Managing Resources for Shared Micromobility: Approximate Optimality in Large-Scale Systems

Online Appendix

Akturk, Deniz

Candogan, Ozan

Gupta, Varun

Appendix A: A Summary of Notation Used

γ : Discount rate (Section 2, Page 8).

n_e : Number of stations at type e (Section 2, Page 9).

$\hat{e}$: Number of types (Section 2, Page 9).

n : Number of stations (Section 2, Page 9).

$p_{e',e}$: Probability that a unit moves from type e' to type e (Section 2, Page 9).

$\boldsymbol{x}^t$: Vector denoting number of units at each station (before rebalancing) (Section 2, Page 10).

$\boldsymbol{X}$: Set of all inventory positions (Section 2, Page 10).

$\boldsymbol{a}^t$: Vector denoting number of units at each station (after rebalancing) (Section 2, Page 10).

$\boldsymbol{D}^t$: Demand vector (Section 2, Page 10).

$\boldsymbol{R}^t$: Trips vector (Section 2, Page 10).

$\boldsymbol{L}$: Vector of lower bounds for action (Section 2, Page 11).

$\boldsymbol{U}$: Vector of upper bounds for action (Section 2, Page 11).

$\boldsymbol{A}$: Set of all actions (Section 2, Page 11).

c : Rebalancing cost function for original model (Section 2, Page 12).

c_h : Holding cost (Section 2, Page 14).

c_p : Penalty cost (Section 2, Page 14).

N : Newsvendor cost function for original model (Section 2, Page 14).

V : Value function for original model (Section 2, Page 14).

$\hat{\boldsymbol{x}}^t$: Inventory vector for mean-field model (Section 3, Page 15).

$\hat{\boldsymbol{X}}$: Set of all mean-field inventory distributions (Section 3, Page 15).

$\hat{\boldsymbol{a}}^t$: Action vector for mean-field model (Section 3, Page 15).

$\hat{\boldsymbol{A}}$: Set of all mean-field action distributions (Section 3, Page 15).

$\hat{R}^e$: Random variable denoting inflow to a station of type e (Section 3, Page 16).

$\hat{c}$: Rebalancing cost function for mean-field model (Section 3, Page 17).

$\hat{N}$: Newsvendor cost function for mean-field model (Section 3, Page 18).

$\hat{V}$: Value function for mean-field model (Section 3, Page 18).

g : Function mapping $\boldsymbol{X} \to \hat{\boldsymbol{X}}$ (Section 3, Page 19).

$\bar{\boldsymbol{a}}^t$: Rounded action vector of mean-field model (Section 3, Page 19).

$\bar{\boldsymbol{A}}$: Set of rounded actions for the mean-field model (Section 3, Page 19).

f : Function mapping $\hat{\boldsymbol{A}} \to \bar{\boldsymbol{A}}$ (Section 3, Page 19).

h : Function mapping $\bar{\boldsymbol{A}} \to \boldsymbol{A}$ (Section 3, Page 20).

$\breve{\boldsymbol{x}}^t$: Inventory vector for mean-field model restricted within thresholds (Section 3, Page 24).

$\hat{c}_2$: Modified rebalancing cost function for mean-field model (Section 3, Page 24).

T : Transient horizon for the control algorithm (Section 4, Page 25).

$\tilde{\boldsymbol{\pi}}$: Policy outputted through the control algorithm (Section 4, Page 25).

Θ : Parameter determining magnitude of demand shocks (Section 5, Page 37).

H : Season length (Section 6, Page 40).

$c_{f,e}$: Fixed cost of moving a unit to/from a station in type e (Section 6, Page 43).

Appendix B: Proofs of Section 2 Results

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

Proposition 2.2. *For all $\boldsymbol{x}^t, \boldsymbol{a}^t$, the rebalancing cost $c(\boldsymbol{x}^t, \boldsymbol{a}^t)$ can be alternatively expressed as:*

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

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

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

$$\sum_{e_1=0}^{2\hat{e}} \sum_{e_2=0}^{2\hat{e}} \sum_{i=1}^{n_{e_1}} \sum_{j=1}^{n_{e_2}} c_{e_1,e_2} y^{e_1,e_2}_{i,j} = \sum_{e_1=0}^{2\hat{e}} \sum_{e_2=0, e_2 \neq e_1}^{2\hat{e}} \sum_{i=1}^{n_{e_1}} \sum_{j=1}^{n_{e_2}} \left(c_{e_1,e_2} - \frac{c_{e_1,e_1}}{2} - \frac{c_{e_2,e_2}}{2} \right) y^{e_1,e_2}_{i,j}$$
$$+ \sum_{e_1=0}^{2\hat{e}} \sum_{e_2=0, e_2 \neq e_1}^{2\hat{e}} \sum_{i=1}^{n_{e_1}} \sum_{j=1}^{n_{e_2}} \frac{c_{e_1,e_1}}{2} y^{e_1,e_2}_{i,j} + \sum_{e_1=0}^{2\hat{e}} \sum_{e_2=0, e_2 \neq e_1}^{2\hat{e}} \sum_{i=1}^{n_{e_1}} \sum_{j=1}^{n_{e_2}} \frac{c_{e_2,e_2}}{2} y^{e_1,e_2}_{i,j} + \sum_{e_1=0}^{2\hat{e}} \sum_{i=1}^{n_{e_1}} \sum_{j=1}^{n_{e_2}} c_{e_1,e_1} y^{e_1,e_2}_{i,j}$$

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

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 (B.1), we obtain

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

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

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

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 $y_{\cdot,i}^{\cdot,e}, y_{i,\cdot}^{e,\cdot}$ are non-negative, we have that:

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

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

$$
y_{\cdot,i}^{\cdot,e} + y_{i,\cdot}^{e,\cdot} = |x_{e,i}^t - a_{e,i}^t| \qquad \forall e, i.
$$

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. $\square$

Appendix C: Relevance of Model to Micromobility Systems

We make several assumptions in formulating the mathematical model to capture the first-order properties of managing micromobility systems. 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 6. In this section, we discuss the remaining assumptions. While they are mainly motivated by tractability concerns, they are also well-aligned with resource management in micromobility systems.

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 it home, charges it, and brings it to an indicated location before a certain deadline (see, e.g., Helling (2022)). While the contractor's payment 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. Recharging is not done on the spot but rather at another location. In the context of our model, it is, therefore, unnecessary to keep track of the exact stations of the depleted units, allowing us to instead focus on depleted units in a type.

Second, we consider a network of stations. We believe that free-floating systems can also be represented through stations.[1] 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.

It is also assumed that there are no demand spillovers between stations: If customers arrive at a station and cannot find a unit there, they do not move to other stations to pick up a unit. While no empirical analysis

[1] A free-floating system is one where customers can drop units on any part of a sidewalk, in contrast to a station-based system where customers have to drop units at docks.

calculating the rate of spillovers has been published 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 station’s unserved users substitute to other stations.”

One assumption that we make is on the linearity of the rebalancing cost function, 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). Considering trucking routes, we assume that the cost of rebalancing is linear on the number of units rebalanced.[2] While a linear cost function is a standard modeling assumption in rebalancing papers such as in He et al. (2020) or Benjaafar et al. (2022), 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 without batching, justifying our linearity assumption. For bike-sharing, while trucking-based rebalancing still provides some firms a large portion of rebalancing operations, crowd-sourcing-based alternatives are rapidly gaining 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 takes significant time handling individual units (withdrawing/inputting units to docks or finding and loading/dropping units for dockless systems), supporting our linearity assumption.

Another assumption we make is that the action at each station is restricted through an upper threshold. We believe this assumption follows the practice for docked and free-floating systems. For a docked system, the total number of docks at a station provides a natural upper threshold.[3] For flee-floating systems, while there is no limit on the total number of units returned to a specific location (customers, barring some

[2] While we do consider fixed costs in Section 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.

[3] While docked systems also place a limit on the total number of units that 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 to eliminate this limit.

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)).

Appendix D: An Alternative State-Space Representation

While proving the results of Section 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 3 (state, cost functions, policies). Then, we will prove that this model is, in fact, equivalent to the original model and use this equivalence to prove the results of Section 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 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:

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

$$\bar{x}_{e,r}^{t+1} = \frac{1}{n}\left(n\bar{a}_{e,r}^t + \sum_{e'=1}^{\hat{e}} \bar{R}_{e',e,r}^t \right).$$

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:

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

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:

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

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

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

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

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

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

$$\sum_{i=1}^{n_e} (x^t_{e,i} - a^t_{e,i}) = n \sum_{b=L^e}^{U^e} b \left(\frac{\sum_{i=1}^{n_e} \mathbb{I}\left\{x^t_{e,i} = b\right\}}{n} - \frac{\sum_{i=1}^{n_e} \mathbb{I}\left\{a^t_{e,i} = b\right\}}{n} \right)$$

$$x^t_{e,r} - a^t_{e,r} = n \left(\frac{x^t_{e,r}}{n} - \frac{a^t_{e,r}}{n} \right) \quad \forall e > \hat{e}.$$

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:

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

Using Lemma D.2, we establish that $\forall e \in [\hat{e}]$,

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

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

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

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 D.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* $x_1^* \le x_2^* \le \ldots \le x_n^*$ *correspond to* $x_1,\ldots,x_n$ *arranged in increasing order and let* $y_1^* \le y_2^* \le \ldots \le y_n^*$ *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 D.1 and prove that the equivalence between the distributional and original models extends to the value functions of the two models.

Corollary D.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 D.1, we have

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

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

using the proof of Proposition D.1,

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

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

Appendix E: Proofs of Section 3 Results

This section focuses of the three results of Section 3, namely Proposition 3.3, Theorem 3.4, and Corollary 3.5. We prove these results by supporting Lemmas and Propositions, which are provided after the proofs of these results.

Proposition 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* (2.1)-(2.2)*, and let $\hat{\boldsymbol{x}}^{t+1}$ be the state evolution of the mean-field dynamics defined in* (3.1)-(3.3) *under the projected mean-field action $g(\boldsymbol{a}^t)$. Then,*

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

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 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 D) 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 D.1, we can establish that

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

Furthermore, through Lemma E.1, we have:

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

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

For the depleted units, we have:

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

$$=n\mathbb{E}_{\{\boldsymbol{D}^t,\bar{\boldsymbol{R}}^t\}}\left[\sum_{e=1}^{\hat{e}}\max(c_{0,\hat{e}+e},c_{\hat{e}+e,0})\left|(1-q)\sum_{e'=1}^{\hat{e}}p_{e',e}\sum_{b=L^{e'}}^{U^{e'}}\bar{a}^t_{e',b}\mathbb{E}\left[\min(b,D^{e'})\right]-\frac{1}{n}\sum_{e'=1}^{\hat{e}}\bar{R}^t_{e',e,r}\right|\middle|\hat{\boldsymbol{a}}^t=\bar{\boldsymbol{a}}^t\right]. \quad \text{(E.2)}$$

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 $\bar{R}^t_{e',e,r}$ and success probabilities $\frac{qp_{e',e}}{n_e}$. Formally,

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

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 $\bar{R}^t_{e',e,r}$ as a sum of Θ values with:

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

Inputting this expression to (E.2), we have:

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

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

$$-\frac{1}{n}\sum_{e'=1}^{\hat{e}}\sum_{j=1}^{\sum_{b=L^{e'}}^{U^{e'}}\sum_{i=\sum_{k=0}^{b-1}n\bar{a}^t_{e,k}+1}^{\sum_{k=0}^{b}n\bar{a}^t_{e,k}}\min(b,D^t_{e',i})}\Theta_{j,e',e,r}\Bigg|\Bigg|\hat{\boldsymbol{a}}^t=\bar{\boldsymbol{a}}^t\Bigg]$$

