in Section 1.3. We define $\bar{\boldsymbol{R}}_{e'}^t = [\bar{\boldsymbol{R}}_{e',1}^t, \bar{\boldsymbol{R}}_{e',2}^t, \cdots, \bar{\boldsymbol{R}}_{e',\hat{e}}^t]$ where $\bar{\boldsymbol{R}}_{e',e}^t = [\bar{R}_{e',e,r}^t, \bar{R}_{e',e,1}^t, \cdots, \bar{R}_{e',e,n_e}^t]$, is a multinomial distribution with $\sum_{b=L^{e'}}^{U^{e'}} \sum_{i=\sum_{k=0}^{b-1} n\bar{a}_{e',k}^t+1}^{\sum_{k=0}^{b} n\bar{a}_{e',k}^t} \min(b, D_{e',i}^t)$ trials and success probability $(1-q)p_{e',e}$ for $\bar{R}_{e',e,r}^t$, and success probabilities $\frac{qp_{e',e}}{n_e}$ for remaining terms. In order to construct the model, we first introduce the inventory dynamics, where given action $\bar{\boldsymbol{a}}^t$ for period t, we express the state at period $t+1$ through the following equations:

$$\begin{aligned}
\bar{x}_{e,d}^{t+1} &= \frac{1}{n} \sum_{b=L^e}^{U^e} \sum_{i=\sum_{k=0}^{b-1} n\bar{a}_{e,k}^t+1}^{\sum_{k=0}^{b} n\bar{a}_{e,k}^t} \mathbb{I}\left\{ b - \min(b, D_{e,i}^t) + \sum_{e'=1}^{\hat{e}} \bar{R}_{e',e,i}^t = d \right\} \qquad \forall e, d, \\
\bar{x}_{e,r}^{t+1} &= \frac{1}{n}\left(n\bar{a}_{e,r}^t + \sum_{e'=1}^{\hat{e}} \bar{R}_{e',e,r}^t \right).
\end{aligned}$$

We use the same newsvendor and rebalancing cost functions as the mean-field model. Then, given the alternative state and action representation and cost functions, we need to define the class of policies for the distributional model. We denote policies for the distributional model by $\bar{\boldsymbol{\pi}} = \{\bar{\pi}^t\}_{t\in\mathbb{N}}$, where $\bar{\pi}^t : \bar{\boldsymbol{X}} \to \bar{\boldsymbol{A}}$ denotes the time t policy which maps the empirical representation of the state to the action. We label the set of all policies $\bar{\boldsymbol{\pi}}$ as $\bar{\boldsymbol{\Pi}}$. For a policy $\bar{\boldsymbol{\pi}}$, $\bar{V}_{\bar{\boldsymbol{\pi}}}(\bar{\boldsymbol{x}}^t)$ denotes the discounted cost-to-go starting with inventory position $\bar{\boldsymbol{x}}^t$ under the distributional model:

$$\bar{V}_{\bar{\boldsymbol{\pi}}}(\bar{\boldsymbol{x}}^t) = \mathbb{E}_{\{\bar{\boldsymbol{D}}^{t+k}, \bar{\boldsymbol{R}}^{t+k}\}_{k=0}^{\infty}} \left[\sum_{s=t}^{\infty} \gamma^{s-t} \Big(\hat{c}(\bar{\boldsymbol{x}}^s, \bar{\pi}^s(\bar{\boldsymbol{x}}^s)) + \hat{N}(\bar{\pi}^s(\bar{\boldsymbol{x}}^s)) \Big) \right].$$

The optimal value function for the distributional model is defined as

$$\bar{V}(\bar{\boldsymbol{x}}^t) = \min_{\bar{\boldsymbol{\pi}} \in \bar{\boldsymbol{\Pi}}} \bar{V}_{\bar{\boldsymbol{\pi}}}(\bar{\boldsymbol{x}}^t).$$

Similar to the previous models, we express the value function as a fixed point of the Bellman recursion:

$$\bar{V}(\bar{\boldsymbol{x}}^t) = \min_{\bar{\boldsymbol{a}}^t \in \bar{\boldsymbol{A}}} \hat{c}(\bar{\boldsymbol{x}}^t, \bar{\boldsymbol{a}}^t) + \hat{N}(\bar{\boldsymbol{a}}^t) + \gamma \mathbb{E}_{\{\bar{\boldsymbol{D}}^t, \bar{\boldsymbol{R}}^t\}} \left[\bar{V}(\bar{\boldsymbol{x}}^{t+1}) \right].$$

For any inventory position $\bar{\boldsymbol{x}}^t$ and period t, the optimal policy of the distributional model, $\bar{\pi}^*$, satisfies

$$\bar{\boldsymbol{a}}^{t^*} \in \arg\min_{\bar{\boldsymbol{a}}^t \in \bar{\boldsymbol{A}}} \hat{c}(\bar{\boldsymbol{x}}^t, \bar{\boldsymbol{a}}^t) + \hat{N}(\bar{\boldsymbol{a}}^t) + \gamma \mathbb{E}_{\{\bar{\boldsymbol{D}}^t, \bar{\boldsymbol{R}}^t\}} \left[\bar{V}(\bar{\boldsymbol{x}}^{t+1}) \right].$$

The next proposition proves the equivalency between the original model and the distributional model.

Proposition 1.11.1. *For any policy $\bar{\boldsymbol{\pi}} \in \bar{\boldsymbol{\Pi}}$ and inventory position $\boldsymbol{x}^t \in \boldsymbol{X}$,*

$$\bar{V}_{\bar{\boldsymbol{\pi}}}(g(\boldsymbol{x}^t)) = V_{h \circ \bar{\boldsymbol{\pi}} \circ g}(\boldsymbol{x}^t).$$

Consequently, the distributional model is the empirical representation of the original model.

Proof. For ease of notation, we let $\bar{\pi}^t(g(\boldsymbol{x}^t)) = \bar{\boldsymbol{a}}^t$, $g(\boldsymbol{x}^t) = \bar{\boldsymbol{x}}^t$, $h(\bar{\pi}^t(g(\boldsymbol{x}^t))) = \boldsymbol{a}^t$. Through the definition of the h function, $\bar{\boldsymbol{a}}^t$ satisfies

$$\bar{a}^t_{e,d} = \frac{\sum_{i=1}^{n_e} \mathbb{I}\left\{ a^t_{e,i} = d \right\}}{n} \quad \forall e, d,$$
$$n\bar{a}^t_{e,r} = a^t_{e,r} \quad \forall e, d.$$

We will show that if we apply the given action of the distributional model to the original model, we get the same cost as the distributional model and observe that the marginals $\bar{\boldsymbol{x}}^{t+1}$ of the next period's inventory position is consistent for both models. Formally, we will prove that the following four statements hold:

1. $c(\boldsymbol{x}^t, \boldsymbol{a}^t) = \hat{c}(\bar{\boldsymbol{x}}^t, \bar{\boldsymbol{a}}^t)$;

2. $N(\boldsymbol{a}^t) = \hat{N}(\bar{\boldsymbol{a}}^t)$;

3. $x_{e,r}^{t+1} = n\bar{x}_{e,r}^{t+1} \quad \forall e$;

4. $\frac{\sum_{i=1}^{n_e} \mathbb{I}\left\{x_{e,i}^{t+1}=d\right\}}{n} = \bar{x}_{e,d}^{t+1} \quad \forall e, d.$

For the rebalancing cost function, we look at the two components separately. First, for the constraints, we have

$$\sum_{i=1}^{n_e}(x_{e,i}^t - a_{e,i}^t) = n\sum_{b=L^e}^{U^e} b\left(\frac{\sum_{i=1}^{n_e}\mathbb{I}\left\{x_{e,i}^t = b\right\}}{n} - \frac{\sum_{i=1}^{n_e}\mathbb{I}\left\{a_{e,i}^t = b\right\}}{n}\right)$$
$$x_{e,r}^t - a_{e,r}^t = n\left(\frac{x_{e,r}^t}{n} - \frac{a_{e,r}^t}{n}\right) \quad \forall e > \hat{e}.$$

As the decision variable of the mean-field rebalancing cost function is also scaled with n, we have that the two optimization problems are equivalent. Second, we have that

$$\begin{aligned}
&\sum_{e=1}^{\hat{e}}\frac{c_{e,e}}{2}\sum_{i=1}^{n_e}\left|a_{e,i}^t - x_{e,i}^t\right| - n\sum_{e=1}^{\hat{e}}\frac{c_{e,e}}{2}\sum_{b=0}^{\infty}\left|\sum_{d=0}^{b}\left(\frac{\sum_{i=1}^{n_e}\mathbb{I}\left\{x_{e,i}^t = b\right\}}{n} - \frac{\sum_{i=1}^{n_e}\mathbb{I}\left\{a_{e,i}^t = b\right\}}{n}\right)\right| \\
&= \sum_{e=1}^{\hat{e}}\frac{c_{e,e}}{2}\sum_{i=1}^{n_e}\left|a_{e,i}^t - x_{e,i}^t\right| - n\sum_{e=1}^{\hat{e}}\frac{c_{e,e}}{2}\sum_{b=0}^{\infty}\left|\frac{1}{n}\sum_{i=1}^{n_e}\left(\mathbb{I}\left\{x_{e,i}^t \le b\right\} - \mathbb{I}\left\{x_{e,i}^t \le b\right\}\right)\right| \\
&= \sum_{e=1}^{\hat{e}}\frac{c_{e,e}}{2}\left(\sum_{i=1}^{n_e}\left|a_{e,i}^t - x_{e,i}^t\right| - n_e\sum_{b=0}^{\infty}\left|\frac{1}{n_e}\sum_{i=1}^{n_e}(\mathbb{I}\left\{x_{e,i}^t \le b\right\} - \mathbb{I}\left\{a_{e,i}^t \le b\right\})\right|\right).
\end{aligned}$$

Under the h function, we rebalance stations by ordering them in terms of their pre-rebalancing inventory, and we assign the highest action to the station with highest inventory, second highest action to the station with second highest inventory and so on. Then, we let $x_{e,\{i\}}^t$ be the inventory of the station with i'th most units which belongs to type e before rebalancing, and let $a_{e,\{i\}}^t$ be the number of units at the station with i'th most units which belongs to type e

after rebalancing. Consequently:

$$\sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{i=1}^{n_e} \left| a_{e,i}^t - x_{e,i}^t \right| - n \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{b=0}^{\infty} \left| \sum_{d=0}^{b} \left(\frac{\sum_{i=1}^{n_e} \mathbb{I}\left\{ x_{e,i}^t = b \right\}}{n} - \frac{\sum_{i=1}^{n_e} \mathbb{I}\left\{ a_{e,i}^t = b \right\}}{n} \right) \right|$$

$$= \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \left(\sum_{i=1}^{n_e} \left| a_{e,\{i\}}^t - x_{e,\{i\}}^t \right| - n_e \sum_{b=0}^{\infty} \left| \frac{\sum_{i=1}^{n_e} \mathbb{I}\left\{ x_{e,\{i\}}^t \leq b \right\}}{n_e} - \frac{\sum_{i=1}^{n_e} \mathbb{I}\left\{ a_{e,\{i\}}^t \leq b \right\}}{n_e} \right| \right).$$

Using Lemma 1.11.2, we establish that $\forall e \in [\hat{e}]$,

$$\sum_{i=1}^{n_e} \left| a_{e,\{i\}}^t - x_{e,\{i\}}^t \right| = n_e \sum_{b=0}^{\infty} \left| \frac{\sum_{i=1}^{n_e} \mathbb{I}\left\{ x_{e,\{i\}}^t \leq b \right\}}{n_e} - \frac{\sum_{i=1}^{n_e} \mathbb{I}\left\{ a_{e,\{i\}}^t \leq b \right\}}{n_e} \right| \Bigg).$$

Consequently, we have shown the equivalence of rebalancing costs. For the equivalence of newsvendor costs, we show that

$$\begin{aligned} N(\boldsymbol{a}^t) - \hat{N}(\bar{\boldsymbol{a}}^t) = {} & c_h \sum_{e=1}^{\hat{e}} \sum_{i=1}^{n_e} a_{e,i}^t + c_h \sum_{e=1}^{\hat{e}} a_{e,r}^t + c_p \sum_{e=1}^{\hat{e}} \sum_{i=1}^{n_e} \mathbb{E}\left[(D_{e,i}^t - a_{e,i}^t)^+ \right] \\ & - n c_h \sum_{e=1}^{\hat{e}} \sum_{b=L^e}^{U^e} b \frac{\sum_{i=1}^{n_e} \mathbb{I}\left\{ a_{e,i}^t = b \right\}}{n} \\ & - n c_h \sum_{e=1}^{\hat{e}} \frac{a_{e,r}^t}{n} - n c_p \sum_{e=1}^{\hat{e}} \sum_{b=L^e}^{U^e} \frac{\sum_{i=1}^{n_e} \mathbb{I}\left\{ a_{e,i}^t = b \right\}}{n} \mathbb{E}\left[(D^e - b)^+ \right], \end{aligned}$$

as $a_{e,i}^t = \sum_{b=L^e}^{U^e} \mathbb{I}\left\{ a_{e,i}^t = b \right\} b$,

$$= 0.$$

For the equivalence of inventory evolution of depleted units, we need to show that $\forall e$, $\boldsymbol{R}_e^t = \bar{\boldsymbol{R}}_e^t$. We have that the success probabilities for both multinomial distributions are equal. For number of trials, we have that

$$\begin{aligned} \sum_{i=1}^{n_{e'}} \min(a_{e',i}^t, D_{e',i}^t) &= \sum_{b=L^{e'}}^{U^{e'}} \sum_{i=1}^{n_{e'}} \mathbb{I}\left\{ a_{e',i}^t = b \right\} \min(b, D_{e',i}^t) \\ &= \sum_{b=L^{e'}}^{U^{e'}} \sum_{i=\sum_{k=0}^{b-1} n\bar{a}_{e',k}^t + 1}^{\sum_{k=0}^{b} n\bar{a}_{e',k}^t} \min(b, D_{e',i}^t) \end{aligned}$$

As a result, the inventory distributions induced for depleted units under the two trip distributions are equivalent with

$$\begin{aligned}
x_{e,r}^{t+1} &= a_{e,r}^{t} + \sum_{e'=1}^{\hat{e}} R_{e',e,r}^{t} \\
&= n\bar{x}_{e,r}^{t+1}.
\end{aligned}$$

For the equivalence of inventory evolution of the stations, we show that

$$\begin{aligned}
\frac{\sum_{i=1}^{n_e} \mathbb{I}\left\{x_{e,i}^{t+1} = d\right\}}{n} &= \frac{1}{n}\sum_{i=1}^{n_e} \mathbb{I}\left\{a_{e,i}^{t} - \min(a_{e,i}^{t}, D_{e,i}^{t}) + \sum_{e'=1}^{\hat{e}} R_{e',e,i}^{t} = d\right\} \\
&= \frac{1}{n}\sum_{b=L^e}^{U^e}\sum_{i=1}^{n_e} \mathbb{I}\left\{b - \min(b, D_{e,i}^{t}) + \sum_{e'=1}^{\hat{e}} R_{e',e,i}^{t} = d\right\}\mathbb{I}\left\{a_{e,i}^{t} = b\right\} \\
&= \frac{1}{n}\sum_{b=L^e}^{U^e}\sum_{i=\sum_{k=0}^{b-1} n\bar{a}_{e,k}^{t}+1}^{\sum_{k=0}^{b} n\bar{a}_{e,k}^{t}} \mathbb{I}\left\{b - \min(b, D_{e,i}^{t}) + \sum_{e'=1}^{\hat{e}} \bar{R}_{e',e,i}^{t} = d\right\} \\
&= \bar{x}_{e,d}^{t+1}.
\end{aligned}$$

Consequently, we proved all four of the statements and hence the Proposition. □

Lemma 1.11.2. *(Bobkov and Ledoux [2019], Theorem 2.9 and Lemma 4.2) Given two collections of real numbers $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$, let F^1 and F^2 be the respective distribution functions. Furthermore, let $x_1^* \leq x_2^* \leq \ldots \leq x_n^*$ correspond to $x_1, \ldots, x_n$ arranged in increasing order and let $y_1^* \leq y_2^* \leq \ldots \leq y_n^*$ correspond to $y_1, \ldots, y_n$ arranged in increasing order. Then:*

$$\int_{-\infty}^{\infty} \left|F^1(t) - F^2(t)\right| dt = \frac{1}{n}\sum_{i=1}^{n}\left|x_i^* - y_i^*\right|.$$

We further extend Proposition 1.11.1 and prove that the equivalence between the distributional and original models extends to the value functions of the two models.

Corollary 1.11.3. *For any inventory position $\boldsymbol{x}^t \in \boldsymbol{X}$,*

$$\bar{V}(g(\boldsymbol{x}^t)) = V(\boldsymbol{x}^t).$$

Proof. Starting from Proposition 1.11.1, we have

$$\begin{aligned}
\bar{V}_{\bar{\boldsymbol{\pi}}}(g(\boldsymbol{x}^t)) &= V_{h\circ\bar{\boldsymbol{\pi}}\circ g}(\boldsymbol{x}^t) \\
\bar{V}(g(\boldsymbol{x}^t)) &= \min_{\bar{\boldsymbol{\pi}}\in\bar{\boldsymbol{\Pi}}} V_{h\circ\bar{\boldsymbol{\pi}}\circ g}(\boldsymbol{x}^t) \\
&= \min_{\bar{\boldsymbol{\pi}}\in\bar{\boldsymbol{\Pi}}} \mathbb{E}_{\{\boldsymbol{D}^{t+k},\boldsymbol{R}^{t+k}\}_{k=0}^{\infty}}\left[\sum_{s=t}^{\infty}\gamma^{s-t}\Big(c(\boldsymbol{x}^s, h(\bar{\pi}^s(g(\boldsymbol{x}^s)))) + N(h(\bar{\pi}^s(g(\boldsymbol{x}^s)))\Big)\right],
\end{aligned}$$

using the proof of Proposition 1.11.1,

$$\begin{aligned}
&= \min_{\bar{\boldsymbol{\pi}}\in\bar{\boldsymbol{\Pi}}} \mathbb{E}_{\{\boldsymbol{D}^{t+k},\bar{\boldsymbol{R}}^{t+k}\}_{k=0}^{\infty}}\left[\sum_{s=t}^{\infty}\gamma^{s-t}\Big(\hat{c}(g(\boldsymbol{x}^s), \bar{\pi}^s(g(\boldsymbol{x}^s))) + \hat{N}(\bar{\pi}^s(g(\boldsymbol{x}^s)))\Big)\right] \\
&= \bar{V}(g(\boldsymbol{x}^t)). \qquad \square
\end{aligned}$$

1.12 Proofs of Section 1.3 Results

This section focuses of the three results of Section 1.3, namely Proposition 1.3.3, Theorem 1.3.4, and Corollary 1.3.5. We prove these results by supporting Lemmas and Propositions, which are provided after the proofs of these results.

Proposition 1.3.3. *For an arbitrary action $\boldsymbol{a}^t$ taken in the original model, let $\boldsymbol{x}^{t+1}$ be the state evolution defined in* (1.2.1)-(1.2.2)*, and let $\hat{\boldsymbol{x}}^{t+1}$ be the state evolution of the mean-field dynamics defined in* (1.3.1)-(1.3.3) *under the projected mean-field action $g(\boldsymbol{a}^t)$. Then,*

$$\mathbb{E}_{\{\boldsymbol{D}^t,\boldsymbol{R}^t\}}\left[\hat{c}(g(\boldsymbol{x}^{t+1}), \hat{\boldsymbol{x}}^{t+1})\big|\hat{\boldsymbol{a}}^t = g(\boldsymbol{a}^t)\right] \leq \mathcal{O}(\sqrt{n}) \quad \forall \boldsymbol{a}^t \in \boldsymbol{A}.$$

Consequently, given the same initial action, the expected rebalancing cost between next period's inventory distribution under the original and mean-field models is of order $\mathcal{O}(\sqrt{n})$.

Proof. The Proof of Proposition 1.3.3 is lengthy. It requires several supporting lemmas and propositions, which show important properties for our rebalancing cost functions and show that our state transitions satisfy the conditions of the Central Limit Theorem (CLT) for Wasserstein distances. In this proof, we will first upper-bound the rebalancing cost function through a Wasserstein Distance (utilizing a feasible rebalancing policy where all units are sourced from the warehouse). Then, through this upper bound, we will evaluate the expected rebalancing cost of moving from next period's inventory distribution under the distributional model (established in Appendix 1.11) to the next period's inventory distribution under the mean-field model (given that the actions two models take are identical). First, we will show that the expected rebalancing cost between depleted units is of order $\mathcal{O}(\sqrt{n})$. Second, to show the same result for charged units, we will require the assistance of another interim system, where the total number of units departing each type are Poisson random variables. Then, using a coupling argument, we will show that the expected rebalancing cost of rebalancing charged units from the distributional model to this interim model is of order $\mathcal{O}(\sqrt{n})$. Lastly, we will conclude our proof by invoking CLT on the expected rebalancing cost between this interim system and the mean-field state transitions.

We start the proof by letting $\bar{\boldsymbol{a}}^t = g(\boldsymbol{a}^t)$. Then, using the proof of Proposition 1.11.1, we can establish that

$$\mathbb{E}_{\{\boldsymbol{D}^t,\boldsymbol{R}^t\}}\Big[\hat{c}(g(\boldsymbol{x}^{t+1}),\hat{\boldsymbol{x}}^{t+1})\big|\hat{\boldsymbol{a}}^t = g(\boldsymbol{a}^t)\Big] = \mathbb{E}_{\{\bar{\boldsymbol{D}}^t,\bar{\boldsymbol{R}}^t\}}\Big[\hat{c}(\bar{\boldsymbol{x}}^{t+1},\hat{\boldsymbol{x}}^{t+1})\big|\hat{\boldsymbol{a}}^t = \bar{\boldsymbol{a}}^t\Big].$$

Furthermore, through Lemma 1.12.1, we have:

$$\begin{aligned}
&\mathbb{E}_{\{\boldsymbol{D}^t,\bar{\boldsymbol{R}}^t\}}\Big[\hat{c}(\bar{\boldsymbol{x}}^{t+1},\hat{\boldsymbol{x}}^{t+1})\big|\hat{\boldsymbol{a}}^t = \bar{\boldsymbol{a}}^t\Big] \\
&\le n\mathbb{E}_{\{\boldsymbol{D}^t,\bar{\boldsymbol{R}}^t\}}\Bigg[\sum_{e=1}^{\hat{e}}\max(c_{0,e},c_{e,0})\sum_{b=0}^{\infty}\Bigg|\sum_{d=0}^{b}(\hat{x}_{e,d}^{t+1}-\bar{x}_{e,d}^{t+1})\Bigg| \\
&\qquad + \sum_{e=1}^{\hat{e}}\max(c_{0,\hat{e}+e},c_{\hat{e}+e,0})\Big|\hat{x}_{e,r}^{t+1}-\bar{x}_{e,r}^{t+1}\Big|\Big|\hat{\boldsymbol{a}}^t = \bar{\boldsymbol{a}}^t\Bigg]
\end{aligned} \tag{1.12.1}$$

For the depleted units, we have:

$$
\begin{aligned}
&n\mathbb{E}_{\{\boldsymbol{D}^t,\bar{\boldsymbol{R}}^t\}}\left[\sum_{e=1}^{\hat{e}}\max(c_{0,\hat{e}+e},c_{\hat{e}+e,0})\left|\hat{x}_{e,r}^{t+1}-\bar{x}_{e,r}^{t+1}\right|\left|\hat{\boldsymbol{a}}^t=\bar{\boldsymbol{a}}^t\right.\right]\\
&=n\mathbb{E}_{\{\boldsymbol{D}^t,\bar{\boldsymbol{R}}^t\}}\Bigg[\sum_{e=1}^{\hat{e}}\max(c_{0,\hat{e}+e},c_{\hat{e}+e,0})\Bigg|(1-q)\sum_{e'=1}^{\hat{e}}p_{e',e}\sum_{b=L^{e'}}^{U^{e'}}\bar{a}_{e',b}^t\mathbb{E}\left[\min(b,D^{e'})\right]\\
&\qquad-\frac{1}{n}\sum_{e'=1}^{\hat{e}}\bar{R}_{e',e,r}^t\Bigg|\left|\hat{\boldsymbol{a}}^t=\bar{\boldsymbol{a}}^t\right.\Bigg].
\end{aligned}
\tag{1.12.2}
$$

We will use categorical random variables to track inflows to each station and bound the above expression. To this end, we let $\{\Theta_{j,e'}\}$ be an infinite sequence of *i.i.d.* random variables distributed categorically (of $n+\hat{e}$ possible categories), with $(1-q)p_{e',e}$ for $\bar{R}_{e',e,r}^t$ and success probabilities $\frac{qp_{e',e}}{n_e}$. Formally,

$$
\Theta_{j,e'}=\mathrm{Cat}\left((1-q)p_{e',e}\big|_{e=1}^{\hat{e}},\frac{qp_{e',e}}{n_e}\big|_{e=1}^{\hat{e}}\big|_{i=1}^{n_e}\right),
$$

where $\Theta_{j,e'}=(\Theta_{j,e',e,r}|_{e=1}^{\hat{e}},\Theta_{j,e',e,i}|_{e=1}^{\hat{e}}|_{i=1}^{n_e})\in\{0,1\}^{n+\hat{e}}$ with only one component equal to 1. We can then express $\bar{R}_{e',e,r}^t$ as a sum of Θ values with:

$$
\bar{R}_{e',e,r}^t=\sum_{j=1}^{\sum_{b=L^{e'}}^{U^{e'}}\sum_{i=\sum_{k=0}^{b-1}n\bar{a}_{e',k}^t+1}^{\sum_{k=0}^{b}n\bar{a}_{e',k}^t}\min(b,D_{e',i}^t)}\Theta_{j,e',e,r}.\quad\forall e',e.
$$

Inputting this expression to (1.12.2), we have:

$$
\begin{aligned}
&n\mathbb{E}_{\{\boldsymbol{D}^t,\bar{\boldsymbol{R}}^t\}}\left[\sum_{e=1}^{\hat{e}}\max(c_{0,\hat{e}+e},c_{\hat{e}+e,0})\left|\hat{x}^{t+1}_{e,r}-\bar{x}^{t+1}_{e,r}\right|\left|\hat{\boldsymbol{a}}^t=\bar{\boldsymbol{a}}^t\right.\right]\\
&\quad= n\mathbb{E}_{\{\boldsymbol{D}^t,\bar{\boldsymbol{R}}^t\}}\Bigg[\sum_{e=1}^{\hat{e}}\max(c_{0,\hat{e}+e},c_{\hat{e}+e,0})\Bigg|(1-q)\sum_{e'=1}^{\hat{e}}p_{e',e}\sum_{b=L^{e'}}^{U^{e'}}\bar{a}^t_{e',b}\mathbb{E}\left[\min(b,D^{e'})\right]\\
&\quad\quad-\frac{1}{n}\sum_{e'=1}^{\hat{e}}\sum_{j=1}^{\sum_{b=L^{e'}}^{U^{e'}}\sum_{i=\sum_{k=0}^{b-1}n\bar{a}^t_{e,k}+1}^{\sum_{k=0}^{b}n\bar{a}^t_{e,k}}\min(b,D^t_{e',i})}\Theta_{j,e',e,r}\Bigg|\left|\hat{\boldsymbol{a}}^t=\bar{\boldsymbol{a}}^t\right.\Bigg]\\
&\leq\sum_{e=1}^{\hat{e}}\sum_{e'=1}^{\hat{e}}\sum_{b=L^{e'}}^{U^{e'}}\max(c_{0,\hat{e}+e},c_{\hat{e}+e,0})\mathbb{E}_{\{\boldsymbol{D}^t,\bar{\boldsymbol{R}}^t\}}\Bigg[\Bigg|(1-q)p_{e',e}n\bar{a}^t_{e,b}\mathbb{E}\left[\min(b,D^e)\right]\\
&\quad-\sum_{j=1}^{\sum_{i=\sum_{k=0}^{b-1}n\bar{a}^t_{e,k}+1}^{\sum_{k=0}^{b}n\bar{a}^t_{e,k}}\min(b,D^t_{e',i})}\Theta_{j,e',e,r}\Bigg|\left|\hat{\boldsymbol{a}}^t=\bar{\boldsymbol{a}}^t\right.\Bigg],
\end{aligned}
$$

by Lemma 1.12.3,

$$\leq\mathcal{O}(\sqrt{n}).$$

Inputting this bound to Equation (1.12.1),

$$
\begin{aligned}
&\mathbb{E}_{\{\boldsymbol{D}^t,\bar{\boldsymbol{R}}^t\}}\left[\hat{c}(\bar{\boldsymbol{x}}^{t+1},\hat{\boldsymbol{x}}^{t+1})\middle|\hat{\boldsymbol{a}}^t=\bar{\boldsymbol{a}}^t\right]\\
&\quad\leq\mathcal{O}(\sqrt{n})+n\mathbb{E}_{\{\boldsymbol{D}^t,\bar{\boldsymbol{R}}^t\}}\left[\sum_{e=1}^{\hat{e}}\max(c_{0,e},c_{e,0})\sum_{b=0}^{\infty}\left|\sum_{d=0}^{b}(\hat{x}^{t+1}_{e,d}-\bar{x}^{t+1}_{e,d})\right|\left|\hat{\boldsymbol{a}}^t=\bar{\boldsymbol{a}}^t\right.\right].
\end{aligned}
$$

As a stepping stone towards bounding the above term, we introduce an interim system, denoted as $\acute{\boldsymbol{x}}^t$, where the total number of units departing each type e', denoted as $\acute{Y}^t_{e'}$, is an independent Poisson random variable with mean $\sum_{b=L^{e'}}^{U^{e'}}n\bar{a}^t_{e',b}\mathbb{E}\left[\min(b,D^{e'})\right]$ (the independence property will later be necessary when applying CLT to the interim system). We define $\acute{\boldsymbol{x}}^t$ on the same probability space $(\Omega,\mathcal{F},\mathcal{P})$ as $\bar{\boldsymbol{x}}^t$ through coupling their trip matrices. Formally:

$$
\begin{aligned}
&n\mathbb{E}_{\{\boldsymbol{D}^t,\bar{\boldsymbol{R}}^t\}}\left[\sum_{e=1}^{\hat{e}}\max(c_{0,e},c_{e,0})\sum_{m=0}^{\infty}\left|\sum_{d=0}^{m}(\hat{x}_{e,d}^{t+1}-\bar{x}_{e,d}^{t+1}\right|\middle|\hat{\boldsymbol{a}}^t=\bar{\boldsymbol{a}}^t\right]\\
&\qquad\le\mathcal{O}(\sqrt{n})+n\mathbb{E}_{\{\boldsymbol{D}^t,\acute{\boldsymbol{R}}^t\}}\left[\sum_{e=1}^{\hat{e}}\max(c_{0,e},c_{e,0})\sum_{m=0}^{\infty}\left|\sum_{d=0}^{m}(\hat{x}_{e,d}^{t+1}-\acute{x}_{e,d}^{t+1})\right|\middle|\hat{\boldsymbol{a}}^t=\bar{\boldsymbol{a}}^t\right]\\
&\qquad\quad+n\mathbb{E}_{\{\bar{\boldsymbol{D}}^t,\bar{\boldsymbol{R}}^t,\acute{\boldsymbol{R}}^t\}}\left[\sum_{e=1}^{\hat{e}}\max(c_{0,e},c_{e,0})\sum_{m=0}^{\infty}\left|\sum_{d=0}^{m}(\acute{x}_{e,d}^{t+1}-\bar{x}_{e,d}^{t+1})\right|\middle|\hat{\boldsymbol{a}}^t=\bar{\boldsymbol{a}}^t\right],
\end{aligned}
\tag{1.12.3}
$$

where

$$
\acute{x}_{e,d}^{t+1}=\frac{1}{n}\sum_{b=L^e}^{U^e}\sum_{i=\sum_{k=0}^{b-1}n\bar{a}_{e,k}^t+1}^{\sum_{k=0}^{b}n\bar{a}_{e,k}^t}\mathbb{I}\left\{b-\min(b,D_{e,i}^t)+\sum_{e'=1}^{\hat{e}}\acute{R}_{e',e,i}^t=d\right\}\qquad\forall e,d,
$$

