Competitive Multi-Operator Reinforcement Learning for Joint Pricing and Fleet Rebalancing in Autonomous Mobility-on-Demand Systems

Author: Emil Kragh Toft (s233791)
Supervisors: Carolin Schmidt and Filipe Rodrigues

Abstract—**Autonomous Mobility-on-Demand (AMoD) systems promise to revolutionize urban transportation by eliminating driver costs and providing affordable on-demand services. However, realistic AMoD markets will be competitive, with multiple operators competing for passengers through strategic pricing and fleet deployment. Existing reinforcement learning approaches for AMoD control focus on single-operator settings and fail to capture competitive market dynamics. We introduce a multi-operator reinforcement learning framework where two operators simultaneously learn joint pricing and fleet rebalancing policies while competing for demand. By integrating discrete choice theory, passenger allocation emerges endogenously from utility-maximizing decisions based on fare price, travel time, and passenger wages. The framework incorporates wage-sensitive demand modeling, enabling pricing strategies to adapt to regional economic differences. Through experiments on real-world data from multiple cities, we demonstrate that competitive dynamics fundamentally alter learned policies compared to monopolistic settings. Operators develop sophisticated strategic behaviors, with competition leading to lower prices and distinct fleet positioning patterns. This work provides insights into competitive autonomous mobility markets and contributes to platform design and policy decisions for multi-operator AMoD systems.**

Index Terms—**Autonomous Mobility-on-Demand, Multi-Operator Reinforcement Learning, Competitive Pricing, Fleet Rebalancing, Discrete Choice Models, Graph Convolutional Networks**

I. INTRODUCTION

URBAN mobility systems worldwide face mounting pressures from population growth, urbanization, and evolving consumer expectations. Traditional transportation modes—private vehicles and public transit—struggle to efficiently meet the diverse mobility needs of modern cities. Private car ownership contributes to congestion, parking scarcity, and emissions, while public transit often lacks the flexibility and coverage demanded by passengers. This gap has fueled the rapid growth of ride-hailing services over the past decade, fundamentally transforming urban mobility patterns [1].

The advent of autonomous vehicle technology presents an opportunity to further revolutionize urban transportation through Autonomous Mobility-on-Demand (AMoD) systems. By eliminating driver compensation, one of the largest expenses for ride-hailing platforms, AMoD could provide affordable, convenient, and sustainable door-to-door transportation at scale [2]. However, unlike traditional public transit systems that are typically operated as monopolies or regulated utilities, AMoD markets are likely to be competitive, with multiple operators deploying fleets and competing for passengers.

In competitive AMoD markets, operators must make strategic decisions about pricing and fleet deployment while anticipating and responding to competitor actions. A price reduction by one operator affects its own demand and simultaneously diverts passengers from competitors. Fleet positioning decisions similarly influence service quality and market share dynamics. These strategic interactions create a complex game-theoretic environment where optimal policies depend on competitor behavior.

Reinforcement Learning (RL) has emerged as a powerful approach for sequential decision-making in complex environments, demonstrating success in games, robotics, and various control problems. For AMoD systems, RL offers the advantage of learning near-optimal policies directly from data without requiring perfect models of demand patterns, traffic conditions, or competitor strategies. Existing research on RL for AMoD systems predominantly focuses on centralized, single-operator settings where the objective is to maximize social welfare or system efficiency through optimal vehicle rebalancing [3]–[7], pricing [8]–[10], or both [11]. While these studies provide valuable insights into operational challenges, they do not address the competitive dynamics inherent in realistic multi-operator market structures.

In this work, we develop a competitive multi-operator reinforcement learning framework for joint pricing and fleet rebalancing in AMoD systems. The approach models a realistic scenario where two operators simultaneously learn and adapt their strategies while competing for passengers. We incorporate a discrete choice model that captures how passengers select between operators based on price, travel time, and passenger wages, reflecting real behavioral patterns observed in transportation economics. This choice model creates a dynamic coupling between operators' actions and market outcomes, enabling the emergence of complex competitive strategies.

The main contributions of this paper are as follows:

- We formulate a competitive, dual-operator AMoD control problem in which two independent operators jointly learn pricing and fleet rebalancing policies using reinforcement learning, extending prior joint-control frameworks beyond the single-operator setting.
- We integrate a passenger choice mechanism into the learning loop, allowing demand allocation between competing operators to emerge endogenously from operator actions rather than being imposed exogenously.
- We provide an empirical analysis of how competition alters learned strategies, service quality, and market efficiency relative to monopolistic control through experiments on real-world data from multiple cities.

The remainder of this paper is organized as follows. Section II reviews related work on AMoD rebalancing and pricing

problems. Section III provides a brief theoretical overview of reinforcement learning and graph neural networks. Section IV presents the AMoD control problem formulation in both single-operator and competitive dual-operator contexts. Section V presents our experimental results and analysis, demonstrating how the framework learns control policies in a competitive dual-operator setting and how these policies differ from those in single-operator scenarios. Section VI concludes with a discussion of future research directions in competitive autonomous mobility systems.

II. LITERATURE REVIEW

Shared mobility and on-demand transportation systems face persistent operational challenges arising from the uneven spatial and temporal distribution of vehicles. Existing literature primarily addresses these inefficiencies through rebalancing strategies and dynamic pricing mechanisms designed to better align supply with demand. Additionally, an emerging body of work examines these issues in multi-operator settings where competition between service providers influences pricing, fleet allocation, and overall system performance.

A. Rebalancing

Early foundational work by Zhang and Pavone [12], [13] modeled Autonomous MoD systems as closed Jackson queueing networks, establishing a mathematical framework for understanding system dynamics and flow conservation constraints. Building on this foundation, Model Predictive Control (MPC) approaches have been extensively applied for real-time fleet optimization [14]–[17]. These methods partition cities into zones and employ discrete-time dynamical models to optimize fleet distribution. Tsao et al. [16] proposed scalable predictive control frameworks, and Warrington et al. [15] developed two-stage stochastic approximation methods that explicitly account for demand uncertainty. While these optimization approaches provide strong theoretical guarantees, they face practical challenges when demand patterns deviate from assumed distributions or when computational requirements exceed real-time constraints [17].

Recognizing that demand prediction is inherently uncertain, researchers have developed robust optimization frameworks that explicitly account for forecast errors and stochastic patterns. Guo et al. [18] introduced robust Matching-Integrated Vehicle Rebalancing (MIVR) models that consider sets of possible demand realizations rather than single forecasts. Data-driven predictive prescription approaches combine real-time forecasting with stochastic programming [19], generating rebalancing actions that explicitly incorporate prediction uncertainty. The integration of robust optimization with MPC has proven particularly effective for handling stochastic urban mobility patterns, enabling controllers to hedge against worst-case demand fluctuations while avoiding overly conservative solutions [17].