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

$$-\sum_{j=1}^{\sum_{i=\sum_{k=0}^{b-1}n\bar{a}^t_{e,k}+1}^{\sum_{k=0}^{b}n\bar{a}^t_{e,k}}\min(b,D^t_{e',i})}\Theta_{j,e',e,r}\Bigg|\Bigg|\hat{\boldsymbol{a}}^t=\bar{\boldsymbol{a}}^t\Bigg],$$

by Lemma E.3,

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

Inputting this bound to Equation (E.1),

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

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

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 $\acute{Y}_{e'}^t$, is an independent Poisson random variable with mean $\sum_{b=L^{e'}}^{U^{e'}} n\bar{a}_{e',b}^t \mathbb{E}\left[\min(b, D^{e'})\right]$ (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:

$$\begin{aligned}
&n\mathbb{E}_{\{\boldsymbol{D}^t, \bar{\boldsymbol{R}}^t\}}\left[\sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{m=0}^{\infty}\left|\sum_{d=0}^{m}(\hat{x}_{e,d}^{t+1} - \bar{x}_{e,d}^{t+1})\right| \Big| \hat{\boldsymbol{a}}^t = \bar{\boldsymbol{a}}^t\right] \\
&\quad\leq \mathcal{O}(\sqrt{n}) + n\mathbb{E}_{\{\boldsymbol{D}^t, \acute{\boldsymbol{R}}^t\}}\left[\sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{m=0}^{\infty}\left|\sum_{d=0}^{m}(\hat{x}_{e,d}^{t+1} - \acute{x}_{e,d}^{t+1})\right| \Big| \hat{\boldsymbol{a}}^t = \bar{\boldsymbol{a}}^t\right] \\
&\qquad + n\mathbb{E}_{\{\boldsymbol{D}^t, \bar{\boldsymbol{R}}^t, \acute{\boldsymbol{R}}^t\}}\left[\sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{m=0}^{\infty}\left|\sum_{d=0}^{m}(\acute{x}_{e,d}^{t+1} - \bar{x}_{e,d}^{t+1})\right| \Big| \hat{\boldsymbol{a}}^t = \bar{\boldsymbol{a}}^t\right],
\end{aligned} \tag{E.3}$$

where

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

$\acute{\boldsymbol{R}}_{e'}^t = [\acute{\boldsymbol{R}}_{e',1}^t, \acute{\boldsymbol{R}}_{e',2}^t, \cdots, \acute{\boldsymbol{R}}_{e',\hat{e}}^t]$, $\acute{\boldsymbol{R}}_{e',e}^t = [\acute{R}_{e',e,r}^t, \acute{R}_{e',e,1}^t, \cdots, \acute{R}_{e',e,n_e}^t]$, $\{\acute{R}_{e',e,i}^t\}|_{i=1}^{n_e}$ 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:

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

$$\acute{R}_{e',e,i}^t = \sum_{j=1}^{\acute{Y}_{e'}^t} \Theta_{j,e',e,i} \qquad \forall e', e, i.$$

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 $\bar{x}_{e,d}^{t+0.5}$ as

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

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

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

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

Using the above equations, we have that

$$n\mathbb{E}_{\{\boldsymbol{D}^t, \bar{\boldsymbol{R}}^t, \acute{\boldsymbol{R}}^t\}}\left[\sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{m=0}^{\infty} \left|\sum_{d=0}^{m} (\acute{x}_{e,d}^{t+1} - \bar{x}_{e,d}^{t+1})\right| \middle| \hat{\boldsymbol{a}}^t = \bar{\boldsymbol{a}}^t\right] = \mathbb{E}_{\{\boldsymbol{D}^t, \bar{\boldsymbol{R}}^t, \acute{\boldsymbol{R}}^t\}}\Bigg[$$
$$\sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{m=0}^{\infty} \left|\sum_{d=0}^{m} \sum_{b=L^e}^{U^e} \sum_{i=\sum_{k=0}^{b-1} n\bar{x}_{e,k}^{t+0.5}+1}^{\sum_{k=0}^{b} n\bar{x}_{e,k}^{t+0.5}} \left(\mathbb{I}\left\{\sum_{e'=1}^{\hat{e}} \acute{R}_{e',e,i}^t = d - b\right\} - \mathbb{I}\left\{\sum_{e'=1}^{\hat{e}} \bar{R}_{e',e,i}^t = d - b\right\}\right)\right| \middle| \hat{\boldsymbol{a}}^t = \bar{\boldsymbol{a}}^t\Bigg] \tag{E.4}$$

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 $\Theta_{1,e'}, \ldots, \Theta_{\min(K_1,K_2),e'}$ 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 $\min(K_1, K_2) + 1, \ldots, \max(K_1, K_2)$, 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 (E.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,

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

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

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

by Lemma E.3,

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

Inserting this bound to Equation (E.3),

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

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

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

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

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

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

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

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

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 E.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 E.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

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

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

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

As a result, we obtain

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

Remark E.2. *The value function of the mean-field model, $\hat{V}(\hat{\boldsymbol{x}}^t)$, is Lipschitz continuous in $\hat{\boldsymbol{x}}^t$ with*[4]

$$\left| \hat{V}(\hat{\boldsymbol{x}}^{t,1}) - \hat{V}(\hat{\boldsymbol{x}}^{t,2}) \right| \le \hat{c}(\hat{\boldsymbol{x}}^{t,1}, \hat{\boldsymbol{x}}^{t,2}) \quad \forall \hat{\boldsymbol{x}}^{t,1}, \hat{\boldsymbol{x}}^{t,2} \in \hat{\boldsymbol{X}}, \tag{E.5}$$

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

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

Lemma E.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*

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

Proof. Through Lemma E.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 E.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 E.4. *(Ross (2010), Chapter 7, 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*

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

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

Lemma E.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 $0 \leq a < \infty$. 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,

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

Proof. We will show that the two conditions provided in Theorem E.6 hold. First,

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

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

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

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 E.7 on the factorial term,

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

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

As the two conditions hold, through Theorem E.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 E.8,

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

□

Theorem E.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*

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

Then, if

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

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

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 E.7. *For any $k \in \mathbb{N}$,*

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

Proof.

$$
\begin{aligned}
\ln(k!) &= \sum_{i=1}^{k} \ln(i) \\
\ln(k!) &\ge \int_1^k \ln(z)dz \\
\ln(k!) &\ge k\ln(k) - k + 1 \\
k! &\ge k^k e^{-k+1}.
\end{aligned}
$$

□

Lemma E.8. *Let $\{Y, Y_i\}$ be a sequence of i.i.d. non-negative random variables with common distribution F. Then:*

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

where B denotes Brownian Bridge.

Proof.

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

□

Theorem 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 D.3 and Proposition D.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 E.9 proves the theorem. □

Proposition E.9. *For any inventory position $\bar{\boldsymbol{x}}^t \in \bar{\boldsymbol{X}}$,*

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

$$2.\ \hat{V}(\bar{\boldsymbol{x}}^t) - \bar{V}(\bar{\boldsymbol{x}}^t) \leq \mathcal{O}(\sqrt{n}).$$

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 3.3, as well as Lemma E.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

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

through applying Lemma E.10 on the rebalancing cost function and the newsvendor costs,

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

through the Lipschitz bound given at (E.5),

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

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 3.3 on the first term and Lemma E.10 on the second term,

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

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 3.3, which is the only step where we observe an additional cost of $\mathcal{O}(\sqrt{n})$ (all three terms in Lemma E.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 3.3 implies that $\exists M > 0, n_0 \in \mathbb{R}_+$ such that $\forall n \ge n_o$,

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

Moreover, $M < \infty$, in both Lemma E.3 and Lemma E.5 (which together give the sub-optimality coefficients used in Proposition 3.3), coefficients of $\sqrt{n}$ are bounded and are independent of γ. Consequently, we can express the total cost as

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

For the second statement,

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

through the Lipschitz bound given at (E.5),

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

by Proposition 3.3,

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

repeating the analysis in the first statement,

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

Lemma E.10. *For any action $\hat{\boldsymbol{a}}^t \in \hat{\boldsymbol{A}}$, function f satisfies*

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

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 E.1:

$$\begin{aligned}
\hat{c}(\hat{\boldsymbol{a}}^t, f(\hat{\boldsymbol{a}}^t)) &\leq \sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{b=0}^{\infty} \left| \sum_{d=0}^{b} n\hat{a}^t_{e,d} - \sum_{d=0}^{b} n\bar{a}^{t^*}_{e,d} \right| + \sum_{e=1}^{\hat{e}} \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0}) \left| n\hat{a}^t_{e,r} - n\bar{a}^{t^*}_{e,r} \right| \\
&\leq \sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{b=L^e}^{U^e} \sum_{d=0}^{b} \left| n\hat{a}^t_{e,d} - n\bar{a}^{t^*}_{e,d} \right| + \sum_{e=1}^{\hat{e}} \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0}) \left| n\hat{a}^t_{e,r} - n\bar{a}^{t^*}_{e,r} \right|, \qquad \text{(E.6)}
\end{aligned}$$

using the Largest Remainder Algorithm,

$$\begin{aligned}
&\leq \sum_{e=1}^{\hat{e}} \max(c_{0,e}, c_{e,0}) \sum_{b=L^e}^{U^e} \sum_{d=L^e}^{b} 1 + \sum_{e=1}^{\hat{e}} \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0}) 0.5 \\
&< \infty.
\end{aligned}$$

For the second statement:

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

We replace $\max(c_{0,e}, c_{e,0})$ with $U^e c_h + c_p \mathbb{E}[D^e]$ and $\max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0})$ with c_h in Equation (E.6). 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 E.1:

$$\begin{aligned}
\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| \\
\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} \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}^t_{e',b} - n\bar{a}^{t^*}_{e',b} \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}^{t^*}_{e,m} - n\hat{a}^t_{e,m}) \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}^{t^*}_{e,m} - n\hat{a}^t_{e,m}) \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}^{t^*}_{e,m} - n\hat{a}^t_{e,m}) \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 $\sum_{m=L^e}^{U^e} (n\bar{a}^{t^*}_{e,m} - n\hat{a}^t_{e,m}) = 0$, we have

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

□

Corollary 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 3.4,

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

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

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:

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

Appendix F: 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 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 $\breve{\boldsymbol{x}}_e^t = \{\{\breve{x}_{e,r}^t, \breve{x}_{e,s}^t, \breve{x}_{e,u}^t, \breve{\boldsymbol{x}}_{e,d}^t\}_{d=0}^{\infty}\}_{e=1}^{\hat{e}} \in \breve{\boldsymbol{X}}$ satisfies:

$$\begin{aligned}
&\breve{x}_{e,d}^t = 0 && \forall e, d > U^e,\\
&\sum_{d=0}^{U^e-1} \breve{x}_{e,d}^t \leq \frac{n_e}{n} && \forall e,\\
&\breve{x}_{e,r}^t \in \left[0, U_r^e + (1-q)\sum_{e'=1}^{\hat{e}} p_{e',e}\frac{n_{e'}}{n}\mathbb{E}\left[\min(U^{e'}, D^{e'})\right]\right] && \forall e.\\
&\breve{x}_{e,s}^t \in \left[0, \sum_{d=U^e+1}^{\infty} d\frac{n_e}{n}\mathbb{P}\left[U^e - \min(U^e, D^e) + \hat{R}^{e,u} = d\right]\right] && \forall e.\\
&\hat{R}^{e,u} \stackrel{d}{=} \text{Pois}\left(q\frac{n}{n_e}\sum_{e'=1}^{\hat{e}} p_{e',e}\frac{n_{e'}}{n}\mathbb{E}\left[\min(U^{e'}, D^{e'})\right]\right) && \forall e\\
&\breve{x}_{e,u}^t \in \left[0, \sum_{d=U^e+1}^{\infty} (d-U^e)\frac{n_e}{n}\mathbb{P}\left[U^e - \min(U^e, D^e) + \hat{R}^{e,u} = d\right]\right] && \forall e.
\end{aligned}$$

Here, both $\breve{x}_{e,s}^t$ and $\breve{x}_{e,u}^t$ are variables indicating the excess amount of units at a type, albeit with different weights. The upper bounds selected for all three of $\breve{x}_{e,s}^t$, $\breve{x}_{e,u}^t$, $\breve{x}_{e,r}^t$ are not arbitrary as they correspond to the highest inventory, which they can observe through their inventory evolution given in Equations (F.1)-(F.5). 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 F.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.*

We must also introduce the new state transition function given a new state. The state transitions under this setting are:

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

$$\breve{x}^{t+1}_{e,d} = 0 \qquad \forall e, t, d > U^e, \tag{F.2}$$

$$\breve{x}^{t+1}_{e,r} = \hat{a}^t_{e,r} + (1-q) \sum_{e'=1}^{\hat{e}} p_{e',e} \sum_{b=L^{e'}}^{U^{e'}} \hat{a}^t_{e',b} \mathbb{E}\left[\min(b, D^{e'})\right] \qquad \forall e, t, \tag{F.3}$$

$$\breve{x}^{t+1}_{e,s} = \sum_{d=U^e+1}^{\infty} d \sum_{b=L^e}^{U^e} \hat{a}^t_{e,b} \mathbb{P}\left[b - \min(b, D^e) + \hat{R}^e = d\right] \qquad \forall e, t, \tag{F.4}$$

$$\breve{x}^{t+1}_{e,u} = \sum_{d=U^e+1}^{\infty} (d - U^e) \sum_{b=L^e}^{U^e} \hat{a}^t_{e,b} \mathbb{P}\left[b - \min(b, D^e) + \hat{R}^e = d\right] \qquad \forall e, t. \tag{F.5}$$

Here, Equation (F.1) preserves the state transitions for stations with less than or equal to U^e units, and stations with more than U^e units are represented in Equations (F.4),(F.5), 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 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 F.2. *The Wasserstein component of the rebalancing cost function of the mean-field model, $\hat{c}$, can be partitioned into two parts as:*

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

Proof.