$\acute{\boldsymbol{R}}_{e'}^t=[\acute{\boldsymbol{R}}_{e',1}^t,\acute{\boldsymbol{R}}_{e',2}^t,\cdots,\acute{\boldsymbol{R}}_{e',\hat{e}}^t]$, $\acute{\boldsymbol{R}}_{e',e}^t=[\acute{R}_{e',e,r}^t,\acute{R}_{e',e,1}^t,\cdots,\acute{R}_{e',e,n_e}^t]$, $\{\acute{R}_{e',e,i}^t\}|_{i=1}^{n_e}$ is a sequence of *i.i.d.* random variables distributed Poisson with mean $\frac{qnp_{e',e}}{n_e}\sum_{b=L^{e'}}^{U^{e'}}\hat{a}_{e',b}^t\mathbb{E}\left[\min(b,D^{e'})\right]$, $\{\acute{R}_{e',e,r}^t\}|_{e=1}^{\hat{e}}$ is a sequence of independent random variables distributed Poisson with mean $(1-q)np_{e',e}\sum_{b=L^{e'}}^{U^{e'}}\hat{a}_{e',b}^t\mathbb{E}\left[\min(b,D^{e'})\right]$, and $\acute{Y}_{e'}^t=\sum_{e=1}^{\hat{e}}\sum_{i=1}^{n_e}\acute{R}_{e',e,i}^t+\sum_{e=1}^{\hat{e}}\acute{R}_{e',e,r}^t$. For $\bar{\boldsymbol{R}}_{e'}^t$, we know that for each unit $1,\ldots,\sum_{b=L^{e'}}^{U^{e'}}\sum_{i=\sum_{k=0}^{b-1}n\bar{a}_{e',k}^t+1}^{\sum_{k=0}^{b}n\bar{a}_{e',k}^t}\min(b,D_{e',i}^t)$, the probability of that unit moving to i'th station of type e charged is $\frac{qp_{e',e}}{n_e}$. For $\acute{\boldsymbol{R}}_{e'}^t$, conditioning on $\acute{Y}_{e'}^t=C$ where C is a constant, the probability of each unit $1,\ldots,C$ moving to i'th station of type e charged is also $\frac{qp_{e',e}}{n_e}$. We build on this analysis to couple $\acute{\boldsymbol{R}}_{e'}^t$ and $\bar{\boldsymbol{R}}_{e'}^t$ by augmenting the probability space with a countably infinite sequence of categorical random variables. To this end, we utilize the categorical random variables $\{\Theta_{j,e'}\}$ previously introduced. Under these categorical random variables, the trip distributions can be expressed as:

$$\bar{R}^t_{e',e,i} = \sum_{j=1}^{\sum_{b=L^{e'}}^{U^{e'}} \sum_{i=\sum_{k=0}^{b-1} n\bar{a}^t_{e',k}+1}^{\sum_{k=0}^{b} n\bar{a}^t_{e',k}} \min(b, D^t_{e',i})} \Theta_{j,e',e,i} \qquad \forall e', e, i,$$

$$\acute{R}^t_{e',e,i} = \sum_{j=1}^{\acute{Y}^t_{e'}} \Theta_{j,e',e,i} \qquad \forall e', e, i.$$

We will prove that the rebalancing cost of moving from $\bar{\boldsymbol{x}}^t$ to $\acute{\boldsymbol{x}}^t$ is bounded by a function of order $\sqrt{n}$. To do that, we introduce $\bar{\boldsymbol{x}}^{t+0.5}$, which denotes the intermediate step in the inventory evolution where demand realizes at stations but no units have been returned yet. We define $\bar{x}^{t+0.5}_{e,d}$ as

$$\bar{x}^{t+0.5}_{e,d} = \frac{1}{n} \sum_{b=L^e}^{U^e} \sum_{i=\sum_{k=0}^{b-1} n\bar{a}^t_{e,k}+1}^{\sum_{k=0}^{b} n\bar{a}^t_{e,k}} \mathbb{I}\left\{b - \min(b, D^t_{e,i}) = d\right\} \qquad \forall e, d.$$

We can use this definition to express the inventory evolution at stations as

$$\bar{x}^{t+1}_{e,d} = \frac{1}{n} \sum_{b=L^e}^{U^e} \sum_{i=\sum_{k=0}^{b-1} n\bar{x}^{t+0.5}_{e,k}+1}^{\sum_{k=0}^{b} n\bar{x}^{t+0.5}_{e,k}} \mathbb{I}\left\{\sum_{e'=1}^{\hat{e}} \bar{R}^t_{e',e,i} = d - b\right\} \qquad \forall e, d,$$

$$\acute{x}^{t+1}_{e,d} = \frac{1}{n} \sum_{b=L^e}^{U^e} \sum_{i=\sum_{k=0}^{b-1} n\bar{x}^{t+0.5}_{e,k}+1}^{\sum_{k=0}^{b} n\bar{x}^{t+0.5}_{e,k}} \mathbb{I}\left\{\sum_{e'=1}^{\hat{e}} \acute{R}^t_{e',e,i} = d - b\right\} \qquad \forall e, d.$$

Using the above equations, we have that

$$n\mathbb{E}_{\{\bar{\boldsymbol{D}}^t,\bar{\boldsymbol{R}}^t,\acute{\boldsymbol{R}}^t\}}\left[\sum_{e=1}^{\hat{e}}\max(c_{0,e},c_{e,0})\sum_{m=0}^{\infty}\left|\sum_{d=0}^{m}(\acute{x}_{e,d}^{t+1}-\bar{x}_{e,d}^{t+1})\right|\middle|\hat{\boldsymbol{a}}^t=\bar{\boldsymbol{a}}^t\right]=\mathbb{E}_{\{\bar{\boldsymbol{D}}^t,\bar{\boldsymbol{R}}^t,\acute{\boldsymbol{R}}^t\}}\Bigg[$$

$$\sum_{e=1}^{\hat{e}}\max(c_{0,e},c_{e,0})\sum_{m=0}^{\infty}\Bigg|\sum_{d=0}^{m}\sum_{b=L^e}^{U^e}\sum_{i=\sum_{k=0}^{b-1}n\bar{x}_{e,k}^{t+0.5}+1}^{\sum_{k=0}^{b}n\bar{x}_{e,k}^{t+0.5}}\Bigg($$

$$\mathbb{I}\left\{\sum_{e'=1}^{\hat{e}}\acute{R}_{e',e,i}^t=d-b\right\}-\mathbb{I}\left\{\sum_{e'=1}^{\hat{e}}\bar{R}_{e',e,i}^t=d-b\right\}\Bigg)\Bigg|\,\Bigg|\hat{\boldsymbol{a}}^t=\bar{\boldsymbol{a}}^t\Bigg] \tag{1.12.4}$$

We use a coupling argument as follows: Let $\boldsymbol{d}^t=[d_{e,1}^t,...,d_{e,n_e}^t]_{e=1}^{\hat{e}}$ be some realization of demand with $\sum_{b=L^{e'}}^{U^{e'}}\sum_{i=\sum_{k=0}^{b-1}n\bar{a}_{e',k}^t+1}^{\sum_{k=0}^{b}n\bar{a}_{e',k}^t}\min(b,d_{e',i}^t)=K_1$, and let $\acute{y}_{e'}^t=K_2$ be some realization of $\acute{Y}_{e'}^t$. We have that any realization of $\Theta_{1,e'},\dots,\Theta_{\min(K_1,K_2),e'}$ is common for the two distributions (the first $\min(K_1,K_2)$ units departing move to the same stations). The difference between the resultant inventory distributions occurs through the units $\min(K_1,K_2)+1,\dots,\max(K_1,K_2)$, which have to go to some station, charged or depleted. Assuming that they all go to stations with the highest rebalancing costs charged, and that differences through the outflow at different types do not cancel, Equation (1.12.4) can be upper bounded as $\big(\max(K_1,K_2)-\min(K_1,K_2)\big)\max_{e\in[\hat{e}]}\max(c_{0,e},c_{e,0})$. Consequently,

$$\begin{aligned}
&n\mathbb{E}_{\{\bar{\boldsymbol{D}}^t,\bar{\boldsymbol{R}}^t,\acute{\boldsymbol{R}}^t\}}\left[\sum_{e=1}^{\hat{e}}\max(c_{0,e},c_{e,0})\sum_{m=0}^{\infty}\left|\sum_{d=0}^{m}(\acute{x}_{e,d}^{t+1}-\bar{x}_{e,d}^{t+1})\right|\middle|\hat{\boldsymbol{a}}^t=\bar{\boldsymbol{a}}^t\right]\\
&\leq\sum_{e=1}^{\hat{e}}\mathbb{E}_{\{\bar{\boldsymbol{D}}^t,\bar{\boldsymbol{R}}^t,\acute{\boldsymbol{R}}^t\}}\Bigg[\Bigg|\sum_{b=L^{e'}}^{U^{e'}}\sum_{i=\sum_{k=0}^{b-1}n\bar{a}_{e',k}^t+1}^{\sum_{k=0}^{b}n\bar{a}_{e',k}^t}\min(b,D_{e',i}^t)-\acute{Y}_{e'}^t\Bigg|\\
&\leq\sum_{e'=1}^{\hat{e}}\mathbb{E}_{\{\bar{\boldsymbol{D}}^t,\bar{\boldsymbol{R}}^t,\acute{\boldsymbol{R}}^t\}}\Bigg[\Bigg|\sum_{b=L^{e'}}^{U^{e'}}\sum_{i=\sum_{k=0}^{b-1}n\bar{a}_{e',k}^t+1}^{\sum_{k=0}^{b}n\bar{a}_{e',k}^t}\min(b,D_{e',i}^t)-\sum_{b=L^{e'}}^{U^{e'}}n\bar{a}_{e',b}^t\mathbb{E}\left[\min(b,D^{e'})\right]\Bigg|\\
&+\Bigg|\sum_{b=L^{e'}}^{U^{e'}}n\bar{a}_{e',b}^t\mathbb{E}\left[\min(b,D^{e'})\right]-\acute{Y}_{e'}^t\Bigg|\Bigg]
\end{aligned}$$

$$\leq \sum_{e'=1}^{\hat{e}} \mathbb{E}_{\{\bar{\boldsymbol{D}}^t, \bar{\boldsymbol{R}}^t, \acute{\boldsymbol{R}}^t\}} \Bigg[\sum_{b=L^{e'}}^{U^{e'}} \Bigg| \sum_{i=\sum_{k=0}^{b-1} n\bar{a}^t_{e',k}+1}^{\sum_{k=0}^{b} n\bar{a}^t_{e',k}} \min(b, D^t_{e',i}) - n\bar{a}^t_{e',b} \mathbb{E}\left[\min(b, D^{e'})\right] \Bigg|$$

$$+ \sum_{b=L^{e'}}^{U^{e'}} \Bigg| n\bar{a}^t_{e',b} \mathbb{E}\left[\min(b, D^{e'})\right] - \sum_{j=1}^{n\bar{a}^t_{e',b}} \mathrm{Pois}(\mathbb{E}\left[\min(b, D^{e'})\right]) \Bigg| \Bigg],$$

by Lemma 1.12.3,

$$\leq \mathcal{O}(\sqrt{n}).$$

Inserting this bound to Equation (1.12.3),

$$\mathbb{E}_{\{\boldsymbol{D}^t, \bar{\boldsymbol{R}}^t\}} \left[\hat{c}(\bar{\boldsymbol{x}}^{t+1}, \hat{\boldsymbol{x}}^{t+1}) \middle| \hat{\boldsymbol{a}}^t = \bar{\boldsymbol{a}}^t \right]$$

$$\leq \mathcal{O}(\sqrt{n}) + n\mathbb{E}_{\{\boldsymbol{D}^t, \acute{\boldsymbol{R}}^t\}} \Bigg[\sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{m=0}^{\infty} \Bigg| \sum_{d=0}^{m} (\hat{x}^{t+1}_{e,d} - \acute{x}^{t+1}_{e,d}) \Bigg| \Bigg| \hat{\boldsymbol{a}}^t = \bar{\boldsymbol{a}}^t \Bigg]$$

$$= \mathcal{O}(\sqrt{n}) + n\mathbb{E}_{\{\boldsymbol{D}^t, \acute{\boldsymbol{R}}^t\}} \Bigg[\sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{m=0}^{\infty} \Bigg| \sum_{b=L^e}^{U^e} \bar{a}^t_{e,b} \mathbb{P}\left[b - \min(b, D^e) + \hat{R}^e \leq m\right]$$

$$- \frac{1}{n} \sum_{b=L^e}^{U^e} \sum_{i=\sum_{k=0}^{b-1} n\bar{a}^t_{e,k}+1}^{\sum_{k=0}^{b} n\bar{a}^t_{e,k}} \mathbb{I}\left\{ b - \min(b, D^t_{e,i}) + \sum_{e'=1}^{\hat{e}} \acute{R}^t_{e',e,i} \leq m \right\} \Bigg| \Bigg]$$

$$\leq \mathcal{O}(\sqrt{n}) + \sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{b=L^e}^{U^e} n\mathbb{E}_{\{\boldsymbol{D}^t, \acute{\boldsymbol{R}}^t\}} \Bigg[\sum_{m=0}^{\infty} \Bigg| \bar{a}^t_{e,b} \mathbb{P}\left[b - \min(b, D^e) + \hat{R}^e \leq m\right]$$

$$- \frac{1}{n} \sum_{i=\sum_{k=0}^{b-1} n\bar{a}^t_{e,k}+1}^{\sum_{k=0}^{b} n\bar{a}^t_{e,k}} \mathbb{I}\left\{ b - \min(b, D^t_{e,i}) + \sum_{e'=1}^{\hat{e}} \acute{R}^t_{e',e,i} \leq m \right\} \Bigg| \Bigg]$$

$$= \mathcal{O}(\sqrt{n}) + \sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{b=L^e, \bar{a}^t_{e,b} \neq 0}^{U^e} n\mathbb{E}_{\{\boldsymbol{D}^t, \acute{\boldsymbol{R}}^t\}} \Bigg[\sum_{m=0}^{\infty} \Bigg| \mathbb{P}\left[b - \min(b, D^e) + \hat{R}^e \leq m\right]$$

$$- \frac{1}{n\bar{a}^t_{e,b}} \sum_{i=\sum_{k=0}^{b-1} n\bar{a}^t_{e,k}+1}^{\sum_{k=0}^{b} n\bar{a}^t_{e,k}} \mathbb{I}\left\{ b - \min(b, D^t_{e,i}) + Z_{e,i} \leq m \right\} \Bigg| \Bigg],$$

where $Z_{e,i}$ is a sequence of *i.i.d.* random variables distributed Poisson with mean $\sum_{e'=1}^{\hat{e}} \frac{qnp_{e',e}}{n_e} \sum_{b=L^{e'}}^{U^{e'}} \hat{a}^t_{e',b} \mathbb{E}\left[\min(b, D^{e'})\right]$. Then, applying Lemma 1.12.5,

$$\mathbb{E}_{\{\boldsymbol{D}^t, \bar{\boldsymbol{R}}^t\}}\left[\hat{c}(\bar{\boldsymbol{x}}^{t+1}, \hat{\boldsymbol{x}}^{t+1}) \middle| \hat{\boldsymbol{a}}^t = \bar{\boldsymbol{a}}^t\right] \leq \mathcal{O}(\sqrt{n}) + \sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{b=L^e, \bar{a}^{t,e}_{e,b} \neq 0}^{U^e} \mathcal{O}(\sqrt{n})$$

$$\leq \mathcal{O}(\sqrt{n}). \qquad \square$$

Lemma 1.12.1. *For any inventory position $\hat{\boldsymbol{x}}^t$ and action $\hat{\boldsymbol{a}}^t$,*

$$\hat{c}(\boldsymbol{x}^t, \boldsymbol{a}^t) \leq n \sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{b=0}^{\infty} \left| \sum_{d=0}^{b} (\hat{x}^t_{e,d} - \hat{a}^t_{e,d}) \right| + n \sum_{e=1}^{\hat{e}} \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0}) |\hat{x}^t_{e,r} - \hat{a}^t_{e,r}|.$$

Proof. To show this bound, we will insert a sub-optimal solution for the mean-field rebalancing optimization problem where we will assume that all units are sourced from the warehouse. Under this solution, $y^{e_1,e_2} = 0$ if both $e_1, e_2 > 0$, at most one of $y^{0,e}, y^{e,0}$ is strictly positive $\forall e \in [2\hat{e}]$, $|y^{e,0} + y^{0,e}| = |\sum_{b=1}^{\infty} b(\hat{x}^t_{e,b} - \hat{a}^t_{e,b})|$ $\forall e \in [\hat{e}]$, and $|y^{e,0} + y^{0,e}| = |\hat{x}^t_{e,r} - \hat{a}^t_{e,r}|$ $\forall e > \hat{e}$. Inserting this solution, we obtain

$$\begin{aligned}\hat{c}(\boldsymbol{x}^t, \boldsymbol{a}^t) \leq n \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{b=0}^{\infty} &\left| \sum_{d=0}^{b} (\hat{x}^t_{e,d} - \hat{a}^t_{e,d}) \right| \\ &+ n \sum_{e=1}^{\hat{e}} \max\left(c_{0,e} - \frac{c_{e,e}}{2}, c_{e,0} - \frac{c_{e,e}}{2}\right) \left| \sum_{b=1}^{\infty} b(\hat{x}^t_{e,b} - \hat{a}^t_{e,b}) \right| \\ &+ n \sum_{e=1}^{\hat{e}} \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0}) |\hat{x}^t_{e,r} - \hat{a}^t_{e,r}|.\end{aligned}$$

Furthermore, through Jensen's inequality ($\left|\mathbb{E}[X] - \mathbb{E}[Y]\right| \leq \mathbb{E}[|X - Y|]$), we have :

$$\left| \sum_{b=1}^{\infty} b(\hat{x}^t_{e,b} - \hat{a}^t_{e,b}) \right| \leq \sum_{b=0}^{\infty} \left| \sum_{d=0}^{b} (\hat{x}^t_{e,d} - \hat{a}^t_{e,d}) \right|.$$

As a result, we obtain

$$\hat{c}(\boldsymbol{x}^t, \boldsymbol{a}^t) \leq n \sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{b=0}^{\infty} \left| \sum_{d=0}^{b} (\hat{x}_{e,d}^t - \hat{a}_{e,d}^t) \right| + n \sum_{e=1}^{\hat{e}} \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0}) |\hat{x}_{e,r}^t - \hat{a}_{e,r}^t|.$$

□

Remark 1.12.2. *The value function of the mean-field model, $\hat{V}(\hat{\boldsymbol{x}}^t)$, is Lipschitz continuous in $\hat{\boldsymbol{x}}^t$ with*[24]

$$\left| \hat{V}(\hat{\boldsymbol{x}}^{t,1}) - \hat{V}(\hat{\boldsymbol{x}}^{t,2}) \right| \leq \hat{c}(\hat{\boldsymbol{x}}^{t,1}, \hat{\boldsymbol{x}}^{t,2}) \quad \forall \hat{\boldsymbol{x}}^{t,1}, \hat{\boldsymbol{x}}^{t,2} \in \hat{\boldsymbol{X}}, \tag{1.12.5}$$

where, by Lemma 1.12.1, $\forall \hat{\boldsymbol{x}}^{t,1}, \hat{\boldsymbol{x}}^{t,2} \in \hat{\boldsymbol{X}}$,

$$\leq n \sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{b=0}^{\infty} \left| \sum_{d=0}^{b} (\hat{x}_{e,d}^{t,1} - \hat{x}_{e,d}^{t,2}) \right| + n \sum_{e=1}^{\hat{e}} \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0}) |\hat{x}_{e,r}^{t,1} - \hat{x}_{e,r}^{t,2}|.$$

Lemma 1.12.3. *Let $\{Z_i\}$ be a sequence of i.i.d. random variables with mean μ, variance σ^2 and let $\{Y_i\}$ be a sequence of i.i.d. random variables with mean λ, variance ν^2 . Then*

$$\mathbb{E}\left[\left| \sum_{i=1}^{\sum_{j=1}^{n} Z_j} Y_i - n\mu\lambda \right| \right] \leq \sqrt{\mu\nu^2 + \lambda^2\sigma^2}\sqrt{n}.$$

Proof. Through Lemma 1.12.4, we know that $\mathbb{E}\left[\sum_{i=1}^{\sum_{j=1}^{n} Z_j} Y_i\right] = n\mu\lambda$. Then, we are interested in the average absolute deviation of $\sum_{i=1}^{\sum_{j=1}^{n} Z_j} Y_i$. For any random variable, the average absolute deviation is always smaller then the standard deviation so we can upper bound the average absolute deviation through the standard deviation. Then, using Lemma

24. Value function of the original model, $V(\boldsymbol{x}^t)$, is also Lipschitz continuous in $\boldsymbol{x}^t$.

1.12.4,

$$\mathrm{Var}\left(\sum_{i=1}^{\sum_{j=1}^n Z_j} Y_i\right) = \mathbb{E}\left[\sum_{i=1}^n Z_i\right]\mathrm{Var}(Y) + (\mathbb{E}\,[Y])^2\mathrm{Var}\left(\sum_{i=1}^n Z_i\right),$$

$$\mathbb{E}\left[\left|\sum_{i=1}^{\sum_{j=1}^n Z_j} Y_i - n\mu\lambda\right|\right] \le \sqrt{\mu\nu^2 + \lambda^2\sigma^2}\sqrt{n}. \qquad \square$$

Lemma 1.12.4. *(Ross [2010], Example 5D/5P). Let $\{Z_i\}$ be a sequence of i.i.d. random variables and let $\{Y, Y_i\}$ be another sequence of i.i.d. random variables . Then*

$$\mathbb{E}\left[\sum_{i=1}^{\sum_{j=1}^n Z_j} Y_i\right] = \mathbb{E}\,[Y]\mathbb{E}\left[\sum_{i=1}^n Z_i\right],$$

$$Var\left(\sum_{i=1}^{\sum_{j=1}^n Z_j} Y_i\right) = \mathbb{E}\left[\sum_{i=1}^n Z_i\right] Var(Y) + (\mathbb{E}\,[Y])^2\, Var\left(\sum_{i=1}^n Z_i\right).$$

Lemma 1.12.5. *Let $\{M, M_i\}$ be a sequence of i.i.d. random variables distributed Poisson with mean $\lambda < \infty$ and $\{N, N_i\}$ be a sequence of i.i.d. non-negative random variables. Furthermore, let*

$$Y = a + M - \min(a, N),$$

$$Y_i = a + M_i - \min(a, N_i),$$

where a is a constant satisfying $0 \le a < \infty$. We let F denote the corresponding cumulative distribution of Y and let F_n denote the empirical distribution based on $Y_1, \ldots, Y_n$, $n \in \mathbb{N}$. Then, letting $\lfloor . \rfloor$ denote the floor function,

$$\lim_{n\to\infty} \sqrt{n}\mathbb{E}\left[\int_{-\infty}^{\infty}\left|F_n(t) - F(t)\right|dt\right] \le a + \left\lfloor \lambda e^2 \right\rfloor + e^{-(\lfloor \lambda e^2 \rfloor + \lambda + 1)} < \infty.$$

Proof. We will show that the two conditions provided in Theorem 1.12.6 hold. First,

$$
\begin{aligned}
\mathbb{E}\left[|Y|^2\right] &\leq \mathbb{E}\left[(a+M)^2\right] \\
&= \lambda + (\lambda + a)^2 \\
&< \infty.
\end{aligned}
$$

Consequently, the first condition holds. For the second condition:

$$
\begin{aligned}
\int_0^\infty \sqrt{\mathbb{P}\left[|Y| > t\right]}dt &\leq \sum_{t=0}^{\infty} \sqrt{\mathbb{P}\left[|a+M| > t\right]} \\
&= \sum_{t=0}^{\infty} \sqrt{\mathbb{P}\left[M > t - a\right]} \\
&= \sum_{t=-a}^{-1} \sqrt{\mathbb{P}\left[M > t\right]} + \sum_{t=0}^{\infty} \sqrt{\mathbb{P}\left[M > t\right]} \\
&= a + \sum_{t=0}^{\infty} \sqrt{\sum_{k=t+1}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!}},
\end{aligned}
$$

we let $\bar{t} = \lfloor \lambda e^2 \rfloor$ and obtain,

$$
\begin{aligned}
&= a + \sum_{t=0}^{\bar{t}-1} \sqrt{\sum_{k=t+1}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!}} + \sum_{t=\bar{t}}^{\infty} \sqrt{\sum_{k=t+1}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!}} \\
&\leq a + \bar{t} + \sum_{t=\bar{t}}^{\infty} \sqrt{\sum_{k=t+1}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!}},
\end{aligned}
$$

applying Lemma 1.12.7 on the factorial term,

$$
\begin{aligned}
&\leq a + \bar{t} + \sum_{t=\bar{t}}^{\infty} \sqrt{\sum_{k=t+1}^{\infty} e^{-\lambda} (\frac{\lambda}{k})^k e^{k-1}} \\
&\leq a + \bar{t} + \sum_{t=\bar{t}}^{\infty} \sqrt{\sum_{k=t+1}^{\infty} e^{-\lambda} e^{-2k} e^{k-1}} \\
&= a + \bar{t} + e^{-\lambda-1} \sum_{t=\bar{t}}^{\infty} \sqrt{\sum_{k=t+1}^{\infty} e^{-k}} \\
&= a + \bar{t} + \frac{1}{e^{\frac{\bar{t}}{2}+\lambda+\frac{1}{2}}(\sqrt{e}-1)(\sqrt{e-1})} \\
&\leq a + \bar{t} + e^{-(\bar{t}+\lambda+1)}
\end{aligned}
$$

$$< \infty.$$

As the two conditions hold, through Theorem 1.12.6,

$$\lim_{n\to\infty} \sqrt{n}\mathbb{E}\left[\int_{-\infty}^{\infty}\Big|F_n(t) - F(t)\Big|dt\right] = \mathbb{E}\left[\int_{-\infty}^{\infty}\Big|B(F(t))\Big|dt\right],$$

by Lemma 1.12.8,

$$\begin{aligned} &\leq \int_0^\infty \sqrt{\mathbb{P}\left[|Y| > t\right]}dt \\ &\leq a + \bar{t} + e^{-(\bar{t}+\lambda+1)} \\ &< \infty. \end{aligned}$$

□

Theorem 1.12.6. *(del Barrio et al. [1999], Theorem 2.4) Let* $\{Y, Y_i\}$ *be a sequence of i.i.d. random variables with common distribution* F*, and let* F_n *denote the empirical distribution based on* $Y_1, \ldots, Y_n$*,* $n \in \mathbb{N}$ *where*

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{Y_i \leq t\} \qquad \forall t.$$

Then, if

$$\int_0^\infty \sqrt{\mathbb{P}\left[|Y| > t\right]}dt < \infty,$$

$$\mathbb{E}\left[|Y|^2\right] < \infty,$$

we have

$$\lim_{n\to\infty} \sqrt{n}\mathbb{E}\left[\int_{-\infty}^{\infty}\Big|F_n(t) - F(t)\Big|dt\right] = \mathbb{E}\left[\int_{-\infty}^{\infty}\Big|B(F(t))\Big|dt\right],$$

where B *denotes Brownian Bridge.*

Lemma 1.12.7. *For any* $k \in \mathbb{N}$*,*

$$k! \geq k^k e^{-k+1}.$$

Proof.

$$\begin{aligned}
\ln(k!) &= \sum_{i=1}^{k} \ln(i) \\
\ln(k!) &\geq \int_{1}^{k} \ln(z)dz \\
\ln(k!) &\geq k\ln(k) - k + 1 \\
k! &\geq k^k e^{-k+1}.
\end{aligned} \qquad \square$$

Lemma 1.12.8. *Let $\{Y, Y_i\}$ be a sequence of i.i.d. non-negative random variables with common distribution F. Then:*

$$\mathbb{E}\left[\int_{-\infty}^{\infty} \Big| B(F(t)) \Big| dt\right] \leq \int_{0}^{\infty} \sqrt{\mathbb{P}\left[\big|Y\big| > t\right]} dt,$$

where B denotes Brownian Bridge.

Proof.

$$\begin{aligned}
\int_{-\infty}^{\infty} \mathbb{E}\left[\Big| B(F(t)) \Big| dt\right] &= \int_{0}^{\infty} \mathbb{E}\left[\Big| B(F(t)) \Big| dt\right] \\
&\leq \int_{0}^{\infty} \sqrt{\mathbb{E}\left[B(F(t))^2\right]} dt \\
&= \int_{0}^{\infty} \sqrt{F(t)(1 - F(t))} dt \\
&\leq \int_{0}^{\infty} \sqrt{\mathbb{P}\left[\big|Y\big| > t\right]} dt.
\end{aligned} \qquad \square$$

Theorem 1.3.4. *Let $\hat{\pi}^*$ be an optimal policy for the mean-field model. Then, the lifted mean-field policy $h \circ f \circ \hat{\pi}^* \circ g$ satisfies*

$$V_{h \circ f \circ \hat{\pi}^* \circ g}(\boldsymbol{x}^t) - V(\boldsymbol{x}^t) \leq \mathcal{O}(\sqrt{n}) \quad \forall \boldsymbol{x}^t \in \boldsymbol{X}.$$

Consequently, the optimality gap of the composite policy $h \circ f \circ \hat{\pi}^* \circ g$ is at most $\mathcal{O}(\sqrt{n})$.

Proof. Through Corollary 1.11.3 and Proposition 1.11.1, respectively,

$$V(\boldsymbol{x}^t) = \bar{V}(g(\boldsymbol{x}^t)),$$
$$V_{h\circ f\circ\hat{\pi}^*\circ g}(\boldsymbol{x}^t) = \bar{V}_{f(\hat{\pi}^*)}(g(\boldsymbol{x}^t)).$$

Then, summing items 1 and 2 in Proposition 1.12.9 proves the theorem. □

Proposition 1.12.9. *For any inventory position* $\bar{\boldsymbol{x}}^t \in \bar{\boldsymbol{X}}$*,*

$$1.\ \bar{V}_{f(\hat{\pi}^*)}(\bar{\boldsymbol{x}}^t) - \hat{V}(\bar{\boldsymbol{x}}^t) \le \mathcal{O}(\sqrt{n}),$$
$$2.\ \hat{V}(\bar{\boldsymbol{x}}^t) - \bar{V}(\bar{\boldsymbol{x}}^t) \le \mathcal{O}(\sqrt{n}).$$

Proof. In order to prove both statements, we will show a recursive structure where the differences in per-period costs of both expressions can be bounded by a function of order $\mathcal{O}(\sqrt{n})$. To prove this bound, we will use Proposition 1.3.3, as well as Lemma 1.12.10, which proves that the additional cost incurred through the discretization via the f function is bounded. Then, through the discounted setting, we show that the total difference in costs are also bounded with $\mathcal{O}(\sqrt{n})$.