Reinforcement learning represents a paradigm shift from model-based optimization to data-driven policy learning. Unlike optimization approaches requiring explicit mathematical models, RL methods learn rebalancing policies directly from observed state transitions and rewards. Early deep RL applications demonstrated passenger waiting time reductions compared to heuristic approaches [20], while Wen et al. [6] introduced deep Q-network methods achieving near-optimal performance with significantly reduced computational requirements. The key advantage of RL lies in its ability to adapt to complex spatiotemporal dynamics without explicit demand models. However, RL faces scalability challenges in large urban networks where state and action spaces grow exponentially with the number of zones.

To address these scalability challenges while leveraging the network structure of urban transportation systems, Gammelli et al. [3] proposed a Graph Convolutional Network (GCN) reinforcement learning framework for autonomous mobility-on-demand systems. This approach models transportation networks as graphs with nodes representing city areas and edges representing connectivity, learning node-wise policies by aggregating information from neighboring nodes through message-passing operations. The GCN architecture captures spatial dependencies in demand and supply patterns, enabling decisions informed by both local conditions and connected area states. The architecture naturally handles expanding service areas and topology changes, making it particularly suitable for real-world deployment. The dual-operator framework presented in this paper builds on this work.

B. Dynamic Pricing and Joint Policy Optimization

While rebalancing addresses supply-demand imbalances by relocating vehicles, dynamic pricing offers a complementary mechanism by influencing demand patterns themselves. Early research on dynamic pricing in MoD systems focused on profit maximization and congestion management, with most approaches relying on equilibrium-based models where system dynamics are modeled as constraints in an optimization framework [21]–[24]. In the equilibrium-based approach, the pricing policy affects the equilibrium, which is subsequently used to calculate revenue. In operations research-based works, demand is typically considered elastic, and the pricing decision is shaped by optimization frameworks and the constraint set [25], [26]. Pricing strategies can reduce demand in oversaturated areas while stimulating requests in underutilized zones, effectively reshaping the spatial distribution of trip requests to better match vehicle availability. However, implementing pricing or rebalancing in isolation can be overly restrictive and fails to account for emerging synergies between the two strategies [11], [27].

The recognition that rebalancing and pricing interact in complex ways has motivated joint optimization frameworks. Proper pricing influences where and when passengers request rides, affecting where vehicles are needed; conversely, effective rebalancing reduces wait times and enables different pricing strategies. Bilevel optimization frameworks have emerged as a principled approach for joint decision-making, formulating operator decisions (pricing, vehicle relocation) at the upper level while modeling passenger responses (demand, route choice) at the lower level [28]. Li et al. [11] developed a reinforcement learning approach for learning joint rebalancing

and dynamic pricing policies for autonomous mobility-on-demand that integrates GCN architectures with hierarchical policy optimization. This framework simultaneously learns rebalancing and pricing policies, leveraging GCN spatial reasoning capabilities to scale across large urban networks. The approach employs a bilevel structure to improve computational tractability: the upper-level GCN policy determines target vehicle distributions per node and node-based pricing decisions, which are then used by a lower-level optimization routine, typically a linear program, to compute minimum-cost origin-destination (OD) rebalancing flows to achieve those target distributions. This hierarchical architecture enables the model to capture both spatial dependencies through GCN message-passing and strategic pricing-rebalancing interactions through joint policy optimization.

Motivated by recent work on joint policy optimization for AMoD systems [11], this work addresses the critical gap of explicitly modeling price-responsive demand through the integration of discrete choice models. In this approach, incoming ride requests are distributed to available operators based on a choice model where prices directly affect the utility of each service option. Passengers evaluate each service option according to a utility function that incorporates price, wage, and travel time. If no option exceeds their reservation utility threshold, they reject all alternatives and exit the system. This choice-based demand adjustment mechanism enables the system to capture elastic demand responses to pricing decisions, providing a more realistic representation of passenger's behavior and allowing the joint rebalancing-pricing policy to actively shape demand patterns rather than treating demand as exogenous. Additionally, this work explores how salary levels in different regions affect passenger choices, capturing heterogeneous price sensitivity across geographic areas and enabling more spatially nuanced pricing strategies that account for local economic conditions.

C. Multi-Operator Environments

While much of the literature focuses on single-operator optimization, real-world mobility markets increasingly feature multiple competing platform operators, each managing its own vehicle fleet and simultaneously optimizing pricing and rebalancing strategies. This competitive setting introduces fundamentally different dynamics compared to centralized optimization, as operators must account for strategic interactions with rivals while pursuing their individual objectives.

Multi-operator competition in ride-hailing and autonomous mobility-on-demand systems has been studied through game-theoretic frameworks that model the strategic interactions between platforms competing for both drivers and passengers. Yang and Ramezani [29] analyze intraday competition in duopoly ride-hailing markets, examining how platforms adjust pricing and service strategies throughout the day in response to competitor actions and fluctuating demand. Research on competing AMoD operators demonstrates that fragmented markets with multiple independent operators can reduce pooling efficiency and overall system performance compared to monopolistic or regulated scenarios [30], [31]. Game-theoretic models reveal that platforms engage in noncooperative pricing games where passenger fares must be strategically adjusted, as passengers can easily switch between platforms [32]. Studies examining quality-of-service competition show that platforms must balance pricing strategies with matching efficiency and wait times, as these factors jointly determine market share in competitive environments [33].

This work addresses joint rebalancing-pricing optimization in a multi-operator competitive setting, where two competing operators simultaneously optimize their strategies while accounting for competitor actions and passenger responses. This introduces fundamentally different challenges compared to single-operator settings [11]: rather than focusing purely on coordination and optimality, multi-operator environments require modeling strategic interactions between operators with potentially conflicting objectives. These game-theoretic considerations and competitive dynamics fundamentally alter the optimization landscape, requiring operators to anticipate and respond to rival strategies while managing their own fleets and pricing policies.

III. PRELIMINARIES

A. Reinforcement Learning and the A2C Algorithm

The problem of optimizing fleet management and pricing is cast as a sequential decision-making task. We formalize the environment as a Markov Decision Process (MDP) denoted by the tuple $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, d_0, \mathcal{R}, \gamma)$. The state space $\mathcal{S}$ consists of states $\mathbf{s} \in \mathcal{S}$ that capture the global configuration of the transportation network. Similarly, $\mathcal{A}$ represents the set of feasible pricing and rebalancing actions $\mathbf{a} \in \mathcal{A}$. The system evolves according to the transition dynamics $P(\mathbf{s}_{t+1}|\mathbf{s}_t, \mathbf{a}_t)$, starting from an initial distribution $d_0(\mathbf{s}_0)$. At each step, the operator receives a scalar reward $r(\mathbf{s}_t, \mathbf{a}_t) \in \mathcal{R}$, with future rewards discounted by a factor $\gamma \in (0, 1]$.