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

Lemma F.2 tells us that we can aggregate $\hat{x}^t_{e,d}$ for $d\geq U^e+1$ into a single variable, which we do by letting:

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

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

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

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

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

We add these two components to form $\hat{c}_2(\breve{\boldsymbol{x}}^t,\hat{\boldsymbol{a}}^t)$, the new rebalancing cost function. Combined, we express the new rebalancing cost function as:

$$\begin{aligned}
\hat{c}_2(\breve{\boldsymbol{x}}^t,\hat{\boldsymbol{a}}^t) = {} & n\sum_{e=1}^{\hat{e}}\frac{c_{e,e}}{2}\breve{x}^t_{e,u} + n\sum_{e=1}^{\hat{e}}\frac{c_{e,e}}{2}\sum_{b=0}^{U^e-1}\left|\sum_{d=0}^{b}(\breve{x}^t_{e,d}-\hat{a}^t_{e,d})\right| \\
& + n\min_{y^{e_1,e_2}}\sum_{e_1=0}^{2\hat{e}}\sum_{e_2=0,e_2\neq e_1}^{2\hat{e}}\left(c_{e_1,e_2}-\frac{c_{e_1,e_1}}{2}-\frac{c_{e_2,e_2}}{2}\right)y^{e_1,e_2} \qquad \text{(F.6)}
\end{aligned}$$

$$\text{s.t.} \quad \breve{x}^t_{e,s} + \sum_{b=1}^{U^e} b(\breve{x}^t_{e,b} - \hat{a}^t_{e,b}) = \sum_{e_1=0}^{2\hat{e}} (y^{e,e_1} - y^{e_1,e}) \qquad \forall e \in [\hat{e}],$$

$$\breve{x}^t_{e,r} - \hat{a}^t_{e,r} = \sum_{e_1=0}^{2\hat{e}} (y^{e+\hat{e},e_1} - y^{e_1,e+\hat{e}}) \qquad \forall e \in [\hat{e}],$$

$$y^{e_1,e_2} \geq 0 \qquad \forall e_1, e_2.$$

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 F.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* (F.1)-(F.5).

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 (F.1)-(F.5) can also be expressed as:

$$\breve{x}^{t+1}_{e,d} = \hat{x}^{t+1}_{e,d} \qquad \forall e, t, d \leq U^e,$$

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

$$\breve{x}^{t+1}_{e,r} = \hat{x}^{t+1}_{e,r} \qquad \forall e, t,$$

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

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

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

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

□

Appendix G: Reformulating the Control Algorithm into a Linear Program

Under the control algorithm provided in Section 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 (F.3) 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

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

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) $\hat{\boldsymbol{s}}^t = \{\hat{s}_{e,b}^t|_{b=0}^{U^e-1}\}_{e=1}^{\hat{e}}$, *(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:

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

$$\text{s.t.} \quad \sum_{b=L^e}^{U^e}\hat{a}^{t+k,e}_{e,b} = \frac{n_e}{n} \quad \forall e,k\in\{0,\dots,T\}, \tag{G.1}$$

$$\hat{a}^{t+k}_{e,b} \geq 0 \quad \forall e,k\in\{0,\dots,T\}, b, \tag{G.2}$$

$$\frac{n_e}{n}U_0 \geq \hat{a}^{t+k}_{e,r} \quad \forall e,k\in\{0,\dots,T\}, \tag{G.3}$$

$$\hat{a}^{t+k}_{e,r} \geq \frac{n_e}{n}L_0 \quad \forall e,k\in\{0,\dots,T\}, \tag{G.4}$$

$$\breve{x}^t_{e,s} + \sum_{d=1}^{U^e} d(\breve{x}^t_{e,d} - \hat{a}^t_{e,d}) = \sum_{e_1=0}^{2\hat{e}}(y^{t,e,e_1} - y^{t,e_1,e}) \quad \forall e\in[\hat{e}], \tag{G.5}$$

$$\breve{x}^t_{e,r} - \hat{a}^t_{e,r} = \sum_{e_1=0}^{2\hat{e}}(y^{t,e+\hat{e},e_1} - y^{t,e_1,e+\hat{e}}) \quad \forall e\in[\hat{e}], \tag{G.6}$$

$$\sum_{d=1}^{U^e} d(\sum_{b=L^e}^{U^e}\hat{a}^{t+k-1}_{e,b}\mathbb{P}\left[b-\min(b,D^e)=d\right] - \hat{a}^{t+k}_{e,d}) = \sum_{e_1=0}^{2\hat{e}}(y^{t+k,e,e_1} - y^{t+k,e_1,e}) \quad \forall e\in[\hat{e}], k\in[T], \tag{G.7}$$

$$\hat{a}^{t+k-1}_{e,r} + \sum_{e'=1}^{\hat{e}} p_{e',e}\sum_{b=L^{e'}}^{U^{e'}}\hat{a}^{t+k-1}_{e',b}\mathbb{E}\left[\min(b,D^{e'})\right] - \hat{a}^{t+k}_{e,r} = \sum_{e_1=0}^{2\hat{e}}(y^{t+k,e+\hat{e},e_1} - y^{t+k,e_1,e+\hat{e}}) \quad \forall e\in[\hat{e}], k\in[T], \tag{G.8}$$

$$\sum_{d=1}^{U^e} d(\sum_{b=L^e}^{U^e}\hat{a}^{t+T}_{e,b}\mathbb{P}\left[b-\min(b,D^e)=d\right] - \hat{a}^{t+T}_{e,d}) = \sum_{e_1=0}^{2\hat{e}}(y^{t+T+1,e,e_1} - y^{t+T+1,e_1,e}) \quad \forall e\in[\hat{e}] \tag{G.9}$$

$$\sum_{e'=1}^{\hat{e}} p_{e',e}\sum_{b=L^{e'}}^{U^{e'}}\hat{a}^{t+T}_{e',b}\mathbb{E}\left[\min(b,D^{e'})\right] = \sum_{e_1=0}^{2\hat{e}}(y^{t+T+1,e+\hat{e},e_1} - y^{t+T+1,e_1,e+\hat{e}}) \quad \forall e\in[\hat{e}] \tag{G.10}$$

$$y^{t+k,e_1,e_2} \geq 0 \quad \forall e_1,e_2,k\in\{0,\dots,T\}, \tag{G.11}$$

$$\hat{s}^t_{e,b} \geq \sum_{d=0}^{b}(\breve{x}^t_{e,d} - \hat{a}^t_{e,d}) \quad \forall e, b\in\{0,\dots,U^e-1\}, \tag{G.12}$$

$$\hat{s}^t_{e,b} \geq -\sum_{d=0}^{b}(\breve{x}^t_{e,d} - \hat{a}^t_{e,d}) \quad \forall e, b\in\{0,\dots,U^e-1\}, \tag{G.13}$$

$$\hat{z}^{t+k}_{e,b} \geq \sum_{d=0}^{b}\left(\sum_{m=L^e}^{U^e}\hat{a}^{t+k-1}_{e,m}\mathbb{P}\left[m-\min(m,D^e)=d\right] - \hat{a}^{t+k}_{e,d}\right) \forall e,k\in[T], b\in\{0,\dots,U^e-1\}, \tag{G.14}$$

$$\hat{z}_{e,b}^{t+k} \geq -\sum_{d=0}^{b}\left(\sum_{m=L^e}^{U^e} \hat{a}_{e,m}^{t+k-1}\mathbb{P}\left[m-\min(m,D^e)=d\right]-\hat{a}_{e,d}^{t+k}\right) \;\forall e,k\in[T], b\in\{0,\ldots,U^e-1\}, \tag{G.15}$$

$$\hat{l}_{e,b}^{t+T} \geq \sum_{d=0}^{b}\left(\sum_{m=L^e}^{U^e} \hat{a}_{e,m}^{t+T}\mathbb{P}\left[m-\min(m,D^e)=d\right]-\hat{a}_{e,d}^{t+T}\right) \;\; \forall e, b\in\{0,\ldots,U^e-1\}, \tag{G.16}$$

$$\hat{l}_{e,b}^{t+T} \geq -\sum_{d=0}^{b}\left(\sum_{m=L^e}^{U^e} \hat{a}_{e,m}^{t+T}\mathbb{P}\left[m-\min(m,D^e)=d\right]-\hat{a}_{e,d}^{t+T}\right) \;\; \forall e, b\in\{0,\ldots,U^e-1\}. \tag{G.17}$$

Here, (G.1), (G.2), (G.3), and (G.4) are the standard action constraints we impose, (G.5), (G.6), (G.7), (G.8), (G.9), (G.10), and (G.11) are the flow constraints of the rebalancing cost function, (G.12) and (G.13) are constraints on dummy variables assigned to the Wasserstein component of the cost term $\hat{c}_2(\breve{\boldsymbol{x}}^t, \hat{\boldsymbol{a}}^t)$, (G.14) and (G.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 (G.16) and (G.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.

Appendix H: Existence and Utilization of Fixed Points

As discussed in Section 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 H.1. *$\breve{\boldsymbol{x}}^f$ is a fixed point of $\hat{V}$ if at $\breve{\boldsymbol{x}}^f$, there exists an optimal 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 H.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* (3.6)*:*

1. *$\breve{\boldsymbol{X}}$ and $\hat{\boldsymbol{A}}$ are compact;*
2. *$\breve{\boldsymbol{X}}$ and $\hat{\boldsymbol{A}}$ are convex;*
3. *The mapping from $\hat{\boldsymbol{a}}^t \rightarrow \breve{\boldsymbol{x}}^{t+1}$ 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 H.3, we verify that for $q=0$, our system satisfies these conditions and that a fixed point exists.

Proposition H.3. *The dynamic program defined in* (3.6) *satisfies the fixed point property if $q=0$.*

Proof. We will prove that $\hat{V}$ satisfies all four conditions stated in Proposition H.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 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 (F.6) and (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 H.4, showing that

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

   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. □

Lemma H.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 H.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:*

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

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 (E.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 H.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 H.5, which we do so in Proposition H.10. Finally, the proof of Theorem 4.2, split into two different Propositions, is presented below.

Theorem 4.2. *The optimality gap of the policy $\tilde{\boldsymbol{\pi}}$ obtained via* (4.1)-(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:

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

Proof. The first statement is proven in Proposition H.6 and the second statement is proven in Proposition H.10.

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 H.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 H.7, H.8, H.9), we can bound the difference in per-period costs and then use recursion to obtain the final bound.

For Proposition H.10, 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. □

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 H.6. *The optimality gap of the policy $\tilde{\boldsymbol{\pi}}$ obtained via* (4.1)-(4.3) *decreases exponentially with respect to length T of the transient horizon:*

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

where

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

Proof. We first express the optimal cost as follows:

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

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

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

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

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

applying the triangle inequality (Lemma H.7) on the first term and the Lipschitz bound provided at (E.5) on the last term,

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

By Lemma H.8,

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

Inserting this expression to Equation (H.1) and simplifying the terms

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

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

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

by Lemma H.9,

$$\le \frac{\gamma^{T+1}}{1-\gamma} 2 \sum_{e=1}^{\hat{e}} \left(n_e \max(c_{0,e}, c_{e,0})(U^e - L^e) n_e + n \max(c_{0,\hat{e}+e}, c_{\hat{e}+e,0})(U^{r,e} - L^{r,e}) \right). \qquad \square$$

Lemma H.7. *The $\hat{c}_2$ function satisfies triangle inequality.*

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

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

For the Wasserstein component:

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

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

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. □

Lemma H.8. *(Akturk (2022), Lemma 2.2) If the $\hat{c}_2$ function satisfies the triangle inequality; for any inventory position $\breve{\boldsymbol{x}}$, there exists an optimal policy satisfying*

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

Lemma H.9. *For any two actions $\hat{\boldsymbol{a}}^1,\hat{\boldsymbol{a}}^2\in\hat{\boldsymbol{A}}$, the rebalancing cost function $\hat{c}_2$ satisfies:*

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

Proof. Through the proof of Proposition F.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 E.1,

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

Proposition H.10. *Let $q=0$. Then, the optimality gap of the policy $\tilde{\boldsymbol{\pi}}$ obtained via* (4.1)-(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

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

Proof. We first express the optimal cost as follows:

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

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 (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 H.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

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

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

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

by Lemma H.5,

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

applying the triangle inequality (Lemma H.7) on the first term and the Lipschitz bound provided at (E.5) on the last term,

$$\leq \gamma^T \left(\hat{c}_2(\breve{\boldsymbol{x}}^{t+T}, \hat{\boldsymbol{a}}^{t+T^*}) + \hat{c}_2(\hat{\boldsymbol{a}}^{t+T^*}, \hat{\boldsymbol{a}}^f) + \hat{V}(\breve{\boldsymbol{a}}^f) \right)$$
$$- \gamma^T \left(\hat{c}_2(\breve{\boldsymbol{x}}^{t+T}, \hat{\boldsymbol{a}}^{t+T^*}) - \hat{c}_2(\hat{\boldsymbol{a}}^{t+T^*}, \hat{\boldsymbol{a}}^f) + \hat{V}(\hat{\boldsymbol{a}}^f) \right)$$
$$\leq 2\gamma^T \hat{c}_2(\hat{\boldsymbol{a}}^{t+T^*}, \hat{\boldsymbol{a}}^f),$$

by Lemma H.9,

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

Remark H.11. *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 $\gamma < 0.5$.*

Appendix I: Supplementary Material for Numerical Analysis

In this section, we provide supplementary material for Section 5, which includes mathematical formulations for certain policies we implement, additional descriptive information to understand the setting we implement our policies, and additional experiments to test our policies performance.