For the first statement, we have

$$\begin{aligned}
\bar{V}_{f(\hat{\pi}^*)}(\bar{\boldsymbol{x}}^t) - \hat{V}(\bar{\boldsymbol{x}}^t) &= \hat{c}(\bar{\boldsymbol{x}}^t, f(\hat{\pi}^*(\bar{\boldsymbol{x}}^t))) + \hat{N}(f(\hat{\pi}^t(\bar{\boldsymbol{x}}^t))) + \gamma\mathbb{E}_{\{\boldsymbol{D}^t,\bar{\boldsymbol{R}}^t\}}\left[\bar{V}_{f(\hat{\pi}^*)}(\bar{\boldsymbol{x}}^{t+1})\right] \\
&\qquad - \hat{c}(\bar{\boldsymbol{x}}^t, \hat{\pi}^*(\bar{\boldsymbol{x}}^t)) - \hat{N}(\hat{\pi}^*(\bar{\boldsymbol{x}}^t)) - \gamma\hat{V}(\hat{\boldsymbol{x}}^{t+1}) \\
&\le \hat{c}(\bar{\boldsymbol{x}}^t, \hat{\pi}^*(\bar{\boldsymbol{x}}^t)) + \hat{c}(\hat{\pi}^*(\bar{\boldsymbol{x}}^t), f(\hat{\pi}^*(\bar{\boldsymbol{x}}^t))) \\
&\qquad + \hat{N}(f(\hat{\pi}^{t,n^*}(\bar{\boldsymbol{x}}^t))) + \gamma\mathbb{E}_{\{\boldsymbol{D}^t,\bar{\boldsymbol{R}}^t\}}\left[\bar{V}_{f(\hat{\pi}^*)}(\bar{\boldsymbol{x}}^{t+1})\right] \\
&\qquad - \hat{c}(\bar{\boldsymbol{x}}^t, \hat{\pi}^*(\bar{\boldsymbol{x}}^t)) - \hat{N}(\hat{\pi}^*(\bar{\boldsymbol{x}}^t)) - \gamma\hat{V}(\hat{\boldsymbol{x}}^{t+1}),
\end{aligned}$$

through applying Lemma 1.12.10 on the rebalancing cost function and the newsvendor costs,

$$\le \mathcal{O}(1) + \gamma\mathbb{E}_{\{\boldsymbol{D}^t,\bar{\boldsymbol{R}}^t\}}\left[\bar{V}_{f(\hat{\pi}^*)}(\bar{\boldsymbol{x}}^{t+1})\right] - \gamma\hat{V}(\hat{\boldsymbol{x}}^{t+1})$$

$$= \mathcal{O}(1) + \gamma \mathbb{E}_{\{\boldsymbol{D}^t, \bar{\boldsymbol{R}}^t\}} \left[\bar{V}_{f(\hat{\pi}^*)}(\bar{\boldsymbol{x}}^{t+1}) - \hat{V}(\bar{\boldsymbol{x}}^{t+1}) + \hat{V}(\bar{\boldsymbol{x}}^{t+1}) \right] - \gamma \hat{V}(\hat{\boldsymbol{x}}^{t+1}),$$

through the Lipschitz bound given at (1.12.5),

$$\begin{aligned}
&\leq \mathcal{O}(1) + \gamma \mathbb{E}_{\{\boldsymbol{D}^t, \bar{\boldsymbol{R}}^t\}} \left[\bar{V}_{f(\hat{\pi}^*)}(\bar{\boldsymbol{x}}^{t+1}) - \hat{V}(\bar{\boldsymbol{x}}^{t+1}) + \hat{c}(\bar{\boldsymbol{x}}^{t+1}, \hat{\boldsymbol{x}}^{t+1}) \right] \\
&\leq \mathcal{O}(1) + \gamma \mathbb{E}_{\{\boldsymbol{D}^t, \bar{\boldsymbol{R}}^t\}} \left[\bar{V}_{f(\hat{\pi}^*)}(\bar{\boldsymbol{x}}^{t+1}) - \hat{V}(\bar{\boldsymbol{x}}^{t+1}) + \hat{c}(\bar{\boldsymbol{x}}^{t+1}, \hat{\boldsymbol{x}}^{t+1,2}) \right] \\
&\qquad + \gamma \hat{c}(\hat{\boldsymbol{x}}^{t+1,2}, \hat{\boldsymbol{x}}^{t+1}),
\end{aligned}$$

where $\hat{\boldsymbol{x}}^{t+1,2}$ is the mean-field evolution of the inventory position when the action taken is $f(\hat{\pi}^{t,*}(\hat{\boldsymbol{x}}^t))$. Then, by applying Proposition 1.3.3 on the first term and Lemma 1.12.10 on the second term,

$$\bar{V}_{f(\hat{\pi}^*)}(\bar{\boldsymbol{x}}^t) - \hat{V}(\bar{\boldsymbol{x}}^t) \leq \mathcal{O}(\sqrt{n}) + \gamma \mathbb{E}_{\{\boldsymbol{D}^t, \bar{\boldsymbol{R}}^t\}} \left[\bar{V}_{f(\hat{\pi}^*)}(\bar{\boldsymbol{x}}^{t+1}) - \hat{V}(\bar{\boldsymbol{x}}^{t+1}) \right].$$

As the equation holds for all $\bar{\boldsymbol{x}}^t$, we can repeat these steps for the next period with the new realized inventory position. As Proposition 1.3.3, which is the only step where we observe an additional cost of $\mathcal{O}(\sqrt{n})$ (all three terms in Lemma 1.12.10 are bounded with the bounds independent of γ), holds for all actions, costs of future periods are also of order $\mathcal{O}(\sqrt{n})$. Formally, Proposition 1.3.3 implies that $\exists M > 0, n_0 \in \mathbb{R}_+$ such that $\forall n \geq n_o$,

$$\max_{\boldsymbol{a}^t \in \boldsymbol{A}} \mathbb{E}_{\{\boldsymbol{D}^t, \boldsymbol{R}^t\}} \left[\hat{c}(g(\boldsymbol{x}^{t+1}), \hat{\boldsymbol{x}}^{t+1}) \middle| \hat{\boldsymbol{a}}^t = g(\boldsymbol{a}^t) \right] \leq M\sqrt{n}.$$

Moreover, $M < \infty$, in both Lemma 1.12.3 and Lemma 1.12.5 (which together give the sub-optimality coefficients used in Proposition 1.3.3), coefficients of $\sqrt{n}$ are bounded and are independent of γ. Consequently, we can express the total cost as

$$\begin{aligned}
\bar{V}_{f(\hat{\pi}^*)}(\bar{\boldsymbol{x}}^t) - \hat{V}(\bar{\boldsymbol{x}}^t) &\leq \frac{1}{1-\gamma} \mathcal{O}(\sqrt{n}) \\
&\leq \mathcal{O}(\sqrt{n}).
\end{aligned}$$

For the second statement,

$$\begin{aligned}\hat{V}(\bar{\boldsymbol{x}}^t) - \bar{V}(\bar{\boldsymbol{x}}^t) \le\ & \hat{c}(\bar{\boldsymbol{x}}^t, \bar{\pi}^*(\bar{\boldsymbol{x}}^t)) + \hat{N}(\bar{\pi}^*(\bar{\boldsymbol{x}}^t)) + \gamma \hat{V}(\hat{\boldsymbol{x}}^{t+1}) \\ & - \hat{c}(\bar{\boldsymbol{x}}^t, \bar{\pi}^*(\bar{\boldsymbol{x}}^t)) - \hat{N}(\bar{\pi}^*(\bar{\boldsymbol{x}}^t)) - \gamma \mathbb{E}_{\{\boldsymbol{D}^t, \bar{\boldsymbol{R}}^t\}} \left[\bar{V}(\bar{\boldsymbol{x}}^{t+1})\right],\end{aligned}$$

through the Lipschitz bound given at (1.12.5),

$$\le \gamma \mathbb{E}_{\{\boldsymbol{D}^t, \bar{\boldsymbol{R}}^t\}} \left[\hat{c}(\bar{\boldsymbol{x}}^{t+1}, \hat{\boldsymbol{x}}^{t+1}) + \hat{V}(\bar{\boldsymbol{x}}^{t+1}) - \bar{V}(\bar{\boldsymbol{x}}^{t+1})\right],$$

by Proposition 1.3.3,

$$\le \mathcal{O}(\sqrt{n}) + \gamma \mathbb{E}_{\{\boldsymbol{D}^t, \bar{\boldsymbol{R}}^t\}} \left[\hat{V}(\bar{\boldsymbol{x}}^{t+1}) - \bar{V}(\bar{\boldsymbol{x}}^{t+1})\right],$$

repeating the analysis in the first statement,

$$\le \mathcal{O}(\sqrt{n}).$$

□

Lemma 1.12.10. *For any action $\hat{\boldsymbol{a}}^t \in \hat{\boldsymbol{A}}$, function f satisfies*

1. $\hat{c}(\hat{\boldsymbol{a}}^t, f(\hat{\boldsymbol{a}}^t)) \le C_1 < \infty$;
2. $\left|\hat{N}(\hat{\boldsymbol{a}}^t) - \hat{N}(f(\hat{\boldsymbol{a}}^t))\right| \le C_2 < \infty$;
3. $\hat{c}(\hat{\boldsymbol{x}}^{t+1,1}, \hat{\boldsymbol{x}}^{t+1,2}) \le C_3 < \infty$,

where $\hat{\boldsymbol{x}}^{t+1,1}$ is the mean-field evolution of the inventory position when the action taken is $f(\hat{\boldsymbol{a}}^{t^*})$, $\hat{\boldsymbol{x}}^{t+1,2}$ is the mean-field evolution of the inventory position when the action taken is $\hat{\boldsymbol{a}}^t$, and C_1, C_2, C_3 are constant values with respect to n.

Proof. Let $\bar{\boldsymbol{a}}^{t^*}$ denote $f(\hat{\boldsymbol{a}}^t)$. For the first statement, through Lemma 1.12.1:

$$\hat{c}(\hat{\boldsymbol{a}}^t, f(\hat{\boldsymbol{a}}^t)) \le \sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{b=0}^{\infty} \left| \sum_{d=0}^{b} n\hat{a}^t_{e,d} - \sum_{d=0}^{b} n\bar{a}^{t^*}_{e,d} \right| \tag{1.12.6}$$

$$+ \sum_{e=1}^{\hat{e}} \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0}) \left| n\hat{a}^t_{e,r} - n\bar{a}^{t^*}_{e,r} \right|$$

$$\le \sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{b=L^e}^{U^e} \sum_{d=0}^{b} \left| n\hat{a}^t_{e,d} - n\bar{a}^{t^*}_{e,d} \right| \tag{1.12.7}$$

$$+\sum_{e=1}^{\hat{e}} \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0}) \left| n\hat{a}^{t}_{e,r} - n\bar{a}^{t^*}_{e,r} \right|, \tag{1.12.8}$$

using the Largest Remainder Algorithm,

$$\leq \sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{b=L^e}^{U^e} \sum_{d=L^e}^{b} 1 + \sum_{e=1}^{\hat{e}} \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0}) 0.5$$

$$< \infty.$$

For the second statement:

$$\begin{aligned}
\left| \hat{N}(\hat{\boldsymbol{a}}^t) - \hat{N}(f(\hat{\boldsymbol{a}}^t)) \right| &= \Bigg| c_h \sum_{e=1}^{\hat{e}} \sum_{b=L^e}^{U^e} b(n\hat{a}^t_{e,b} - n\bar{a}^{t^*}_{e,b}) + \sum_{e=1}^{\hat{e}} c_h (n\hat{a}^t_{e,r} - n\bar{a}^{t^*}_{e,r}) \\
&\quad + c_p \sum_{e=1}^{\hat{e}} \sum_{b=L^e}^{U^e} (n\hat{a}^t_{e,b} - n\bar{a}^{t^*}_{e,b}) \mathbb{E}\left[(D^e - b)^+\right] \Bigg| \\
&\leq \left| \sum_{e=1}^{\hat{e}} \sum_{b=L^e}^{U^e} (bc_h + c_p \mathbb{E}\left[(D^e - b)^+\right])(n\hat{a}^t_{e,b} - n\bar{a}^{t^*}_{e,b}) \right| \\
&\quad + \sum_{e=1}^{\hat{e}} c_h \left| n\hat{a}^t_{e,r} - n\bar{a}^{t^*}_{e,r} \right| \\
&\leq \sum_{e=1}^{\hat{e}} (U^e c_h + c_p \mathbb{E}\left[D^e\right]) \sum_{b=L^e}^{U^e} \left| n\hat{a}^t_{e,b} - n\bar{a}^{t^*}_{e,b} \right| + \sum_{e=1}^{\hat{e}} c_h \left| n\hat{a}^t_{e,r} - n\bar{a}^{t^*}_{e,r} \right|.
\end{aligned}$$

We replace $\max(c_{0,e}, c_{e,0})$ with $U^e c_h + c_p \mathbb{E}\left[D^e\right]$ and $\max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0})$ with c_h in Equation (1.12.8). Repeating the following steps,

$$\left| \hat{N}(\hat{\boldsymbol{a}}^t) - \hat{N}(f(\hat{\boldsymbol{a}}^t)) \right| < \infty.$$

For the third statement, through Lemma 1.12.1:

$$\hat{c}(\hat{\boldsymbol{x}}^{t+1,1}, \hat{\boldsymbol{x}}^{t+1,2})$$

$$\leq \sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{b=0}^{\infty} \left| \sum_{d=0}^{b} \sum_{m=L^e}^{U^e} (n\bar{a}^{t^*}_{e,m} - n\hat{a}^t_{e,m}) \mathbb{P}\left[m - \min(m, D^e) + \hat{R}^e = d\right] \right| + \sum_{e=1}^{\hat{e}}$$

$$\max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0}) \left| n\hat{a}^t_{e,r} - n\bar{a}^{t^*}_{e,r} + (1-q) \sum_{e'=1}^{\hat{e}} p_{e',e} \sum_{b=L^{e'}}^{U^{e'}} (n\hat{a}^t_{e',b} - n\bar{a}^{t^*}_{e',b}) \mathbb{E}\left[\min(b, D^{e'})\right] \right|$$

$$
\begin{aligned}
&\leq \sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{b=0}^{\infty}\left|\sum_{d=0}^{b} \sum_{m=L^e}^{U^e} (n\bar{a}_{e,m}^{t^*} - n\hat{a}_{e,m}^{t})\mathbb{P}\left[m - \min(m, D^e) + \hat{R}^e = d\right]\right| \\
&+ \sum_{e=1}^{\hat{e}} \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0})\left|n\hat{a}_{e,r}^{t} - n\bar{a}_{e,r}^{t^*}\right| \\
&+ \sum_{e=1}^{\hat{e}} \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0})(1-q) \sum_{e'=1}^{\hat{e}} p_{e',e} \sum_{b=L^{e'}}^{U^{e'}} \mathbb{E}\left[\min(b, D^{e'})\right]\left|n\hat{a}_{e',b}^{t} - n\bar{a}_{e',b}^{t^*}\right| \\
&\leq \sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{b=0}^{\infty}\left|\sum_{m=L^e}^{U^e} (n\bar{a}_{e,m}^{t^*} - n\hat{a}_{e,m}^{t})\mathbb{P}\left[m - \min(m, D^e) + \hat{R}^e \leq b\right]\right| \\
&+ \sum_{e=1}^{\hat{e}} \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0})0.5 + \sum_{e=1}^{\hat{e}} \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0})(1-q) \sum_{e'=1}^{\hat{e}} p_{e',e} \sum_{b=L^{e'}}^{U^{e'}} \mathbb{E}\left[\min(b, D^{e'})\right] \\
&\leq \sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{b=0}^{\infty}\left|\sum_{m=L^e}^{U^e} (n\bar{a}_{e,m}^{t^*} - n\hat{a}_{e,m}^{t})\mathbb{P}\left[m - \min(m, D^e) + \hat{R}^e > b\right]\right| \\
&+ \sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{b=0}^{\infty}\left|\sum_{m=L^e}^{U^e} (n\bar{a}_{e,m}^{t^*} - n\hat{a}_{e,m}^{t})\right| \\
&+ \sum_{e=1}^{\hat{e}} \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0})0.5 + \sum_{e=1}^{\hat{e}} \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0})(1-q) \sum_{e'=1}^{\hat{e}} p_{e',e} \sum_{b=L^{e'}}^{U^{e'}} \mathbb{E}\left[\min(b, D^{e'})\right],
\end{aligned}
$$

as $\sum_{m=L^e}^{U^e}(n\bar{a}_{e,m}^{t^*} - n\hat{a}_{e,m}^{t}) = 0$, we have

$$
\begin{aligned}
&\leq \sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{m=L^e}^{U^e} \sum_{b=0}^{\infty} \mathbb{P}\left[m - \min(m, D^e) + \hat{R}^e > b\right]\left|n\bar{a}_{e,m}^{t^*} - n\hat{a}_{e,m}^{t}\right| \\
&+ \sum_{e=1}^{\hat{e}} \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0})0.5 + \sum_{e=1}^{\hat{e}} \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0})(1-q) \sum_{e'=1}^{\hat{e}} p_{e',e} \sum_{b=L^{e'}}^{U^{e'}} \mathbb{E}\left[\min(b, D^{e'})\right] \\
&\leq \sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{m=L^e}^{U^e} \mathbb{E}\left[m - \min(m, D^e) + \hat{R}^e\right] \\
&+ \sum_{e=1}^{\hat{e}} \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0})0.5 + \sum_{e=1}^{\hat{e}} \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0})(1-q) \sum_{e'=1}^{\hat{e}} p_{e',e} \sum_{b=L^{e'}}^{U^{e'}} \mathbb{E}\left[\min(b, D^{e'})\right] \\
&< \infty. \qquad \square
\end{aligned}
$$

Corollary 1.3.5. *For any inventory position $\boldsymbol{x}^t \in \boldsymbol{X}$,*

$$\lim_{n\to\infty} \frac{V_{h\circ f\circ\hat{\pi}^*\circ g}(\boldsymbol{x}^t)}{V(\boldsymbol{x}^t)} = 1.$$

Thus, the mean-field model provides an asymptotically optimal policy for the original stochastic model.

Proof. By Theorem 1.3.4,

$$\begin{aligned} V_{h\circ f\circ\hat{\pi}^*\circ g}(\boldsymbol{x}^t) - V(\boldsymbol{x}^t) &\leq \mathcal{O}(\sqrt{n}) \\ \frac{V_{h\circ f\circ\hat{\pi}^*\circ g}(\boldsymbol{x}^t)}{V(\boldsymbol{x}^t)} &\leq 1 + \frac{\mathcal{O}(\sqrt{n})}{V(\boldsymbol{x}^t)}. \end{aligned}$$

We have that $V(\boldsymbol{x}^t) = \Omega(n)$ as $\forall \boldsymbol{a}^t \in \boldsymbol{A}, N(\boldsymbol{a}^t)$ grows linearly with n. Consequently:

$$\lim_{n\to\infty} \frac{V_{h\circ f\circ\hat{\pi}^*\circ g}(\boldsymbol{x}^t)}{V(\boldsymbol{x}^t)} = 1.$$ □

1.13 Reformulating the Mean-Field Model through a Compact State-Space

In this subsection, we will provide the necessary mathematical details in order to provide an alternative state-space for the mean-field model in a compact set. In Section 1.3.4, we introduced the new state space $\breve{\boldsymbol{X}}$. In order for this state space to be compact, $\breve{\boldsymbol{x}}$ has to be finite-dimensional, closed, and bounded (Bolzano-Weitrass Theorem). As a result, we define the new state-space such that $\breve{\boldsymbol{x}}^t_e = \{\{\breve{x}^t_{e,r}, \breve{x}^t_{e,s}, \breve{x}^t_{e,u}, \breve{\boldsymbol{x}}^t_{e,d}\}_{d=0}^{\infty}\}_{e=1}^{\hat{e}} \in \breve{\boldsymbol{X}}$ satisfies:

$$
\begin{aligned}
&\breve{x}^t_{e,d} = 0 && \forall e, d > U^e,\\
&\sum_{d=0}^{U^e-1} \breve{x}^t_{e,d} \leq \frac{n_e}{n} && \forall e,\\
&\breve{x}^t_{e,r} \in \left[0, U^e_r + (1-q)\sum_{e'=1}^{\hat{e}} p_{e',e}\frac{n_{e'}}{n}\mathbb{E}\left[\min(U^{e'}, D^{e'})\right]\right] && \forall e.\\
&\breve{x}^t_{e,s} \in \left[0, \sum_{d=U^e+1}^{\infty} d\frac{n_e}{n}\mathbb{P}\left[U^e - \min(U^e, D^e) + \hat{R}^{e,u} = d\right]\right] && \forall e.\\
&\hat{R}^{e,u} \stackrel{d}{=} \mathrm{Pois}\left(q\frac{n}{n_e}\sum_{e'=1}^{\hat{e}} p_{e',e}\frac{n_{e'}}{n}\mathbb{E}\left[\min(U^{e'}, D^{e'})\right]\right) && \forall e\\
&\breve{x}^t_{e,u} \in \left[0, \sum_{d=U^e+1}^{\infty} (d-U^e)\frac{n_e}{n}\mathbb{P}\left[U^e - \min(U^e, D^e) + \hat{R}^{e,u} = d\right]\right] && \forall e.
\end{aligned}
$$

Here, both $\breve{x}^t_{e,s}$ and $\breve{x}^t_{e,u}$ are variables indicating the excess amount of units at a type, albeit with different weights. The upper bounds selected for all three of $\breve{x}^t_{e,s}$, $\breve{x}^t_{e,u}$, $\breve{x}^t_{e,r}$ are not arbitrary as they correspond to the highest inventory, which they can observe through their inventory evolution given in Equations (1.3.6)-(1.3.10). We obtain these expressions by assuming that each station of type e has a post-rebalancing inventory of U^e units. By implementing these upper bounds, we impose a finite bound necessary for compactness while preserving the affine state transitions of the inventory variables.

Remark 1.13.1. *The implicit assumption that we are going to make is that the mapping of the initial inventory $\hat{\boldsymbol{x}}^t$ to this state-space is within these bounds. This assumption, however, is not restrictive as these bounds can be replaced with a maximum function of their resultant values through the initial inventory/the current maximum values, and all results extend.*

As we are given an initial inventory position in $\hat{\boldsymbol{X}}$, we have to map this to an inventory position in $\breve{\boldsymbol{X}}$. To understand how to construct this mapping, we provide a discussion of the variables used. First, the variable $\breve{x}^t_{e,d}$ has the same interpretation as $\hat{x}^t_{e,d}$ for $d \leq U^e$: the proportion of stations of type e and inventory position d, at time t. However, the variable $\breve{\boldsymbol{x}}^t$ differs from $\hat{\boldsymbol{x}}^t$ as $\breve{x}^t_{e,d}$ can only take strictly positive values if d is less than or equal to the assigned upper threshold U^e, while the constraint on the summation for inventory distributions of a certain type e is relaxed from being equal to $\frac{n_e}{n}$ to being less than or equal to $\frac{n_e}{n}$. We now explain why it suffices to only carry the state information in non-aggregate form for $d \leq U^e$. The cost $\hat{c}$ incurred at time t depends on both the action $\hat{\boldsymbol{a}}^t$ and current inventory $\hat{\boldsymbol{x}}^t$. In the following lemma, we prove that the Wasserstein component of $\hat{c}$ can be decomposed into two parts, one which only depends on the pre-rebalancing state $\hat{\boldsymbol{x}}^t$ through $\hat{x}^t_{e,d}$ for $d \geq U^e + 1$ for each e (and is independent of the action), and the second which depends on both the state and the action but only depends on $\hat{\boldsymbol{x}}^t$ through $\hat{x}^t_{e,d}$ for $d \leq U^e - 1$ for each e. Intuitively, the first component can be thought of as rebalancing all stations of type e with more than U^e units to U^e, and the optimal action only depends on the exact proportions of stations with number of units at most $U^e - 1$ through the second part(for the Wasserstein component), for which we can use the alternative compact $\breve{\boldsymbol{x}}^t$ state representation.

Lemma 1.13.2. *The Wasserstein component of the rebalancing cost function of the mean-field model, $\hat{c}$, can be partitioned into two parts as:*

$$
\begin{aligned}
&n\sum_{e=1}^{\hat{e}}\frac{c_{e,e}}{2}\sum_{b=0}^{\infty}\left|\sum_{d=0}^{b}(\hat{x}^t_{e,d}-\hat{a}^t_{e,d})\right| \\
&\qquad = n\sum_{e=1}^{\hat{e}}\frac{c_{e,e}}{2}\sum_{b=U^e+1}^{\infty}(b-U^e)\hat{x}^t_{e,b} + n\sum_{e=1}^{\hat{e}}\frac{c_{e,e}}{2}\sum_{b=0}^{U^e-1}\left|\sum_{d=0}^{b}(\hat{x}^t_{e,d}-\hat{a}^t_{e,d})\right|,
\end{aligned}
$$

Proof.

$$
\begin{aligned}
&n\sum_{e=1}^{\hat{e}}\frac{c_{e,e}}{2}\sum_{b=0}^{\infty}\left|\sum_{d=0}^{b}(\hat{x}^t_{e,d}-\hat{a}^t_{e,d})\right| \\
&= n\sum_{e=1}^{\hat{e}}\frac{c_{e,e}}{2}\sum_{b=0}^{U^e-1}\left|\sum_{d=0}^{b}(\hat{x}^t_{e,d}-\hat{a}^t_{e,d})\right| + n\sum_{e=1}^{\hat{e}}\frac{c_{e,e}}{2}\sum_{b=U^e}^{\infty}\left|\sum_{d=0}^{b}(\hat{x}^t_{e,d}-\hat{a}^t_{e,d})\right|
\end{aligned}
$$

$$
\begin{aligned}
&= n\sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{b=0}^{U^e-1} \left|\sum_{d=0}^{b} (\hat{x}_{e,d}^t - \hat{a}_{e,d}^t)\right| + n\sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{b=U^e}^{\infty} \left|\sum_{d=0}^{b} \hat{x}_{e,d}^{t,e} - \frac{n_e}{n}\right| \\
&= n\sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{b=0}^{U^e-1} \left|\sum_{d=0}^{b} (\hat{x}_{e,d}^t - \hat{a}_{e,d}^t)\right| + n\sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{b=U^e}^{\infty} \sum_{d=b+1}^{\infty} \hat{x}_{e,d}^t \\
&= n\sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{b=U^e+1}^{\infty} (b - U^e)\hat{x}_{e,b}^t + n\sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{b=0}^{U^e-1} \left|\sum_{d=0}^{b} (\hat{x}_{e,d}^t - \hat{a}_{e,d}^t)\right|. \qquad \square
\end{aligned}
$$

Lemma 1.13.2 tells us that we can aggregate $\hat{x}_{e,d}^t$ for $d \geq U^e + 1$ into a single variable, which we do by letting:

$$
\breve{x}_{e,u}^t = \sum_{d=U^e+1}^{\infty} (d - U^e)\hat{x}_{e,d}^t \quad \forall e.
$$

As for the flow component of $\hat{c}$, we observe that the constraints already aggregate the inventory variables. As a result, letting

$$
\breve{x}_{e,s}^t = \sum_{d=U^e+1}^{\infty} d\hat{x}_{e,d}^t \quad \forall e,
$$

the flow component of $\hat{c}$ can be expressed as:

$$
\begin{aligned}
&n \min_{y^{e_1,e_2}} \sum_{e_1=0}^{2\hat{e}} \sum_{e_2=0, e_2\neq e_1}^{2\hat{e}} \left(c_{e_1,e_2} - \frac{c_{e_1,e_1}}{2} - \frac{c_{e_2,e_2}}{2}\right) y^{e_1,e_2} \\
&\text{s.t.} \qquad y^{e_1,e_2} \geq 0 \quad \forall e_1, e_2 \\
&\breve{x}_{e,s}^t + \sum_{b=1}^{U^e} b(\breve{x}_{e,b}^t - \hat{a}_{e,b}^t) = \sum_{e_1=0}^{2\hat{e}} (y^{e,e_1} - y^{e_1,e}) \quad \forall e \in [\hat{e}] \\
&\breve{x}_{e,r}^t - \hat{a}_{e,r}^t = \sum_{e_1=0}^{2\hat{e}} (y^{e,e_1} - y^{e_1,e}) \quad \forall e > \hat{e}
\end{aligned}
$$

We add these two components to form $\hat{c}_2(\breve{\boldsymbol{x}}^t, \hat{\boldsymbol{a}}^t)$, the new rebalancing cost function. Lastly, we prove the main result of this section, where we can re-express the value function of the mean-field model through these new inventory variables.

Proposition 1.13.3. *The value function for the mean-field model can be expressed as:*

$$\hat{V}(\hat{\boldsymbol{x}}^t) = \min_{\{\hat{\boldsymbol{a}}^s \in \hat{\boldsymbol{A}}\}_{s=t}^{\infty}} \sum_{s=t}^{\infty} \gamma^{s-t} \Big(\hat{c}_2(\breve{\boldsymbol{x}}^s, \hat{\boldsymbol{a}}^s) + \hat{N}(\hat{\boldsymbol{a}}^s) \Big) \qquad \forall \hat{\boldsymbol{x}}^t \in \hat{\boldsymbol{X}},$$

where,

$$\breve{x}^t_{e,d} = \hat{x}^t_{e,d} \qquad \forall e, d \leq U^e,$$

$$\breve{x}^t_{e,d} = 0 \qquad \forall e, d > U^e,$$

$$\breve{x}^t_{e,r} = \hat{x}^t_{e,r} \qquad \forall e,$$

$$\breve{x}^t_{e,s} = \sum_{d=U^e+1}^{\infty} d\hat{x}^{t+1}_{e,d} \qquad \forall e,$$

$$\breve{x}^t_{e,u} = \sum_{d=U^e+1}^{\infty} (d - U^e)\hat{x}^{t+1}_{e,d} \qquad \forall e,$$

and $\forall s > t$*,* $\breve{\boldsymbol{x}}^s$ *satisfies Equations* (1.3.6)-(1.3.10).

Proof. Let

$$\breve{x}^t_{e,d} = \hat{x}^t_{e,d} \qquad \forall e, d \leq U^e,$$

$$\breve{x}^t_{e,d} = 0 \qquad \forall e, d > U^e,$$

$$\breve{x}^t_{e,r} = \hat{x}^t_{e,r} \qquad \forall e,$$

$$\breve{x}^t_{e,s} = \sum_{d=U^e+1}^{\infty} d\hat{x}^{t}_{e,d} \qquad \forall e,$$

$$\breve{x}^t_{e,u} = \sum_{d=U^e+1}^{\infty} (d - U^e)\hat{x}^{t}_{e,d} \qquad \forall e.$$

Under these equations, we previously showed that $\hat{c}_2(\breve{\boldsymbol{x}}^t, \hat{\boldsymbol{a}}^t) = \hat{c}(\hat{\boldsymbol{x}}^t, \hat{\boldsymbol{a}}^t)$. Furthermore, Equations (1.3.6)-(1.3.10) can also be expressed as:

$$
\begin{aligned}
\breve{x}_{e,d}^{t+1} &= \hat{x}_{e,d}^{t+1} && \forall e, t, d \leq U^e, \\
\breve{x}_{e,d}^{t+1} &= 0 && \forall e, t, d > U^e, \\
\breve{x}_{e,r}^{t+1} &= \hat{x}_{e,r}^{t+1} && \forall e, t, \\
\breve{x}_{e,s}^{t+1} &= \sum_{d=U^e+1}^{\infty} d\hat{x}_{e,d}^{t+1} && \forall e, t, \\
\breve{x}_{e,u}^{t+1} &= \sum_{d=U^e+1}^{\infty} (d - U^e)\hat{x}_{e,d}^{t+1} && \forall e, t.
\end{aligned}
$$

As a result, we have

$$\hat{c}_2(\breve{\boldsymbol{x}}^{t+s}, \hat{\boldsymbol{a}}^{t+s}) = \hat{c}(\hat{\boldsymbol{x}}^{t+s}, \hat{\boldsymbol{a}}^{t+s}) \quad \forall s \geq 0,$$

and therefore

$$
\begin{aligned}
\hat{V}(\hat{\boldsymbol{x}}^t) &= \min_{\{\hat{\boldsymbol{a}}^s \in \hat{\boldsymbol{A}}\}_{s=t}^{\infty}} \sum_{s=t}^{\infty} \gamma^{s-t} \Big(\hat{c}(\hat{\boldsymbol{x}}^{t+s}, \hat{\boldsymbol{a}}^{t+s}) + \hat{N}(\hat{\boldsymbol{a}}^s) \Big), \\
&= \min_{\{\hat{\boldsymbol{a}}^s \in \hat{\boldsymbol{A}}\}_{s=t}^{\infty}} \sum_{s=t}^{\infty} \gamma^{s-t} \Big(\hat{c}_2(\breve{\boldsymbol{x}}^{t+s}, \hat{\boldsymbol{a}}^{t+s}) + \hat{N}(\hat{\boldsymbol{a}}^s) \Big). \qquad \square
\end{aligned}
$$

1.14 Reformulating the Control Algorithm into a Linear Program

Under the control algorithm provided in Section 1.4, we solve for

$$
\begin{aligned}
\{\tilde{\boldsymbol{a}}^{t+k}\}_{k=0}^{T} \in \arg \min_{\{\hat{\boldsymbol{a}}^{t+k} \in \hat{\boldsymbol{A}}\}_{k=0}^{T}} & \sum_{k=0}^{T-1} \gamma^k \Big(\hat{c}_2(\breve{\boldsymbol{x}}^{t+k}, \hat{\boldsymbol{a}}^{t+k}) + \hat{N}(\hat{\boldsymbol{a}}^{t+k}) \Big) \\
& + \gamma^T \hat{c}_2(\breve{\boldsymbol{x}}^{t+T}, \hat{\boldsymbol{a}}^{t+T}) + \frac{\gamma^T}{1-\gamma} \Big(\gamma \hat{c}_2(\breve{\boldsymbol{x}}^{t+T+1}, \hat{\boldsymbol{a}}^{t+T}) + \hat{N}(\hat{\boldsymbol{a}}^{t+T}) \Big).
\end{aligned}
$$

For $q > 0$, this program cannot be reformulated as a linear program as the mean of $\hat{R}^e$, distributed Poisson, depends on the action taken at other stations. In this section, we provide the exact linear reformulation for the case $q = 0$ as, under this case, all picked-up units become depleted, where under Equation (1.3.8) we observe that the next period's depleted units are linear in action. Nevertheless, the resultant program is not immediately linear as $\hat{c}_2(\breve{\boldsymbol{x}}^{t+k}, \hat{\boldsymbol{a}}^{t+k})$ contains absolute values. In this subsection, we will reformulate this problem as a linear program. First, we introduce the rebalancing cost function $\hat{c}_3(\hat{\boldsymbol{a}}^t, \hat{\boldsymbol{a}}^{t+1})$, which gives the rebalancing cost in period $t+1$ as a function of the actions taken in periods t and $t+1$ where

$$\hat{c}_3(\hat{\boldsymbol{a}}^t, \hat{\boldsymbol{a}}^{t+1}) = \hat{c}_2(\hat{\boldsymbol{x}}^{t+1}, \hat{\boldsymbol{a}}^{t+1}).$$

Using the $\hat{c}_3$ function, we can rewrite the optimization program as

$$\begin{aligned}\{\tilde{\boldsymbol{a}}^{t+k}\}_{k=0}^{T} :\in \arg \min_{\{\hat{\boldsymbol{a}}^{t+k} \in \hat{\boldsymbol{A}}\}_{k=0}^{T}} & \hat{c}_2(\breve{\boldsymbol{x}}^t, \hat{\boldsymbol{a}}^t) + \sum_{k=0}^{T-1} \gamma^k \hat{N}(\hat{\boldsymbol{a}}^{t+k}) \\ & + \sum_{k=1}^{T} \gamma^k \hat{c}_3(\hat{\boldsymbol{a}}^{t+k-1}, \hat{\boldsymbol{a}}^{t+k}) + \frac{\gamma^T}{1-\gamma}\left(\gamma \hat{c}_3(\hat{\boldsymbol{a}}^{t+T}, \hat{\boldsymbol{a}}^{t+T}) + \hat{N}(\hat{\boldsymbol{a}}^{t+T})\right).\end{aligned}$$