The operator acts according to a stochastic policy $\pi(\mathbf{a}_t|\mathbf{s}_t)$, which maps the current system configuration to a probability distribution over actions. This policy induces a distribution $p_\pi(\tau)$ over trajectories $\tau = (\mathbf{s}_0, \mathbf{a}_0, \ldots, \mathbf{s}_H, \mathbf{a}_H)$. The optimization goal is to find a policy that maximizes the expected discounted return over the episode horizon H:

$$J(\pi) = \mathbb{E}_{\tau \sim p_\pi(\tau)} \left[\sum_{t=0}^{H} \gamma^t r(\mathbf{s}_t, \mathbf{a}_t) \right] \quad (1)$$

To tackle the high-dimensional nature of the AMoD control problem, we employ the Advantage Actor-Critic (A2C) algorithm. This approach utilizes two distinct GCNs to approximate the value and policy functions, allowing the operator to leverage the topological structure of the transportation network.

The critic network, parameterized by ϕ, serves to reduce the variance of the learning process by estimating the state-value function $V_\phi(\mathbf{s}_t)$. This function represents the expected return from state $\mathbf{s}_t$ when following the current policy:

$$V_\phi(\mathbf{s}_t) = \mathbb{E}_{\tau \sim p_\pi(\tau|\mathbf{s}_t)} \left[\sum_{t'=t}^{H} \gamma^{t'-t} r(\mathbf{s}_{t'}, \mathbf{a}_{t'}) \right] \quad (2)$$

During training, the critic parameters are updated to minimize the mean squared error between the estimated value $V_\phi(\mathbf{s}_t)$ and the actual realized return $R_t = \sum_{t'=t}^{H} \gamma^{t'-t} r(\mathbf{s}_{t'}, \mathbf{a}_{t'})$:

$$L_{critic}(\phi) = \frac{1}{2} \left(R_t - V_\phi(\mathbf{s}_t) \right)^2 \tag{3}$$

The actor network, parameterized by θ, defines the policy $\pi_\theta(\mathbf{a}_t | \mathbf{s}_t)$. It is optimized via gradient ascent on the objective $J(\pi_\theta)$. To compute the gradient, we utilize the advantage estimator $\hat{A}(\mathbf{s}_t, \mathbf{a}_t)$, which uses the critic's value estimate as a baseline $b(\mathbf{s}_t) = V_\phi(\mathbf{s}_t)$ to determine the relative quality of the chosen action:

$$\hat{A}(\mathbf{s}_t, \mathbf{a}_t) = \sum_{t'=t}^{H} \gamma^{t'-t} r(\mathbf{s}_{t'}, \mathbf{a}_{t'}) - b(\mathbf{s}_t) \tag{4}$$

The policy gradient is then estimated by averaging over the trajectory:

$$\nabla_\theta J(\pi_\theta) = \mathbb{E}_{\tau \sim p_{\pi_\theta}(\tau)} \left[\sum_{t=0}^{H} \gamma^t \nabla_\theta \log \pi_\theta(\mathbf{a}_t | \mathbf{s}_t) \hat{A}(\mathbf{s}_t, \mathbf{a}_t) \right] \tag{5}$$

B. Graph Convolutional Networks

Recent advances in deep learning have demonstrated strong performance in domains where data exhibit compositional and spatial structure. Convolutional neural networks (CNNs), originally introduced for visual pattern recognition, exploit local connectivity and parameter sharing to efficiently learn from grid-structured data [34]. However, many real-world systems, including urban transportation networks, are more naturally represented as graphs rather than regular grids. In such settings, the learning architecture must respect the fact that the indexing of nodes is arbitrary. In particular, if the nodes of a graph are relabeled, the output of the model should remain unchanged. This property, known as permutation invariance, is essential in transportation networks, where decisions should depend on spatial relationships and node attributes rather than on an imposed ordering of regions. GCNs address this challenge by extending the notion of convolution to non-Euclidean domains [35].

Let the transportation system be modeled as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_i\}_{i=1}^{N}$ denotes the set of nodes corresponding to spatial regions or stations, and $\mathcal{E}$ denotes the set of edges encoding travel connectivity. Each node v_i is associated with a feature vector $\mathbf{x}_i \in \mathbb{R}^D$, and the node features are collected in a matrix $\mathbf{X} \in \mathbb{R}^{N \times D}$. To learn a permutation-invariant representation of the network, we employ a GCN, whose layer-wise propagation rule is given by

$$\mathbf{H}^{(l+1)} = \sigma\left(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{H}^{(l)} \mathbf{W}^{(l)} \right), \tag{6}$$

where $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ is the adjacency matrix with self-loops, $\tilde{\mathbf{D}}$ is the corresponding diagonal degree matrix, $\mathbf{W}^{(l)}$ is a trainable weight matrix, $\sigma(\cdot)$ is a nonlinear activation function, and $\mathbf{H}^{(0)} = \mathbf{X}$. This propagation mechanism aggregates information from neighboring nodes through a shared local filter, ensuring that the update of each node depends only on its local neighborhood and not on the ordering of nodes in $\mathcal{V}$. The resulting node embeddings capture spatial correlations such as demand imbalances and vehicle movements across regions and provide a compact, permutation-invariant state representation suitable for integration with reinforcement learning frameworks for vehicle rebalancing and dynamic pricing in city-scale mobility-on-demand systems [36].

IV. MULTI-OPERATOR AMOD CONTROL

We represent the AMoD environment as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the vertex set $\mathcal{V}$ corresponds to N_v spatial regions, each centered around a designated station for pickups and drop-offs. The framework considers a dual-operator architecture, where the total fleet is partitioned into two distinct sets of M_0 and M_1 autonomous vehicles, each managed by its own independent operator. These operators act within a synchronized discrete time horizon $\mathcal{T} = \{1, \ldots, T\}$ of interval length ΔT. Movement between regions i and j follows the shortest path, requiring $\tau_{ij} \in \mathbb{Z}_+$ time steps and incurring a time-variant operational cost c_{ij}^t for the respective fleet.