I.1. Formulating the Myopic Policy

The optimization program provided in Equation (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 D, we are able to exactly solve for the myopic policy through a more manageable mixed-integer reformulation. Specifically, we utilize Proposition D.1 to state that $\exists \boldsymbol{a}_M^t$ such that given inventory position $\boldsymbol{x}^t$,

$$g(\boldsymbol{a}_M^t) \in \arg\min_{\bar{\boldsymbol{a}}^t \in \bar{\boldsymbol{A}}} \hat{c}(g(\boldsymbol{x}^t), \bar{\boldsymbol{a}}^t) + \hat{N}(\bar{\boldsymbol{a}}^t). \tag{I.1}$$

Furthermore, again through Proposition D.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 (I.1), which is a mixed-integer optimization problem (due to the constraints $n\bar{a}_{e,d}^t \in \mathbb{N}_0 \ \forall e, d$ and $n\bar{a}_{e,r}^t \in \mathbb{N}_0 \, \forall e$). We then project this action to the original stochastic problem using the h function.

I.2. Formulating the Large Market Policy

The large market policy uses a different fluid approximation to obtain a policy. Under the large market approximation, the state at period $t+1$ is given by the following equations:

$$x_{e,i}^{t+1} = a_{e,i}^t - \min(a_{e,i}^t, \mu_e) + \sum_{e'=1}^{\hat{e}} \sum_{j=1}^{n_{e'}} \frac{q p_{e',e}}{n_e} \min(a_{e',j}^t, \mu_{e'}) \quad \forall i \in [n], \tag{I.2}$$

$$x_{e,r}^{t+1} = a_{e,r}^t + (1-q) \sum_{e'=1}^{\hat{e}} p_{e',e} \sum_{j=1}^{n_{e'}} \min(a_{e',j}^t, \mu_{e'}) \quad \forall e. \tag{I.3}$$

Remark I.1. *As the resultant inventory positions are fractional, we extend the action space $\boldsymbol{A}$ to allow for fractional actions, which we then round when implementing the policy.*

The resultant value function can then be expressed as

$$V^L(\boldsymbol{x}^t) = \min_{\boldsymbol{a}^t \in \boldsymbol{A}} c(\boldsymbol{x}^t, \boldsymbol{a}^t) + c_h \sum_{e=1}^{\hat{e}} \sum_{i=1}^{n_e} a_{e,i}^t + c_h \sum_{e=1}^{\hat{e}} a_{e,r}^t + c_p \sum_{e=1}^{\hat{e}} \sum_{i=1}^{n_e} (\mu_e - a_{e,i}^t)^+ + \gamma V^L(\boldsymbol{x}^{t+1}). \tag{I.4}$$

Furthermore, 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 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.

While formulating the large market policy, we established that the optimal policy obtained through Equation (I.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:

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

$$\mu_e \geq a^t_{e,i} - z^t_{e,i}M \qquad \forall e, i.$$

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 (I.2), (I.3) as

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

$$x^{t+1}_{e,r} = a^t_{e,r} + (1-q) \sum_{e'=1}^{\hat{e}} p_{e',e} \sum_{j=1}^{n_{e'}} \left(z^t_{e',j}\mu_{e'} + (1 - z^t_{e',j}) a^t_{e',j} \right) \quad \forall e,$$

and Equation (I.4) as

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

I.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.[5] 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 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 their testing. As we expect usage to be 56.6 minutes daily, we assign $q = 0$.[6]

Lastly, for completeness, we provide the state transition probabilities and the mean values for the demand distributions of the synthetic experiments. These values are provided in Table 1 and Table 2, respectively.

[5] 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.

[6] In Appendix I.6, we look at the performance of our policy for different q values.

Transition Probability	1	2	3	4
1	0.4	0.3	0.2	0.1
2	0.3	0.3	0.1	0.3
3	0.2	0.4	0.3	0.1
4	0.1	0.3	0.3	0.3

Table 1 Transition Probabilities for Synthetic Experiment

Type	1	2	3	4
Mean Demand	4	7	5	6

Table 2 Mean Demand for Synthetic Experiment

I.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.

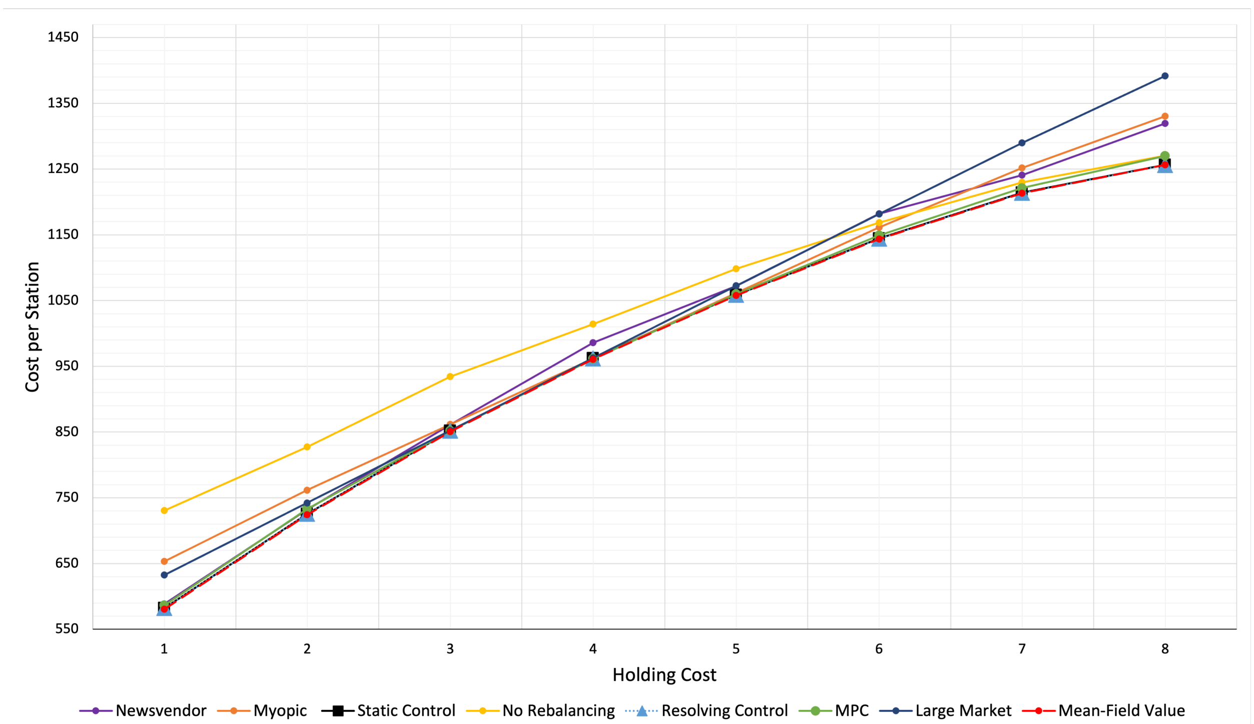


Figure 1 Impact of the Holding Cost on the Performance of Policies

We first observe that mean-field-based policies are effective at all holding cost values. These policies can reduce or increase the fleet size according to the holding cost value.

I.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.

I.5.1. Adding a Constant to all Rebalancing/Recharging/Sourcing costs We add a constant k (k varies in the x-axis) to all rebalancing/recharging/sourcing costs for this experiment. The results are shown in Figure 2.

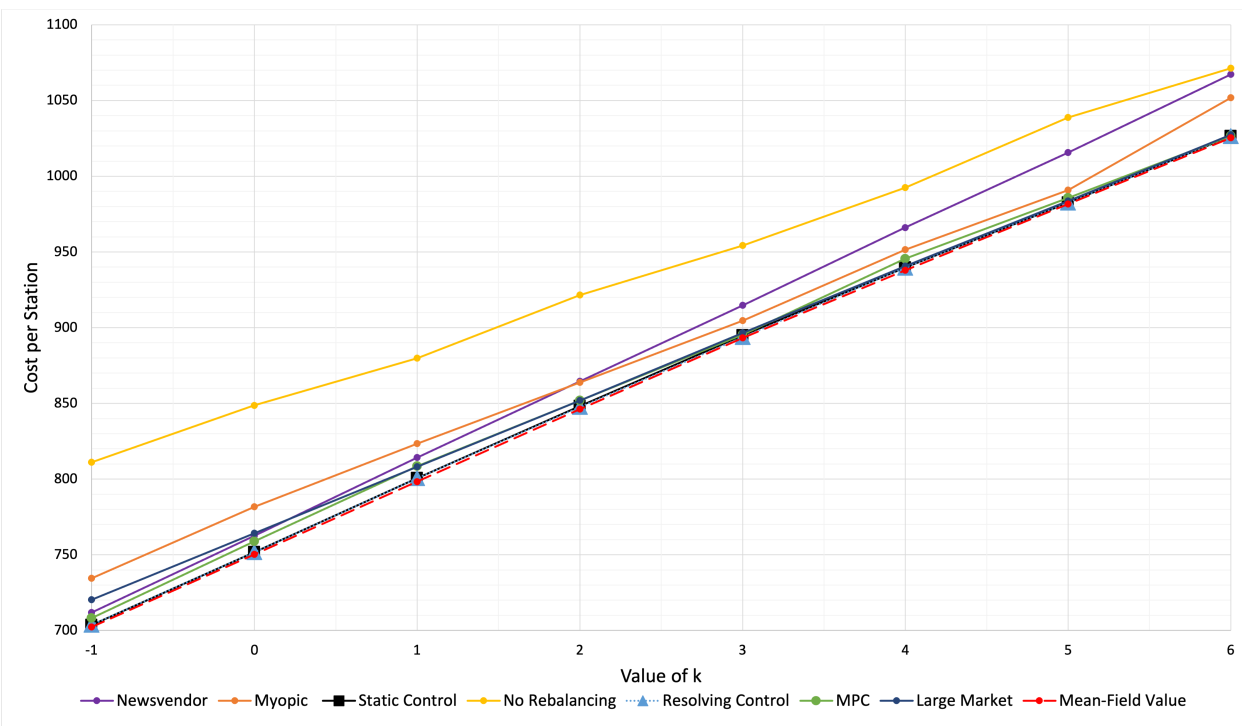
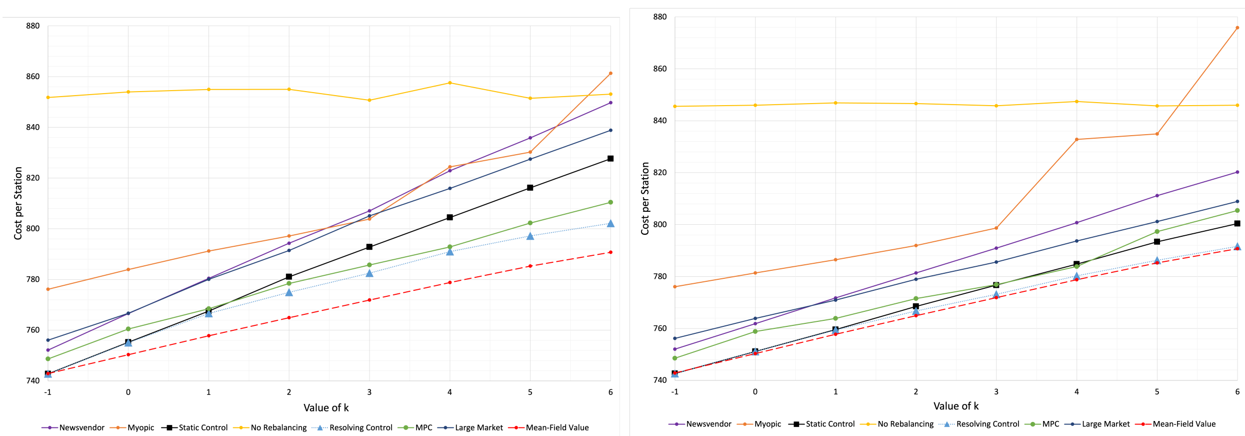


Figure 2 Impact of k on the Performance of Policies

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.

I.5.2. 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 3.