It is established in linear programming that a nonlinear optimization program in form of $\min_x |r(x)|$ can be expressed in form $\min_{(z \geq r(x)),(z \geq -r(x))} z$, which is a linear program if and only if r is a linear function. We will use this reformulation on the control algorithm, where we introduce dummy variables (i) $\hat{\boldsymbol{s}}^t = \{\hat{s}^t_{e,b}|_{b=0}^{U^e-1}\}_{e=1}^{\hat{e}}$, (ii) $\hat{\boldsymbol{z}}^{t+k}|_{k=0}^{T} = \{\hat{z}^{t+k}_{e,b}|_{b=0}^{U^e-1}\}_{e=1}^{\hat{e}}|_{k=0}^{T}$, (iii) $\hat{\boldsymbol{l}}^{t+T} = \{\hat{l}^{t+T,e}_{e,b}|_{b=0}^{U^e-1}\}_{e=1}^{\hat{e}}$, in order to replace the absolute value expressions of the rebalancing cost functions. Using these variables, we get the linear program:

$$\begin{aligned}\min_{\{\hat{\boldsymbol{a}}^{t+k}|_{k=0}^{T}, \hat{\boldsymbol{y}}^{t+k}|_{k=0}^{T}, \hat{\boldsymbol{s}}^t, \hat{\boldsymbol{z}}^{t+k}|_{k=0}^{T}, \hat{\boldsymbol{l}}^{t+T}\}} & \sum_{e=1}^{\hat{e}} \sum_{b=0}^{U^e-1} \frac{c_{e,e}}{2} \hat{s}^t_{b,e} \\ & + \sum_{k=0}^{T-1} \sum_{e_1=0}^{2\hat{e}} \sum_{e_2=0, e_2 \neq e_1}^{2\hat{e}} \left(c_{e_1,e_2} - \frac{c_{e_1,e_1}}{2} - \frac{c_{e_2,e_2}}{2}\right) y^{t+k,e_1,e_2} + \sum_{k=0}^{T-1} \sum_{e_1=0}^{\hat{e}} \gamma^k c_h \hat{a}^{t+k}_{e,r} \\ & + \sum_{k=0}^{T-1} \sum_{e=1}^{\hat{e}} \sum_{b=L^e}^{U^e} \hat{a}^{t+k}_{e,b} \gamma^k (c_h b + c_p \mathbb{E}\left[(D^e - b)^+\right]) + \sum_{k=1}^{T} \sum_{e=1}^{\hat{e}} \sum_{b=0}^{U^e-1} \gamma^k \frac{c_{e,e}}{2} \hat{z}^{t+k}_{e,b} \\ & + \sum_{e=1}^{\hat{e}} \sum_{b=0}^{U^e-1} \frac{\gamma^{T+1}}{1-\gamma} \frac{c_{e,e}}{2} \hat{l}^{t+T}_{e,b} + \frac{\gamma^{T+1}}{1-\gamma} \sum_{e_1=0}^{2\hat{e}} \sum_{e_2=0, e_2 \neq e_1}^{2\hat{e}} \left(c_{e_1,e_2} - \frac{c_{e_1,e_1}}{2} - \frac{c_{e_2,e_2}}{2}\right) y^{t+T,e_1,e_2}\end{aligned}$$

$$+\frac{\gamma^T}{1-\gamma}c_h\sum_{e=1}^{\hat{e}}\hat{a}_{e,r}^{t+T}+\sum_{e=1}^{\hat{e}}\sum_{b=L^e}^{U^e}\hat{a}_{e,b}^{t+T}\frac{\gamma^T}{1-\gamma}(c_h b+c_p\mathbb{E}\left[(D^e-b)^+\right]),$$

$$\text{s.t.}\qquad \sum_{b=L^e}^{U^e}\hat{a}_{e,b}^{t+k,e}=\frac{n_e}{n}\quad\forall e,k\in\{0,\dots,T\},\tag{1.14.1}$$

$$\hat{a}_{e,b}^{t+k}\geq 0\quad\forall e,k\in\{0,\dots,T\},b,\tag{1.14.2}$$

$$\frac{n_e}{n}U_0\geq\hat{a}_{e,r}^{t+k}\quad\forall e,k\in\{0,\dots,T\},\tag{1.14.3}$$

$$\hat{a}_{e,r}^{t+k}\geq\frac{n_e}{n}L_0\quad\forall e,k\in\{0,\dots,T\},\tag{1.14.4}$$

$$\breve{x}_{e,s}^t+\sum_{d=1}^{U^e}d(\breve{x}_{e,d}^t-\hat{a}_{e,d}^t)=\sum_{e_1=0}^{2\hat{e}}(y^{t,e,e_1}-y^{t,e_1,e})\quad\forall e\in[\hat{e}],\tag{1.14.5}$$

$$\breve{x}_{e,r}^t-\hat{a}_{e,r}^t=\sum_{e_1=0}^{2\hat{e}}(y^{t,e+\hat{e},e_1}-y^{t,e_1,e+\hat{e}})\quad\forall e\in[\hat{e}],\tag{1.14.6}$$

$$\sum_{d=1}^{U^e}d(\sum_{b=L^e}^{U^e}\hat{a}_{e,b}^{t+k-1}\mathbb{P}\left[b-\min(b,D^e)=d\right]-\hat{a}_{e,d}^{t+k})=\sum_{e_1=0}^{2\hat{e}}(y^{t+k,e,e_1}-y^{t+k,e_1,e})\quad\forall e\in[\hat{e}],k\in[T],\tag{1.14.7}$$

$$\begin{aligned}&\hat{a}_{e,r}^{t+k-1}+\sum_{e'=1}^{\hat{e}}p_{e',e}\sum_{b=L^{e'}}^{U^{e'}}\hat{a}_{e',b}^{t+k-1}\mathbb{E}\left[\min(b,D^{e'})\right]-\hat{a}_{e,r}^{t+k}\\&\qquad=\sum_{e_1=0}^{2\hat{e}}(y^{t+k,e+\hat{e},e_1}-y^{t+k,e_1,e+\hat{e}})\quad\forall e\in[\hat{e}],k\in[T],\end{aligned}\tag{1.14.8}$$

$$\sum_{d=1}^{U^e}d(\sum_{b=L^e}^{U^e}\hat{a}_{e,b}^{t+T}\mathbb{P}\left[b-\min(b,D^e)=d\right]-\hat{a}_{e,d}^{t+T})=\sum_{e_1=0}^{2\hat{e}}(y^{t+T+1,e,e_1}-y^{t+T+1,e_1,e})\quad\forall e\in[\hat{e}]\tag{1.14.9}$$

$$\sum_{e'=1}^{\hat{e}}p_{e',e}\sum_{b=L^{e'}}^{U^{e'}}\hat{a}_{e',b}^{t+T}\mathbb{E}\left[\min(b,D^{e'})\right]=\sum_{e_1=0}^{2\hat{e}}(y^{t+T+1,e+\hat{e},e_1}-y^{t+T+1,e_1,e+\hat{e}})\quad\forall e\in[\hat{e}]\tag{1.14.10}$$

$$y^{t+k,e_1,e_2}\geq 0\quad\forall e_1,e_2,k\in\{0,\dots,T\},\tag{1.14.11}$$

$$\hat{s}_{e,b}^t\geq\sum_{d=0}^{b}(\breve{x}_{e,d}^t-\hat{a}_{e,d}^t)\quad\forall e,b\in\{0,\dots,U^e-1\},\tag{1.14.12}$$

$$\hat{s}_{e,b}^t\geq-\sum_{d=0}^{b}(\breve{x}_{e,d}^t-\hat{a}_{e,d}^t)\quad\forall e,b\in\{0,\dots,U^e-1\},\tag{1.14.13}$$

$$\hat{z}_{e,b}^{t+k}\geq\sum_{d=0}^{b}\left(\sum_{m=L^e}^{U^e}\hat{a}_{e,m}^{t+k-1}\mathbb{P}\left[m-\min(m,D^e)=d\right]-\hat{a}_{e,d}^{t+k}\right)\ \forall e,k\in[T],b\in\{0,\dots,U^e-1\},\tag{1.14.14}$$

$$\hat{z}_{e,b}^{t+k} \geq -\sum_{d=0}^{b}\left(\sum_{m=L^e}^{U^e} \hat{a}_{e,m}^{t+k-1}\mathbb{P}\left[m-\min(m,D^e)=d\right]-\hat{a}_{e,d}^{t+k}\right) \;\forall e,k\in[T], b\in\{0,\ldots,U^e-1\}, \tag{1.14.15}$$

$$\hat{l}_{e,b}^{t+T} \geq \sum_{d=0}^{b}\left(\sum_{m=L^e}^{U^e} \hat{a}_{e,m}^{t+T}\mathbb{P}\left[m-\min(m,D^e)=d\right]-\hat{a}_{e,d}^{t+T}\right) \quad \forall e, b\in\{0,\ldots,U^e-1\}, \tag{1.14.16}$$

$$\hat{l}_{e,b}^{t+T} \geq -\sum_{d=0}^{b}\left(\sum_{m=L^e}^{U^e} \hat{a}_{e,m}^{t+T}\mathbb{P}\left[m-\min(m,D^e)=d\right]-\hat{a}_{e,d}^{t+T}\right) \quad \forall e, b\in\{0,\ldots,U^e-1\}. \tag{1.14.17}$$

Here, (1.14.1), (1.14.2), (1.14.3), and (1.14.4) are the standard action constraints we impose, (1.14.5), (1.14.6), (1.14.7), (1.14.8), (1.14.9), (1.14.10), and (1.14.11) are the flow constraints of the rebalancing cost function, (1.14.12) and (1.14.13) are constraints on dummy variables assigned to the Wasserstein component of the cost term $\hat{c}_2(\breve{\boldsymbol{x}}^t, \hat{\boldsymbol{a}}^t)$, (1.14.14) and (1.14.15) are constraints on dummy variables assigned to the Wasserstein components of the cost term $\sum_{k=1}^{T}\gamma^k\hat{c}_3(\hat{\boldsymbol{a}}^{t+k-1}, \hat{\boldsymbol{a}}^{t+k})$, and (1.14.16) and (1.14.17) are constraints on dummy variables assigned to the Wasserstein component of the cost term $\hat{c}_3(\hat{\boldsymbol{a}}^{t+T}, \hat{\boldsymbol{a}}^{t+T})$. In the LP formulation, the terms $\hat{R}^e, \breve{x}_{e,s}^{t+k}, \breve{x}_{e,s}^{t+k}$ for $k > 0$ are not included. The reason is that as $q = 0$, there is no inflow of charged units to the stations and therefore the resultant number of charged units at a station/type cannot increase through state transition.

1.15 Existence and Utilization of Fixed Points

As discussed in Section 1.4, we can use the existence of a fixed point/steady state to derive an alternative worst-case bound for the algorithm. Formally, a fixed point is defined as:

Definition 1.15.1. $\breve{\boldsymbol{x}}^f$ *is a fixed point of* $\hat{V}$ *if at* $\breve{\boldsymbol{x}}^f$*, there exists an* <u>*optimal*</u> *action* $\hat{\boldsymbol{a}}^f$ *such that when action* $\hat{\boldsymbol{a}}^f$ *is taken at, say, period* t*, then* $\breve{\boldsymbol{x}}^{t+1} = \breve{\boldsymbol{x}}^f$*. We label* $\hat{\boldsymbol{a}}^f$ *as the corresponding fixed point action.*

Conditions under which deterministic dynamic programs possess a fixed point have been investigated before in the literature.

Proposition 1.15.2 ([Flynn, 1979, Theorem 7.1.2]). *The following conditions, together, are sufficient for the existence of a fixed point for the deterministic dynamic program defined in* (1.3.12)*:*

1. $\breve{\boldsymbol{X}}$ *and* $\hat{\boldsymbol{A}}$ *are compact;*
2. $\breve{\boldsymbol{X}}$ *and* $\hat{\boldsymbol{A}}$ *are convex;*
3. *The mapping from* $\hat{\boldsymbol{a}}^t \to \breve{\boldsymbol{x}}^{t+1}$ *is continuous, affine;*
4. *The per-stage cost function is convex.*

While conditions 1, 2, and 4 are satisfied for all possible values of q, condition 3 is satisfied only for $q = 0$. In Proposition 1.15.3, we verify that for $q = 0$, our system satisfies these conditions and that a fixed point exists.

Proposition 1.15.3. *The dynamic program defined in* (1.3.12) *satisfies the fixed point property if* $q = 0$.

Proof. We will prove that $\hat{V}$ satisfies all four conditions stated in Proposition 1.15.2:

1. Both $\breve{\boldsymbol{X}}$ and $\hat{\boldsymbol{A}}$ are closed as they contain all of their boundary points. Furthermore, both $\breve{\boldsymbol{X}}$ and $\hat{\boldsymbol{A}}$ are bounded. As they are both finite dimensional, they are compact.

2. For any two inventory/action positions selected, all the positions in the line segment connecting them are feasible, so $\breve{\boldsymbol{X}}$ and $\hat{\boldsymbol{A}}$ are convex.

3. Under the assumption that $q = 0$, the state transitions in Remark 1.3.1 are affine and hence continuous.

4. The per-stage cost function is $\hat{c}_2(\breve{\boldsymbol{x}}^t, \hat{\boldsymbol{a}}^t) + \hat{N}(\hat{\boldsymbol{a}}^t)$, where the components are defined in Equations (1.3.11) and (1.3.4). For $\hat{N}(\hat{\boldsymbol{a}}^t)$, we observe that all components are linear in $\hat{\boldsymbol{a}}^t$, hence the function is convex in $\hat{\boldsymbol{a}}^t$. Furthermore, since the function $|x - a|$ is jointly convex in x and a, $n \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \breve{x}_{e,u}^t + n \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{b=0}^{U^e-1} \left| \sum_{d=0}^{b} (\breve{x}_{e,d}^t - \hat{a}_{e,d}^t) \right|$ is jointly convex in $\breve{\boldsymbol{x}}^t, \hat{\boldsymbol{a}}^t$. For the flow component of the rebalancing cost, through

Lemma 1.15.4, showing that

$$
\begin{aligned}
& n \min_{y^{e_1,e_2}} \sum_{e_1=0}^{2\hat{e}} \sum_{e_2=0, e_2 \neq e_1}^{2\hat{e}} \left(c_{e_1,e_2} - \frac{c_{e_1,e_1}}{2} - \frac{c_{e_2,e_2}}{2} \right) y^{e_1,e_2} \\
\text{s.t.} \quad & -y^{e_1,e_2} \leq 0 \quad \forall e_1, e_2 \\
& \breve{x}^t_{e,s} + \sum_{b=1}^{U^e} b(\breve{x}^t_{e,b} - \hat{a}^t_{e,b}) = \sum_{e_1=0}^{2\hat{e}} (y^{e,e_1} - y^{e_1,e}) \quad \forall e \in [\hat{e}] \\
& \breve{x}^t_{e,r} - \hat{a}^t_{e,r} = \sum_{e_1=0}^{2\hat{e}} (y^{e,e_1} - y^{e_1,e}) \quad \forall e > \hat{e}
\end{aligned}
$$

is convex in $\boldsymbol{y}$ is sufficient (as the function always provides non-negative values). As the y^{e,e_1} values are linear $\forall e, e_1$ for the objective as well as the constraints, the resultant optimization problem is jointly convex in $\breve{\boldsymbol{x}}^t, \hat{\boldsymbol{a}}^t$ (Boyd and Vandenberghe [2004], Chapter 4.2.1). Consequently, the per-stage cost function is convex. $\square$

Lemma 1.15.4. *(Boyd and Vandenberghe [2004], Chapter 3.2.5). If $f(x,z)$ is convex in (x,z) and C is a convex non-empty set, then the function:*

$$g(x) = \inf_{z \in C} f(x,z)$$

is convex in x, provided $g(x) > -\infty \, \forall x$.

The significance of fixed points for our analysis is that they allow us to obtain an lower bound on the value function. Since once we reach the fixed point, the optimal action returns the state to the fixed point, the optimal cost at some inventory position $\breve{\boldsymbol{x}}^t$ can be lower bounded by taking the optimal fixed point action indefinitely, which in turn has a simple expression, as we state below.

Lemma 1.15.5. *Assume that $\breve{\boldsymbol{x}}^f$ is a fixed point of $\hat{V}$ and $\hat{\boldsymbol{a}}^f$ is the fixed point action. Then:*

$$\hat{V}(\breve{\boldsymbol{x}}^t) \geq \frac{1}{1-\gamma}\left(\gamma \hat{c}_2(\breve{\boldsymbol{x}}^f, \hat{\boldsymbol{a}}^f) + \hat{N}(\hat{\boldsymbol{a}}^f)\right) - \hat{c}_2(\breve{\boldsymbol{x}}^t, \hat{\boldsymbol{a}}^f) \qquad \forall \breve{\boldsymbol{x}}^t \in \breve{\boldsymbol{X}},$$

where the bound is tight if $\breve{\boldsymbol{x}}^t = \hat{\boldsymbol{a}}^f$.

Proof.

$$\hat{V}(\breve{\boldsymbol{x}}^f) = \min_{\hat{\boldsymbol{a}}^t \in \hat{\boldsymbol{A}}} \hat{c}_2(\breve{\boldsymbol{x}}^f, \hat{\boldsymbol{a}}^t) + \hat{N}(\hat{\boldsymbol{a}}^t) + \gamma \hat{V}(\breve{\boldsymbol{x}}^{t+1}).$$

Using the definition of a fixed point,

$$\begin{aligned}
\hat{V}(\breve{\boldsymbol{x}}^f) &= \hat{c}_2(\breve{\boldsymbol{x}}^f, \hat{\boldsymbol{a}}^f) + \hat{N}(\hat{\boldsymbol{a}}^f) + \gamma \hat{V}(\breve{\boldsymbol{x}}^f) \\
&= \frac{1}{1-\gamma}\Big(\hat{c}_2(\breve{\boldsymbol{x}}^f, \hat{\boldsymbol{a}}^f) + \hat{N}(\hat{\boldsymbol{a}}^f)\Big) \\
\hat{V}(\hat{\boldsymbol{a}}^f) &= \frac{1}{1-\gamma}\Big(\gamma \hat{c}_2(\breve{\boldsymbol{x}}^f, \hat{\boldsymbol{a}}^f) + \hat{N}(\hat{\boldsymbol{a}}^f)\Big).
\end{aligned}$$

Furthermore, through the Lipschitz bound provided at (1.12.5), we know that

$$\hat{V}(\breve{\boldsymbol{x}}^t) \geq \hat{V}(\hat{\boldsymbol{a}}^f) - \hat{c}_2(\breve{\boldsymbol{x}}^t, \hat{\boldsymbol{a}}^f).$$

Combining the two equations, we obtain

$$\hat{V}(\breve{\boldsymbol{x}}^t) \geq \frac{1}{1-\gamma}\Big(\gamma \hat{c}_2(\breve{\boldsymbol{x}}^f, \hat{\boldsymbol{a}}^f) + \hat{N}(\hat{\boldsymbol{a}}^f)\Big) - \hat{c}_2(\breve{\boldsymbol{x}}^t, \hat{\boldsymbol{a}}^f) \qquad \forall \breve{\boldsymbol{x}}^t \in \breve{\boldsymbol{X}}. \qquad \square$$

One difficulty in utilizing Lemma 1.15.5 is that we do not know which inventory position corresponds to the fixed point. Identification of the fixed point requires partially solving for the optimal policy, which we already argued is difficult. Nevertheless, we know that there exists at least one fixed point in $\breve{\boldsymbol{X}}$ if $q = 0$ and this information alone allows us to utilize Lemma 1.15.5, which we do so in Proposition 1.15.9. Finally, the proof of Theorem 1.4.2, split into two different Propositions, is presented below.

Theorem 1.4.2. *The optimality gap of the policy* $\tilde{\boldsymbol{\pi}}$ *obtained via* (1.4.1)-(1.4.3) *decreases exponentially with respect to length* T *of the transient horizon:*

$$\hat{V}_{\tilde{\boldsymbol{\pi}}}(\breve{\boldsymbol{x}}^t) - \hat{V}(\breve{\boldsymbol{x}}^t) \leq \frac{\gamma}{1-\gamma} C \gamma^T \quad \forall \breve{\boldsymbol{x}}^t \in \breve{\boldsymbol{X}},$$

where

$$C = 2\sum_{e=1}^{\hat{e}} \left(n_e \max(c_{0,e}, c_{e,0})(U^e - L^e) + n \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0})(U^{r,e} - L^{r,e}) \right).$$

Furthermore, we can derive an alternative bound for $q = 0$ with:

$$\hat{V}_{\tilde{\boldsymbol{\pi}}}(\breve{\boldsymbol{x}}^t) - \hat{V}(\breve{\boldsymbol{x}}^t) \leq C\gamma^T \quad \forall \breve{\boldsymbol{x}}^t \in \breve{\boldsymbol{X}}.$$

Proof. The first statement is proven in Proposition 1.15.6 and the second statement is proven in Proposition 1.15.9.

The initial step of the two proofs is identical, where we utilize the observation that the first T actions that an optimal policy take is feasible for the control algorithm. The proofs differ in action taken by the control algorithm in the $T+1$'th period. For Proposition 1.15.6, which proves the general case, we do not know if a fixed point exists, so the control algorithm takes the same action in period $T+1$ as an optimal policy. The difference between the two policies occurs in period $T+2$, where the control algorithm repeats the same action, but the optimal policy takes a different action. Using fundamental properties of the value function, such as Lipschitzness and triangle inequality of the rebalancing cost function (proven in Lemmas 1.15.7 and 1.15.8), we can bound the difference in per-period costs and then use recursion to obtain the final bound.

For Proposition 1.15.9, which provides a bound under the additional assumption $q = 0$, we take the fixed action at period $T+1$. Through the definition of the fixed action, there exists an inventory distribution for which it is optimal to take that action such that the fixed action moves to that inventory distribution next period. Consequently, we invoke Lipschitzness between that inventory distribution (and the ending inventory distribution at period T), providing the bound. $\square$

In the following proofs, we use the notation $\hat{c}_2(\hat{\boldsymbol{a}}^1, \hat{\boldsymbol{a}}^2)$ to calculate the rebalancing cost between two actions, for simplicity. For our analysis, $\hat{c}_2(\hat{\boldsymbol{a}}^1, \hat{\boldsymbol{a}}^2)$ is equal to $\hat{c}_2(\breve{\boldsymbol{x}}^1, \hat{\boldsymbol{a}}^2)$ with

$$
\begin{aligned}
\breve{x}^1_{e,u} &= 0 && \forall e, \\
\breve{x}^1_{e,s} &= 0 && \forall e, \\
\breve{x}^1_{e,r} &= \hat{a}^1_{e,r} && \forall e, \\
\breve{x}^1_{e,d} &= \hat{a}^1_{e,d} && \forall e, d.
\end{aligned}
$$

Proposition 1.15.6. *The optimality gap of the policy $\tilde{\boldsymbol{\pi}}$ obtained via* (1.4.1)-(1.4.3) *decreases exponentially with respect to length T of the transient horizon:*

$$
\hat{V}_{\tilde{\boldsymbol{\pi}}}(\breve{\boldsymbol{x}}^t) - \hat{V}(\breve{\boldsymbol{x}}^t) \leq \frac{\gamma}{1-\gamma} C \gamma^T \qquad \forall \breve{\boldsymbol{x}}^t \in \breve{\boldsymbol{X}},
$$

where

$$
C = 2 \sum_{e=1}^{\hat{e}} \Big(n_e \max(c_{0,e}, c_{e,0})(U^e - L^e) + n \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0})(U^{r,e} - L^{r,e}) \Big).
$$

Proof. We first express the optimal cost as follows:

$$
\begin{aligned}
\hat{V}(\breve{\boldsymbol{x}}^t) = \min_{\{\hat{\boldsymbol{a}}^{t+k} \in \hat{\boldsymbol{A}}\}_{k=0}^T} & \sum_{k=0}^{T} \gamma^k \Big(\hat{c}_2(\breve{\boldsymbol{x}}^{t+k}, \hat{\boldsymbol{a}}^{t+k}) + \hat{N}(\hat{\boldsymbol{a}}^{t+k}) \Big) \\
& + \gamma^{T+1} \Big(\hat{c}_2(\breve{\boldsymbol{x}}^{t+T+1}, \hat{\boldsymbol{a}}^{t+T+1}) + \hat{V}(\hat{\boldsymbol{a}}^{t+T+1}) \Big),
\end{aligned}
$$

and let $\hat{\boldsymbol{a}}^{t^*}, \hat{\boldsymbol{a}}^{t+1^*}, \ldots$ denote a sequence of optimal actions for $\hat{V}(\breve{\boldsymbol{x}}^t)$. The first $T+1$ terms of this sequence is feasible for the program provided in (1.4.1). Without a guarantee that this sequence of actions is optimal for the control algorithm, we have

$$
\hat{V}_{\tilde{\boldsymbol{\pi}}}(\breve{\boldsymbol{x}}^t) \leq \sum_{k=0}^{T} \gamma^k \Big(\hat{c}_2(\breve{\boldsymbol{x}}^{t+k}, \hat{\boldsymbol{a}}^{t+k^*}) + \hat{N}(\hat{\boldsymbol{a}}^{t+k^*})) \Big) + \frac{\gamma^{T+1}}{1-\gamma} \Big(\hat{c}_2(\breve{\boldsymbol{x}}^{t+T+1^*}, \hat{\boldsymbol{a}}^{t+T^*}) + \hat{N}(\hat{\boldsymbol{a}}^{t+T^*}) \Big).
$$

Subtracting the two functions, costs until the $T+2$'th period cancel out, leaving,

$$\begin{aligned}\hat{V}^t_{\tilde{\boldsymbol{\pi}}}(\breve{\boldsymbol{x}}^t) - \hat{V}(\breve{\boldsymbol{x}}^t) \leq \frac{\gamma^{T+1}}{1-\gamma}\bigg(&\hat{c}_2(\breve{\boldsymbol{x}}^{t+T+1^*}, \hat{\boldsymbol{a}}^{t+T^*}) + \hat{N}(\hat{\boldsymbol{a}}^{t+T^*})\bigg) \\ &- \gamma^{T+1}\bigg(\hat{c}_2(\breve{\boldsymbol{x}}^{t+T+1}, \hat{\boldsymbol{a}}^{t+T+1^*}) + \hat{V}(\hat{\boldsymbol{a}}^{t+T+1^*})\bigg),\end{aligned}$$

applying the triangle inequality (Lemma 1.15.7) on the first term and the Lipschitz bound provided at (1.12.5) on the last term,

$$\begin{aligned}&\leq \frac{\gamma^{T+1}}{1-\gamma}\bigg(\hat{c}_2(\breve{\boldsymbol{x}}^{t+T+1}, \hat{\boldsymbol{a}}^{t+T+1^*}) + \hat{c}_2(\hat{\boldsymbol{a}}^{t+T+1^*}, \hat{\boldsymbol{a}}^{t+T^*}) + \hat{N}(\hat{\boldsymbol{a}}^{t+T^*})\bigg) \\ &\quad - \gamma^{T+1}\bigg(\hat{c}_2(\breve{\boldsymbol{x}}^{t+T+1}, \hat{\boldsymbol{a}}^{t+T+1^*}) - \hat{c}_2(\hat{\boldsymbol{a}}^{t+T+1^*}, \hat{\boldsymbol{a}}^{t+T^*}) + \hat{V}(\hat{\boldsymbol{a}}^{t+T^*})\bigg).\end{aligned} \tag{1.15.1}$$

Then, using the triangle inequality, we have

$$\hat{V}(\hat{\boldsymbol{a}}^{t+T^*}) = \hat{N}(\hat{\boldsymbol{a}}^{t+T^*}) + \gamma\hat{V}(\breve{\boldsymbol{x}}^{t+T+1}).$$

Inserting this expression to Equation (1.15.1) and simplifying the terms

$$\begin{aligned}\hat{V}^t_{\tilde{\boldsymbol{\pi}}}(\breve{\boldsymbol{x}}^t) - \hat{V}(\breve{\boldsymbol{x}}^t) &\leq \gamma^{T+1} 2\hat{c}_2(\hat{\boldsymbol{a}}^{t+T+1^*}, \hat{\boldsymbol{a}}^{t+T^*}) + \frac{\gamma^{T+2}}{1-\gamma}\bigg(\hat{c}_2(\breve{\boldsymbol{x}}^{t+T+1}, \hat{\boldsymbol{a}}^{t+T+1^*}) \\ &\quad + \hat{c}_2(\hat{\boldsymbol{a}}^{t+T+1^*}, \hat{\boldsymbol{a}}^{t+T^*}) + \hat{N}(\hat{\boldsymbol{a}}^{t+T^*})\bigg) - \gamma^{T+2}\bigg(\hat{V}(\breve{\boldsymbol{x}}^{t+T+1})\bigg), \\ &= \gamma^{T+1} 2\hat{c}_2(\hat{\boldsymbol{a}}^{t+T+1^*}, \hat{\boldsymbol{a}}^{t+T^*}) + \frac{\gamma^{T+2}}{1-\gamma}\bigg(\hat{c}_2(\breve{\boldsymbol{x}}^{t+T+1}, \hat{\boldsymbol{a}}^{t+T+1^*}) \\ &\quad + \hat{c}_2(\hat{\boldsymbol{a}}^{t+T+1^*}, \hat{\boldsymbol{a}}^{t+T^*}) + \hat{N}(\hat{\boldsymbol{a}}^{t+T^*})\bigg) \\ &\quad - \gamma^{T+2}\bigg(\hat{c}_2(\breve{\boldsymbol{x}}^{t+T+1}, \hat{\boldsymbol{a}}^{t+T+1^*}) + \hat{V}(\hat{\boldsymbol{a}}^{t+T+1^*})\bigg).\end{aligned}$$

Consequently, we observe a recurrent structure where the sub-optimality gap at each period is bounded by $2\hat{c}_2(\hat{\boldsymbol{a}}^{t+T+1^*}, \hat{\boldsymbol{a}}^{t+T^*})$. We can then express the sub-optimality gap as

$$\hat{V}^t_{\tilde{\boldsymbol{\pi}}}(\breve{\boldsymbol{x}}^t) - \hat{V}(\breve{\boldsymbol{x}}^t) \leq \frac{\gamma^{T+1}}{1-\gamma} 2\hat{c}_2(\hat{\boldsymbol{a}}^{t+T+1^*}, \hat{\boldsymbol{a}}^{t+T^*}),$$

by Lemma 1.15.8,

$$\leq \frac{\gamma^{T+1}}{1-\gamma} 2 \sum_{e=1}^{\hat{e}} \Bigg(n_e \max(c_{0,e}, c_{e,0})(U^e - L^e) n_e$$

$$+ n \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0})(U^{r,e} - L^{r,e}) \Bigg). \qquad \square$$

Lemma 1.15.7. *The $\hat{c}_2$ function satisfies triangle inequality.*

Proof. We want to show that $\forall \breve{\boldsymbol{x}} \in \breve{\boldsymbol{X}}, \hat{\boldsymbol{a}}, \hat{\boldsymbol{z}} \in \hat{\boldsymbol{A}}$,

$$\hat{c}_2(\breve{\boldsymbol{x}}, \hat{\boldsymbol{a}}) \leq \hat{c}_2(\breve{\boldsymbol{x}}, \hat{\boldsymbol{z}}) + \hat{c}_2(\hat{\boldsymbol{z}}, \hat{\boldsymbol{y}}).$$

For the Wasserstein component:

$$\begin{aligned}
& n \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \breve{x}^t_{e,u} + n \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{b=0}^{U^e-1} \left| \sum_{d=0}^{b} (\breve{x}^t_{e,d} - \hat{a}^t_{e,d}) \right| \\
&= n \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \breve{x}^t_{e,u} + n \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{b=0}^{U^e-1} \left| \sum_{d=0}^{b} (\breve{x}^t_{e,d} + \hat{z}^t_{e,d} - \hat{z}^t_{e,d} - \hat{a}^t_{e,d}) \right| \\
&\leq n \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \breve{x}^t_{e,u} + n \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{b=0}^{U^e-1} \left| \sum_{d=0}^{b} (\breve{x}^t_{e,d} - \hat{z}^t_{e,d}) \right| + n \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{b=0}^{U^e-1} \left| \sum_{d=0}^{b} (\hat{z}^t_{e,d} - \hat{a}^t_{e,d}) \right|.
\end{aligned}$$