The available supply at any station i is defined by the local density of idle vehicles from each fleet, denoted as $m_{i,0}^t$ and $m_{i,1}^t$. On the demand side, potential passengers arrive at origin i and evaluate the available service providers—including both AMoD operators and an alternative transportation mode—using a discrete choice model. Specifically, a utility value U is calculated for each option as a function of the trip price, the passenger's salary, and the estimated travel time. These utilities are then mapped to a probability distribution, from which the passenger's final choice is sampled via a categorical distribution. Once a provider is selected, passengers enter a first-come, first-served (FCFS) queue. To reflect realistic behavior, we impose a maximum waiting threshold ω; if the assigned operator cannot provide a vehicle within this window, the passenger exits the system, representing a loss of potential demand.

Based on this structural foundation, the subsequent sections formalize the interaction between the two operators. We first cast the pricing and rebalancing task as a Markov Decision Process (MDP) and then detail a GCN architecture designed to handle the spatial dependencies inherent in the dual-operator learning process.

A. Three-step Control Architecture

In this paper, we study an explicitly dual-operator AMoD setting with two independent operators, each seeking to maximize its own profit while interacting with the same environment. The control architecture follows the bi-level formulation introduced in [3] but is adapted here to a competitive setting in which each operator controls its own fleet. The bi-level formulation consists of: a first level where the actor outputs the desired share of idle vehicles per region, and a second level where a minimal-cost rebalancing problem determines the number of vehicles to rebalance from and to each region to achieve the desired distribution. In the implementation, this
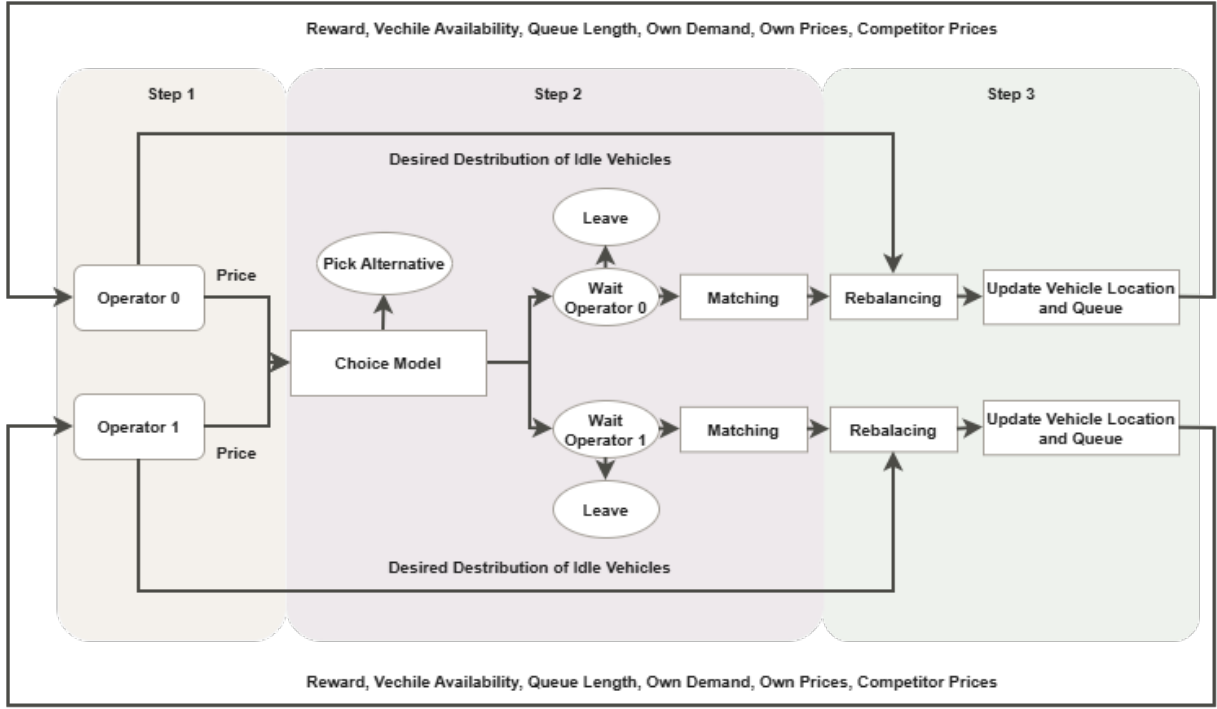
Fig. 1. Three-step policy loop for dual-operator AMoD control. Step 1: operators formulate pricing and desired idle-vehicle distribution policies. Step 2: passenger assignment via choice model, queueing, and matching. Step 3: idle-vehicle rebalancing and update of vehicle positions and queues.

bi-level approach induces a three-step control architecture, illustrated in Fig. 1.

At each time step t, the policy first produces a joint control for each operator consisting of a node-based pricing decision and a node-based desired idle-vehicle distribution (first level). The node-level price scalars are transformed into OD-level prices by multiplying them with OD base prices estimated from historical data, yielding the OD fares used in the environment. These OD-based prices are then passed to the environment, where the stochastic choice model generates passenger requests and assigns them to one of the two operators via the underlying price-dependent choice model, resulting in operator-specific passenger streams. The generated passengers enter the queue at their corresponding departure region and wait to be served by vehicles of the chosen operator; service is restricted to vehicles located in the same region.

In the second step, passengers wait subject to a maximum waiting time, assumed homogeneous across the system, and vehicles are matched to waiting passengers in a first-come, first-served manner. Following matching, the node-based desired distribution produced by each operator's policy is mapped to executable rebalancing actions by solving a minimal rebalancing-cost problem, constrained by the desired vehicle distribution (second level), which returns a set of rebalancing flows for each operator. The formulation of the minimal rebalancing-cost problem is presented in Appendix A. The rebalancing flows are executed separately for each operator, and the vehicle states and queues are updated accordingly. The system then transitions to the next time step, and the new state together with the per-operator rewards produced by the

transition are returned to the two operators.

B. The AMoD Control Problem as an MDP

As discussed in Section III, the control problem of an AMoD system can be formulated as a Markov Decision Process (MDP). In this section, we describe in detail the four fundamental components of the MDP: the state space $\mathcal{S}$, the action space $\mathcal{A}$, the system dynamics $\mathcal{P}$, and the reward function $\mathcal{R}$. We specify how each element is defined in the context of multi-operator AMoD control.