(a) Results for n=16 **(b) Results for n=160**

Figure 3 Impact of Cross Rebalancing Cost 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 than static control, whereas, for $n=160$, static control performs better for all k.

I.6. Impact of the Depletion Probability

For this experiment, we vary the depletion probability q (assumed to be 0 previously). Due to the long computation time of MPC and resolving control for the $q>0$ case, we reduce the number of repetitions.We also remove the mean-field value due to the computational cost of solving the model for a very large T. The results are shown in Figure 4.

Through Figure 4, 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 observation is that as q changes, the relative

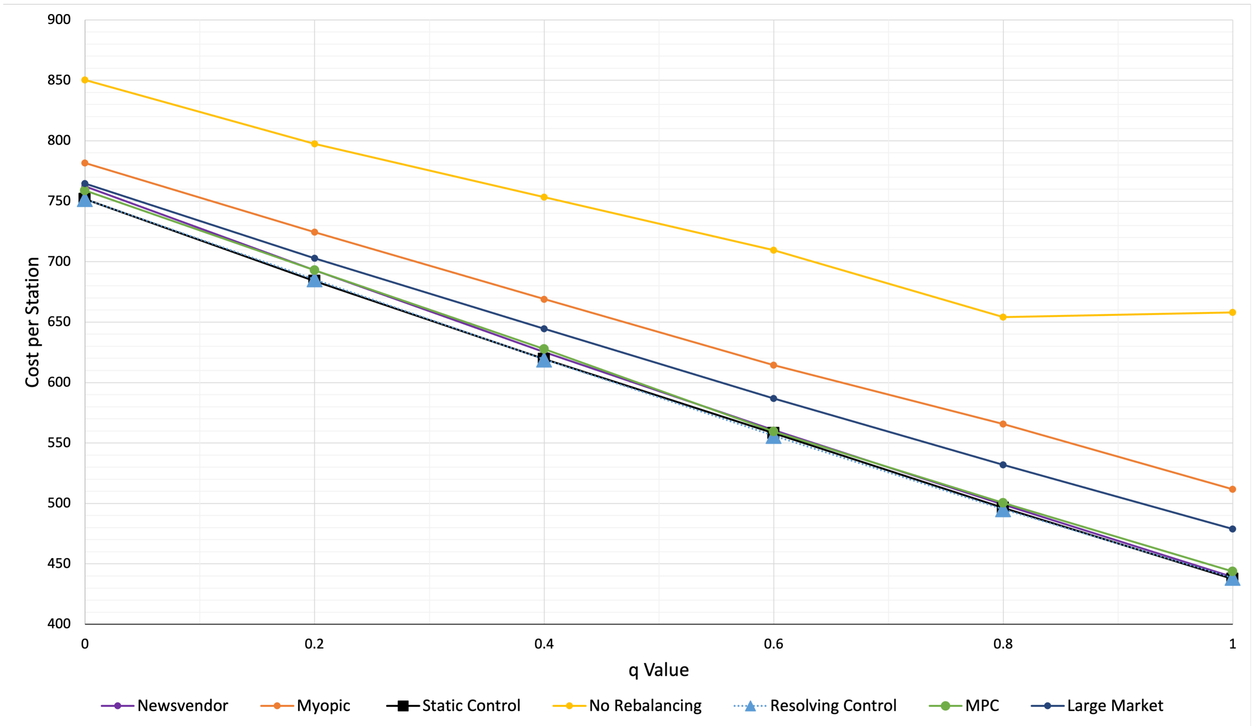


Figure 4 **Impact of the Depletion Probability on the Performance of Policies**

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.[7] 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.

I.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 5.

The main observation is that the upper threshold does not impact the cost of policies (except the hybrid no rebalancing policy, where thresholds do play an important role) past a specific value. The reason is that policies themselves will restrict the number of charged units at a station through either a restriction on the fleet size or through rebalancing the excessive accumulation of charged units at a station. As a result, the upper threshold restriction is redundant.

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

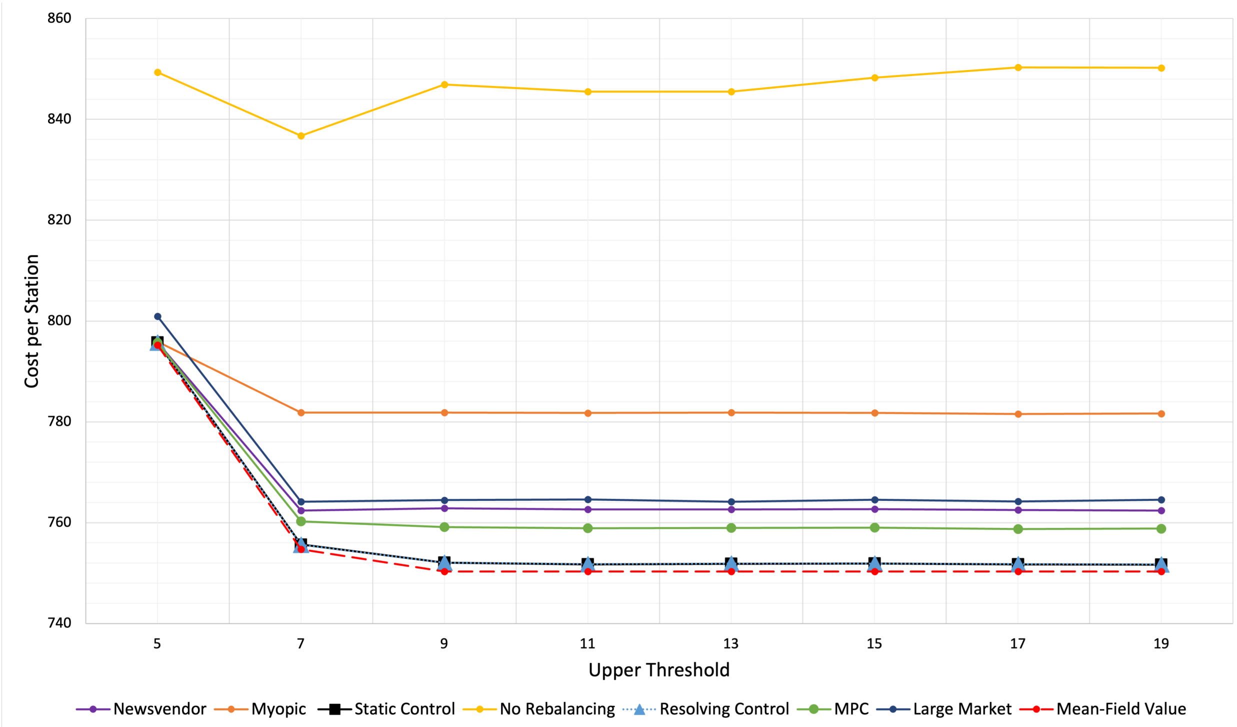


Figure 5 Impact of the Upper Threshold on the Performance of Policies

In contrast, most policies are severely restricted in their actions under a low upper threshold value. Hence, we observe similar performance for low upper threshold values.

I.8. Increasing the Number of Types

We also conduct an additional experiment to see if the control algorithms can outperform other policies given a larger number of types. To this end, we construct a 20-type network by letting $\hat{e} = 20$, with $n_e = 4\ \forall e$. We keep the remaining parameters the same as our other synthetic experiments, assuming Poisson demand at stations and an imbalanced network (with 20 types instead of 4). We provide the results (the values correspond to cost per station under the given policy) in Table 3.[8]

Newsvendor	Myopic	Static Control	Resolving Control	MPC	Large Market	Mean-Field Value
772.2	790.9	762.7	762.6	767.7	775.0	760.6

Table 3 Results for 20 Types

[8] We haven't included the Modified No Rebalancing Policy in the results due to the computational challenge of obtaining a good fleet size with 20 types.

Table 3 shows that control algorithms continue to outperform the other benchmarks for a larger number of types, similar to the low-n data points in Figure 3. While there is a slight gap between the control policies' performance and the mean-field lower bound, this gap is due to the low number of stations per type.

I.9. Descriptive Statistics for the Austin Scooter-Sharing Data-set

As mentioned in Section 5, for the numerical experiment, we will focus on the 203,810 scooter trips obtained from the Austin Shared Micromobility Vehicle Trips dataset (City of Austin Transportation Department (2022)). Figure 6 provides a snapshot of scooter usage in Austin.

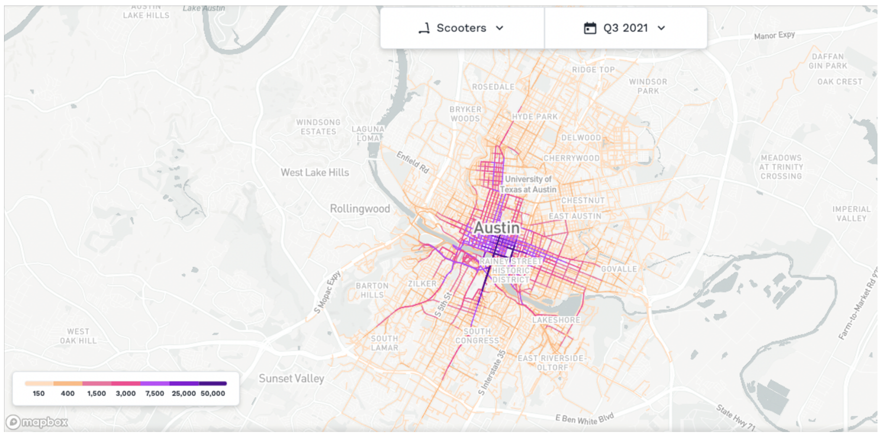


Figure 6 **Overview of Scooter Usage in Austin (City of Austin Transportation Department (2022))**

The dataset provides identifying information about each ride, including

1. Device ID
2. Date/Time of Pick-up
3. Date/Time of Drop off
4. Trip Distance
5. Origin and Destination Census Tract of the Ride
6. Origin and Destination Council District of the Ride

Each ride originates from one of 122 Census Tracks and ten council districts in Austin. First, as Census Tracks are smaller, they offer more granularity and will be used instead of Council Districts. Second, we observe that some of these tracts have only a few rides originating from them. To this end, we remove all rides with incomplete information and those originating from census tracts with less than 28 pickups (at least two per day on average) to filter the data. After this filtering, we are left with 203,385 rides and 63 census tracts.

Lastly, we provide the distribution of the duration of each trip (in seconds) and the distance covered by the trips (in meters) in Figures 7 and 8, respectively.

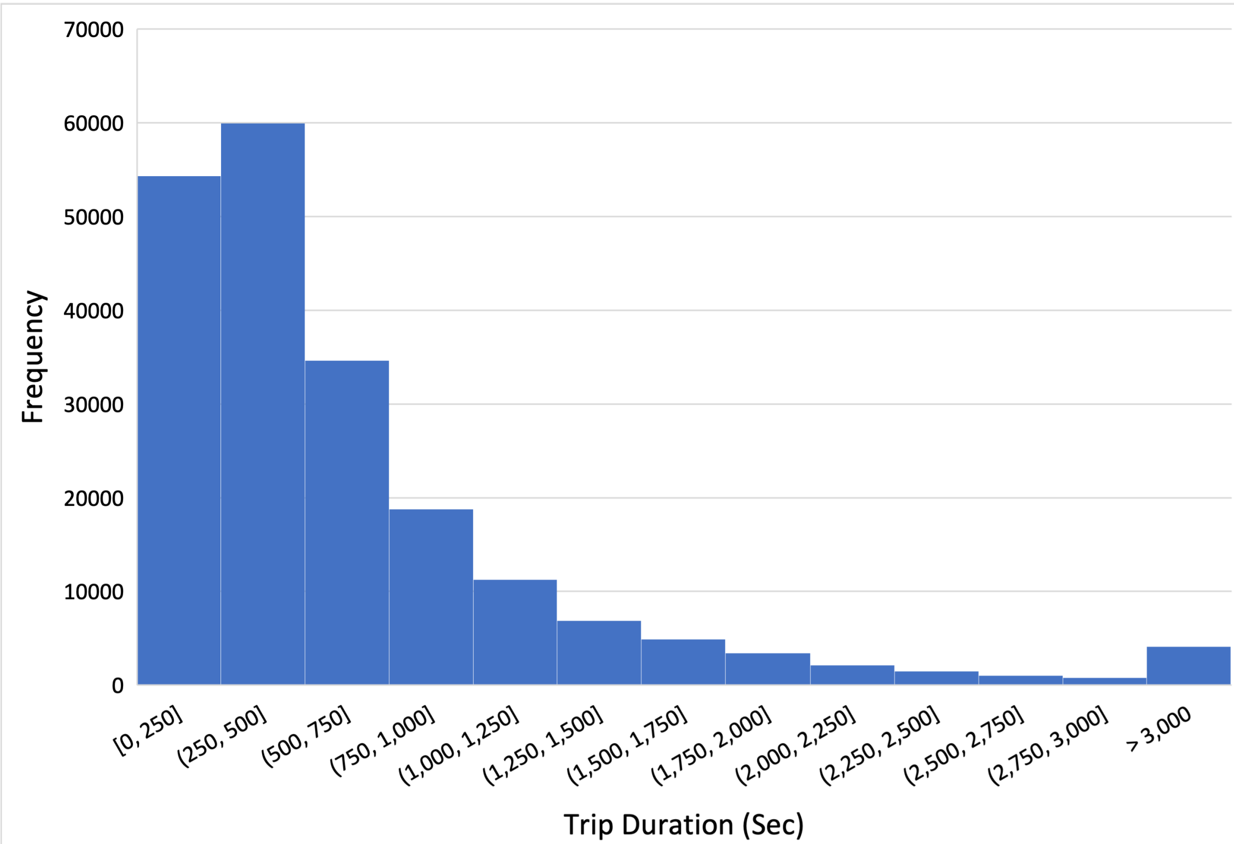


Figure 7 **Duration of Trips**

I.10. Forming Clusters From the Austin Scooter-Sharing Data-set

This subsection will detail grouping Census Tracts into clusters. Mathematically, the challenge is to assign each of the 63 census tracts to one of N possible clusters. As mentioned in Section 5, we will use a two-step algorithm where we first use K-means clustering to cluster census tracks according to their coordinates and then adjust these interim clusters according to their probability transition values.