For the flow component, let $y^{e_1,e_2,1}|_{e_1=1}^{2\hat{e}+1}|_{e_2=1}^{2\hat{e}+1}$ be an optimal solution of $\hat{c}_2(\breve{\boldsymbol{x}}, \hat{\boldsymbol{z}})$ and let $y^{e_1,e_2,2}|_{e_1=1}^{2\hat{e}+1}|_{e_2=1}^{2\hat{e}+1}$ be an optimal solution of $\hat{c}_2(\hat{\boldsymbol{z}}, \hat{\boldsymbol{y}})$. Then, $y^{e_1,e_2,1} + y^{e_1,e_2,2}|_{e_1=1}^{2\hat{e}+1}|_{e_2=1}^{2\hat{e}+1}$ satisfies the constraints of $\hat{c}_2(\breve{\boldsymbol{x}}, \hat{\boldsymbol{a}})$ and gives the same flow component cost as $\hat{c}_2(\breve{\boldsymbol{x}}, \hat{\boldsymbol{z}}) + \hat{c}_2(\hat{\boldsymbol{z}}, \hat{\boldsymbol{y}})$. With no guarantee of optimality for the flow component of $\hat{c}_2(\breve{\boldsymbol{x}}, \hat{\boldsymbol{a}})$, we prove the lemma. $\square$

Lemma 1.15.8. *For any two actions $\hat{\boldsymbol{a}}^1, \hat{\boldsymbol{a}}^2 \in \hat{\boldsymbol{A}}$, the rebalancing cost function $\hat{c}_2$ satisfies:*

$$\hat{c}_2(\hat{\boldsymbol{a}}^1, \hat{\boldsymbol{a}}^2) \leq \sum_{e=1}^{\hat{e}} \Bigg(n_e \max(c_{0,e}, c_{e,0})(U^e - L^e) + n \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0})(U^{r,e} - L^{r,e}) \Bigg).$$

Proof. Through the proof of Proposition 1.13.3, we have

$$\hat{c}_2(\hat{\boldsymbol{a}}^1, \hat{\boldsymbol{a}}^2) = \hat{c}(\hat{\boldsymbol{a}}^1, \hat{\boldsymbol{a}}^2),$$

by Lemma 1.12.1,

$$\begin{aligned}
&\leq n \sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \Bigg(\sum_{b=0}^{L^e-1} \Bigg| \sum_{d=0}^{b} (\hat{a}^1_{e,d} - \hat{a}^2_{e,d}) \Bigg| + \sum_{b=L^e}^{U^e-1} \Bigg| \sum_{d=0}^{b} (\hat{a}^1_{e,d} - \hat{a}^2_{e,d}) \Bigg| \Bigg) \\
&\qquad + n \sum_{e=1}^{\hat{e}} \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0}) |\hat{a}^1_{e,r} - \hat{a}^2_{e,r}| \\
&= n \sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{b=L^e}^{U^e-1} \Bigg| \sum_{d=0}^{b} (\hat{a}^1_{e,d} - \hat{a}^2_{e,d}) \Bigg| + n \sum_{e=1}^{\hat{e}} \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0}) |\hat{a}^1_{e,r} - \hat{a}^2_{e,r}| \\
&\leq n \sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{b=L^e}^{U^e-1} \frac{n_e}{n} + n \sum_{e=1}^{\hat{e}} \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0}) (U^{r,e} - L^{r,e}) \\
&= \sum_{e=1}^{\hat{e}} \Bigg(n_e \max(c_{0,e}, c_{e,0}) (U^e - L^e) + n \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0}) (U^{r,e} - L^{r,e}) \Bigg).
\end{aligned}$$

□

Proposition 1.15.9. *Let $q = 0$. Then, the optimality gap of the policy $\tilde{\boldsymbol{\pi}}$ obtained via* (1.4.1)-(1.4.3) *decreases exponentially with respect to length T of the transient horizon:*

$$\hat{V}_{\tilde{\boldsymbol{\pi}}}(\breve{\boldsymbol{x}}^t) - \hat{V}(\breve{\boldsymbol{x}}^t) \leq C\gamma^T \quad \forall \breve{\boldsymbol{x}}^t \in \breve{\boldsymbol{X}},$$

where

$$C = \sum_{e=1}^{\hat{e}} \Bigg(n_e \max(c_{0,e}, c_{e,0}) (U^e - L^e) + n \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0}) (U^{r,e} - L^{r,e}) \Bigg).$$

Proof. We first express the optimal cost as follows:

$$\hat{V}(\breve{\boldsymbol{x}}^t) = \min_{\{\hat{\boldsymbol{a}}^{t+k} \in \hat{\boldsymbol{A}}\}_{k=0}^{T}} \sum_{k=0}^{T-1} \gamma^k \Big(\hat{c}_2(\breve{\boldsymbol{x}}^{t+k}, \hat{\boldsymbol{a}}^{t+k}) + \hat{N}(\hat{\boldsymbol{a}}^{t+k}) \Big) + \gamma^T \Big(\hat{c}_2(\breve{\boldsymbol{x}}^{t+T}, \hat{\boldsymbol{a}}^{t+T}) + \hat{V}(\hat{\boldsymbol{a}}^{t+T}) \Big),$$

and let $\hat{\boldsymbol{a}}^{t^*}, \hat{\boldsymbol{a}}^{t+1^*}, \ldots$ denote a sequence of optimal actions for $\hat{V}(\breve{\boldsymbol{x}}^t)$. The first T terms of this sequence is feasible for the program provided in (1.4.1). For the $T+1$'th action, we assign:

$$\hat{\boldsymbol{a}}^{t+T} = \hat{\boldsymbol{a}}^{f},$$

where $\breve{\boldsymbol{x}}^f$ is a fixed point of $\hat{V}$ and $\hat{\boldsymbol{a}}^f$ is the corresponding fixed point action. As $q = 0$, through Proposition 1.15.3, we can establish that such a fixed point/fixed point action pair always exists.

Without a guarantee that this sequence of actions is optimal for the control algorithm, we have

$$\begin{aligned}\hat{V}_{\tilde{\boldsymbol{\pi}}}(\breve{\boldsymbol{x}}^t) \leq \sum_{k=0}^{T-1} \gamma^k \Big(\hat{c}_2(\breve{\boldsymbol{x}}^{t+k}, \hat{\boldsymbol{a}}^{t+k^*}) + \hat{N}(\hat{\boldsymbol{a}}^{t+k^*})) \Big) + \gamma^T \hat{c}_2(\breve{\boldsymbol{x}}^{t+T}, \hat{\boldsymbol{a}}^f) \\ + \frac{\gamma^T}{1-\gamma} \Big(\gamma \hat{c}_2(\breve{\boldsymbol{x}}^f, \hat{\boldsymbol{a}}^f) + \hat{N}(\hat{\boldsymbol{a}}^f) \Big).\end{aligned}$$

Subtracting the two functions, costs until the T'th period cancel out, leaving,

$$\begin{aligned}\hat{V}^t_{\tilde{\boldsymbol{\pi}}}(\breve{\boldsymbol{x}}^t) - \hat{V}(\breve{\boldsymbol{x}}^t) \leq \gamma^T \hat{c}_2(\breve{\boldsymbol{x}}^{t+T}, \hat{\boldsymbol{a}}^f) + \frac{\gamma^T}{1-\gamma} \Big(\gamma \hat{c}_2(\breve{\boldsymbol{x}}^f, \hat{\boldsymbol{a}}^f) + \hat{N}(\hat{\boldsymbol{a}}^f) \Big) \\ - \gamma^T \Big(\hat{c}_2(\breve{\boldsymbol{x}}^{t+T}, \hat{\boldsymbol{a}}^{t+T}) + \hat{V}(\hat{\boldsymbol{a}}^{t+T^*}) \Big),\end{aligned}$$

by Lemma 1.15.5,

$$= \gamma^T \Big(\hat{c}_2(\breve{\boldsymbol{x}}^{t+T}, \hat{\boldsymbol{a}}^f) + \hat{V}(\breve{\boldsymbol{a}}^f) \Big) - \gamma^T \Big(\hat{c}_2(\breve{\boldsymbol{x}}^{t+T}, \hat{\boldsymbol{a}}^{t+T^*}) + \hat{V}(\hat{\boldsymbol{a}}^{t+T^*}) \Big),$$

applying the triangle inequality (Lemma 1.15.7) on the first term and the Lipschitz bound provided at (1.12.5) on the last term,

$$\begin{aligned}&\leq \gamma^T \Big(\hat{c}_2(\breve{\boldsymbol{x}}^{t+T}, \hat{\boldsymbol{a}}^{t+T^*}) + \hat{c}_2(\hat{\boldsymbol{a}}^{t+T^*}, \hat{\boldsymbol{a}}^f) + \hat{V}(\breve{\boldsymbol{a}}^f) \Big) \\ &\quad - \gamma^T \Big(\hat{c}_2(\breve{\boldsymbol{x}}^{t+T}, \hat{\boldsymbol{a}}^{t+T^*}) - \hat{c}_2(\hat{\boldsymbol{a}}^{t+T^*}, \hat{\boldsymbol{a}}^f) + \hat{V}(\hat{\boldsymbol{a}}^f) \Big) \\ &\leq 2\gamma^T \hat{c}_2(\hat{\boldsymbol{a}}^{t+T^*}, \hat{\boldsymbol{a}}^f),\end{aligned}$$

by Lemma 1.15.8,

$$\leq 2\gamma^T \sum_{e=1}^{\hat{e}} \Big(n_e \max(c_{0,e}, c_{e,0})(U^e - L^e) n_e$$

$$+ n \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0})(U^{r,e} - L^{r,e})\Bigg). \qquad \square$$

Remark 1.15.10. *For $q = 0$, our approach is different from papers such as Flynn [1979], which focus on optimal policies in the limit $\gamma \to 1$. In such settings, it is possible to solve for the fixed point, and any policy which moves to a fixed point in finite time is guaranteed to have a finite optimality gap in the limit. In addition to covering this special case, our policy is applicable for all discount rates. Furthermore, we do not impose the condition that the action taken at $(T + 1)$st period is a fixed point action as in settings where the discount rate is low, it may be possible to obtain a lower cost through a different action. This is also observed when comparing the two bounds, with $\frac{\gamma}{1-\gamma} < 1$ if $\gamma < 0.5$.*

1.16 Transforming Non-Stationary MDP's through Composite Periods

As highlighted in Subsection 1.6.2, our goal is to re-define the problem in a stationary setting. We do this through using composite periods, where given the initial period t, a composite period covers periods $t, t + 1, \ldots, t + H - 1$. The intuition here is that by modifying the action and the per-period cost function, we will only define inventory at periods that are multiples of H, hence observing identical demand distributions and trip probabilities in each composite period. In accordance, we introduce the composite action $\hat{\boldsymbol{a}}^{H,t} = [\hat{\boldsymbol{a}}^t, \hat{\boldsymbol{a}}^{t+1}, \ldots, \hat{\boldsymbol{a}}^{t+H-1}] \in \hat{\boldsymbol{A}}^{H,t}$, where $\hat{\boldsymbol{A}}^{H,t} = \cup_{k=0}^{H-1} \hat{\boldsymbol{A}}^{t+k}$. In order to re-define the problem and only define inventory at periods which are multiples of H, we refer back to the rebalancing cost function $\hat{c}_3(\hat{\boldsymbol{a}}^t, \hat{\boldsymbol{a}}^{t+1})$, introduced in Appendix 1.14. Under the periodic structure, for any two consecutive periods $t, t + 1$:

$$\hat{c}_3(\hat{\boldsymbol{a}}^t, \hat{\boldsymbol{a}}^{t+1}) = n \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{d=U^{t,e}+1}^{\infty} (d - U^{t,e}) \sum_{b=L^{t,e}}^{U^{t,e}} \hat{a}^t_{e,b} \mathbb{P}\left[b - \min(b, D^{t,e}) + \hat{R}^{t,e} = d\right]$$

$$
\begin{aligned}
&+ n\sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{b=0}^{U^{t+1,e}-1} \left| \sum_{d=0}^{b} (\sum_{k=L^{t,e}}^{U^{t,e}} \hat{a}^{t}_{e,k} \mathbb{P}\left[k - \min(k, D^{t,e}) + \hat{R}^{t,e} = k\right] - \hat{a}^{t+1}_{e,d}) \right| \\
&+ n \min_{y^{e_1,e_2}} \sum_{e_1=0}^{2\hat{e}} \sum_{e_2=0, e_2\neq e_1}^{2\hat{e}} \left(c_{e_1,e_2} - \frac{c_{e_1,e_1}}{2} - \frac{c_{e_2,e_2}}{2} \right) y^{e_1,e_2} \\
\text{s.t.} & \sum_{d=U^{t,e}+1}^{\infty} d \sum_{b=L^{t,e}}^{U^{t,e}} \hat{a}^{t}_{e,b} \mathbb{P}\left[b - \min(b, D^{t,e}) + \hat{R}^{t,e} = d\right] \\
&\quad + \sum_{d=1}^{U^{t+1,e}} d(\sum_{b=L^{t,e}}^{U^{t,e}} \hat{a}^{t}_{e,b} \mathbb{P}\left[b - \min(b, D^{t,e}) + \hat{R}^{t,e} = d\right] - \hat{a}^{t+1}_{e,d}) \\
&\quad = \sum_{e_1=0}^{2\hat{e}} (y^{e,e_1} - y^{e_1,e}) \quad \forall e \in [\hat{e}], \\
& \hat{a}^{t}_{e,r} + (1-q) \sum_{e'=1}^{\hat{e}} p_{e',e} \sum_{b=L^{t,e'}}^{U^{t,e'}} \bar{a}^{t}_{e',b} \mathbb{E}\left[\min(b, D^{t,e'})\right] - \hat{a}^{t+1}_{e,r} \\
&\quad = \sum_{e_1=0}^{2\hat{e}} (y^{e,e_1} - y^{e_1,e}) \quad \forall e > \hat{e}, \\
& y^{e_1,e_2} \geq 0 \quad \forall e_1, e_2.
\end{aligned}
$$

We input this rebalancing cost function as well as composite actions to the mean-field model defined in (1.3.12) to obtain:

$$
\hat{V}^t(\breve{\boldsymbol{x}}^t) = \min_{\hat{\boldsymbol{a}}^{H,t} \in \hat{\boldsymbol{A}}^{H,t}} \hat{c}_2(\breve{\boldsymbol{x}}^t, \hat{\boldsymbol{a}}^t) + \sum_{k=1}^{H-1} \gamma^k \hat{c}_3(\hat{\boldsymbol{a}}^{t+k-1}, \hat{\boldsymbol{a}}^{t+k}) + \sum_{k=0}^{H-1} \gamma^k \hat{N}^{t+k}(\hat{\boldsymbol{a}}^{t+k}) + \gamma^H \hat{V}^t(\breve{\boldsymbol{x}}^{t+H}). \tag{1.16.1}
$$

Consequently, $\hat{V}^t(\breve{\boldsymbol{x}}^t)$ is a stationary deterministic dynamic program with state $\breve{\boldsymbol{x}}^t$, action $\hat{\boldsymbol{a}}^{H,t}$, per-period cost $\hat{c}_2(\breve{\boldsymbol{x}}^t, \hat{\boldsymbol{a}}^t) + \sum_{k=1}^{H-1} \gamma^k \hat{c}_3(\hat{\boldsymbol{a}}^{t+k-1}, \hat{\boldsymbol{a}}^{t+k}) + \sum_{k=0}^{H-1} \gamma^k \Big(\hat{N}^{t+k}(\hat{\boldsymbol{a}}^{t+k}) + \hat{y}^{t+k}(\hat{\boldsymbol{a}}^{t+k}) \Big)$, discount rate γ^H, and state transitions $\breve{\boldsymbol{x}}^t \to \breve{\boldsymbol{x}}^{t+H}$.

It is then straightforward to show that the model defined in (1.16.1) also satisfies the conditions of Proposition 1.15.2 if $q = 0$. What remains is to modify the control algorithm and the worst-case bound. Accordingly, given an initial inventory distribution $\breve{\boldsymbol{x}}^t$ and a parameter $T \geq 0$, we now choose $T+1$ composite actions such that the $(T+1)$'st composite action is repeated for all future periods. We also solve for:

$$\begin{aligned}
\{\tilde{\boldsymbol{a}}^{H,t+kH}\}_{k=0}^{T} \in \arg \min_{\{\hat{\boldsymbol{a}}^{H,t+kH} \in \hat{\boldsymbol{A}}^{H,t}\}_{k=0}^{T}} & \sum_{k=0}^{T-1} \gamma^{kH} \Big(\hat{c}_2(\breve{\boldsymbol{x}}^{t+kH}, \hat{\boldsymbol{a}}^{t+kH}) \\
& + \sum_{r=1}^{H-1} \gamma^r \hat{c}_3(\hat{\boldsymbol{a}}^{t+kH+r-1}, \hat{\boldsymbol{a}}^{t+kH+r}) + \sum_{r=0}^{H-1} \gamma^r \hat{N}^{t+r}(\hat{\boldsymbol{a}}^{t+kH+r}) \Big) \\
& + \gamma^{TH} \hat{c}_2(\breve{\boldsymbol{x}}^{t+TH}, \hat{\boldsymbol{a}}^{t+TH}) + \frac{\gamma^{TH}}{1-\gamma^H} \Big(\gamma \hat{c}_2(\breve{\boldsymbol{x}}^{t+TH+H}, \hat{\boldsymbol{a}}^{t+TH}) \\
& + \sum_{r=1}^{H-1} \gamma^r \hat{c}_3(\hat{\boldsymbol{a}}^{t+TH+r-1}, \hat{\boldsymbol{a}}^{t+TH+r}) + \sum_{r=0}^{H-1} \gamma^r \hat{N}^{t+r}(\hat{\boldsymbol{a}}^{t+TH+r}) \Big).
\end{aligned}$$

The worst-case bound can be updated as follows:

Theorem 1.16.1. *The optimality gap of the policy $\tilde{\boldsymbol{\pi}}^t$ obtained via* (1.4.1)-(1.4.3) *decreases exponentially with respect to length T of the transient horizon:*

$$\hat{V}^t_{\tilde{\boldsymbol{\pi}}^t}(\breve{\boldsymbol{x}}^t) - \hat{V}^t(\breve{\boldsymbol{x}}^t) \leq \frac{\gamma^H}{1-\gamma^H} C^t \gamma^{TH} \quad \forall \breve{\boldsymbol{x}}^t \in \breve{\boldsymbol{X}}^t,$$

where

$$C^t = 2 \sum_{e=1}^{\hat{e}} \Big(n_e \max(c_{0,e}, c_{e,0})(U^{t,e} - L^{t,e}) + n \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0})(U^{t,r,e} - L^{t,r,e}) \Big).$$

Furthermore, we can derive an alternative bound for $q = 0$ with:

$$\hat{V}^t_{\tilde{\boldsymbol{\pi}}^t}(\breve{\boldsymbol{x}}^t) - \hat{V}^t(\breve{\boldsymbol{x}}^t) \leq C^t \gamma^{TH} \quad \forall \breve{\boldsymbol{x}}^t \in \breve{\boldsymbol{X}}^t.$$

The proof for the updated bound directly follows the proof of Theorem 1.4.2 and is omitted.

Remark 1.16.2. *Under the updated worst-case bound, C^t is dependent on the initial period t. We can utilize this dependence by first solving for the t value which minimizes C^t, which we label as t'. We can then construct the equivalent stationary representation for the value* *function at t', and add $t'-t$ ($H+t'-t$ if $t'-t<0$) periods to the current $T\cdot H$ period transient horizon so that we move to the stationary model on a desirable period. By doing so, we can replace C^t with $\min_t C^t$ which is useful for applications where the upper and lower thresholds differ between periods.*

1.17 Supplementary Material for Numerical Analysis

In this section, we provide supplementary material for Section 1.5, which include mathematical formulations for certain policies we implement, additional descriptive information in order to understand the setting we implement our policies, and additional experiments to test our policies performance.

1.17.1 Formulating the Myopic Policy

The optimization program provided in Equation (1.5.1) provides difficulties in computation for two main reasons: (i) it requires minimizing the function $c_p \sum_{e=1}^{\hat{e}} \sum_{i=1}^{n_e} \mathbb{E}\left[(D^e - a_{e,i}^t)^+\right]$, (i) the number of variables linearly increase with the number of stations. Nevertheless, using Appendix 1.11, we are able to exactly solve for the myopic policy through a more manageable mixed-integer reformulation. Specifically, we utilize Proposition 1.11.1 to state that $\exists \boldsymbol{a}_M^t$ such that given inventory position $\boldsymbol{x}^t$,

$$g(\boldsymbol{a}_M^t) \in \arg\min_{\bar{\boldsymbol{a}}^t \in \bar{\boldsymbol{A}}} \hat{c}(g(\boldsymbol{x}^t), \bar{\boldsymbol{a}}^t) + \hat{N}(\bar{\boldsymbol{a}}^t). \tag{1.17.1}$$

Furthermore, again through Proposition 1.11.1, we also have that

$$h(g(\boldsymbol{a}_M^t)) \in \arg\min_{\boldsymbol{a}^t \in \boldsymbol{A}} c(\boldsymbol{x}^t, \boldsymbol{a}^t) + N(\boldsymbol{a}^t).$$

Consequently, at each period t, given inventory $\boldsymbol{x}^t$, we solve for an optimal myopic action through (1.17.1), which is a mixed-integer optimization problem (due to the constraints

$n\bar{a}^t_{e,d} \in \mathbb{N}_0\ \forall e, d$ and $n\bar{a}^t_{e,r} \in \mathbb{N}_0\ \forall e$). We then project this action to the original stochastic problem using the h function.

1.17.2 Formulating the Large Market Policy

While formulating the large market policy, we established that the optimal policy obtained through Equation (1.5.4) requires solving an infinite-dimensional program. Thus we adopt the control algorithm, which only requires solving a $T+1$ period finite horizon problem. Another difficulty of solving for the large market policy is that both the objective function and state transition equations involve piece-wise elements. In this section, we will reformulate these expressions to more manageable quadratic equations using binary variables. Specifically, we will introduce a new binary variable $z^t_{e,i} \in \{0, 1\}$ which takes value 0 if $a^t_{e,i} < \mu_e$ and value 1 otherwise. To ensure this mapping, we use the Big-M method where for some $M >> \max_e \mu_e$, we add the constraints:

$$\begin{aligned} a^t_{e,i} + (1 - z^t_{e,i})M \geq \mu_e & \qquad \forall e, i, \\ \mu_e \geq a^t_{e,i} - z^t_{e,i}M & \qquad \forall e, i. \end{aligned}$$

These constraints ensure that given the interval which $a^t_{e,i}$ is in, the right value for $z^t_{e,i}$ is assigned. Then, we can re-express Equations (1.5.2), (1.5.3) as

$$\begin{aligned} x^{t+1}_{e,i} &= z^t_{e,i}(a^t_{e,i} - \mu_e) + \sum_{e'=1}^{\hat{e}} \sum_{j=1}^{n_{e'}} \frac{q p_{e',e}}{n_e} \left(z^t_{e',j}\mu_{e'} + (1 - z^t_{e',j}) a^t_{e',j} \right) \qquad \forall i \in [n], \\ x^{t+1}_{e,r} &= a^t_{e,r} + (1 - q) \sum_{e'=1}^{\hat{e}} p_{e',e} \sum_{j=1}^{n_{e'}} \left(z^t_{e',j}\mu_{e'} + (1 - z^t_{e',j}) a^t_{e',j} \right) \quad \forall e, \end{aligned}$$

and Equation (1.5.4) as

$$V^M(\boldsymbol{x}^t) = \min_{\boldsymbol{a}^t \in \boldsymbol{A}} c(\boldsymbol{x}^t, \boldsymbol{a}^t) + c_h \sum_{e=1}^{\hat{e}} \sum_{i=1}^{n_e} a^t_{e,i} + c_h \sum_{e=1}^{\hat{e}} a^t_{e,r} + c_p \sum_{e=1}^{\hat{e}} \sum_{i=1}^{n_e} (\mu_e - a^t_{e,i})^+ + \gamma V^M(\boldsymbol{x}^{t+1}).$$

1.17.3 Calibrating Cost Parameters

In this section, we describe how the cost parameters are calibrated.

First, for the rebalancing costs, we use Holyoak [2021], which indicates a price of $2 for moving units and around $4.5 for recharging units in Austin.

For the discount rate, we set $\gamma = 0.95$. We make this selection to balance the trade-off between discounting of future periods with the uncertainty of future demand and routing distributions.

Next, for bike-sharing and scooter-sharing, holding cost corresponds to the cost (wear/tear, depreciation) of having a unit in circulation. Here, we separate the damage a unit receives from usage (included in the penalty cost calculations) from the cost unit receives from staying outside, often under unfavorable weather conditions. To calibrate the holding cost, we start with the unit's purchase price, 800 dollars for the Segway Ninebot ES4 Electric scooter, one of the most common scooters used by Lime Scooter (Strobel [2021]). Based on Hayes [2022], we assume that a scooter's lifespan under circulation is one year, which is less than the estimated 3-5 years for personal usage, but more than the 1-5 months estimated for rental electric scooters. Letting one period consist of a single day, we obtain $c_h = 2.2$.

To calibrate the penalty cost, we calculate the revenue obtained by a scooter per day, minus the damage the scooter observes from usage per day[25]. First, we look at the Austin scooter-sharing data to calculate the revenue per ride, obtaining an average of 11.32 minutes per ride. Using the 2021 Lime pricing of 1 dollar starting price with 29 cents added per minute, we obtain a revenue of 4.48 dollars. Based on Griswold [2018], we assume that the average scooter does five trips per day, giving us a per-period revenue of 22.4 dollars. Finally, we assume a 2-month lifespan for the scooter under constant usage, which, subtracting the holding cost component, gives us $\frac{800-(2.2*60)}{60} = 11.13$. Subtracting the two values, we obtain

25. Several other factors affect the penalty cost, such as fees/taxes paid on this revenue and additional societal welfare from customers using these systems. Nevertheless, as many of these factors are subjective, we do not include them in our calculation and separately conduct an experiment where we vary the penalty cost.

a rounded penalty cost of $c_p = 11.3$.

For depletion probability q, we use Somerville [2021], which reports that the Ninebot ES4 has a battery range of 50 minutes under top speed (longer battery ranges are expected for slower speeds). As we expect usage to be 56.6 minutes per day, we assign $q = 0$[26].

1.17.4 Impact of the Holding Cost

For this experiment, we vary the holding cost at all stations (assumed to be 2.2 previously). The results are shown in Figure 1.5.

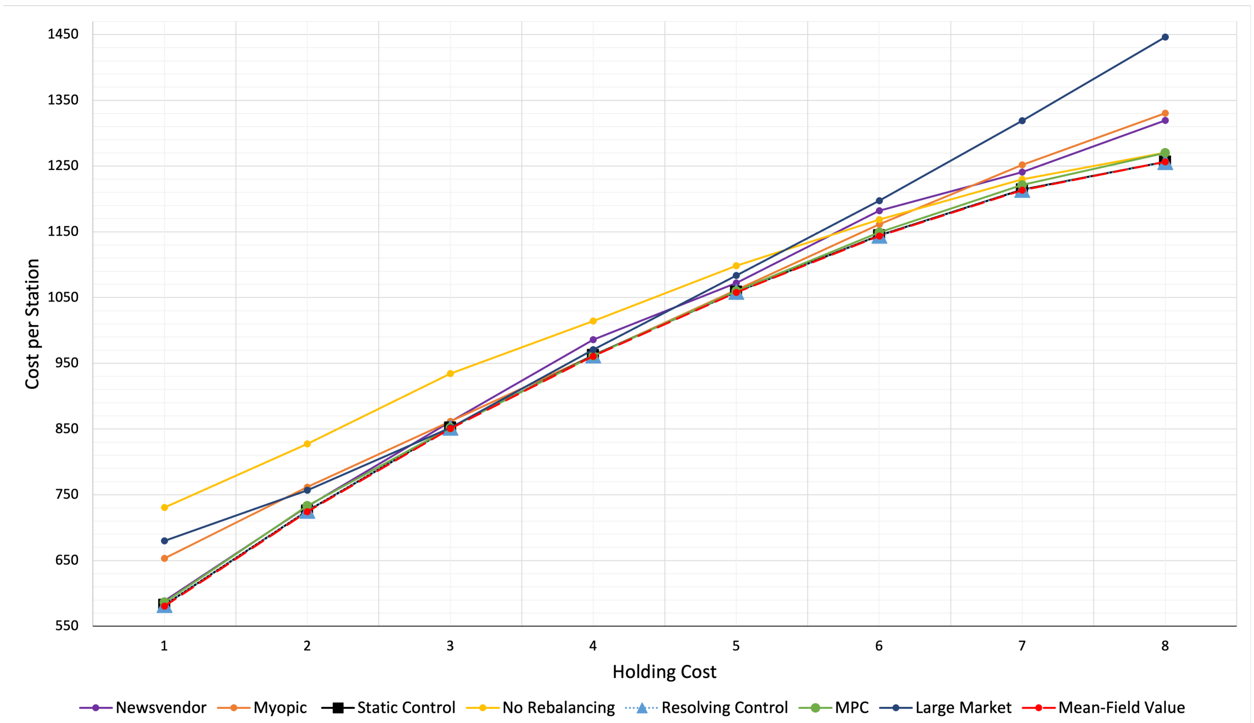


Figure 1.5: Impact of the Holding Cost on the Performance of Policies

We first observe that mean-field-based policies are effective at all holding cost values for the results. These policies can effectively reduce or increase the fleet size according to the holding cost value. One important observation is the poor effectiveness of the large market policy for high holding costs. This is stemmed from the deterministic demand assumption,

26. In Appendix 1.17.6, we look at the performance of our policy for different q values.

which provides the same issues as discussed in Section 1.5.3.

1.17.5 Impact of the Rebalancing Cost

This section will conduct two experiments where we change the rebalancing cost matrix. During our experiments, we assigned a rebalancing cost of \$1 for rebalancing units within the own type, \$2 for rebalancing units to other types, and \$3 for rebalancing units (providing a cost of \$4 or \$5 dollars depending on which type the unit is dropped) and a sourcing/withdrawing cost of \$6. In the first experiment, we will add a constant to these values. In the second experiment, we will only change the cost of rebalancing/recharging/sourcing units to other types.

Adding a Constant to all Rebalancing/Recharging/Sourcing costs

For this experiment, we add a constant k (the value of k varies in the x-axis) to all rebalancing/recharging/sourcing costs. The results are shown in Figure 1.6.

We observe that as the rebalancing costs increase, the performance of the no rebalancing policies improves, while the performance of the newsvendor policy deteriorates. This is as expected, as even though the no rebalancing policy also recharges units, the newsvendor policy requires excessive rebalancing and is more effective in systems with low rebalancing/recharging costs. Lastly, we observe that the mean-field based policies adapt to the cost settings and provide near-optimal results.

Increasing the Cost of Rebalancing Units between types

For this experiment, we add a constant k to the cost of moving units between types. Specifically, we let the cost of rebalancing a unit between types be $2 + k$, recharging a unit and dropping it to a different type be $4 + k$, and the sourcing cost is $6 + k$. The remaining costs are not changed. The results are shown in Figure 1.7.