State-space ($\mathcal{S}$): The state encodes the complete information set that enables operators to compute prices and rebalancing flows. This includes information regarding the transportation network as well as region-level data on vehicle supply and ride demand. Specifically, at any given time $t \in \mathcal{T}$, the state $\mathbf{s}_{t,o}$ for a representative operator o includes:

- The network adjacency matrix $\mathbf{A}$.
- The current number of the operator's own idle vehicles in each region, $m_{i,o}^t \in [0, M_o]$ for all $i \in \mathcal{V}$.
- The number of vehicles en route to each region over a planning horizon H_p, denoted by $\{m_{i,o}^{t'}\}_{t'=\tau,\dots,\tau+H_p}$.
- The operator's own current prices, $p_{i,j,o}^t$ for all $i,j \in \mathcal{V}$.
- The competitor's current prices, $p_{i,j,o'}^t$ for all $i,j \in \mathcal{V}$ (where $o' \neq o$).
- The length of the operator's queue in each region, $q_{i,o}^t$ for all $i \in \mathcal{V}$.
- The operator's own current demand at each region, $d_{i,o}^t$ for all $i \in \mathcal{V}$.

Notably, operators do not share demand or vehicle location data, but they are able to observe the competitor's current prices across all regions.

Action-space ($\mathcal{A}$): As shown in Fig. 1, we consider an AMoD control problem that involves joint rebalancing of vehicles and trip pricing. An action from operator o specifies both a ride origin price scalar $p_{i,o}^t \in (0, 1]$ and the desired share of idle vehicles for each region, defined as $w_{i,o}^t \in [0, 1]$ per region i, bounded by the constraint $\sum_{i=1}^{N_v} w_{i,o}^t = 1$, where N_v is the total number of regions in the graph. It is this desired distribution of idle vehicles that is later used to solve a minimum-cost rebalancing problem to arrive at the specific vehicle flow.

We adopt an origin-based price scaling approach for two principal reasons. First, empirical evidence from both our experimental results (Appendix C) and prior work in single-operator settings [11] demonstrates that origin-based scaling achieves comparable performance to origin-destination (OD)-based scaling. Second, the origin-based approach substantially reduces the dimensionality of the action space, thereby enhancing the scalability of the proposed framework. The joint action at time t by operator o is given by $\mathbf{a}_{t,o} = [\mathbf{w}_{t,o}, \mathbf{p}_{t,o}]$, where $\mathbf{w}_{t,o} = [w_{1,o}^t, \ldots, w_{N_v,o}^t]$ and $\mathbf{p}_{t,o} = [p_{1,o}^t, \ldots, p_{N_v,o}^t]$. The origin price scalar is used to specify the trip price for operator o from region i to j via

$$p_{i,j,o}^t = \beta \cdot p_{i,o}^t \cdot \overline{p}_{i,j}^t, \tag{7}$$

where $\overline{p}_{i,j}^t$ is a historical reference price shared by both operators, and β is an upper bound on the price level. In our experiments, we set $\beta = 2$, effectively resulting in $p_{i,j,o}^t \in (0, 2\overline{p}_{i,j}^t]$. The vector $\mathbf{w}_{t,o}$ is used to specify the minimal rebalancing-cost problem solved to determine vehicle movement between regions, as described in Section IV-A.

Reward ($\mathcal{R}$): The reward is defined from the perspective of each operator separately, whose objective is to maximize its own profits. Consequently, operators act in their own best interest without regard for the competitor's performance. The reward for operator o is defined as the revenue earned from trips minus the cost of operations:

$$r_o = \sum_{i,j \in \mathcal{V}} x_{i,j,o}^t (p_{i,j,o}^t - c_{i,j,o}^t) - \sum_{(i,j) \in \mathcal{E}} y_{i,j,o}^t c_{i,j,o}^t, \tag{8}$$

where $x_{i,j,o}^t$ denotes the number of passengers served at time t from region i to j, $y_{i,j,o}^t$ denotes the number of vehicles rebalanced at time t from region i to j, and $c_{i,j,o}^t$ denotes the cost of moving a vehicle (with or without a passenger) between these regions.

Dynamics ($\mathcal{P}$): The system dynamics $\mathcal{P}$ characterize the evolution of the state space in response to the joint pricing and rebalancing actions of the operators. This transition process is driven by the stochastic nature of passenger demand, the update logic of regional queue lengths, and the conservation of vehicle availability across the network. The network adjacency matrix, while included in the state representation, remains exogenous to the operators' actions and is therefore treated as a static component of the dynamics.

Demand Generation and Mode Choice: The passenger demand process is grounded in historical data to ensure realistic simulation behavior. For each OD pair (i, j) and time t, a reference demand $\overline{d}_{ij}^t$ and reference price $\overline{p}_{ij}^t$ are sampled. To model a competitive market, we define a potential demand pool by scaling the reference demand to $2\overline{d}_{ij}^t$. Each potential passenger k chooses between operator 1, operator 2, or an outside option (rejecting both), modeled via a Multinomial Logit (MNL) framework.

The utility of passenger k at time t for a trip offered by operator o on OD pair (i, j) depends on travel time, the passenger's wage, and the trip price:

$$U_{k,i,j,o}^t = \beta_0 - \beta_t \cdot w_k \cdot \tau_{i,j}^t - \frac{\overline{w}}{w_k} p_{i,j,o}^t, \tag{9}$$

where β_0 represents a baseline preference parameter, β_t captures the marginal disutility of travel time, w_k denotes the passenger's hourly wage, and $\overline{w}$ is the average hourly wage across all passengers in the scenario. The variables $\tau_{i,j}^t$ and $p_{i,j,o}^t$ represent the travel time and price charged by operator o, respectively. We set $\beta_t = 0.71$, which corresponds to a 29% undervaluation of time relative to direct monetary cost. This parameterization is consistent with findings from other studies in the literature [37]. The price term is scaled by $\frac{\overline{w}}{w_k}$ to capture income effects, such that a given price has a relatively larger impact on lower-wage passengers and a smaller impact on higher-wage passengers. The utility of the outside option is normalized to zero: $U_{k,i,j,\emptyset}^t = 0$.

Passenger wage data and average wages are derived from income data from the US Census Bureau [38]–[43]. For the San Francisco scenario, wage values are adjusted using the national US inflation rate from 2008 to 2011 to obtain 2008-equivalent estimates from 2011 data, that is the closest data point available. Travel times between regions are based on historical averages from taxi trip data [44], [45].

The probability that passenger k selects operator o is given by

$$P_{k,i,j,o}^t = \frac{e^{U_{k,i,j,o}^t}}{e^{U_{k,i,j,\emptyset}^t} + \sum_{o' \in \{0,1\}} e^{U_{k,i,j,o'}^t}}. \tag{10}$$

Individual choices are simulated by drawing from a categorical distribution over the set of options $\{0, 1, \emptyset\}$ based on the calculated probabilities.