For the first step, we use the longitude and latitude values of the census tracks provided in the 2010 US Census.[9] Based on this information, we will follow a standard approach of weighted K-means clustering on

[9] Scooters were first introduced at Austin in 2018, so the dataset tracks rides according to the 2010 Census.

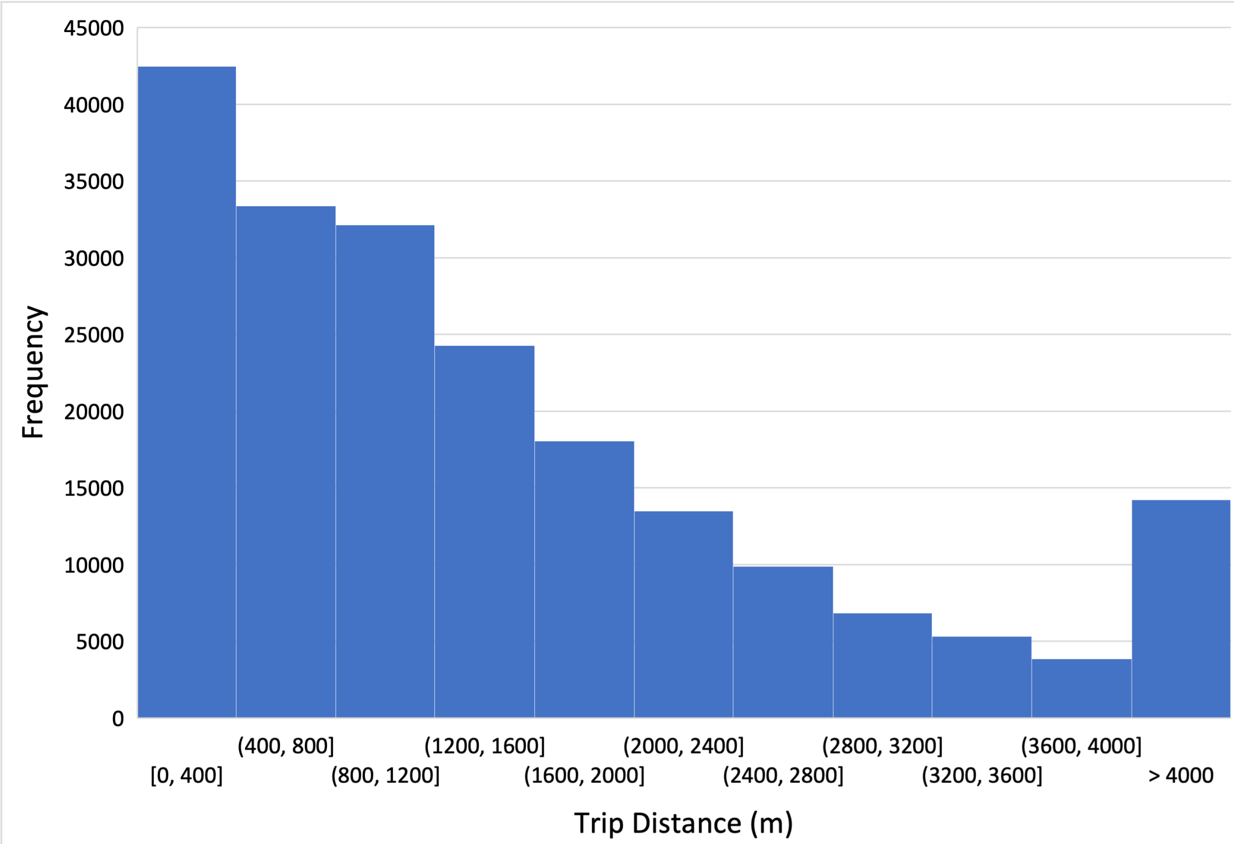


Figure 8 **Trip Distance**

latitude and longitude values. We use the elbow method to determine the best number of clusters, which tells us when the returns diminish. Based on Figure 9, we determine the number of clusters to be 5.

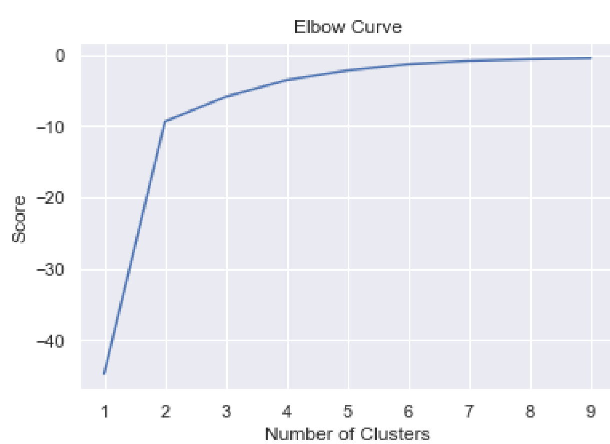


Figure 9 **Elbow Curve for Weighted K-Means Clustering**

We then run the weighted K-means algorithm to assign each of the 63 census tracts to one of five clusters (we run the algorithm several times to minimize any uncertainty with respect to the initial conditions), completing the first step of our overall approach.

For the second step, we first compute the transition probabilities between these newly formed clusters, so we calculate the probability that a ride originating in cluster i will end up in cluster j. We do the same for each census tract, where we compute the probability that a ride originating in that census track will end up in each cluster j. Formally, letting $C_1 : \{1, \cdots, 5\}, C_2 : \{1, \cdots, 63\}$, we compute $\alpha_{i,j}$, $i, j \in C_1$ for clusters (probability transition values between clusters) and $\beta_{k,j}\, j \in C_1, k \in C_2$ for census tracts (probability transition values from the census tract to each of the clusters). Having computed the probabilities using the trip data, we compute the least square error between the probability transition values of the census tract and clusters. We re-assign the census tract to the cluster with the minimum least squares error. So, formally, for each census tract in $k \in C_2$, we solve for:

$$i^* \in \arg\min_{i \in C_1} \sum_{j \in C_1} (\beta_{k,j} - \alpha_{i,j})^2$$

We assign each census track $k \in C_2$ to the cluster i^*, minimizing the computed error.

After completing this process, we allocate census tracts to clusters and construct the types network and the final probability transition matrix. Nevertheless, we still need to calibrate the size of each cluster (in terms of the number of stations) and the mean demand. The issue is that we are only given the total number of pickups from each cluster, corresponding to $E[D^e]n_e$. We use information from the 2010 U.S. Census (U.S. Census Bureau (2011)) to calibrate the individual parameters. Specifically, we calculate the population density of census tracts by dividing the tract population by the tract land area. We then calculate the weighted density for each cluster by taking the weighted average of tract densities, with weights given by the proportion of pickups happening from that tract. Here, we observe that four of the five clusters have similar population densities, ranging between 4000 and 6000 per square mile, while one cluster is an outlier with a density of close to 22,000 per square mile. Assuming a total of 4500 possible pick-up locations across the city and assigning only two pick-ups per day on average at the lowest density areas (same as our cutoff value for a tract to be included), we assign $E[D^e]$ and n_e values for each of the types/clusters. In the next subsection, we list the parameters obtained through our analysis.

I.11. Resultant Parameters Obtained From Clustering

We identify Census Tracts to their GEOIDs, which are 11-digit numbers highlighting the tract's location. For our experiment, all GEOIDs have the same first seven digits, as all 63 census tracts are in Travis County, Texas (with GEOID 4845300). Below, we provide the census tracts which are placed in each of the five clusters, listed by the last four digits (so having an entry 0700 next to cluster/type 1 indicates that we placed the census tract with GEOID 48453000700 in type 1).

Type 1: 0700, 1100 (2 Tracts)

Type 2: 0306, 0402, 0801, 0802, 0803, 0804, 0901, 0902, 1000, 2109, 2110, 2111, 2304, 2312, 2313, 2314, 2315, 2316, 2317, 2318 (20 Tracts)

Type 3: 1303, 1304, 1305, 1307, 1308, 1401, 1402, 1403, 1901, 2002, 2003, 2004, 2005, 2307, 2308, 2402, 2403, 2409, 2410 (19 Tracts)

Type 4: 0101, 0203, 0204, 0205, 0206, 0302, 0304, 0305, 0307, 0401, 0500, 0601, 0603, 0604, 1503, 1603, 2105 (17 Tracts)

Type 5: 1200, 1602, 1605, 1910, 1911 (5 Tracts)

One interesting observation is that more than 50% of all scooters are picked up from just two tracks: 48453000700 and 48453001100. This is because these two tracts cover most of the core downtown area in Austin, which, as seen in Figure 6, corresponds to a very high level of usage of scooters.

Second, in Table 4, we provide the transition probabilities between the types we formed. The transition probabilities indicate a very imbalanced network, where most picked-up units stay in the same type or are returned to type 1 (consisting of the core downtown area).

Transition Probability	1	2	3	4	5
1	0.74863009	0.08665178	0.09260059	0.03050405	0.0416135
2	0.2735699	0.66770423	0.03596873	0.0122432	0.01051394
3	0.27495372	0.03441474	0.63263634	0.00552335	0.05247185
4	0.09850834	0.01143887	0.00825442	0.86893489	0.01286349
5	0.37813685	0.02482398	0.15047843	0.02491424	0.42164651

Table 4 Transition Probabilities for Formed Types

Type	1	2	3	4	5
Mean Demand	3	3	3	11	2

Table 5 Mean Demand for Formed Types

Type	1	2	3	4	5
Proportion	0.52	0.17	0.2	0.03	0.08

Table 6 Proportion of Each Type

Third, in Table 5, we provide the mean demand at each type. The values are similar, except for type 4. The reason why type 4 has a high mean demand is that it includes several high-density census tracts.

Fourth, in Table 6, we provide the proportion ($\frac{n_e}{n}$) of each type e. As previously discussed, type 1 includes more than 50% of formed stations/pickup locations as it covers the core downtown area where scooter usage is very high.

I.12. Austin Results with Different Penalty Cost Values

This subsection presents the resultant cost values for various penalty costs. The resultant figure can be seen in Figure 10.

Figure 10 closely resembles Figure 2 for good performing policies. Most importantly, we observe that the policies based on the control algorithm perform well for all penalty cost values. For the other policies, we observe that the newsvendor policy is near-optimal for high penalty cost values, and the hybrid no rebalancing policy is near-optimal for low penalty cost values.

I.13. Impact of Mean-Field Policies on Fleet Sizes

As introduced in Section 1, our approach jointly optimizes fleet sizing and rebalancing/recharging. Based on the cost parameters, underlying network structure, and demand distributions, our proposed policies control fleet size by increasing or reducing it to the intended value. An important property of the mean-field-based control policies is that this change is not sudden and is done over multiple periods, allowing customers to deplete and reposition some units. To see how this performs in practice, we track the number of units in circulation under the static control and the resolving control policies for the Austin experiment (with $c_p = 11.3$). The resultant values are provided in Table 7.