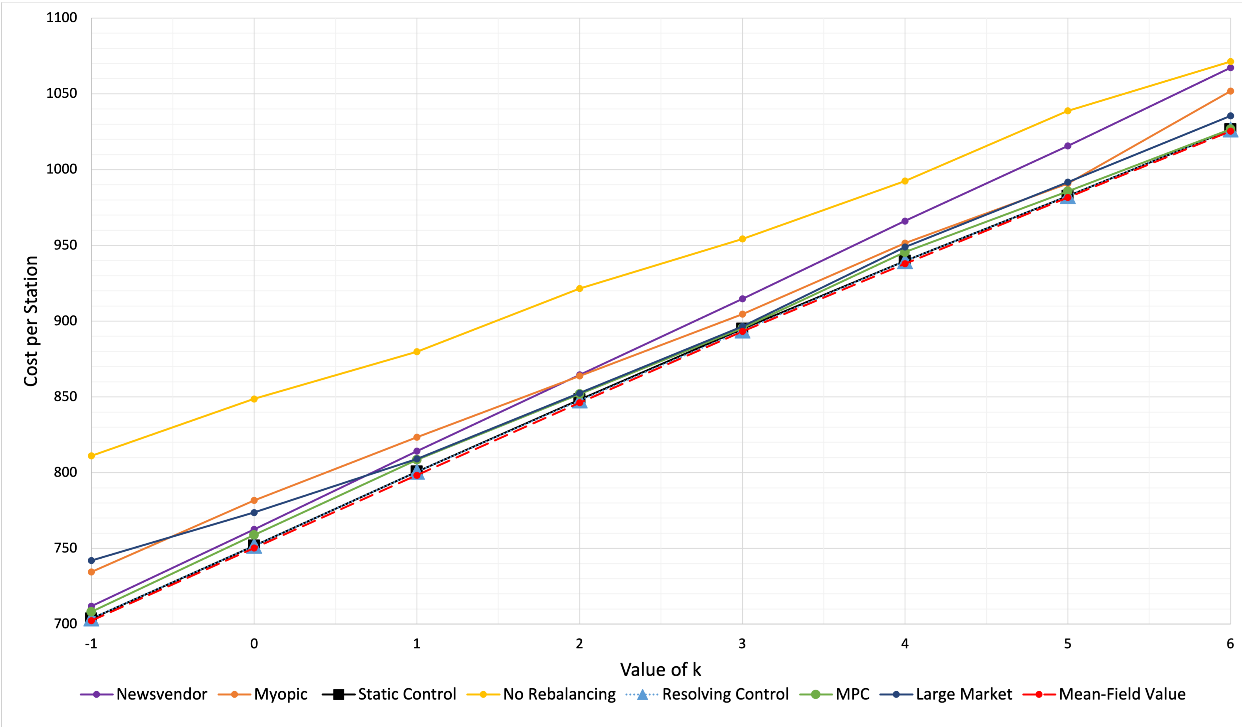


Figure 1.6: Impact of k on the Performance of Policies

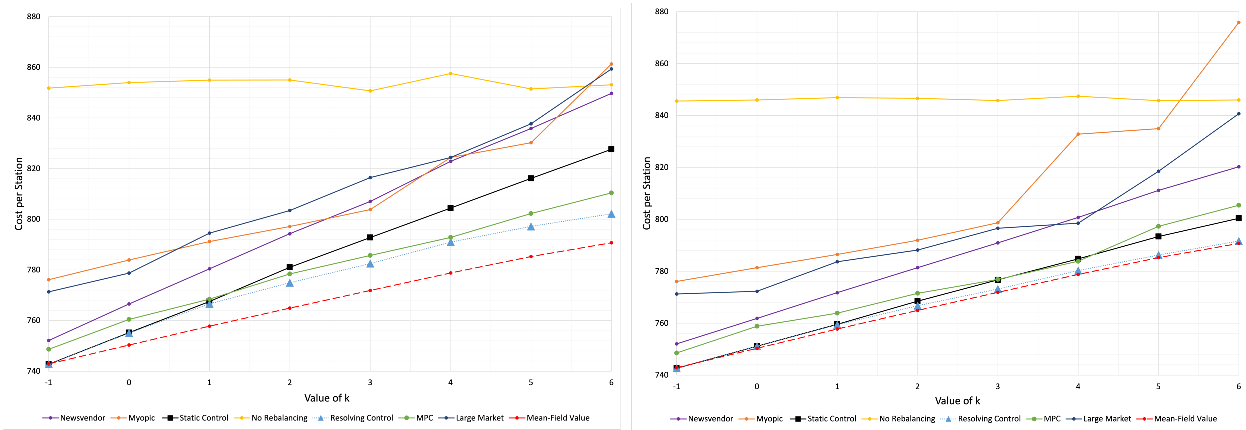
As k increases, the gap between the static and resolving control policies increases. For $n = 16$ and $k > 3$, we see that MPC performs better then static control whereas for $n = 160$, static control performs better for all k. We detail the reasons for the performance of the static policy in Remark 1.5.2.

1.17.6 Impact of the Depletion Probability

For this experiment, we vary the depletion probability q stations (assumed to be 0 previously). Due to the long computation time of MPC and resolving control for the $q > 0$ case, we reduce the discount rate γ to 0.9[27]. Furthermore, we reduce the number of repetitions for the above computational reasons. The results are shown in Figure 1.8.

Through Figure 1.8, we observe that the cost of all policies decreases as q increases, which is intuitive as fewer units are depleted and require recharging. However, the main

27. By reducing γ, we can reduce the simulation horizon of each repetition



(a) Results for n=16

(b) Results for n=160

Figure 1.7: Impact of Cross Rebalancing Cost on the Performance of Policies

observation is that as q changes, the relative performance of policies does not change. This is especially important considering that for $q > 0$, both MPC and resolving control require an excessive computational resource, as they require solving a non-standard non-linear optimization program each period[28]. Consequently, the near-optimal performance of static control is even more critical as the difference in computational resources between static and resolving control is much larger.

1.17.7 Impact of the Upper Threshold

For this experiment, we vary the upper threshold at all stations (assumed to be 17 previously). The results are shown in Figure 1.9.

The main observation is that we see no impact of the upper threshold on the cost of near-optimal policies past a specific value. The reason is that these policies will not take actions that high; hence the upper threshold restriction is redundant.

In contrast, under a low upper threshold value, both the optimal and other policies are severely restricted in their actions. Hence we observe similar performance.

28. While static control policy also requires solving the same optimization program as resolving control, the computation cost is one-off.

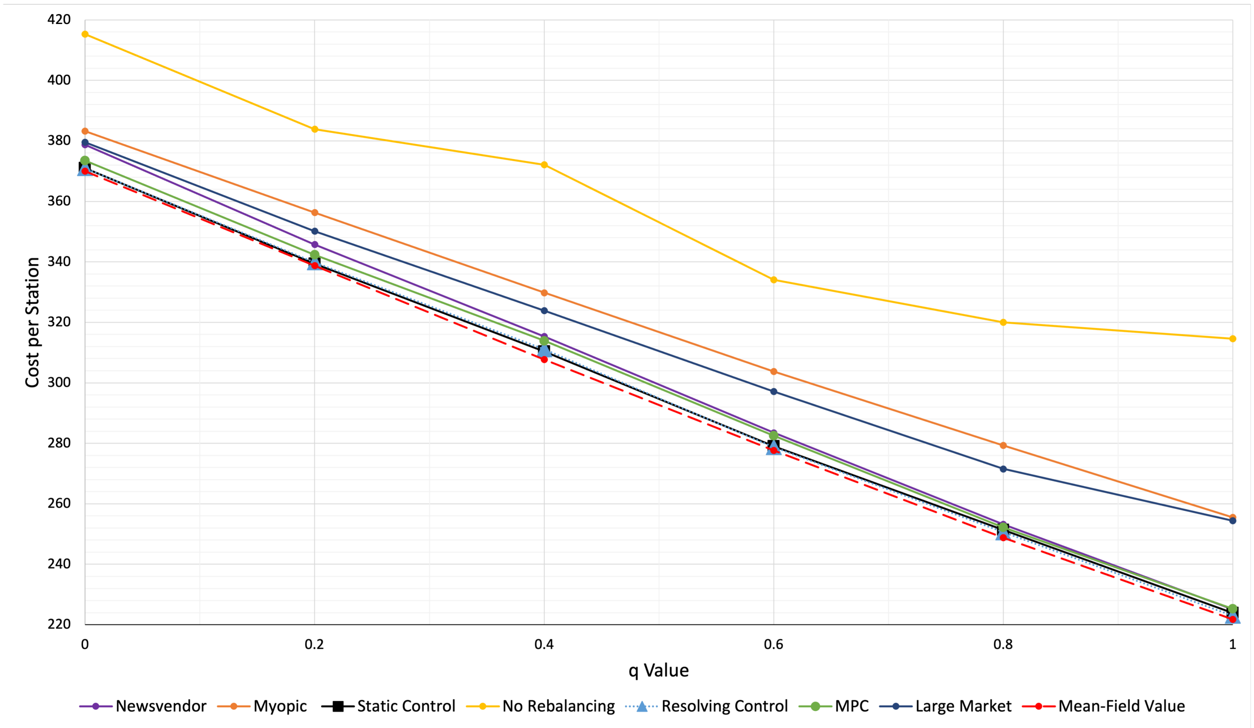


Figure 1.8: Impact of the Depletion Probability on the Performance of Policies

1.18 Proofs of Section 1.6 Results

In this section, we prove that the fixed cost component of the mean-field rebalancing cost can be upper bounded through a linear scaling of the Wasserstein component.

Lemma 1.6.1. *Fixed cost component of the mean-field rebalancing cost satisfies*

$$\sum_{e=1}^{\hat{e}} c_{f,e} \sum_{b=0}^{\infty} \frac{1}{2} \left| \hat{x}_{e,b}^{t} - \hat{a}_{e,b}^{t} \right| \leq \sum_{e=1}^{\hat{e}} c_{f,e} \sum_{b=0}^{\infty} \left| \sum_{d=0}^{b} (\hat{x}_{e,d}^{t} - \hat{a}_{e,d}^{t}) \right| \quad \forall \hat{\boldsymbol{x}}^{t} \in \hat{X}, \hat{\boldsymbol{a}}^{t} \in \hat{A}.$$

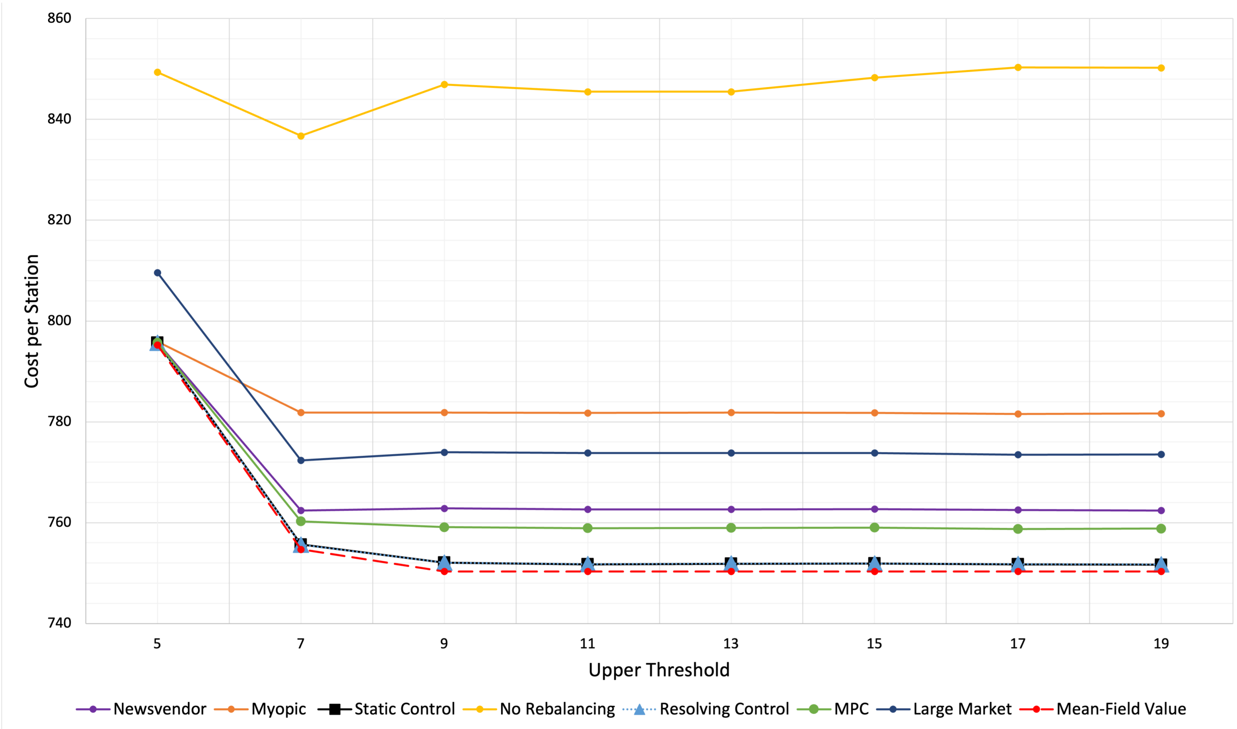


Figure 1.9: Impact of the Upper Threshold on the Performance of Policies

Proof.

$$\begin{aligned}\sum_{e=1}^{\hat{e}} c_{f,e} \sum_{b=0}^{\infty} \frac{1}{2}\left|\hat{x}_{e,b}^{t} - \hat{a}_{e,b}^{t}\right| &= \sum_{e=1}^{\hat{e}} c_{f,e} \sum_{b=0}^{\infty} \frac{1}{2}\left|\sum_{d=0}^{b} \hat{x}_{e,d}^{t} - \sum_{d=0}^{b-1} \hat{x}_{e,d}^{t} - \sum_{d=0}^{b} \hat{a}_{e,d}^{t} + \sum_{d=0}^{b-1} \hat{a}_{e,d}^{t}\right| \\ &\le \sum_{e=1}^{\hat{e}} c_{f,e} \sum_{b=0}^{\infty} \frac{1}{2}\left(\left|\sum_{d=0}^{b} \hat{x}_{e,d}^{t} - \sum_{d=0}^{b} \hat{a}_{e,d}^{t}\right| + \left|\sum_{d=0}^{b-1} \hat{a}_{e,d}^{t} - \sum_{d=0}^{b-1} \hat{x}_{e,d}^{t}\right|\right) \\ &= \sum_{e=1}^{\hat{e}} c_{f,e} \sum_{b=0}^{\infty} \left|\sum_{d=0}^{b} (\hat{x}_{e,d}^{t} - \hat{a}_{e,d}^{t})\right|. \qquad \square\end{aligned}$$

CHAPTER 2

MANAGING INVENTORY IN A NETWORK: PERFORMANCE BOUNDS FOR SIMPLE POLICIES

2.1 Introduction

Recent years have seen a rapid increase in the prevalence of vehicle sharing systems. Such systems, which allow customers to choose where they start and end a ride, suffer from imbalances in inventory accumulation, resulting in difficulties in managing inventory. This problem of customers not returning reusable resources to their original pick-up locations is prevalent in many other applications of service systems and sharing economies. Many industries address this imbalance by repositioning units (rebalancing) or sourcing units from outside sources. Rebalancing is a significant cost driver, to the extent that in industries such as bike-sharing, many firms report rebalancing to be the single most significant operating expense Andersen [2016]. Motivated by this, a growing number of papers look at the rebalancing problem with customer induced relocations. Such papers' primary objective is to characterize optimal policies and provide methodologies to obtain near-optimal policies under restrictive cost assumptions and small network sizes. The problem with this objective is that it may not reflect the realities of the applications, where such assumptions do not hold, and proposed methodologies are computationally intractable in large-scale settings. Consequently, in practice, decision-making is done by using heuristics for which there are no optimality guarantees available in the literature. Motivated by this gap between industry realities and the current state of the research literature, our objective in this paper is to evaluate and bound the performance of the most common policies used in practice.

In our paper, we consider the joint optimization of rebalancing (moving units between stations of a system) and sourcing (moving units from/to stations to/from exterior sources) decisions in the lost-sales setting with customer induced relocations, while allowing for correlations in demand distributions of different stations. Within this framework, we analyze three classes of commonly implemented policies. We look at (*i*) *fixed target policies*, where the system moves to a fixed inventory position (independent of the starting inventory position) at each period, (*ii*) *no action policy*, which never rebalances/sources/withdraws units and allows inventory to evolve naturally, and (*iii*) *myopic/greedy policy*, which solves for the optimal action through neglecting continuation payoffs. We introduce a new coupling technique to obtain worst-case performance bounds for the class of fixed target policies. We provide further insight on cost regimes where such policies are effective. For the no action policy, we provide a methodology to obtain a worst-case performance bound and address cost settings where this policy performs well. We provide tight instances for the worst-case bounds of fixed target policies and the no action policy. Lastly, we compare these policies with the myopic/greedy policy.

Our work is inspired by the transshipment literature, which looks at simultaneously rebalancing and sourcing units to meet demand. For the transshipment problem with two stations, Abouee-Mehrizi et al. [2015] show that optimal policies have a threshold structure. Chen et al. [2015] extend the proof of such results using discrete convexity. A critical difference from this stream of work to ours is that customers consume units instead of relocating them.

Our work is also closely related to proactive rebalancing, where the system rebalances before observing demand, and excessive demand is lost. Here, Zhao et al. [2020] characterize optimal policies for a setting with fixed costs and propose a tractable heuristic. Benjaafar et al. [2022a] provide a characterization of optimal policies and also provides a cutting plane algorithm to obtain an optimal policy under cost assumptions. Another paper in this stream, He et al. [2020], look at a distributionally robust optimization model and shows the strong performance of the proposed policy through numerical examples. One major difference of our work from these is that in contrast to characterizing the optimal policy, we work on evaluating and bounding commonly used policies in practice. Other significant differences in our setting are that these papers assume the total number of units in circulation is fixed and only allow rebalancing between stations and make restrictive assumptions on the underlying cost parameters to prove structural properties such as convexity.

2.2 Problem Definition

We model the joint rebalancing/sourcing problem as an infinite horizon, discounted, discrete-time Markov Decision Process with a discrete state and action space. We let γ denote the discount rate (where $\gamma < 1$) and consider a general network with n stations (nodes). At the start of each period, the central planner observes state $\boldsymbol{x} = [x_1, \cdots, x_n]$, which corresponds to the number of units in each station. We assume that the initial total number of units is bounded, with $\sum_{i=1}^{n} x_i^{(1)} \leq \bar{U} < \infty$. Then, the planner decides on the number of units that will move from station i to station j and the number of units that will move to/from the stations from/to an exterior warehouse with an infinite capacity of units. The planner pays a linear cost of $c_{i,j}$ for each unit moved from station i to station j, a sourcing cost of $c_{s,i}$ for each unit moved from the warehouse to station i, and a withdrawal cost of $c_{m,i}$ for each unit moved from station i to the warehouse. Given these costs, we define a rebalancing/sourcing cost function $c(\boldsymbol{x}, \boldsymbol{y})$, corresponding to the minimum cost to reach target inventory position $\boldsymbol{y} = [y_1, \cdots, y_n]$ from initial inventory position $\boldsymbol{x} = [x_1, \cdots, x_n]$ (where $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{N}_0^n$). Computed through a linear program, $c(\boldsymbol{x}, \boldsymbol{y})$ is expressed as

$$
\begin{aligned}
c(\boldsymbol{x}, \boldsymbol{y}) = \min & \sum_{i=1}^{n}\sum_{j=1}^{n} c_{i,j} w_{i,j} + \sum_{i=1}^{n} c_{s,i} s_i + \sum_{i=1}^{n} c_{m,i} m_i, \\
\text{s.t} \qquad & \sum_{j=1}^{n} w_{j,i} - \sum_{j=1}^{n} w_{i,j} + s_i - m_i = y_i - x_i \quad \forall i, \\
& w_{i,j}, s_i, m_i \geq 0 \quad \forall i, j,
\end{aligned}
$$

where $w_{i,j}$ denotes the number of units rebalanced from station i to station j, s_i denotes the number of units sourced to station i (from the warehouse), and m_i denotes the number of units withdrawn from station i (to the warehouse). We then show that the c function satisfies the following fundamental property (all proofs are deferred to the appendix):

Lemma 2.2.1. *The c function satisfies triangle inequality.*

After rebalancing/sourcing is completed and the system moves to a new inventory position $\boldsymbol{y}$, a holding cost (accounting for depreciation of units) of $c_{h,i}$ is paid per unit in station $i \in 1, \cdots, n$, corresponding to a total holding cost of $\sum_{i=1}^{n} c_{h,i} y_i$. Then, we move to the last stage, where customers start arriving at stations, pick up (if a unit is available) and relocate units (one per customer) to randomly sampled destinations. We assume customers return their units by the end of the period. Here, we let $\boldsymbol{D} = [D_1, \cdots, D_n]$ denote the vector of non-negative, discrete valued random variables with finite mean and variance, corresponding to the number of customers arriving at each station (demand), where $\min(y_i, D_i)$ units depart station i. Furthermore, a lost-sales penalty of $c_{p,i}0$ is paid for each unit of excess demand in station i, corresponding to a total penalty cost of $\sum_{i=1}^{n} c_{p,i}(D_i - y_i)^+$. We define a function $N(\boldsymbol{y})$ to collect the holding and penalty costs. We label it the newsvendor function because it resembles the classic inventory order problem under uncertain demand. We formally express $N(\boldsymbol{y})$ as:

$$N(\boldsymbol{y}) = \sum_{i=1}^{n} N_i(y_i) = \sum_{i=1}^{n} \Bigg(c_{h,i} y_i + c_{p,i} \mathbb{E}\left[(D_i - y_i)^+\right] \Bigg).$$

To denote the stochasticity in the destination stations of these departing units, we introduce the routing vector $\boldsymbol{R} = [\boldsymbol{R}_1, \cdots, \boldsymbol{R}_n]$ with $\boldsymbol{R}_i = [R_{i,1}, \cdots, R_{i,n}]$, where $\boldsymbol{R}_i$ is a multinomial random variable with the number of trials equal to $\min(y_i, D_i)$ and success probabilities are given by $[p_{i,1}, \cdots, p_{i,n}]$. Here, $p_{i,j}$ denotes the probability that a customer who picks up a unit from station i drops the unit to station j. Furthermore, we assume that demand distributions and $p_{i,j}$ values are perfectly known and stationary with $\boldsymbol{R}^t \overset{d}{=} \boldsymbol{R}$ and $\boldsymbol{D}^t \overset{d}{=} \boldsymbol{D}$ for all time periods $t \in \mathbb{N}$. Most notably, we do not assume independence of demand across stations.

Through this formulation, we capture the stochasticity in both customer arrivals and the destination of arriving customers. As a result, the initial inventory of the next period, as a function of the inventory position moved to in the current period, can be expressed as:

$$x_i^{(2)} = y_i - \min(y_i, D_i) + \sum_{j=1}^{n} R_{j,i} \qquad \forall i.$$

In this paper, we only focus on stationary policies, where a stationary policy $\pi : \mathbb{N}_0^n \to \mathbb{N}_0^n$ is a mapping from the set of non-negative inventory positions to the set of non-negative target inventory positions. We assume that all policies satisfy:

$$\sum_{i=1}^{n} \pi_i(\boldsymbol{x}) \leq \bar{U} \qquad \forall \boldsymbol{x}.$$

Given a policy π, we denote the discounted cost-to-go, given initial inventory position $\boldsymbol{x}$, as $V_\pi(\boldsymbol{x})$. Formally:

$$V_\pi(\boldsymbol{x}) = \mathbb{E}_{\{\boldsymbol{D}^t, \boldsymbol{R}^t\}_{t=1}^{\infty}} \left[\sum_{t=1}^{\infty} \gamma^{t-1} \Big(c(\boldsymbol{x}^{(t)}, \pi(\boldsymbol{x}^{(t)})) + N(\pi(\boldsymbol{x}^{(t)})) \Big) \right],$$

where $\boldsymbol{x}^{(1)} = \boldsymbol{x}$. The corresponding value function is defined as

$$V(\boldsymbol{x}) = \min_{\pi} V_\pi(\boldsymbol{x}).$$

We can also define the value function in terms of the Bellman recursion as:

$$V(\boldsymbol{x}) = \min_{\boldsymbol{y}} c(\boldsymbol{x}, \boldsymbol{y}) + N(\boldsymbol{y}) + \gamma \mathbb{E}_{\{\boldsymbol{D}, \boldsymbol{R}\}} \left[V(\boldsymbol{x}^{(2)}) \right].$$

Lastly, given an initial inventory position $\boldsymbol{x}$, an optimal policy π^* satisfies $\pi^*(\boldsymbol{x}) = \boldsymbol{y}$ with

$$\boldsymbol{y} \in \arg\min c(\boldsymbol{x}, \boldsymbol{y}) + N(\boldsymbol{y}) + F(\boldsymbol{y}).$$

The following basic property of an optimal policy will become important in our analysis:

Lemma 2.2.2. *For any inventory position $\boldsymbol{x}$, there exists an optimal policy satisfying*

$$\pi^*(\pi^*(\boldsymbol{x})) = \pi^*(\boldsymbol{x}).$$

2.3 Fixed Target Policies

As discussed in Section 2.1, most of the studies on inventory rebalancing focus on the structural properties of optimal policies. This paper takes a different path to gain insight into the policies used in practice through coupling analysis. In this section, we study one of the most common of such policies as *fixed target* policies, where the system selects a target inventory position $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_n]$ and moves to this inventory position regardless of the initial inventory position. Such policies are popular because, given an $\boldsymbol{\alpha}$, online decision making reduces to solving the linear program of finding the cheapest way to move to $\boldsymbol{\alpha}$ from the current inventory position. To this end, we let $V_{\boldsymbol{\alpha}}(\boldsymbol{x})$ denote the discounted cost under target inventory position $\boldsymbol{\alpha}$ and show that it has a closed-form expression given by the following lemma:

Lemma 2.3.1. *For any initial inventory position $\boldsymbol{x}$,*

$$V_{\boldsymbol{\alpha}}(\boldsymbol{x}) = c(\boldsymbol{x}, \boldsymbol{\alpha}) + \frac{1}{1-\gamma}\left(N(\boldsymbol{\alpha}) + \gamma \mathbb{E}_{\{\boldsymbol{D},\boldsymbol{R}\}}\left[c(\boldsymbol{x}^{(2)}, \boldsymbol{\alpha})\right]\right).$$

While we used the recurrent structure of a fixed target policy to express its cost through a closed-form solution, the resultant value does not tell us anything about its performance, which is analogous to understanding how worse this policy is with respect to an optimal policy. This motivates a primary result of our paper, developing a proof methodology to bound the performance of any fixed target policy.

2.3.1 Proving a Worst-Case Bound Through Coupling

In this subsection, we are concerned with a single objective, finding a worst-case bound to the ratio $\frac{V_{\boldsymbol{\alpha}}(\boldsymbol{x})}{V(\boldsymbol{x})}$ for any $\boldsymbol{\alpha}$ and inventory position $\boldsymbol{x}$. The mathematical difficulty here is that optimal policies and the fixed target policy move to different inventory positions, inducing different inventory positions in later periods. We take the primary step to overcome this issue in the following theorem, where we prove that we can construct this bound through a coupling technique that uses the structure of a fixed target policy to express the cost of the fixed target policy at the sequence of states visited by an optimal policy. Using this technique, we upper bound the ratio of the costs under the two policies through a maximization problem which solves for an inventory position maximizing the ratio of modified per-period costs of the two policies. We present the formal theorem below:

Theorem 2.3.2. *For any $\boldsymbol{\alpha}$ and inventory position $\boldsymbol{x}$, discounted cost $V_{\boldsymbol{\alpha}}(\boldsymbol{x})$ satisfies:*

$$\frac{V_{\boldsymbol{\alpha}}(\boldsymbol{x})}{V(\boldsymbol{x})} \leq \max_{\boldsymbol{z}\in\mathcal{X}_{\boldsymbol{x}}} \frac{c(\boldsymbol{z},\boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}}}{c(\boldsymbol{z},\pi^*(\boldsymbol{z})) + N(\pi^*(\boldsymbol{z}))},$$
$$C_{\boldsymbol{\alpha}} = \mathbb{E}_{\{\boldsymbol{D},\boldsymbol{R}\}}\left[c(\boldsymbol{x}^{(2,1)},\boldsymbol{\alpha})\right],$$

$\boldsymbol{x}^{(2,1)}$ *denotes the random vector for the inventory position after demand and routing realization when the action taken is $\boldsymbol{\alpha}$ and $\mathcal{X}_{\boldsymbol{x}}$ is expressed as*

$$\mathcal{X}_{\boldsymbol{x}} = \{\boldsymbol{z} : \exists T, \pi^* \ s.t\ \mathbb{P}\left[\boldsymbol{x}^{(T)} = z\right] > 0 \vee \mathbb{P}\left[\pi^*(\boldsymbol{x}^{(T)}) = z\right] > 0\}.$$

While Theorem 2.3.2 is an essential step in bounding the performance of fixed target policies, it presents difficulties in using it in its current form. The optimization problem providing the bound requires evaluating an optimal policy, which provides a dilemma as we are working with fixed target policies because of the difficulty of solving for an optimal policy. We solve this problem through the assistance of the following proposition, which provides an important property for this optimization problem.

Proposition 2.3.3. *There $\exists z^*$ satisfying:*

$$z^* \in \arg\max_{\boldsymbol{z}\in\mathcal{X}_{\boldsymbol{x}}} \frac{c(\boldsymbol{z},\boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}}}{c(\boldsymbol{z},\pi^*(\boldsymbol{z})) + N(\pi^*(\boldsymbol{z}))},$$
$$z^* = \pi^*(z^*).$$

The proposition allows us to add the constraint $\boldsymbol{z} = \pi^*(\boldsymbol{z})$ to the optimization problem provided in Theorem 2.3.2 without loss. Adding the constraint and using Lemma 2.2.2, we obtain:

$$\frac{V_{\boldsymbol{\alpha}}(\boldsymbol{x})}{V(\boldsymbol{x})} \leq \max_{\boldsymbol{z}\in\mathcal{X}_{\boldsymbol{x}},\boldsymbol{z}=\pi^*(\boldsymbol{z})} \frac{c(\boldsymbol{z},\boldsymbol{\alpha})+N(\boldsymbol{\alpha})+\gamma C_{\boldsymbol{\alpha}}}{N(\boldsymbol{z})},$$

relaxing the bound by dropping the optimal policy constraint, we get

$$\frac{V_{\boldsymbol{\alpha}}(\boldsymbol{x})}{V(\boldsymbol{x})} \leq \max_{\boldsymbol{z}\in\mathcal{X}_{\boldsymbol{x}}} \frac{c(\boldsymbol{z},\boldsymbol{\alpha})+N(\boldsymbol{\alpha})+\gamma C_{\boldsymbol{\alpha}}}{N(\boldsymbol{z})}.$$

Given the above bound, we can derive tight cases for instances where we can compute elements of $\mathcal{X}_{\boldsymbol{x}}$. One such instance is the following: a single station system where demand is deterministic and equal to λ, $\alpha = \lambda$, the initial inventory position x satisfies $x > \lambda$, $\gamma = 0$, and $c_p > c_m > c_h$.

We solve for the optimal action with these parameters, obtaining $\pi^*(x) = x$. As we are dealing with a single station problem, all λ units departing return to the station by the end of the period, providing $X_{\boldsymbol{x}} = \{x\}$. Computing the numerator of the bound gives $(x-\lambda)c_m + \lambda c_h$. Inputting these values to the bound, we show that it is tight with

$$\frac{V_{\boldsymbol{\alpha}}(\boldsymbol{x})}{V(\boldsymbol{x})} = \max_{\boldsymbol{z}\in\mathcal{X}_{\boldsymbol{x}}} \frac{c(\boldsymbol{z},\boldsymbol{\alpha})+N(\boldsymbol{\alpha})+\gamma C_{\boldsymbol{\alpha}}}{N(\boldsymbol{z})} = \frac{(x-\lambda)c_m+\lambda c_h}{x c_h}.$$

In most realistic settings, however, solving for the elements of $\mathcal{X}_{\boldsymbol{x}}$ is computationally intractable, so we relax the constraint $\boldsymbol{z} \in \mathcal{X}_{\boldsymbol{x}}$ to obtain our final bound; providing

$$\frac{V_{\boldsymbol{\alpha}}(\boldsymbol{x})}{V(\boldsymbol{x})} \leq \max_{\boldsymbol{z}\in\mathbb{N}_0^n,\sum_{i=1}^n z_i\leq\bar{U}} \frac{c(\boldsymbol{z},\boldsymbol{\alpha})+N(\boldsymbol{\alpha})+\gamma C_{\boldsymbol{\alpha}}}{N(\boldsymbol{z})}. \tag{2.3.1}$$

Under Equation (2.3.1), we quantify the two differences a fixed target policy has from an optimal policy. First, the per-period costs will differ from an optimal policy under a fixed target policy, as the actions are different. The term $c(\boldsymbol{z},\boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) - N(\boldsymbol{z})$ represents this component, where the negative terms were canceled out in the process. Second, as discussed previously, the system will reach different inventory positions under different policies, and we bound the cost of the system moving to states induced by the fixed target policy rather than an optimal policy through $\gamma C_{\boldsymbol{\alpha}}$.

Equation (2.3.1) also implies that the difference between fixed target policies and the optimal policy is driven by the rebalancing/sourcing necessary for fixed target policies, which is intuitive as fixed target policies require excessive rebalancing/sourcing. Imbalanced networks further exacerbate this cost difference, as excessive rebalancing/sourcing is required to move the system back to the same inventory position each period. Consequently, fixed target policies perform better in systems where moving units are cheap; because, in that case, optimal policies also excessively move units to minimize newsvendor costs (we study a policy to this end in the following subsection). However, optimal policies seldom move units in systems with high rebalancing/sourcing costs, causing the difference in costs.

While discussing the implications of Equation (2.3.1), one important observation is that this bound holds for all fixed target policies and consequently provides large values for ineffective selections of $\boldsymbol{\alpha}$, as these policies perform poorly. As a result, we defer the discussion on computing values for the worst-case bound to Subsection 2.3.3 after introducing a method for selecting $\boldsymbol{\alpha}$.