Queue Dynamics: The queue state for operator o in region i is updated according to the following flow conservation equation:

$$q_{i,o}^t = q_{i,o}^{t-1} + \sum_{j \in \mathcal{V} \setminus \{i\}} (d_{i,j,o}^t - x_{i,j,o}^t). \tag{11}$$

Here, the current queue length $q_{i,o}^t$ is determined by the previous state $q_{i,o}^{t-1}$, incremented by the volume of new passenger arrivals $\sum_{j \in \mathcal{V} \setminus \{i\}} d_{i,j,o}^t$, and decremented by the number of successful vehicle and passenger matches $\sum_{j \in \mathcal{V} \setminus \{i\}} x_{i,j,o}^t$ performed by operator o.

Vehicle Dynamics: In the context of autonomous vehicle control, we assume full compliance with system directives. Unlike human drivers, who may reject matched trips based on pricing or personal preference, autonomous vehicles strictly
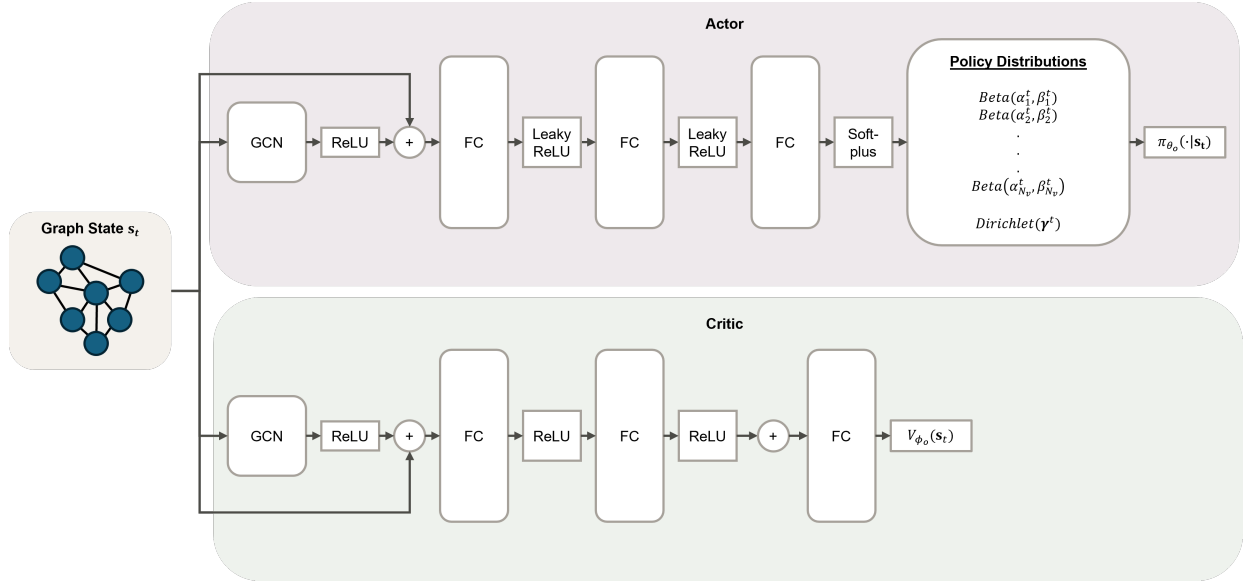
Fig. 2. The Actor-Critic architecture employed by the operators. Each operator maintains independent actor and critic networks.

adhere to the controller's dispatch and rebalancing decisions. Consequently, the available number of idle autonomous vehicles for operator $o \in \{0, 1\}$ in region i at time t is updated according to the following conservation law:

$$m_{i,o}^t = m_{i,o}^{t-1} + \sum_{j \in \mathcal{V} \setminus \{i\}} v_{j,i,o}^{\text{arr},t} - \sum_{j \in \mathcal{V} \setminus \{i\}} \left(x_{i,j,o}^{t-1} + y_{i,j,o}^{t-1} \right). \quad (12)$$

In this formulation, the current vehicle availability $m_{i,o}^t$ is determined by the idle pool from the previous time step $m_{i,o}^{t-1}$, incremented by the total number of incoming vehicles $v_{j,i,o}^{\text{arr},t}$ arriving at region i from all other regions $j \in \mathcal{V}$ at time t. This arrival term accounts for vehicles completing both passenger trips and rebalancing maneuvers initiated in previous intervals. The availability is then decremented by the total outflow of vehicles that departed region i at time $t-1$, which includes both vehicles matched with passengers $x_{i,j,o}^{t-1}$ and those dispatched for rebalancing $y_{i,j,o}^{t-1}$ to other regions.

C. The Model Architecture

We model the two competing fleet operators as independent, indexed by o, each equipped with its own parameterized policy (actor) and value function (critic). The policy of operator o is represented by a neural network $\pi_{\theta_o}(\cdot \mid \mathbf{s}_t)$ with parameters θ_o, while the corresponding value function is approximated by a neural network $V_{\phi_o}(\mathbf{s}_t)$ with parameters ϕ_o. The neural architecture for both π_{θ_o} and V_{ϕ_o} is grounded in the Graph Convolutional Network (GCN) reinforcement learning framework proposed by [3] and [11]. Both networks consist of a GCN layer followed by three fully connected (FC) layers. The input to both networks is the state representation s_t, which was presented in Section IV-B. A schematic illustration of the architecture is shown in Figure 2.

Critic Architecture: The value function is approximated by the parameterized critic network $V_{\phi_o}(\mathbf{s}_t)$. Given the input state $\mathbf{s}_t$, node-level features are first encoded using a GCN layer with ReLU activation. A residual connection is applied by adding the encoded features to the original node states. The resulting node embeddings are then passed through two FC layers with hidden size h and ReLU activations. To estimate the operator-specific global value function, the node-level representations are aggregated via a global summation operator across all regions, yielding a system-level embedding. This embedding is then mapped through a final FC layer to produce a scalar value $V_{\phi_o}(\mathbf{s}_t)$.

Actor Architecture: The policy network $\pi_{\theta_o}(\cdot \mid \mathbf{s}_t)$ follows the same initial encoding procedure as the critic. The input state $\mathbf{s}_t$ is first processed by a GCN layer with a residual connection. The resulting node embeddings are then passed through two FC layers with hidden size h and LeakyReLU activations. The output layer of π_{θ_o} depends on the control mode. For joint pricing and rebalancing, the final FC layer outputs three values per region, which are transformed via a Softplus activation to ensure strictly positive parameters. The first two outputs per region, α_i^t and β_i^t, parameterize a Beta distribution at time step t:

$$p_{i,o}^t \sim \text{Beta}(\alpha_i, \beta_i),$$

from which the origin-based price scalar for region i, at time t for operator o is sampled. The third output forms a concentration vector $\gamma^t \in \mathbb{R}^{N_v}$ that parameterize a Dirichlet distribution

$$\mathbf{w}_o^t \sim \text{Dirichlet}(\gamma^t),$$

from which the desired fraction of idle vehicles that should be rebalanced to each region at time t for operator o is sampled. For pricing-only control, π_{θ_o} outputs only the Beta distribution parameters, while for rebalancing-only control, it outputs only the Dirichlet concentration parameters.