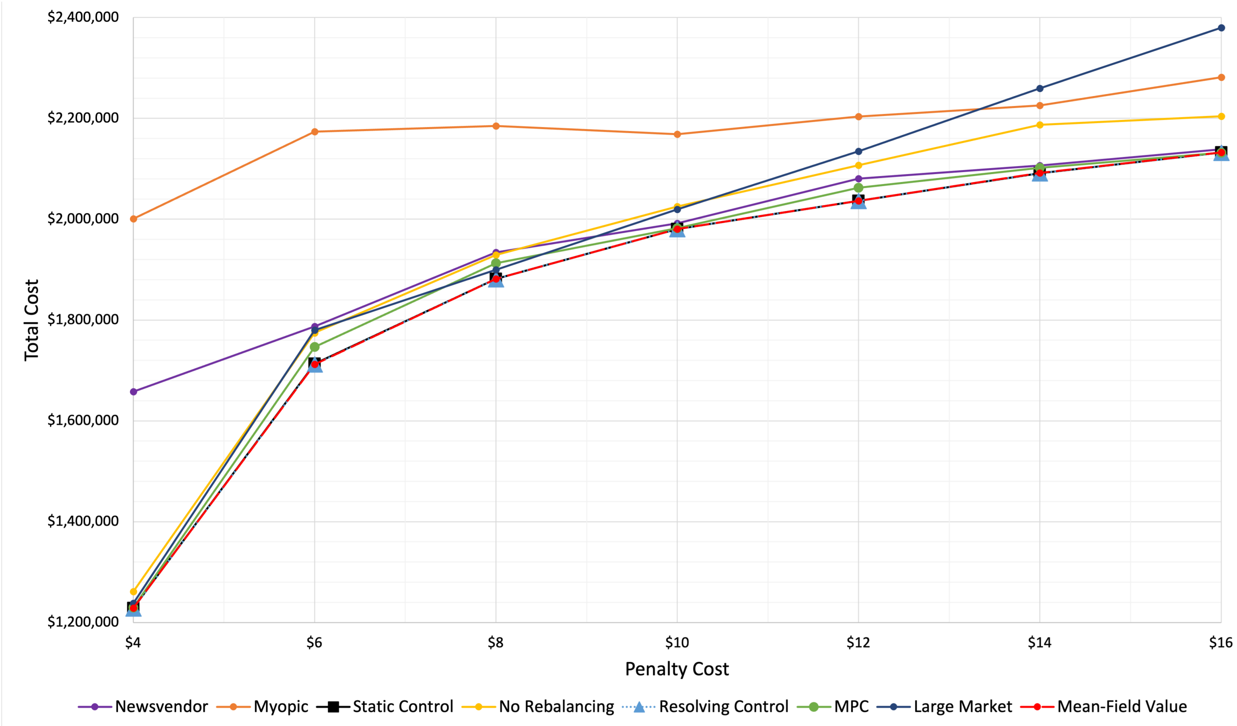


Figure 10 **Policy Costs under Different Penalty Costs**

Period	Static Control	Resolving Control
1	22500	22500
2	19082	19047
3	18882	18868
4	18855	18868
5-	18855	18855

Table 7 **Number of Units in Circulation Under Different Mean-Field Policies**

Table 7 shows that the two policies (and the other benchmark policies) initially have 22500 units in circulation. This value is due to our initial parameter assumption, where we assume that each station initially holds five charged units and zero depleted units. The mean-field based control policies, observing that this initial fleet size is too high, gradually decrease the number of units by sending some depleted units back to the warehouse instead of recharging them. Here, we observe that the resolving policy takes longer than the static control policy. This follows the static control policy structure, which has to take the same mean-field action distribution starting from period 4, resulting in the same fixed fleet size.

The most important observation is that both policies eventually reach the same fleet size and never deviate from this inventory position. This is due to the high cost of sourcing units and both policies solving the same optimization problem when making rebalancing/recharging/sourcing decisions. In our model, it takes 6 dollars to source/withdraw a unit, compared to 5 dollars, which includes recharging a unit and repositioning it to a different type. Consequently, once the two policies reach a fleet size, managing the available fleet is cheaper than introducing additional units.

Appendix J: Supplementary Material for Extensions

This section provides the theoretical analysis and discussions for extending our results to incorporate non-stationary systems with seasonality, multiple charge levels, travel times, and fixed costs.

J.1. Transforming Non-Stationary MDP's through Composite Periods

As highlighted in Subsection 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 G. Under the periodic structure, for any two consecutive periods $t, t+1$:

$$
\begin{aligned}
\hat{c}_3(\hat{\boldsymbol{a}}^t, \hat{\boldsymbol{a}}^{t+1}) = {} & n \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{d=U^{t,e}+1}^{\infty} (d - U^{t,e}) \sum_{b=L^{t,e}}^{U^{t,e}} \hat{a}^t_{e,b} \mathbb{P}\left[b - \min(b, D^{t,e}) + \hat{R}^{t,e} = d\right] \\
& + n \sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{b=0}^{U^{t+1,e}-1} \left| \sum_{d=0}^{b} \Big(\sum_{k=L^{t,e}}^{U^{t,e}} \hat{a}^t_{e,k} \mathbb{P}\left[k - \min(k, D^{t,e}) + \hat{R}^{t,e} = k\right] - \hat{a}^{t+1}_{e,d} \Big) \right| \\
& + n \min_{y^{e_1,e_2}} \sum_{e_1=0}^{2\hat{e}} \sum_{e_2=0, e_2 \neq e_1}^{2\hat{e}} \left(c_{e_1,e_2} - \frac{c_{e_1,e_1}}{2} - \frac{c_{e_2,e_2}}{2} \right) y^{e_1,e_2} \\
\text{s.t.} \quad & \sum_{d=U^{t,e}+1}^{\infty} d \sum_{b=L^{t,e}}^{U^{t,e}} \hat{a}^t_{e,b} \mathbb{P}\left[b - \min(b, D^{t,e}) + \hat{R}^{t,e} = d\right] \\
& + \sum_{d=1}^{U^{t+1,e}} d \Big(\sum_{b=L^{t,e}}^{U^{t,e}} \hat{a}^t_{e,b} \mathbb{P}\left[b - \min(b, D^{t,e}) + \hat{R}^{t,e} = d\right] - \hat{a}^{t+1}_{e,d} \Big) = \sum_{e_1=0}^{2\hat{e}} (y^{e,e_1} - y^{e_1,e}) \qquad \forall e \in [\hat{e}], \\
& \hat{a}^t_{e,r} + (1-q) \sum_{e'=1}^{\hat{e}} p_{e',e} \sum_{b=L^{t,e'}}^{U^{t,e'}} \bar{a}^t_{e',b} \mathbb{E}\left[\min(b, D^{t,e'})\right] - \hat{a}^{t+1}_{e,r} = \sum_{e_1=0}^{2\hat{e}} (y^{e,e_1} - y^{e_1,e}) \qquad \forall e > \hat{e},
\end{aligned}
$$

$$y^{e_1,e_2} \geq 0 \qquad \forall e_1, e_2.$$

We input this rebalancing cost function as well as composite actions to the mean-field model defined in (3.6) to obtain:

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

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 $\hat{c}_2(\breve{\boldsymbol{x}}^t, \hat{\boldsymbol{a}}^t) + \sum_{k=1}^{H-1}\gamma^k \hat{c}_3(\hat{\boldsymbol{a}}^{t+k-1}, \hat{\boldsymbol{a}}^{t+k}) + \sum_{k=0}^{H-1}\gamma^k \left(\hat{N}^{t+k}(\hat{\boldsymbol{a}}^{t+k}) + \hat{y}^{t+k}(\hat{\boldsymbol{a}}^{t+k})\right)$, 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 (J.1) also satisfies the conditions of Proposition H.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:

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

The worst-case bound can be updated as follows:

Theorem J.1. *The optimality gap of the policy $\tilde{\boldsymbol{\pi}}^t$ obtained via* (4.1)-(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

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

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 4.2 and is omitted.

Remark J.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.*

J.2. Extending the Depletion Probability and Multiple Charge Levels

As we stated in Section 6.3, we can extend our model to consider heterogeneous depletion probabilities 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

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

Nevertheless, our methodology for obtaining asymptotic optimality and 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 affect any of the steps.

Second, as stated in Section 6.3, we can extend our model to allow for different recharging costs within the same type. To do this, we will split the depletion probability 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 $x^t_{e,r,1}, ..., x^t_{e,r,M}$, $a^t_{e,r,1}, ..., a^t_{e,r,M}$ 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. Resultant mean-field state transitions for the depleted unit groups can be expressed as

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

where we can show that splitting depleted units arriving at 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, our model does not explicitly define battery charge level as an attribute. 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 considered different products) version of our problem, where different products are available at a station, and arriving customers can substitute between products based on availability.[10]

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 captures the total number of units depleted more accurately. 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).

J.3. Travel Times

As introduced in Section 6.5, our model can be extended to incorporate travel times through the usage of dummy types. To this end, assume that it takes exactly M periods for a unit picked up at type e to be dropped in type e'. Then, we introduce dummy types $e_1, e_2, ..., e_{M-1}$ to extend our model to incorporate multi-period travel (corresponding to regions where the unit must move through to reach the final destination). For these dummy types, we assign:

$$p_{e,e_1} = p_{e,e'}, \tag{J.2}$$

$$p_{e_{i-1},e_i} = 1 \qquad i \in \{2, \ldots, M-1\}, \tag{J.3}$$

$$p_{e_{M-1},e'} = 1, \tag{J.4}$$

$$q_{e,e_1} = 1, \tag{J.5}$$

$$q_{e_{i-1},e_i} = 1 \qquad i \in \{2, \ldots, M-1\}, \tag{J.6}$$

$$q_{e_{M-1},e'} = q_{e,e'}. \tag{J.7}$$

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

Equations (J.2), (J.3),(J.4) allow us to preserve the probability distribution of the underlying network while also ensuring that units are not dropped at another type while in-transit at types $e_1, e_2, ..., e_{M-1}$. Equations (J.5), (J.6), and (J.7), which use the results of Section 6.3 to introduce heterogeneous q values, ensure that units are not depleted while in transit, and the depletion probability is preserved for the final segment of the trip.

In addition to Equations (J.2)-(J.7), we also require that no rebalancing/sourcing occurs for units in transit. To this end, at the dummy types, we first assume that demand is deterministic to ensure that all units travel past the dummy types. Furthermore, we have:

$$n_{e_i} = \max(n_e, n_{e'}) \qquad \forall i \in \{1, \dots, M-1\},$$

$$a^t_{e_i,r} = x^t_{e_i,r} = 0 \qquad \forall i \in \{1, \dots, M-1\}, \tag{J.8}$$

$$a^t_{e_i,d} = x^t_{e_i,d} \qquad \forall i \in \{1, \dots, M-1\}, d \in \{1, \dots, n_{e_i}\}, \tag{J.9}$$

$$D^{e_i} = U^{e_i} \gg U^e \qquad \forall i \in \{1, \dots, M-1\},$$

$$L^{e_i} = 0 \qquad \forall i \in \{1, \dots, M-1\},$$

$$c_{p,e_i} = 0 \qquad \forall i \in \{1, \dots, M-1\}.$$

In the mean-field model, Equations (J.8), (J.9) will correspond to

$$\hat{a}^t_{e_i,r} = \hat{x}^t_{e_i,r} = 0 \qquad \forall i \in \{1, \dots, M-1\},$$

$$\hat{a}^t_{e_i,d} = \hat{x}^t_{e_i,d} \qquad \forall i \in \{1, \dots, M-1\}, d \in \{0, \dots, U^{e_i}\}.$$

By assuming large and deterministic demand at the dummy types, we ensure that all units leave the dummy types every period.