2.3.2 Newsvendor Policy

Whether the rebalancing/sourcing costs are low or high, the selection of $\boldsymbol{\alpha}$ is important for the performance of a fixed target policy. Using Lemma 2.3.1, we can solve for the cost minimizing $\boldsymbol{\alpha}$ through:

$$\boldsymbol{\alpha}^* \in \arg\min_{\boldsymbol{\alpha}} c(\boldsymbol{x}, \boldsymbol{\alpha}) + \frac{1}{1-\gamma}\left(N(\boldsymbol{\alpha}) + \mathbb{E}_{\{\boldsymbol{D},\boldsymbol{R}\}}\left[c(\boldsymbol{x}^{(2)}, \boldsymbol{\alpha})\right]\right). \tag{2.3.2}$$

Unfortunately, for most systems, solving (2.3.2) is computationally intractable. While it is possible to simulate $\mathbb{E}_{\{\boldsymbol{D},\boldsymbol{R}\}}\left[c(\boldsymbol{x}^{(2)}, \boldsymbol{\alpha})\right]$ for a specific $\boldsymbol{\alpha}$, the number of simulations necessary to accurately capture the $2n$ dimensional probability transition matrix from $\boldsymbol{\alpha}$ to all possible realizations of $\boldsymbol{x}^{(2)}$ is computationally intractable. Building over the earlier observation that fixed target policies are effective when the rebalancing/sourcing costs are low, we propose another fixed target policy aimed at such cost regimes by choosing an $\boldsymbol{\alpha}$ satisfying (assuming $\exists \boldsymbol{K}$ s.t. $\sum_{i=1}^{n} K_i \leq \bar{U}$):

$$\boldsymbol{K} \in \arg\min_{\boldsymbol{\alpha}} N(\boldsymbol{\alpha}).$$

Under the newsvendor policy, the system moves to $\boldsymbol{K}$ (newsvendor solution) at every period. If the sourcing costs $\boldsymbol{c_s} = [c_{s,1}, \cdots, c_{s,n}]$ and withdrawal costs $\boldsymbol{c_m} = [c_{m,1}, \cdots, c_{m,n}]$ are equal to $\mathbf{0}$; $c(\boldsymbol{x}, \boldsymbol{y}) = 0 \ \forall \boldsymbol{x}, \boldsymbol{y}$ and this policy is optimal as Equation (2.3.1) provides

$$\frac{V_{\boldsymbol{K}}(\boldsymbol{x})}{V(\boldsymbol{x})} \leq \max_{\boldsymbol{z}} \frac{N(\boldsymbol{K})}{N(\boldsymbol{z})} = 1.$$

Easy to compute, the newsvendor policy is beneficial when rebalancing/sourcing is cheap but ill-equipped when rebalancing/sourcing is expensive. To approach such instances, we consider another established policy, a direct opposite to the newsvendor policy, in Section 2.4.

2.3.3 Interpreting the Bound

While we proved a worst-case bound for the class of fixed target policies, the bound is hard to interpret due to the underlying optimization program providing its value. To provide insight, we consider a more simplified setting where all stations have identical holding and penalty costs, $c_s = c_{s,i} = c_{m,i} \, \forall i$ (system has a homogeneous sourcing cost) and $\frac{c_{i,j}}{2} > c_s \, \forall i, j$ (all units move through sourcing). We only use a single cost parameter for rebalancing/sourcing through these assumptions. Furthermore, we can establish that:

$$\lim_{\boldsymbol{z} \to \boldsymbol{\infty}} \frac{c(\boldsymbol{z}, \boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}}}{N(\boldsymbol{z})} = \frac{c_s}{c_h}.$$

This expression holds as when $z_i \to \infty \ \forall i$, $c(\boldsymbol{z}, \boldsymbol{\alpha})$ increases with rate nc_s and $N(\boldsymbol{z})$ increases with rate nc_h. For large $\bar{U}$, this value provides a good approximation for the maximization program when sourcing costs are high. While this indicates a high magnitude for the bound for large c_s, as discussed earlier, the optimal policy does not rebalance when rebalancing/-sourcing costs are high. Excluding extreme cases where fixed target policies do not require rebalancing/sourcing (such as if $\boldsymbol{\alpha} = \boldsymbol{x} = \mathbf{0}$), cost under a fixed target policy is linear in c_s through Lemma 2.3.1. As a result, the ratio of the two costs becomes linear in c_s for large c_s, which the bound captures.

One of our main goals through this bound will be to compute lower-bounds on cost regimes where fixed target policies can be considered, which we do so through a numerical experiment. We further assume that $\boldsymbol{\alpha} = \boldsymbol{K}, n = 10, \gamma = 0.9, \bar{U} = 10,000, c_h = 1, c_p = 2$, $p_{i,j} = \frac{1}{n}\ \forall i, j$, and demand at each station is an i.i.d. Poisson random variable with mean $\lambda = 6$. Given these values, we compute the bound for different c_s. To do that, we first simulate $\frac{C_{\boldsymbol{K}}}{c_s}$ by simulating 10,000,000 instances of demand and routing realizations and computing the average number of units sourced/withdrawn. Inputting $\frac{C_{\boldsymbol{K}}}{c_s}$, for each c_s value, we compute the bound and provide the results in Figure 2.1.

As seen in Figure 2.1, value of the bound is approximately equal to c_s/c_h for $c_s \geq 2$. As c_s increases, all fixed target policies will approach the $\frac{c_s}{c_h}$ line due to the above reasons. Most importantly, the bound indicates that fixed target policies are near-optimal for $0 < c_s < c_p$. This cost regime is critical to address as it is feasible in real-life applications and where fixed target policies are thought to be effective. We quantify this performance through this bound, which can be helpful to practitioners who now have a better theoretical sense of a simple heuristic's performance (or loss of performance) in a complex problem structure. Furthermore, while we have assumed that units only move through sourcing to simplify the analysis, cheaper local rebalancing options are available in real-life settings. While not demonstrated in Figure 2.1, the bound also considers such options and can show that the newsvendor policy performs even better at such settings (as the rebalancing cost function is only at the numerator).

To approach settings with high rebalancing/sourcing costs, we consider another estab-

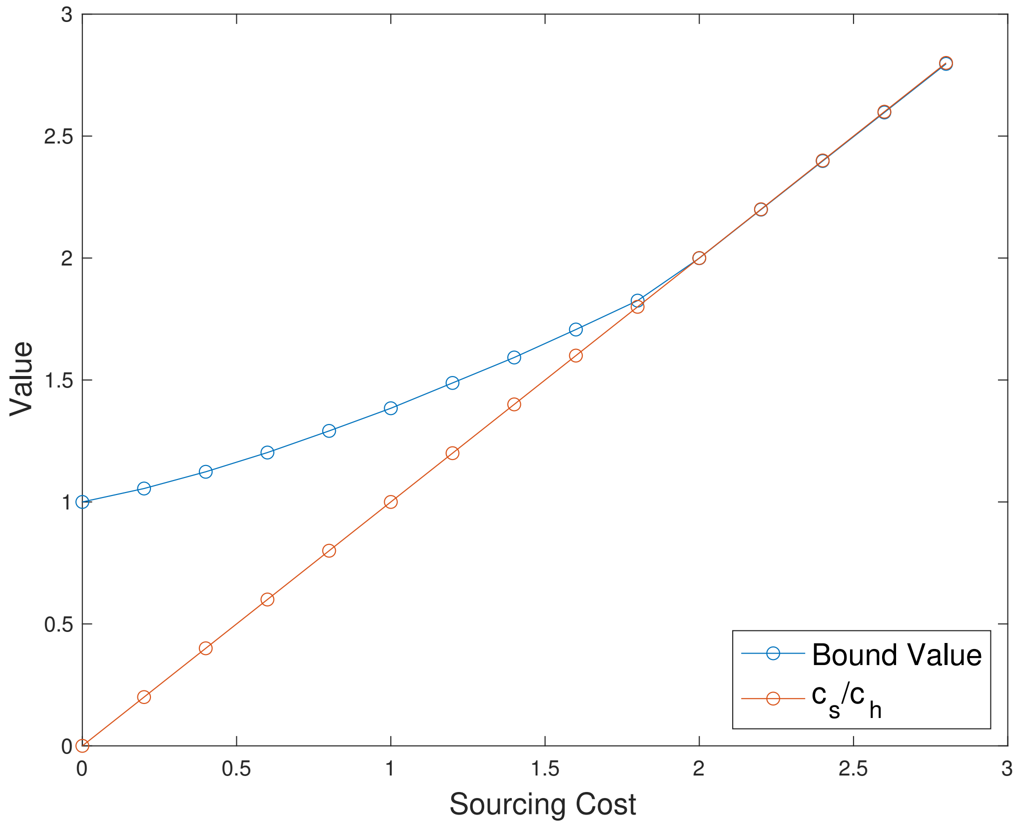


Figure 2.1: Computing the Worst-Case Bound for Newsvendor Policy

lished policy, directly opposite to the newsvendor policy, in the next section.

2.4 No Action Policy

In settings where rebalancing/sourcing is expensive, an optimal policy transitions from one which excessively rebalances/sources units to a policy with minimal rebalancing/sourcing. Useful in such settings, we look at the no action policy, which, as the name implies, never moves units. Formally, letting π^N denote the no action policy, $\pi^N(\boldsymbol{x}) = \boldsymbol{x}$ for any inventory position $\boldsymbol{x}$.

As in the previous section, our main concern will be to generalize the performance of this policy in different cost/network settings through a worst-case bound.

2.4.1 Proving a Worst-Case Bound

As in Section 2.3.1, we are concerned with a single goal, finding a worst-case bound to the ratio $\frac{V_{\pi^N}(\boldsymbol{x})}{V(\boldsymbol{x})}$ for any inventory position $\boldsymbol{x}$. The mathematical difficulty is the same as the no action policy reaches different inventory positions than an optimal policy. However, our approach to resolving this difficulty is different. Instead of coupling over the inventory trajectory of an optimal policy, we will first generate the bound by looking at an alternative setting where rebalancing is free. Then, using our results from the previous section to establish an optimal policy in that setting, we will prove a worst-case bound for that setting that can be generalized. Finally, we present the theorem below:

Theorem 2.4.1. *Given any initial inventory position $\boldsymbol{x}$, the no action policy π^N satisfies:*

$$\frac{V_{\pi^N}(\boldsymbol{x})}{V(\boldsymbol{x})} \leq \frac{N(\mathbf{0})}{N(\boldsymbol{K})} + \frac{1}{N(\boldsymbol{K})} \max_{j \in \{1,\cdots,n\}} \left(N_j(\sum_{i=1}^{n} x_i) - N_j(0) \right).$$

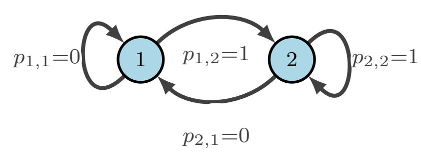


Figure 2.2: Two Station Feedforward Network

The worst-case bound provided in Theorem 2.4.1 follows intuition as due to the lack of rebalancing/sourcing, most units collect over time in stations with the highest net inflow rates. The bound merely considers the worst possible scenario where the station that starts with all the units keeps these units while also having the highest increase in the newsvendor cost. We formalize this scenario by providing an instance where the worst-case bound is tight. We look at the network defined in Figure 2.2 and assume that demand at both stations is identical, holding and penalty costs are homogeneous, $x_1 = 0$, and $\boldsymbol{c_s}, \boldsymbol{c_m} = \mathbf{0}$.

First, through Theorem 2.4.1, we have

$$\frac{V_{\pi^N}(\boldsymbol{x})}{V(\boldsymbol{x})} \leq \frac{N_1(0) + N_2(x_2)}{N(\boldsymbol{K})}.$$

Second, as $\boldsymbol{c_s}, \boldsymbol{c_m} = \mathbf{0}$, newsvendor policy is optimal and

$$V(\boldsymbol{x}) = \sum_{t=1}^{\infty} \gamma^{t-1} N(\boldsymbol{K}).$$

Third, for $V_{\pi^N}(\boldsymbol{x})$, due to the feed-forward structure of the network, all x_2 units in station 2 will remain there, providing

$$V_{\pi^N}(\boldsymbol{x}) = \sum_{t=1}^{\infty} \gamma^{t-1} \Big(N_1(0) + N_2(x_2) \Big).$$

Taking the ratio, we show tightness with

$$\frac{V_{\pi^N}(\boldsymbol{x})}{V(\boldsymbol{x})} = \frac{N_1(0) + N_2(x_2)}{N(\boldsymbol{K})}.$$

2.4.2 Interpreting the Bound

Similar to the discussion for the worst-case bound of fixed target policies, the bound for the no action policy depends on the cost parameters, albeit with different correlations. To show this dependence, we focus on a system with n stations, identical deterministic demand of λ at each station, and identical penalty and holding costs (with $c_p \geq c_h$). Under this setting, $\boldsymbol{K} = \boldsymbol{\lambda}$, and we can express the worst-case bound of the no action policy as:

$$\frac{V_{\pi^N}(\boldsymbol{x})}{V(\boldsymbol{x})} \leq \frac{n\lambda c_p + c_p(\lambda - \sum_{i=1}^n x_i)^+ + c_h \sum_{i=1}^n x_i - \lambda c_p}{n\lambda c_h},$$

letting $\sum_{i=1}^n x_i = \lambda$ (which also minimizes the above expression with respect to $\sum_{i=1}^n x_i$),

$$\begin{aligned} &= \frac{n\lambda c_p + c_h\lambda - \lambda c_p}{n\lambda c_h} \\ &= \frac{1}{n} + \frac{n-1}{n}\frac{c_p}{c_h}. \end{aligned} \tag{2.4.1}$$

This expression follows intuition as the no action policy is effective when paying the penalty costs to avoid moving units is cheaper. In the case where $c_h = c_p$, $\frac{1}{n} + \frac{n-1}{n}\frac{c_p}{c_h} = 1$ as in this case, it is optimal to not rebalance/source units. Conversely, as the penalty cost increases, the optimal policy transitions to a newsvendor policy and excessively rebalances/sources units to avoid paying penalty costs, with the ratio of the costs depending on the penalty cost.

To compute the no action policy worst-case bound in the stochastic setting, we construct at a similar setup to above (n stations, identical penalty and holding costs) but assume $\sum_{i=1}^{n} x_i = K$ and that demand at each station is an i.i.d. Poisson random variable with mean λ. Given this setting, the realization of K depends on the values of c_h, c_p, and λ. Furthermore, we can express the worst-case bound of the no action policy as:

$$\frac{V_{\pi^N}(\boldsymbol{x})}{V(\boldsymbol{x})} \leq \frac{n\lambda c_p + c_p \mathbb{E}\left[(D-K)^+\right] + c_h K - \lambda c_p}{nKc_h + nc_p \mathbb{E}\left[(D-K)^+\right]}. \tag{2.4.2}$$

We let $c_h = 1$ and compare, for different realizations of c_p, no action worst-case bound values for the deterministic system (provided in Equation (2.4.1)) and the stochastic system (provided in Equation (2.4.2)). We provide the results in Figure 2.3.

As seen in Figure 2.3, the bound provides significantly lower results when there is variability in demand, as in that case, the optimal policy also suffers from paying penalty costs. The worst-case bound incorporates the effectiveness of the no action policy when the penalty costs are low but, more importantly, includes the stochasticity of the system through the expected penalty cost parameter in the denominator. While we expect an optimal policy to have higher costs under higher demand variability, quantifying this is difficult, especially in networked settings. The worst-case bound does quantify this effect and informs practitioners of the costs faced when the system does not actively reposition units given demand uncertainty. It also allows us to understand the difficulty of rebalancing/sourcing and increasing costs a system suffers under rising variability better.

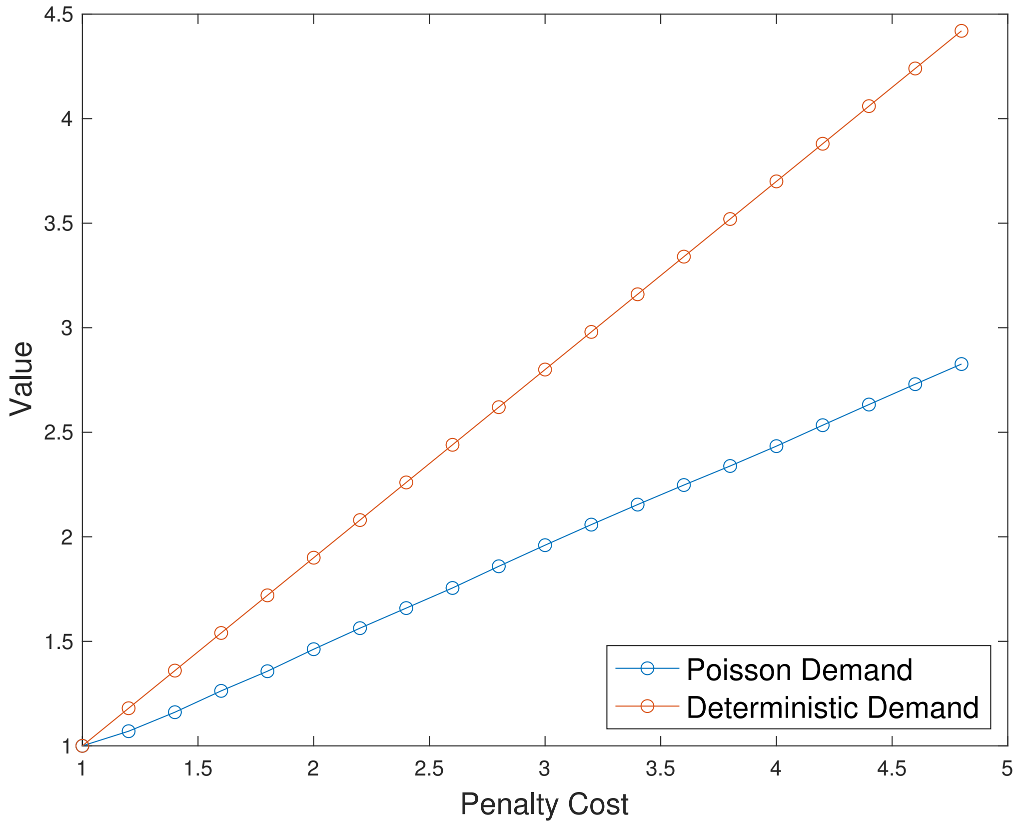


Figure 2.3: Computing the Worst-Case Bound for the No Action Policy

2.5 Myopic Policy

We proved a worst-case performance bound for fixed target policies in Section 2.3 and a worst-case performance bound for the no action policy in Section 2.4. In both cases, the bounds are optimization problems dependent on system parameters and indicate that these policies perform well in some cost regimes and poorly in others. We now show that some standard policies can perform arbitrarily bad. To this end, we evaluate the myopic/greedy policy, which solves the problem assuming that $\gamma = 0$. Formally, the myopic policy, for any initial inventory position $\boldsymbol{x}$, solves

$$\pi^M(\boldsymbol{x}) \in \arg\min_{\boldsymbol{y}} c(\boldsymbol{x}, \boldsymbol{y}) + N(\boldsymbol{y}).$$

The myopic policy is a popular benchmark in inventory management papers. Comparing the myopic policy with the newsvendor policy, we observe that the myopic policy solves the problem at each period, assuming that $\gamma = 0$. In contrast, the newsvendor policy solves the problem assuming that $\boldsymbol{c_s}, \boldsymbol{c_m} = \mathbf{0}$. Even though $\boldsymbol{c_s}, \boldsymbol{c_m} = \mathbf{0}$ is a stronger assumption (as it indirectly disregards both rebalancing costs and continuation payoffs where $\gamma = 0$ only disregards continuation payoffs), we provide an example where the myopic policy performs arbitrarily worse than the newsvendor policy.

For this example, we use the network given in Figure 2.2. For this example, we will relax our previous restriction and let $\bar{U} = \infty$, as our goal is to show a setting where the total number of units in circulation under the myopic policy explodes. We further assume that demand at both stations is deterministic with mean $\lambda > 0$ while holding and penalty costs are homogeneous. We also assume that problem parameters satisfy: (i) $\frac{c_{1,2}}{2} = \frac{c_{2,1}}{2} > c_s = c_m$, (ii) $c_p > 2c_s > 2c_h$, (iii) $x_1 = x_2 = \lambda$.

Lemma 2.5.1. *Given the above network structure, demand distribution, and cost assumptions,*

$$\lim_{\gamma \to 1^-} \frac{V_{\pi^M}([\lambda, \lambda])}{V_{\boldsymbol{K}}([\lambda, \lambda])} = \infty.$$

The poor performance of the myopic policy for γ close to 1 can be generalized to other networks and demand distributions. When penalty costs are high, holding costs are low, and it is cheaper to move units from the warehouse rather than through stations, the myopic policy moves units to the system to avoid penalty costs but does not withdraw units. This causes an inflow of units at some stations with no outflow in others. Consequently, the number of units and incurred holding costs explode (unless constrained by $\bar{U}$). Policies such as the no action policy and fixed target policies, where a constant number of units are in circulation, do not suffer from this problem.

2.6 Concluding Remarks

In this work, we considered an inventory rebalancing/sourcing problem and evaluated/bounded the performance of commonly practiced policies. Thus, we took a step in bridging the gap between the current set of papers, which focus on the characterization of near-optimal policies that can be identified only possible under restrictive assumptions, and industry practice of using simple heuristics for which no performance guarantees were known as yet.

An important future direction of our work would be using the policies and performance bounds we provided to construct more complex policies with theoretical guarantees. One way to do so would be to construct hybrid policies that use fixed target policies at some stations and the no action policy at others.

2.7 Proofs of Results

Lemma 2.2.1. *The c function satisfies triangle inequality.*

Proof. We want to show that $\forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$,

$$c(\boldsymbol{x}, \boldsymbol{y}) \leq c(\boldsymbol{x}, \boldsymbol{z}) + c(\boldsymbol{z}, \boldsymbol{y}).$$

Let $\bar{w}_{i,j}|_{i=1}^{n}|_{j=1}^{n}, \bar{s}_i|_{i=1}^{n}, \bar{m}_i|_{i=1}^{n}$ be an optimal solution of $c(\boldsymbol{x}, \boldsymbol{z})$ and let $\hat{w}_{i,j}|_{i=1}^{n}|_{j=1}^{n}, \hat{s}_i|_{i=1}^{n}, \hat{m}_i|_{i=1}^{n}$ be an optimal solution of $c(\boldsymbol{z}, \boldsymbol{y})$. Then, $\bar{w}_{i,j} + \hat{w}_{i,j}|_{i=1}^{n}|_{j=1}^{n}, \bar{s}_i + \hat{s}_i|_{i=1}^{n}, \bar{m}_i + \hat{m}_i|_{i=1}^{n}$ is clearly a feasible solution for $c(\boldsymbol{x}, \boldsymbol{y})$ with cost equal to $c(\boldsymbol{x}, \boldsymbol{z}) + c(\boldsymbol{z}, \boldsymbol{y})$. With no guarantee that this solution is optimal for $c(\boldsymbol{x}, \boldsymbol{y})$, we prove the lemma. $\square$

Lemma 2.2.2. *For any inventory position $\boldsymbol{x}$, there exists an optimal policy satisfying*

$$\pi^*(\pi^*(\boldsymbol{x})) = \pi^*(\boldsymbol{x}).$$

Proof. We use proof by contradiction where we assume that $\exists \bar{\boldsymbol{x}}$ such that for all optimal policies,

$$\pi^*(\pi^*(\bar{\boldsymbol{x}})) \neq \pi^*(\bar{\boldsymbol{x}}).$$

First, we let $F(\boldsymbol{y})$ denote the continuation payoff with

$$F(\boldsymbol{y}) = \gamma \mathbb{E}_{\{\boldsymbol{D},\boldsymbol{R}\}}\left[V(\boldsymbol{x}^{(2)})\right],$$

to differentiate between the continuation payoffs of two actions. From the optimality of $\pi^*(\bar{\boldsymbol{x}})$ at inventory position $\bar{\boldsymbol{x}}$:

$$c(\bar{\boldsymbol{x}}, \pi^*(\bar{\boldsymbol{x}})) + N(\pi^*(\bar{\boldsymbol{x}})) + F(\pi^*(\bar{\boldsymbol{x}})) \leq c(\bar{\boldsymbol{x}}, \pi^*(\pi^*(\bar{\boldsymbol{x}}))) + N(\pi^*(\pi^*(\bar{\boldsymbol{x}}))) + F(\pi^*(\pi^*(\bar{\boldsymbol{x}}))).$$

From the optimality of $\pi^*(\pi^*(\bar{\boldsymbol{x}}))$ at inventory position $\pi^*(\bar{\boldsymbol{x}})$:

$$c(\pi^*(\bar{\boldsymbol{x}}), \pi^*(\pi^*(\bar{\boldsymbol{x}})))+N(\pi^*(\pi^*(\bar{\boldsymbol{x}})))+F(\pi^*(\pi^*(\bar{\boldsymbol{x}}))) < c(\pi^*(\bar{\boldsymbol{x}}), \pi^*(\bar{\boldsymbol{x}}))+N(\pi^*(\bar{\boldsymbol{x}}))+F(\pi^*(\bar{\boldsymbol{x}})).$$

Summing the two equations and canceling out:

$$c(\bar{\boldsymbol{x}}, \pi^*(\bar{\boldsymbol{x}})) + c(\pi^*(\bar{\boldsymbol{x}}), \pi^*(\pi^*(\bar{\boldsymbol{x}}))) < c(\bar{\boldsymbol{x}}, \pi^*(\pi^*(\bar{\boldsymbol{x}}))).$$

Through Lemma 2.2.1, we have reached a contradiction and thus proved the given lemma. □

Lemma 2.3.1. *For any initial inventory position $\boldsymbol{x}$,*

$$V_{\boldsymbol{\alpha}}(\boldsymbol{x}) = c(\boldsymbol{x}, \boldsymbol{\alpha}) + \frac{1}{1-\gamma}\left(N(\boldsymbol{\alpha}) + \gamma \mathbb{E}_{\{\boldsymbol{D},\boldsymbol{R}\}}\left[c(\boldsymbol{x}^{(2)}, \boldsymbol{\alpha})\right]\right).$$

Proof. We have that

$$V_{\boldsymbol{\alpha}}(\boldsymbol{x}) = c(\boldsymbol{x}, \boldsymbol{\alpha}) + V_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}),$$
$$V_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) = N(\boldsymbol{\alpha}) + \gamma \mathbb{E}_{\{\boldsymbol{D},\boldsymbol{R}\}}\left[c(\boldsymbol{x}^{(2)}, \boldsymbol{\alpha})\right] + \gamma V_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}).$$

Solving the equations simultaneously, we obtain:

$$V_{\boldsymbol{\alpha}}(\boldsymbol{x}) = c(\boldsymbol{x}, \boldsymbol{\alpha}) + \frac{1}{1-\gamma}\left(N(\boldsymbol{\alpha}) + \gamma \mathbb{E}_{\{\boldsymbol{D},\boldsymbol{R}\}}\left[c(\boldsymbol{x}^{(2)}, \boldsymbol{\alpha})\right]\right). \qquad \square$$

Theorem 2.3.2. *For any $\boldsymbol{\alpha}$ and inventory position $\boldsymbol{x}$, discounted cost $V_{\boldsymbol{\alpha}}(\boldsymbol{x})$ satisfies:*

$$\frac{V_{\boldsymbol{\alpha}}(\boldsymbol{x})}{V(\boldsymbol{x})} \leq \max_{\boldsymbol{z} \in \mathcal{X}_{\boldsymbol{x}}} \frac{c(\boldsymbol{z}, \boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}}}{c(\boldsymbol{z}, \pi^*(\boldsymbol{z})) + N(\pi^*(\boldsymbol{z}))},$$
$$C_{\boldsymbol{\alpha}} = \mathbb{E}_{\{\boldsymbol{D},\boldsymbol{R}\}}\left[c(\boldsymbol{x}^{(2,1)}, \boldsymbol{\alpha})\right],$$

$\boldsymbol{x}^{(2,1)}$ denotes the random vector for the inventory position after demand and routing realization when the action taken is $\boldsymbol{\alpha}$ and $\mathcal{X}_{\boldsymbol{x}}$ is expressed as

$$\mathcal{X}_{\boldsymbol{x}} = \{\boldsymbol{z} : \exists T, \pi^* \ s.t\ \mathbb{P}\left[\boldsymbol{x}^{(T)} = z\right] > 0 \vee \mathbb{P}\left[\pi^*(\boldsymbol{x}^{(T)}) = z\right] > 0\}.$$

Proof. First, we have that

$$V_{\boldsymbol{\alpha}}(\boldsymbol{x}) - V(\boldsymbol{x}) = c(\boldsymbol{x}, \boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma \mathbb{E}_{\{\boldsymbol{D},\boldsymbol{R}\}}\left[V_{\boldsymbol{\alpha}}(\boldsymbol{x}^{(2,1)}) - V(\boldsymbol{x}^{(2,2)})\right] - c(\boldsymbol{x}, \pi^*(\boldsymbol{x})) - N(\pi^*(\boldsymbol{x})), \tag{2.7.1}$$

where $\boldsymbol{x}^{(2,1)}$ denotes the random vector for the inventory position after demand realization and routing when the action taken is $\boldsymbol{\alpha}$ and $\boldsymbol{x}^{(2,2)}$ denotes the random vector for the inventory position after demand realization and routing when the action taken is $\pi^*(\boldsymbol{x})$. Furthermore,

$$V_{\boldsymbol{\alpha}}(\boldsymbol{x}^{(2,1)}) = c(\boldsymbol{x}^{(2,1)}, \boldsymbol{\alpha}) + V_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}),$$
$$V_{\boldsymbol{\alpha}}(\boldsymbol{x}^{(2,2)}) = c(\boldsymbol{x}^{(2,2)}, \boldsymbol{\alpha}) + V_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}),$$
$$V_{\boldsymbol{\alpha}}(\boldsymbol{x}^{(2,1)}) = c(\boldsymbol{x}^{(2,1)}, \boldsymbol{\alpha}) - c(\boldsymbol{x}^{(2,2)}, \boldsymbol{\alpha}) + V_{\boldsymbol{\alpha}}(\boldsymbol{x}^{(2,2)}) \tag{2.7.2}$$

For ease of notation, let $\mathbb{E}_{\{\boldsymbol{D},\boldsymbol{R}\}}\left[c(\boldsymbol{x}^{(2,1)}, \boldsymbol{\alpha})\right] = C_{\boldsymbol{\alpha}}$. Inputting Equation (2.7.2) into Equation (2.7.1), we obtain

$$
\begin{aligned}
V_{\boldsymbol{\alpha}}(\boldsymbol{x}) - V(\boldsymbol{x}) = c(\boldsymbol{x}, \boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}} - \gamma \mathbb{E}_{\{\boldsymbol{D},\boldsymbol{R}\}} \left[c(\boldsymbol{x}^{(2,2)}, \boldsymbol{\alpha}) \right] \\
+ \gamma \mathbb{E}_{\{\boldsymbol{D},\boldsymbol{R}\}} \left[V_{\boldsymbol{\alpha}}(\boldsymbol{x}^{(2,2)}) - V(\boldsymbol{x}^{(2,2)}) \right] - c(\boldsymbol{x}, \pi^*(\boldsymbol{x})) - N(\pi^*(\boldsymbol{x})) \\
\leq c(\boldsymbol{x}, \boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}} \\
+ \gamma \mathbb{E}_{\{\boldsymbol{D},\boldsymbol{R}\}} \left[V_{\boldsymbol{\alpha}}(\boldsymbol{x}^{(2,2)}) - V(\boldsymbol{x}^{(2,2)}) \right] - c(\boldsymbol{x}, \pi^*(\boldsymbol{x})) - N(\pi^*(\boldsymbol{x})).
\end{aligned}
$$