V. EXPERIMENTS

Having established the theoretical foundations and model architecture for competitive multi-operator AMoD control, we now present a comprehensive empirical evaluation of the proposed framework. The experimental investigation is designed to address several fundamental questions about competitive autonomous mobility markets: How do learned policies differ between monopolistic and competitive settings? What strategic behaviors emerge when operators compete for passengers? How do market dynamics respond to varying fleet sizes, pricing information asymmetries, and regional economic heterogeneity?

To answer these questions, we structure the experiments along three dimensions. First, we evaluate the robustness and generalizability of the framework across diverse urban environments—San Francisco, Washington DC, and Southern Manhattan—comparing single-operator monopolistic control against competitive dual-operator scenarios under different control modes (rebalancing only, pricing only, and joint control). This cross-city analysis establishes baseline performance characteristics and reveals how competitive dynamics fundamentally alter operator behavior compared to centralized optimization.

Second, we conduct a detailed analysis of competitive dynamics in the Southern Manhattan environment, examining the policies learned by competing operators and investigating how information asymmetry affects strategic behavior. Specifically, we analyze scenarios where operators cannot observe competitor pricing decisions, forcing them to adapt their strategies under incomplete information—a realistic constraint in many competitive markets.

Third, we perform sensitivity analyses to understand how market structure parameters influence equilibrium outcomes. We systematically vary fleet sizes to examine capacity constraints and market power, explore asymmetric fleet distributions between competing operators, and introduce region-specific wage heterogeneity to capture how economic disparities across neighborhoods affect pricing strategies and demand allocation.

Throughout these experiments, we employ real-world data and realistic simulation parameters to ensure the findings provide actionable insights for platform designers, policymakers, and researchers working on competitive autonomous mobility systems.

A. Experimental Setup and Simulation Environment

All experiments are conducted within a discrete-time simulation environment, utilizing an adapted version of the simulator proposed by [44]. The simulation horizon is set to one hour, spanning the peak evening period from 19:00 to 20:00. This horizon is discretized into 20 time steps, with each step representing a 3-minute interval. Passenger demand prior to the choice model is generated using a Poisson distribution derived from real-world historical data, capturing the stochastic nature of urban mobility. The operators are trained using the Advantage Actor-Critic (A2C) algorithm, employing the Graph Convolutional Network (GCN)-based actor and critic architecture detailed in Section IV-C. Each model is trained to convergence.

B. Multi-City Performance Analysis: Monopolistic vs. Competitive Settings

To evaluate the robustness and generalizability of the framework, we conduct experiments across three distinct urban environments: San Francisco, Washington DC, and NYC Manhattan South. Each city presents unique operational challenges characterized by different network topologies, demand patterns, and economic contexts. To quantify spatial demand heterogeneity, we employ the coefficient of variation (CV), calculated by dividing the standard deviation of the demand across all regions by the average regional demand. This metric reflects the degree of variation in demand distribution across the urban network.

Table I summarizes the key characteristics of these scenarios. San Francisco exhibits the highest demand variability (CV=1.31) with a relatively small 10-node network and 374 vehicles serving 5,490 hourly trip requests, indicating highly uneven demand distribution across regions. Washington DC represents a larger-scale operation with 18 nodes, 1,096 vehicles, and 16,881 requests, though with comparable demand variability (CV=1.26). NYC Manhattan South operates at the highest absolute demand level (21,270 requests) but with notably lower spatial variability (CV=0.69), suggesting more predictable and balanced demand patterns across the network. The three cities also differ in their economic contexts, with hourly wages ranging from $17.76 in San Francisco to $25.26 in Washington DC.

TABLE I
CHARACTERISTICS OF THE SCENARIOS USED IN COMPARATIVE ANALYSIS. FOR THE BENCHMARK GAPS, THE FIRST COLUMN CORRESPONDS TO THE REWARD GAP AND THE SECOND COLUMN CORRESPONDS TO THE SERVED PASSENGER NUMBER GAP. A POSITIVE NUMBER INDICATES THAT THE EQUAL-DISTRIBUTION BENCHMARK HAS A HIGHER VALUE COMPARED TO THE NO-CONTROL BENCHMARK.

City	Date	Hourly Wage	Nodes	Veh.	Demand	Reward Gap	Served Gap	CV
San Francisco	2008-06-06	$17.76	10	374	5490	46.74%	62.96%	1.31
Washington DC	2019-03-12	$25.26	18	1096	16881	15.06%	48.14%	1.26
NYC Man. South	2013-03-08	$22.77	12	650	21270	10.00%	20.83%	0.69

Note: Veh.: No. Vehicles; CV: Coefficient of Variation of Demand; All data from 19:00–20:00.

1) Single-Operator Monopolistic Performance: We first examine the monopolistic setting where a single operator controls the entire fleet and serves all demand. Table II presents the total rewards achieved by different control strategies across the three cities. The joint pricing and rebalancing policy consistently outperforms all baselines and single-mode policies across all environments, demonstrating the synergistic benefits of coordinated control. In San Francisco, joint control achieves a reward of 12,447.47, representing a 23.0% improvement over rebalancing alone and a 43.5% improvement over pricing alone. This substantial gain suggests that in highly variable demand environments, the coordination between pricing and fleet positioning is particularly valuable.

The magnitude of improvement from joint control, however, varies considerably across cities. In Washington DC, joint control achieves 16,574.46, only 2.9% above rebalancing alone,

TABLE II
PERFORMANCE OF THE TRAINING POLICIES IN THE THREE SCENARIOS.
WE PERFORM 10 TESTS FOR EACH POLICY AND REPORT THE AVERAGE
PERFORMANCE WITH STANDARD DEVIATIONS IN PARENTHESES. BOLD
INDICATES THE BEST-PERFORMING POLICIES. IN THE COLUMNS, THE
POLICY "NC" REPRESENTS NO CONTROL, "UD" REPRESENTS UNIFORM
DISTRIBUTION, "REB." REPRESENTS REBALANCING, AND "JOINT"
REPRESENTS THE JOINT PRICING AND REBALANCING POLICY.