In summary, we use the dummy variable technique for both extensions in Section 6.4 and 6.5. As a result, we do not expand the action space but the state space in both extensions. While this expansion will result in additional memory required for computation, the impact on the computation time is limited. This can be best observed in the Linear Program formulation of the Control Algorithm in Appendix G.[11] Equations (J.8) and (J.9) restrict the inventory position at these dummy stations such that the post-rebalancing inventory equals the pre-rebalancing inventory. We also ensure that all arriving units to a dummy type move to the next dummy type next period through the above constraints. As a result, through Equations (G.5)-(G.10) in

[11] While Appendix G assumes $q = 0$, our following analysis can be applied to the $q > 0$ case, where the inequalities are non-linear.

the LP, we can deduce that cost-minimizing y values observe no inflow/outflow in the dummy types. As for the remaining decision variables, through (G.12)-(G.17), we have that the only feasible value for the related s, z, l variables is 0. Consequently, while there can be a slight increase in the computation time, specifically due to adding additional constraints and decision variables, our assumptions ensure that these constraints and decision variables are redundant and dropped in the pre-processing step.[12]

J.4. Fixed Costs

As introduced in Section 6.6, we will show that we can extend our results to incorporate the fixed costs of visiting a station. To this end, the requisite function for the rebalancing cost, which includes this fixed cost component, is given by:

$$\begin{aligned}
\hat{c}(\hat{\boldsymbol{x}}^t, \hat{\boldsymbol{a}}^t) = n\sum_{e=1}^{\hat{e}} c_{f,e} \sum_{b=0}^{\infty} \frac{1}{2}\left|\hat{x}^t_{e,b} - \hat{a}^t_{e,b}\right| + n\sum_{e=1}^{\hat{e}} \frac{c_{e,e}}{2} \sum_{b=0}^{\infty}\left|\sum_{d=0}^{b}(\hat{x}^t_{e,d} - \hat{a}^t_{e,d})\right| \\
+ n \min_{y^{e_1,e_2}} \sum_{e_1=0}^{2\hat{e}} \sum_{e_2=0, e_2\neq e_1}^{2\hat{e}} \left(c_{e_1,e_2} - \frac{c_{e_1,e_1}}{2} - \frac{c_{e_2,e_2}}{2}\right) y^{e_1,e_2} \qquad \text{(J.10)}\\
\text{s.t.} \quad \sum_{b=1}^{\infty} b(\hat{x}^t_{e,b} - \hat{a}^t_{e,b}) = \sum_{e_1=0}^{2\hat{e}} (y^{e,e_1} - y^{e_1,e}) \qquad \forall e \in [\hat{e}], \\
\hat{x}^t_{e,r} - \hat{a}^t_{e,r} = \sum_{e_1=0}^{2\hat{e}} (y^{e,e_1} - y^{e_1,e}) \qquad \forall \hat{e} < e, \\
y^{e_1,e_2} \geq 0 \qquad \forall e_1, e_2.
\end{aligned}$$

To see why the above is true, we first need to modify the function h (previously defined in (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, which incorporates the added fixed cost component, is given by:

$$\begin{aligned}
h(\bar{\boldsymbol{a}}) \in \arg\min_{\boldsymbol{a}\in\boldsymbol{A}} \sum_{e=1}^{\hat{e}} \sum_{i=1}^{n_e} c_{f,e} \mathbb{I}\left\{x^t_{e,i} \neq a^t_{e,i}\right\} + \sum_{e=1}^{\hat{e}} \sum_{i=1}^{n_e} \left|a^t_{e,i} - x^t_{e,i}\right| + \sum_{e=1}^{\hat{e}} \left|a^t_{e,r} - x^t_{e,r}\right| \qquad \text{(J.11)}\\
\text{s.t.} \quad \frac{\sum_{i=1}^{n_e} \mathbb{I}\left\{a^t_{e,i} = d\right\}}{n} = \bar{a}^t_{e,d} \quad \forall e, d, \\
a^t_{e,r} = n\bar{a}^t_{e,r} \quad \forall e.
\end{aligned}$$

We want to show that there exists an ordering of $\boldsymbol{a}$ which satisfies the marginals $\bar{\boldsymbol{a}}^t$ while also minimizing the total fixed cost. Before introducing the new ordering, we will prove that the total variation distance

[12] Through these extensions, the optimal actions at non-dummy types may change (units are held for multiple periods meaning higher holding costs).

component (first term) of (J.10) is a function of the number of distinct elements of $\boldsymbol{x}^t$ and $\boldsymbol{a}^t$ at each type e. Formally, we prove:

Lemma J.3. *Let* $\hat{a}^t_{e,d} = \frac{\sum_{i=1}^{n_e} \mathbb{I}\left\{a^t_{e,i}=d\right\}}{n}$ *and* $\hat{x}^t_{e,d} = \frac{\sum_{i=1}^{n_e} \mathbb{I}\left\{x^t_{e,i}=d\right\}}{n}$. *Then, there exists an ordering of the vector* $\boldsymbol{a}^t$ *such that:*

$$n\sum_{e=1}^{\hat{e}} c_{f,e}\sum_{b=0}^{\infty}\frac{1}{2}\left|\hat{x}^t_{e,b}-\hat{a}^t_{e,b}\right| = \min_{\boldsymbol{a}\in\boldsymbol{A}}\sum_{e=1}^{\hat{e}}\sum_{i=1}^{n_e} c_{f,e}\mathbb{I}\left\{x^t_{e,i}\neq a^t_{e,i}\right\}.$$

Proof.

$$n\sum_{e=1}^{\hat{e}} c_{f,e}\sum_{b=0}^{\infty}\frac{1}{2}\left|\hat{x}^t_{e,b}-\hat{a}^t_{e,b}\right| = \min_{\boldsymbol{a}\in\boldsymbol{A}}\sum_{e=1}^{\hat{e}}\sum_{i=1}^{n_e} c_{f,e}\mathbb{I}\left\{x^t_{e,i}\neq a^t_{e,i}\right\}$$

$$\sum_{e=1}^{\hat{e}} c_{f,e}\sum_{b=0}^{\infty}\frac{1}{2}\left|\sum_{i=1}^{n_e}\mathbb{I}\left\{x^t_{e,i}=b\right\}-\mathbb{I}\left\{a^t_{e,i}=b\right\}\right| = \min_{\boldsymbol{a}\in\boldsymbol{A}}\sum_{e=1}^{\hat{e}}\sum_{i=1}^{n_e} c_{f,e}\mathbb{I}\left\{x^t_{e,i}\neq a^t_{e,i}\right\}.$$

We will show that this equality holds for each type e with

$$\sum_{b=0}^{\infty}\frac{1}{2}\left|\sum_{i=1}^{n_e}\mathbb{I}\left\{x^t_{e,i}=b\right\}-\mathbb{I}\left\{a^t_{e,i}=b\right\}\right| = \min_{\boldsymbol{a}\in\boldsymbol{A}}\sum_{i=1}^{n_e}\mathbb{I}\left\{x^t_{e,i}\neq a^t_{e,i}\right\}.$$

We first observe that the minimum expression allows us to shuffle the elements of $\boldsymbol{a}^t_e$ (as all shuffling of elements will give the same marginals). Then, the right-hand side of the above equation, $\min_{\boldsymbol{a}\in\boldsymbol{A}}\sum_{i=1}^{n_e}\mathbb{I}\left\{x^t_{e,i}\neq a^t_{e,i}\right\}$, calculates the number of distinct elements between the sets $\{x^t_{e,1},\cdots,x^t_{e,n_e}\}$ and $\{a^t_{e,1},\cdots,a^t_{e,n_e}\}$. Now, assume that there are $n_e\geq C\geq 0$ distinct elements, meaning n_e-C elements are common in both sets. For the left-hand side, we observe that for the common n_e-C elements, we have that $\mathbb{I}\left\{x^t_{e,i}=b\right\}-\mathbb{I}\left\{a^t_{e,i}=b\right\}=0$. The difference is that for the C distinct elements, we will observe $\mathbb{I}\left\{x^t_{e,i}=b\right\}-\mathbb{I}\left\{a^t_{e,i}=b\right\}=1$ twice; once for the distinct element in x and once for the distinct element in c. Therefore, we have that $\sum_{b=0}^{\infty}\left|\sum_{i=1}^{n_e}\mathbb{I}\left\{x^t_{e,i}=b\right\}-\mathbb{I}\left\{a^t_{e,i}=b\right\}\right|=2C$, and the Lemma holds. □

Lemma J.3 shows that when modifying the h function to move from the current to target inventory position by changing inventory at as few stations as possible, the requisite function for the rebalancing cost gives us the same cost as the modified original rebalancing cost. This, however, presents a new issue. 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. With the existence of fixed costs, we need to introduce a new matching that does that while ensuring that inventory is changed in as few stations as possible.

To achieve this, we consider the following coupling: Focus on stations of type e. Let k^e_b be the smaller of $n\bar{a}^e_b$ and the number of type e stations with $x_{e,i}=b$. Then k^e_b 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 (J.10), while not changing the remaining cost terms.

One last issue is that in the proof of Proposition 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, our new rebalancing cost is not equivalent to a Wasserstein distance, so we cannot directly utilize the CLT for Wasserstein distances. To resolve this issue, we first prove the following lemma:

Lemma J.4. *Fixed cost component of the mean-field rebalancing cost satisfies*

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

Proof.

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

Using this lemma, we can show that the fixed cost component can also be upper bounded 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 $q = 0$).

J.4.1. Impact of Adding Fixed Costs Due to the great importance of fixed costs for some of our applications, we conduct a separate numerical experiment incorporating 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.[13] The results are shown in Figure 11.

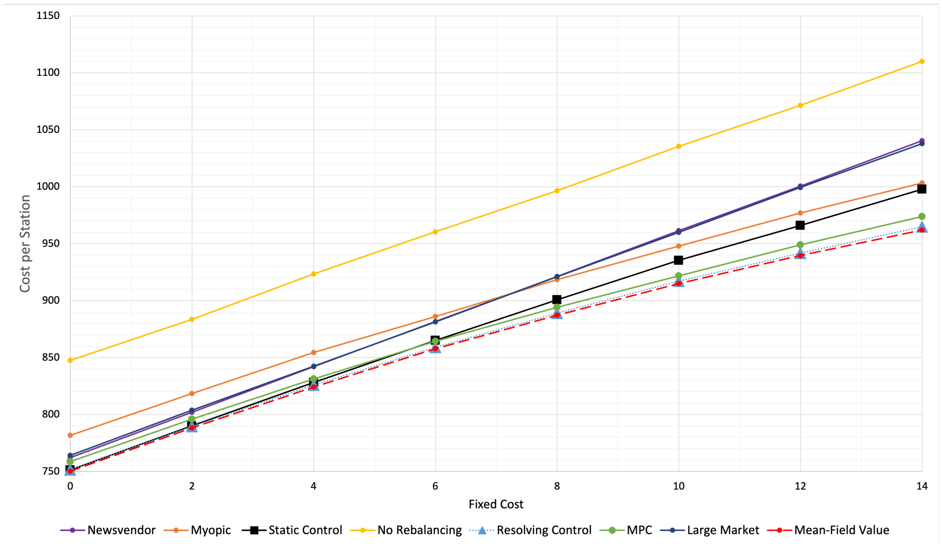


Figure 11 Impact of Adding Fixed Costs on the Performance of Policies

Through Figure 11, we observe that as fixed costs increase, the performance of the static control policy decreases compared to re-solving-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). Resolving control can adapt to the demand realization and thus is near-optimal for all fixed costs, whereas 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 J.5. *The main managerial implication is that re-solving policies are 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 on initial inventory decreases) or if there are many stations (as state transitions become closer to what static policies expect).*

[13] Due to the large market policy structure, incorporating the indicator function providing the fixed costs requires adding an excessive number of binary variables.

References

Akturk D (2022) Managing inventory in a network: Performance bounds for simple policies. *Operations Research Letters* 50(3):315–321.

Benjaafar S, Jiang D, Li X, Li X (2022) Dynamic inventory repositioning in on-demand rental networks. *Management Science* 68(11):7861–7878.

Bobkov S, Ledoux M (2019) *One-Dimensional Empirical Measures, Order Statistics, and Kantorovich Transport Distances*. Memoirs of the American Mathematical Society (American Mathematical Society).

Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press).

City of Austin Transportation Department (2022) Shared micromobility vehicle trips. URL `http://dx.doi.org/10.26000/030.000003`, accessed on May 21, 2022.

City of Los Angeles Department of Transportation (2018) Dockless on-demand personal mobility conditional permit. URL `http://clkrep.lacity.org/onlinedocs/2017/17-1125_misc_11-08-2018.pdf`.

del Barrio E, Giné E, Matrán C (1999) Central limit theorems for the wasserstein distance between the empirical and the true distributions. *The Annals of Probability* 27(2):1009–1071.

Flynn J (1979) Steady state policies for deterministic dynamic programs. *SIAM Journal on Applied Mathematics* 37(1):128–147.

Griswold A (2018) Simple math shows how scooters could make big money. URL `https://qz.com/1325064/scooters-might-actually-have-good-unit-economics/`, accessed on May 4, 2024.

Hayes N (2022) How long do electric scooters last? URL `https://scooter.guide/how-long-do-electric-scooters-last/`, accessed on May 4, 2024.

He L, Hu Z, Zhang M (2020) Robust repositioning for vehicle sharing. *Manufacturing & Service Operations Management* 22(2):241–256.

Helling B (2022) The comprehensive guide to bird scooter charging. URL `https://www.ridester.com/bird-scooter-charger/`, accessed on May 21, 2022.

Holyoak B (2021) I made $1,000 charging lime scooters — an honest review. URL `https://enduringfinances.com/i-made-1000-charging-lime-scooters-an-honest-review/`, accessed on May 4, 2024.

Kabra A, Belavina E, Girotra K (2020) Bike-share systems: Accessibility and availability. *Management Science* 66(9):3803–3824.

Ross S (2010) *A First Course in Probability* (Prentice Hall), 8th edition.

Somerville P (2021) Segway ninebot es4 review: With dual suspension, is it better than the ninebot max? URL `https://electric-scooter.guide/reviews/segway-ninebot-es4-review/`, accessed on May 4, 2024.

Strobel P (2021) What kind of e-scooters do bird & lime use? URL `https://eridehero.com/what-scooters-do-bird-lime-use/`, accessed on May 21, 2022.

US Census Bureau (2011) 2010 census. U.S. Department of Commerce, accessed on May 4, 2024.