We can repeat the same step for the future inventory positions reached to obtain

$$
\begin{aligned}
V_{\boldsymbol{\alpha}}(\boldsymbol{x}) - V(\boldsymbol{x}) \leq \mathbb{E}_{\{\boldsymbol{D}^t,\boldsymbol{R}^t\}_{t=1}^{\infty}} \Bigg[\sum_{t=1}^{\infty} \gamma^{t-1} \Big(c(\boldsymbol{x}^{(t)}, \boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) \\
+ \gamma C_{\boldsymbol{\alpha}} - c(\boldsymbol{x}^{(t)}, \pi^*(\boldsymbol{x}^{(t)})) - N(\pi^*(\boldsymbol{x}^{(t)})) \Big) \Bigg],
\end{aligned}
$$

where the stochastic process $\boldsymbol{x}^{(t)}$ is the inventory position under an optimal policy. Using this bound, we have:

$$
\begin{aligned}
&\frac{V_{\boldsymbol{\alpha}}(\boldsymbol{x})}{V(\boldsymbol{x})} = 1 + \frac{V_{\boldsymbol{\alpha}}(\boldsymbol{x}) - V(\boldsymbol{x})}{V(\boldsymbol{x})} \\
&\leq 1+ \\
&\frac{\mathbb{E}_{\{\boldsymbol{D}^t,\boldsymbol{R}^t\}_{t=1}^{\infty}} \left[\sum_{t=1}^{\infty} \gamma^{t-1} \Big(c(\boldsymbol{x}^{(t)}, \boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}} - c(\boldsymbol{x}^{(t)}, \pi^*(\boldsymbol{x}^{(t)})) - N(\pi^*(\boldsymbol{x}^{(t)})) \Big) \right]}{\mathbb{E}_{\{\boldsymbol{D}^t,\boldsymbol{R}^t\}_{t=1}^{\infty}} \left[\sum_{t=1}^{\infty} \gamma^{t-1} \Big(c(\boldsymbol{x}^{(t)}, \pi^*(\boldsymbol{x}^{(t)})) + N(\pi^*(\boldsymbol{x}^{(t)})) \Big) \right]} \\
&= 1 + \frac{\sum_{k=1}^{\infty} m_k}{\sum_{k=1}^{\infty} n_k},
\end{aligned}
$$

where:

$$
\begin{aligned}
m_1 &= c(\boldsymbol{x}^{(1)}, \boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}} - c(\boldsymbol{x}^{(1)}, \pi^*(\boldsymbol{x}^{(1)})) - N(\pi^*(\boldsymbol{x}^{(1)})) \\
m_k &= \gamma^{k-1} \mathbb{E}_{\{\boldsymbol{D}^t,\boldsymbol{R}^t\}_{t=1}^{k-1}} \left[c(\boldsymbol{x}^{(k)}, \boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}} - c(\boldsymbol{x}^{(k)}, \pi^*(\boldsymbol{x}^{(k)})) - N(\pi^*(\boldsymbol{x}^{(k)})) \right] \forall k > 1 \\
n_1 &= c(\boldsymbol{x}^{(1)}, \pi^*(\boldsymbol{x}^{(1)})) + N(\pi^*(\boldsymbol{x}^{(1)})) \\
n_k &= \gamma^{k-1} \mathbb{E}_{\{\boldsymbol{D}^t,\boldsymbol{R}^t\}_{t=1}^{k-1}} \left[c(\boldsymbol{x}^{(k)}, \pi^*(\boldsymbol{x}^{(k)})) + N(\pi^*(\boldsymbol{x}^{(k)})) \right] \quad \forall k > 1.
\end{aligned}
$$

Furthermore, let $\boldsymbol{X}^k$ be the set of all inventory positions that the system can be in at period k, with non-zero probability (there exists a sequence of feasible demand and routing probabilities where the system reaches this inventory position in period k). As both the numerator and the denominator follow an optimal policy (we can assume they follow the same optimal policy in the case where the optimal policy is not unique without loss), the set of inventory positions $\boldsymbol{X}^k$, as well as the corresponding probabilities of reaching these inventory positions, will be identical. Then,

$$\frac{m_k}{n_k} = \frac{\sum_{\boldsymbol{x}_1^{(k)} \in \boldsymbol{X}^k} \mathbb{P}\left[\boldsymbol{x}^{(k)} = \boldsymbol{x}_1^{(k)}\right]\left(c(\boldsymbol{x}_1^{(k)}, \boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}} - c(\boldsymbol{x}_1^{(k)}, \pi^*(\boldsymbol{x}_1^{(k)})) - N(\pi^*(\boldsymbol{x}_1^{(k)}))\right)}{\sum_{\boldsymbol{x}_1^{(k)} \in \boldsymbol{X}^k} \mathbb{P}\left[\boldsymbol{x}^{(k)} = \boldsymbol{x}_1^{(k)}\right]\left(c(\boldsymbol{x}_1^{(k)}, \pi^*(\boldsymbol{x}_1^{(k)})) + N(\pi^*(\boldsymbol{x}_1^{(k)}))\right)}.$$

Through this construction, we can further expand all m_k, n_k values through assigning an index to each inventory position $\boldsymbol{x}_e^{(k)} \in \boldsymbol{X}^k$, with

$$\begin{aligned}
m_k &= \sum_{e=1}^{|\boldsymbol{X}^k|} m_{k,e}, \\
m_{k,e} &= \gamma^{k-1}\mathbb{P}\left[\boldsymbol{x}^{(k)} = \boldsymbol{x}_e^{(k)}\right]\left(c(\boldsymbol{x}_e^{(k)}, \boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}} - c(\boldsymbol{x}_e^{(k)}, \pi^*(\boldsymbol{x}_e^{(k)})) - N(\pi^*(\boldsymbol{x}_e^{(k)}))\right), \\
n_k &= \sum_{e=1}^{|\boldsymbol{X}^k|} n_{k,e}, \\
n_{k,e} &= \gamma^{k-1}\mathbb{P}\left[\boldsymbol{x}^{(k)} = \boldsymbol{x}_e^{(k)}\right]\left(c(\boldsymbol{x}_e^{(k)}, \pi^*(\boldsymbol{x}_e^{(k)})) + N(\pi^*(\boldsymbol{x}_e^{(k)}))\right),
\end{aligned}$$

where $|\boldsymbol{X}^k|$ denotes the cardinality of set $\boldsymbol{X}^k$. Then, we have:

$$\frac{V_{\boldsymbol{\alpha}}(\boldsymbol{x})}{V(\boldsymbol{x})} \leq 1 + \frac{\sum_{k=1}^{\infty}\sum_{e=1}^{|\boldsymbol{X}^k|} m_{k,e}}{\sum_{k=1}^{\infty}\sum_{e=1}^{|\boldsymbol{X}^k|} n_{k,e}}.$$

Letting

$$z^* = \max_{k,e} \frac{m_{k,e}}{n_{k,e}},$$

through the definition of a maximum,

$$
\begin{aligned}
z^* &\geq \frac{m_{k,e}}{n_{k,e}} \qquad\qquad \forall k, e \\
z^* n_{k,e} &\geq m_{k,e} \\
z^* \sum_{k=1}^{\infty} \sum_{e=1}^{|\boldsymbol{X}^k|} n_{k,e} &\geq \sum_{k=1}^{\infty} \sum_{e=1}^{|\boldsymbol{X}^k|} m_{k,e} \\
z^* &\geq \frac{\sum_{k=1}^{\infty} \sum_{e=1}^{|\boldsymbol{X}^k|} m_{k,e}}{\sum_{k=1}^{\infty} \sum_{e=1}^{|\boldsymbol{X}^k|} n_{k,e}}.
\end{aligned}
$$

As the probabilities and the discount rates cancel out, we can bound the ratio of the policies through solving for an inventory position maximizing the ratio of modified per-period costs of the policies. Consequently:

$$
\begin{aligned}
\frac{V_{\boldsymbol{\alpha}}(\boldsymbol{x})}{V(\boldsymbol{x})} &\leq 1 + \max_{\boldsymbol{z} \in \mathcal{X}_{\boldsymbol{x}}} \frac{c(\boldsymbol{z}, \boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}} - c(\boldsymbol{z}, \pi^*(\boldsymbol{z})) - N(\pi^*(\boldsymbol{z}))}{c(\boldsymbol{z}, \pi^*(\boldsymbol{z})) + N(\pi^*(\boldsymbol{z}))} \\
&= \max_{\boldsymbol{z} \in \mathcal{X}_{\boldsymbol{x}}} \frac{c(\boldsymbol{z}, \boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}}}{c(\boldsymbol{z}, \pi^*(\boldsymbol{z})) + N(\pi^*(\boldsymbol{z}))},
\end{aligned}
$$

where

$$
\mathcal{X}_{\boldsymbol{x}} = \{\boldsymbol{z} : \exists T, \pi^* \text{ s.t } \mathbb{P}\left[\boldsymbol{x}^{(T)} = z\right] > 0 \vee \mathbb{P}\left[\pi^*(\boldsymbol{x}^{(T)}) = z\right] > 0\}.
$$

Here, the set $\mathcal{X}_{\boldsymbol{x}}$ is expanded to include the set of inventory positions for which $\mathbb{P}\left[\boldsymbol{x}^{(T)} = z\right] = 0$ but $\mathbb{P}\left[\pi^*(\boldsymbol{x}^{(T)}) = z\right] > 0$. The motivation for this will be clear in the next proposition.

Before concluding the proof, we also need to establish an important property for the set $\mathcal{X}_{\boldsymbol{x}}$ and therefore the resultant maximization program. As we are working with infinite horizon programs, T (the period which the system can move to z^*) is unbounded with $T \in \mathbb{N}$. However, we can show that $|\mathcal{X}_{\boldsymbol{x}}| < \infty$ (all optimal policies visit some states infinitely often) as the number of units post-rebalancing/sourcing is bounded with $\bar{U}$ and the total number of units in circulation do not change through state-transitions. Letting $\boldsymbol{X}$ be the set of all feasible inventory positions, as there are n stations and at most $\bar{U}$ units in circulation,

$|\boldsymbol{X}| = \frac{(n+\bar{U}+1)!}{\hat{U}!(n-1)!} < \infty$. Furthermore, as some of these inventory positions may be unreachable through an optimal policy, we have $\mathcal{X}_{\boldsymbol{x}} \subset \boldsymbol{X}$ and consequently $|\mathcal{X}_{\boldsymbol{x}}| < \infty$. We also know that the newsvendor and rebalancing cost functions output finite values for bounded inventory positions/actions. Consequently, as the number of inventory positions in $\mathcal{X}_{\boldsymbol{x}}$ is finite and the costs are bounded (as the inventory positions/actions are bounded), an optimal solution for the maximization problem always exists. $\square$

Proposition 2.3.3. *There* $\exists z^*$ *satisfying:*

$$
\begin{aligned}
&z^* \in \arg\max_{\boldsymbol{z} \in \mathcal{X}_{\boldsymbol{x}}} \frac{c(\boldsymbol{z}, \boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}}}{c(\boldsymbol{z}, \pi^*(\boldsymbol{z})) + N(\pi^*(\boldsymbol{z}))}, \\
&z^* = \pi^*(z^*).
\end{aligned}
$$

Proof. We prove the proposition through contradiction where we assume all optimal solutions $\boldsymbol{z}^*$ of the optimization problem provided in Theorem 2.3.2 satisfy (proof of Theorem 2.3.2 establishes that an optimal solution always exists):

$$
\boldsymbol{z}^* \neq \pi^*(\boldsymbol{z}^*). \tag{2.7.3}
$$

Through the extended definition of $\mathcal{X}_{\boldsymbol{x}}$, for any inventory position $\boldsymbol{z}$, if $\boldsymbol{z} \in \mathcal{X}_{\boldsymbol{x}}$, $\exists \pi^*$ such that $\pi^*(\boldsymbol{z}) \in \mathcal{X}_{\boldsymbol{x}}$. As a result, for any optimal solution $\boldsymbol{z}^*$, Equation (2.7.3) is equivalent to:

$$
\frac{c(\boldsymbol{z}^*, \boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}}}{c(\boldsymbol{z}^*, \pi^*(\boldsymbol{z}^*)) + N(\pi^*(\boldsymbol{z}^*))} > \frac{c(\pi^*(\boldsymbol{z}^*), \boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}}}{c(\pi^*(\boldsymbol{z}^*), \pi^*(\pi^*(\boldsymbol{z}^*))) + N(\pi^*(\pi^*(\boldsymbol{z}^*)))},
$$

by Lemma 2.2.2,

$$
\frac{c(\boldsymbol{z}^*, \boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}}}{c(\boldsymbol{z}^*, \pi^*(\boldsymbol{z}^*)) + N(\pi^*(\boldsymbol{z}^*))} > \frac{c(\pi^*(\boldsymbol{z}^*), \boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}}}{N(\pi^*(\boldsymbol{z}^*))},
$$

by Lemma 2.2.1,

$$\begin{aligned}
&\frac{c(\boldsymbol{z}^*, \pi^*(\boldsymbol{z}^*)) + c(\pi^*(\boldsymbol{z}^*), \boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}}}{c(\boldsymbol{z}^*, \pi^*(\boldsymbol{z}^*)) + N(\pi^*(\boldsymbol{z}^*))} > \frac{c(\pi^*(\boldsymbol{z}^*), \boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}}}{N(\pi^*(\boldsymbol{z}^*))} \\
&\frac{c(\boldsymbol{z}^*, \pi^*(\boldsymbol{z}^*))}{c(\boldsymbol{z}^*, \pi^*(\boldsymbol{z}^*)) + N(\pi^*(\boldsymbol{z}^*))} \\
&\qquad > \left(c(\pi^*(\boldsymbol{z}^*), \boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}}\right)\left(\frac{1}{N(\pi^*(\boldsymbol{z}^*))} - \frac{1}{c(\boldsymbol{z}^*, \pi^*(\boldsymbol{z}^*)) + N(\pi^*(\boldsymbol{z}^*))}\right) \\
&1 > \left(c(\pi^*(\boldsymbol{z}^*), \boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}}\right)\frac{1}{N(\pi^*(\boldsymbol{z}^*))} \\
&0 > c(\pi^*(\boldsymbol{z}^*), \boldsymbol{\alpha}) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}} - N(\pi^*(\boldsymbol{z}^*)),
\end{aligned}$$

by Lemma 2.2.1,

$$0 > c(\boldsymbol{z}^*, \boldsymbol{\alpha}) - c(\boldsymbol{z}^*, \pi^*(\boldsymbol{z}^*)) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}} - N(\pi^*(\boldsymbol{z}^*)).$$

This contradicts with the proof of Theorem 2.3.2, where it was clearly implied that

$$\begin{aligned}
0 &\leq \max_{\boldsymbol{z} \in \mathcal{X}_{\boldsymbol{x}}} \frac{c(\boldsymbol{z}, \boldsymbol{\alpha}) - c(\boldsymbol{z}, \pi^*(\boldsymbol{z})) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}} - N(\pi^*(\boldsymbol{z}))}{c(\boldsymbol{z}, \pi^*(\boldsymbol{z})) + N(\pi^*(\boldsymbol{z}))} \\
0 &\leq \max_{\boldsymbol{z} \in \mathcal{X}_{\boldsymbol{x}}} c(\boldsymbol{z}, \boldsymbol{\alpha}) - c(\boldsymbol{z}, \pi^*(\boldsymbol{z})) + N(\boldsymbol{\alpha}) + \gamma C_{\boldsymbol{\alpha}} - N(\pi^*(\boldsymbol{z})).
\end{aligned}$$

Consequently, we reached a contradiction and proved the given proposition. □

Theorem 2.4.1. *Given any initial inventory position $\boldsymbol{x}$, the no action policy π^N satisfies:*

$$\frac{V_{\pi^N}(\boldsymbol{x})}{V(\boldsymbol{x})} \leq \frac{N(\mathbf{0})}{N(\boldsymbol{K})} + \frac{1}{N(\boldsymbol{K})} \max_{j \in \{1, \cdots, n\}} \left(N_j(\sum_{i=1}^{n} x_i) - N_j(0)\right).$$

Proof. First, we have

$$\mathbf{0} \in \arg \max_{\{\boldsymbol{c_s}, \boldsymbol{c_m}\}} \frac{V_{\pi^N}(\boldsymbol{x})}{V(\boldsymbol{x})} \quad \forall \boldsymbol{x},$$

as $V_{\pi^N}(\boldsymbol{x})$ is constant in $\boldsymbol{c_s}, \boldsymbol{c_m}$ and $V(\boldsymbol{x})$ is clearly nondecreasing in $\boldsymbol{c_s}, \boldsymbol{c_m}$. This implies that we can bound $\frac{V_{\pi^N}(\boldsymbol{x})}{V(\boldsymbol{x})}$ assuming that $\boldsymbol{c_s}, \boldsymbol{c_m} = \mathbf{0}$, where we already established the optimality of the newsvendor policy. Consequently,

$$\frac{V_{\pi^N}(\boldsymbol{x})}{V(\boldsymbol{x})} \leq \mathbb{E}_{\{\boldsymbol{D}^t, \boldsymbol{R}^t\}_{t=1}^{\infty}} \left[\frac{\sum_{t=1}^{\infty} \gamma^{t-1} N(\boldsymbol{x}^{(t)})}{\sum_{t=1}^{\infty} \gamma^{t-1} N(\boldsymbol{K})} \right],$$

where the numerator is the exact expression for $V_{\pi^N}(\boldsymbol{x})$ and the denominator is the exact expression for $V(\boldsymbol{x})$, when $\boldsymbol{c_s}, \boldsymbol{c_m} = \boldsymbol{0}$. As in the proof of Theorem 2.3.2, the $\boldsymbol{z}$ value maximizing $\frac{N(\boldsymbol{z})}{N(\boldsymbol{K})}$ will upper-bound the above expression, providing

$$\begin{aligned} \frac{V_{\pi^N}(\boldsymbol{x})}{V(\boldsymbol{x})} &\leq \frac{1}{N(\boldsymbol{K})} \max_{\boldsymbol{z}} N(\boldsymbol{z}) \\ \text{s.t} \qquad & \sum_{i=1}^{n} z_i = \sum_{i=1}^{n} x_i. \end{aligned}$$

We then claim that $N(\boldsymbol{z})$ is convex in $\boldsymbol{z}$, which holds as both $c_{h,i} z_i$ and $c_{p,i}\mathbb{E}\left[(D_i - z_i)^+\right]$ is convex in z_i $\forall i$. Consequently, the worst-case bound is solved through a convex maximization problem. Furthermore, we can relax the domain of the problem to allow for non-integer z values, with

$$\max_{\boldsymbol{z} \in \mathbb{N}_0^n, \sum_{i=1}^n z_i = \sum_{i=1}^n x_i} N(\boldsymbol{z}) \leq \max_{\boldsymbol{z} \in \mathbb{R}_+^n, \sum_{i=1}^n z_i = \sum_{i=1}^n x_i} N(\boldsymbol{z}).$$

As the feasible region of this maximization problem is compact, convex, and non-empty, and $N(\boldsymbol{z})$ is convex in $\boldsymbol{z}$, an optimal solution is an extreme point of the feasible set. As a result, one station has $\sum_{i=1}^n x_i$ units under this optimal solution and the other $n-1$ stations have 0 units. Then:

$$\frac{V_{\pi^N}(\boldsymbol{x})}{V(\boldsymbol{x})} \leq \frac{N(\boldsymbol{0})}{N(\boldsymbol{K})} + \frac{1}{N(\boldsymbol{K})} \max_{j \in \{1, \cdots, n\}} \left(N_j(\sum_{i=1}^{n} x_i) - N_j(0) \right). \qquad \square$$

Lemma 2.5.1. *Given the above network structure, demand distribution, and cost assumptions,*

$$\lim_{\gamma \to 1^-} \frac{V_{\pi^M}([\lambda, \lambda])}{V_{\boldsymbol{K}}([\lambda, \lambda])} = \infty.$$

Proof. We first solve for the myopic policy. To do that, we first let $x_i < \lambda$ for station i. Increasing x_i by 1 would change myopic cost by $c_s + c_h - c_p < 0$ (as moving units through the warehouse is cheaper than rebalancing between stations), so it is optimal to increase x_i up to λ. At $x_i = \lambda$, decreasing one unit will change cost by $c_m + c_p - c_h > 0$, and increasing one unit will change cost by $c_s + c_h > 0$, so it is optimal to stay at x_i. Lastly, for $x_i > \lambda$, decreasing x_i by 1 would change myopic cost by $c_m - c_h > 0$, and increasing x_i by 1 would change myopic cost by $c_s + c_h > 0$ so staying at x_i is optimal. As cost parameters and demand distributions are identical for both stations (transition probabilities do not affect the myopic policy as it is single-staged), the myopic policy is a threshold policy at both stations where inventory is moved up to λ if below λ and unchanged if already above or equal to λ. The resultant cost is then given by

$$V_{\pi^M}([\lambda,\lambda]) = 2\lambda c_h + \gamma\lambda c_s + \gamma 3\lambda c_h + \gamma^2\lambda c_s + \gamma^2 4\lambda c_h + \cdots = 2\lambda c_h + \sum_{t=1}^{\infty}\gamma^t\big((2+t)\lambda c_h + \lambda c_s\big).$$

State transitions that provide these cost values are as follows: First period, all λ units in station 1 move to station 2, leaving 0 units in station 1 and 2λ units in station 2. Next period, according to the myopic policy, λ units are sent to station 1 from the warehouse. These units also move to station 2 at the end of the period leaving 0 units in station 1 and 3λ units in station 2. As a result, the number of units in station 2 increases by λ each subsequent period.

Now, consider the newsvendor policy where we have $K_1 = K_2 = \lambda$ (directly follows from $c_p > c_h$ and deterministic demand). The resultant cost is given by

$$V_{\boldsymbol{K}}([\lambda,\lambda]) = 2\lambda c_h + \gamma 2\lambda c_s + \gamma 2\lambda c_h + \gamma^2 2\lambda c_s + \gamma^2 2\lambda c_h + \cdots = 2\lambda c_h + \sum_{t=1}^{\infty}\gamma^t(2\lambda c_h + 2\lambda c_s).$$

State transitions that provide these cost values are as follows: First period, all λ units in station 1 move to station 2, leaving 0 units in station 1 and 2λ units in station 2. Next period, λ units move to station 1 from the warehouse, and λ units move from station 2 to the warehouse. Then, all λ units in station 1 again move to station 2, leaving 0 units in station 1 and 2λ units in station 2, providing the cost expression.

Taking the ratio of the costs under the two policies, we obtain:

$$\frac{V_{\pi^M}([\lambda,\lambda])}{V_{\boldsymbol{K}}([\lambda,\lambda])} = \frac{2\lambda c_h + \sum_{t=1}^{\infty} \gamma^t \big((2+t)\lambda c_h + \lambda c_s\big)}{2\lambda c_h + \sum_{t=1}^{\infty} \gamma^t (2\lambda c_h + 2\lambda c_s)}.$$

As γ approaches 1, the above ratio explodes with $\lim_{\gamma\to 1^-} \frac{V_{\pi^M}([\lambda,\lambda])}{V_{\boldsymbol{K}}([\lambda,\lambda])} = \infty$. □

CHAPTER 3

CONCLUSION AND FUTURE WORK

In this dissertation, we consider the problem of managing resources for the sharing economy. A common feature in these applications is the imbalanced traffic pattern leading to unequal inflow and outflow rates of units at the network nodes. Labeled as the spatial imbalance of resources, in the first chapter, we introduced a mean-field-based policy to manage these resources while minimizing the cost of the platform. We proved that the mean-field-based policy is asymptotically optimal in the number of stations and also provided a near-optimal approximation algorithm to solve the mean-field model. We then showed the strong performance of the algorithm through a numerical experiment using Austin scooter-sharing data and other synthetic experiments. In the second chapter, we looked at simple policies already used in practice and evaluated their performance in managing these resources. The two chapters provide two perspectives on the same problem: Managing resources effectively to minimize platform costs while maintaining customer service availability.

There are important directions for extending the mathematical framework we have provided in this dissertation and new applications that can be considered through the mean-field approach. First, we have not considered dynamic pricing as a lever to manage resources, which ride-sharing companies use. Micromobility systems, in contrast, use mostly fixed prices with subscription models (for bike-sharing) or fixed per-minute fees (for scooter-sharing systems). Nevertheless, we believe that dynamic pricing can be used effectively with rebalancing/recharging/sourcing for effective management of resources in the sharing economy. Furthermore, under some standard assumptions (such as assuming a finite set of admissible prices at nodes), the mean-field model we presented in Chapter 1 is compatible with also incorporating pricing decisions, where the central planner can charge different prices at each node for each period. The reason for this is the flexibility of our mean-field formulation in networks, which only requires that nodes be split into a finite group of sub-types with identical properties and the same post-rebalancing inventory. Specifically, we previously let

$\hat{a}^t_{e,d}$ denote the proportion of type-e stations that have d units after rebalancing at period t. To incorporate dynamic pricing, we can introduce $\hat{a}^t_{e,d,p}$, which denotes the proportion of type-e stations that have d units after rebalancing and a price of p per unit, at period t. Then,

$$\sum_{p \in \mathbb{P}} \hat{a}^t_{e,d,p} = \hat{a}^t_{e,d},$$

where $\mathbb{P}$ denotes the set of all admissible prices $p_1, p_2, ...$ With $|\mathbb{P}| < \infty$, our results will extend to incorporate dynamic pricing.

Extending to dynamic pricing also offers new applications in revenue and inventory management. Specifically, we can extend to applications where customers consume inventory (instead then repositioned as we looked at in this dissertation). While dynamic pricing in this setting has been extensively studied, a limited number of papers look at dynamic pricing jointly with repositioning, especially considering problem properties such as fixed replenishment costs. The mean-field formulation in Chapter 1 can be beneficial in understanding these systems.

In Chapter 2, we looked at the performance of two well-known policies, newsvendor and no action. While we focused on bike-sharing as an application, these policies have been applied in a wide range of applications. Extending our coupling analysis into these applications to see how these policies perform will be an important extension of work. Furthermore, in some settings, combining the two and constructing hybrid policies may be easy to implement and perform strongly.

Lastly, both chapters focus on micromobility systems and bike-sharing as the primary application. "Despite the promising role that micromobility may play in sustainable transportation, most regulatory approaches have not explicitly encouraged integration into transport networks" Yanocha and Allan [2021]. As a result, both chapters consider the current structure of micromobility systems in North America, where decisions are made based on the constraints and parameters obtained from the regulatory contract signed between the micromobility platform and the city transportation agency. Nevertheless, in the coming years, many city administrations are expected to move to "Shift focus from regulating micromobility to using micromobility to fill in gaps in transportation systems and focus on integration as a path to expanded access" ITDP [2021]. While not explicitly considered in this dissertation, the mathematical framework provided in this dissertation may be utilized to integrate micromobility systems with public transportation. One example is multi-modal transportation, where people can access public transportation through bike-sharing and scooter-sharing systems. Specifically, we can consider a set of bike/scooter-sharing stations connected to bus stops, configuring a network within each other. Then, the mean-field analysis presented in Chapter 1 can be used to make decisions on both rebalancing/recharging/sourcing at micromobility stations and capacity decisions for bus stops.

BIBLIOGRAPHY

Hossein Abouee-Mehrizi, Oded Berman, and Shrutivandana Sharma. Optimal joint replenishment and transshipment policies in a multi-period inventory system with lost sales. *Operations Research*, 63(2):342–350, 2015.

Daniel Adelman. Price-directed control of a closed logistics queueing network. *Operations Research*, 55(6):1022–1038, 2007.

Michael Andersen. How much does each bike share ride cost a system? Lets do the math, 2016. URL `https://betterbikeshare.org/2016/08/16/much-bike-share-ride-cost-system-lets-math/`. Accessed on March 4, 2023.

Santiago R. Balseiro, David B. Brown, and Chen Chen. Dynamic pricing of relocating resources in large networks. *Management Science*, 67(7):4075–4094, 2021.

Siddhartha Banerjee, Daniel Freund, and Thodoris Lykouris. Pricing and optimization in shared vehicle systems: An approximation framework. *Operations Research*, 70(3):1783–1805, 2022.

Saif Benjaafar, Daniel Jiang, Xiang Li, and Xiaobo Li. Dynamic inventory repositioning in on-demand rental networks. *Management Science*, 68(11):7861–7878, 2022a.

Saif Benjaafar, Shining Wu, Hanlin Liu, and Einar Bjarki Gunnarsson. Dimensioning on-demand vehicle sharing systems. *Management Science*, 68(2):1218–1232, 2022b.

Kostas Bimpikis, Ozan Candogan, and Daniela Saban. Spatial pricing in ride-sharing networks. *Operations Research*, 67(3):744–769, 2019.

G. Birkhoff. House monotone apportionment schemes. *Proceedings of the National Academy of Sciences*, 73(3):684–686, 1976.

S. Bobkov and M. Ledoux. *One-Dimensional Empirical Measures, Order Statistics, and Kantorovich Transport Distances*. Memoirs of the American Mathematical Society. American Mathematical Society, 2019.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

Anton Braverman, J. G. Dai, Xin Liu, and Lei Ying. Empty-car routing in ridesharing systems. *Operations Research*, 67(5):1437–1452, 2019.

Xin Chen, Xiangyu Gao, and Zhenyu Hu. A new approach to two-location joint inventory and transshipment control via L convexity. *Operations Research Letters*, 43(1):65–68, 2015.

City of Austin Transportation Department. Shared micromobility vehicle trips, 2022. URL `https://data.austintexas.gov/Transportation-and-Mobility/Shared-Micromobility-Vehicle-Trips/7d8e-dm7r`.

City of Chicago. City of chicago requirements for scooter sharing emerging business permit pilot program, 2019. URL https://www.chicago.gov/content/dam/city/depts/cdot/Misc/EScooters/EScootersPilotProgramTerms_06-07-19.pdf.

City of Los Angeles Department of Transportation. Dockless on-demand personal mobility conditional permit, 2018. URL http://clkrep.lacity.org/onlinedocs/2017/17-1125_misc_11-08-2018.pdf.

Pengcheng Dai, Changxiong Song, Huiping Lin, Pei Jia, and Zhipeng Xu. Cluster-based destination prediction in bike sharing system. In *Proceedings of the 2018 Artificial Intelligence and Cloud Computing Conference*, 2018.

Eustasio del Barrio, Evarist Giné, and Carlos Matrán. Central limit theorems for the wasserstein distance between the empirical and the true distributions. *The Annals of Probability*, 27(2):1009–1071, 1999.

James Flynn. Steady state policies for deterministic dynamic programs. *SIAM Journal on Applied Mathematics*, 37(1):128–147, 1979.

Stephen Gossett. Bike-sharing rebalancing is a classic data challenge that just got a lot harder, 2021. URL https://builtin.com/data-science/bike-share-rebalancing. Accessed on March 25, 2023.

Lacy Greening and Alan Erera. Effective heuristics for distributing vehicles in free-floating micromobility systems. Working paper, 2021.

Alison Griswold. Simple math shows how scooters could make big money, 2018. URL https://qz.com/1325064/scooters-might-actually-have-good-unit-economics/. Accessed on March 25, 2023.

Nely Hayes. How long do electric scooters last?, 2022. URL https://scooter.guide/how-long-do-electric-scooters-last/. Accessed on March 25, 2023.

Long He, Zhenyu Hu, and Meilin Zhang. Robust repositioning for vehicle sharing. *Manufacturing & Service Operations Management*, 22(2):241–256, 2020.

Brent Helling. The comprehensive guide to bird scooter charging, 2022. URL https://www.ridester.com/bird-scooter-charger/. Accessed on March 25, 2023.

Becky Holyoak. I made $1,000 charging lime scooters — an honest review, 2021. URL https://enduringfinances.com/i-made-1000-charging-lime-scooters-an-honest-review/. Accessed on March 25, 2023.

Mahsa Hosseini, Joseph Milner, and Gonzalo Romero. Dynamic relocations in car-sharing networks. Working paper, 2022.

ITDP. Maximizing potential by connecting micromobility and transit, 2021. URL https://www.itdp.org/2021/06/30/maximizing-potential-by-connecting-micromobility-and-transit/. Accessed on March 4, 2023.

Ben Jay. Its not your imagination something is seriously wrong with citi bike right now, 2018. URL https://nyc.streetsblog.org/2018/09/26/its-not-your-imagination-something-is-seriously-wrong-with-citi-bike-right-now/. Accessed on March 25, 2023.

Ashish Kabra, Elena Belavina, and Karan Girotra. Bike-share systems: Accessibility and availability. *Management Science*, 2020.

Eoin O'Mahony and David B. Shmoys. Data analysis and optimization for (citi)bike sharing. In *AAAI'15 Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

Jesus Osorio, Chao Lei, and Yanfeng Ouyang. Optimal rebalancing and on-board charging of shared electric scooters. *Transportation Research Part B: Methodological*, 147:197–219, 2021.

Sheldon Ross. *A First Course in Probability*. Prentice Hall, 8th edition, 2010.

Jia Shu, Mabel C. Chou, Qizhang Liu, Chung-Piaw Teo, and I-Lin Wang. Models for effective deployment and redistribution of bicycles within public bicycle-sharing systems. *Operations Research*, 61(6):1346–1359, 2013.

Paul Somerville. Segway ninebot es4 review: With dual suspension, is it better than the ninebot max?, 2021. URL https://electric-scooter.guide/reviews/segway-ninebot-es4-review/. Accessed on March 25, 2023.

Paul Strobel. What kind of e-scooters do bird & lime use?, 2021. URL https://erideher o.com/what-scooters-do-bird-lime-use/. Accessed on March 4, 2023.

Huseyin Topaloglu and Warren B. Powell. Dynamic-programming approximations for stochastic time-staged integer multicommodity-flow problems. *INFORMS Journal on Computing*, 18(1):31–42, 2006.

Jan A. Van Mieghem and Nils Rudi. Newsvendor networks: Inventory management and capacity investment with discretionary activities. *Manufacturing & Service Operations Management*, 4(4):313–335, 2002.

Yihang Yang, Yimin Yu, Qian Wang, and Junming Liu. Fleet repositioning for vehicle sharing systems: the optimality of balanced myopic policy. Working paper, 2022.

Dana Yanocha and Mackenzie Allan. Maximizing micromobility: Unlocking opportunities to integrate micromobility and public transportation, 2021. URL https://www.itdp.org/wp-content/uploads/2021/06/ITDP_MaximizingMicromobility_2021_singlepage.pdf. Accessed on March 4, 2023.

Ling Zhao, Zhixue Liu, and Peng Hu. Dynamic repositioning for vehicle sharing with setup costs. *Operations Research Letters*, 48(6):792–797, 2020.