City	NC	UD	Reb.	Pricing	Joint
San Francisco	6345.49 (251.60)	9444.98 (318.89)	10116.42 (362.19)	8675.68 (130.66)	**12447.47 (159.63)**
Washington DC	13153.97 (324.58)	15612.59 (383.28)	16099.46 (364.72)	14118.33 (313.26)	**16574.46 (334.12)**
NYC Man. South	16283.02 (487.67)	18224.77 (260.39)	18470.56 (304.30)	18290.21 (391.38)	**18662.76 (289.01)**

while in NYC Manhattan South, the improvement shrinks to merely 1.0% (18,662.76 vs. 18,470.56). This diminishing marginal benefit of pricing in lower-variability environments suggests that when demand patterns are more predictable and spatially balanced, strategic rebalancing alone can achieve near-optimal performance, and the additional complexity of dynamic pricing provides limited incremental value.

Table III provides detailed operational metrics for the three active control modes. A notable pattern emerges in pricing behavior: the learned price scalars vary systematically with demand characteristics. In San Francisco, the joint policy sets conservative prices (scalar 0.86), significantly below historical reference levels, effectively using pricing to stimulate demand in a highly variable environment. Conversely, in Washington DC and NYC Manhattan South, the joint policy sets prices near or above reference levels (1.08 and 1.01 respectively), extracting higher revenue from more stable demand streams. Interestingly, joint control achieves lower rebalancing costs compared to rebalancing-only policies across all cities (e.g., 667.80 vs. 919.98 in San Francisco), indicating that strategic pricing reduces the need for costly vehicle repositioning by shaping demand patterns.

The service quality metrics reveal important trade-offs between different control strategies. In San Francisco, pure rebalancing achieves the lowest waiting times (0.26 minutes) by maintaining excellent fleet positioning but serves relatively modest demand (971.50 passengers). Joint control accepts moderately higher waiting times (0.51 minutes) but dramatically increases served demand to 1,308.90 passengers by using pricing to attract additional trips. This 34.7% increase in served demand illustrates how pricing can expand market size beyond what rebalancing alone achieves. Washington DC exhibits a different pattern: joint control delivers both the lowest waiting times (0.19 minutes) and high served demand (3,748.70), suggesting that in larger networks with more vehicles, the framework can simultaneously optimize service quality and demand capture. In NYC Manhattan South, the high baseline queue lengths (14.20 for rebalancing) indicate capacity constraints, limiting the effectiveness of operational improvements.

2) Dual-Operator Competitive Performance: We now examine how competitive dynamics alter system performance and learned strategies. Table IV presents total rewards for

the dual-operator setup, revealing a striking departure from monopolistic patterns. Unlike the single-operator case where joint control uniformly dominates, the optimal strategy in competitive markets varies by city. In San Francisco, rebalancing-only achieves the highest total reward (10,294.4), narrowly edging out joint control (10,205.1). Washington DC similarly favors rebalancing (15,743.2), while NYC Manhattan South exhibits a remarkable result: pricing-only control achieves the highest total reward (18,879.6), surpassing both rebalancing and joint control.

This heterogeneity in optimal competitive strategies reflects fundamental differences in market structure. In high-variability environments (San Francisco, Washington DC), competition intensifies the challenge of matching supply to fluctuating demand, making efficient fleet positioning critical. Joint control's coordination mechanism, optimized for monopolistic operation, appears less effective when facing a strategic competitor. Conversely, in the stable, high-density NYC Manhattan South environment, pricing becomes the dominant strategic weapon, as operators compete directly for market share through price-based passenger acquisition rather than through spatial positioning advantages.

Table V reveals that competitive dynamics drive systematic changes in pricing behavior and service quality. In San Francisco, dual operators converge to even lower prices (0.67 for both operators under joint control) compared to the monopolistic case (0.86), consistent with Bertrand competition driving prices downward. However, this price competition comes at a substantial service quality cost: waiting times under dual-operator joint control increase to 1.35 minutes (Operator 0) and 1.19 minutes (Operator 1), compared to just 0.51 minutes in the monopolistic case. Despite this, dual-operator joint control achieves slightly higher total served demand (1,385.9 vs. 1,308.90), suggesting that lower prices stimulate enough additional demand to offset the operational inefficiencies of fragmented fleet management.

In NYC Manhattan South, the joint control scenario demonstrates balanced performance across both operators. The total reward reaches 17,685.7, with Operator 0 achieving 8,815.3 and Operator 1 achieving 8,870.4—a near-perfect split indicating convergence to a symmetric Nash equilibrium. Both operators adopt similar pricing strategies (price scalar 0.97), slightly below the monopolistic level (1.01), reflecting moderate competitive pressure. Rebalancing costs are well-balanced (708.8 vs. 671.0), with total costs of 1,379.7 remaining comparable to the monopolistic joint control (1,539.90). The operators serve nearly identical demand volumes (1,781.1 vs. 1,791.4 passengers), with total served demand of 3,572.5, representing a modest 0.4% increase over monopolistic joint control (3,557.5). Waiting times are reasonable (1.43 vs. 1.45 minutes), though elevated compared to monopolistic pricing-only (0.40 minutes). Queue lengths remain moderate (10.3 vs. 9.98), significantly better than the monopolistic rebalancing scenario (14.20). The total demand of 5,151.0 are split nearly evenly (2,568.0 vs. 2,583.0), demonstrating effective market sharing. This balanced outcome suggests that in stable, high-density environments, competitive joint control can achieve near-monopolistic efficiency with equitable market division.

TABLE III
PERFORMANCE METRICS OF THE THREE POLICIES IN SAN FRANCISCO, WASHINGTON DC, AND NYC MANHATTAN SOUTH. THE NUMBERS IN PARENTHESES INDICATE THE STANDARD DEVIATIONS OF EACH METRIC FOR 10 TEST RUNS. "PRICE" IS THE AVERAGE PRICE SCALAR SET BY THE OPERATOR, AND "WAIT/MINS" IS THE WAITING TIME OF THE SERVED PASSENGERS IN MINUTES. BOLD INDICATES THE BEST-PERFORMING POLICIES. "REB." REPRESENTS REBALANCING, "PRICING" IS THE PRICING POLICY, AND "JOINT" IS THE JOINT PRICING AND REBALANCING POLICY. ALL VALUES ARE AVERAGED ACROSS ALL THE REGIONS.

Policy	San Francisco			Washington DC			NYC Man. South		
	Reb.	Pricing	Joint	Reb.	Pricing	Joint	Reb.	Pricing	Joint
Reward	10116.42 (362.19)	8675.68 (130.66)	**12447.47 (159.63)**	16099.46 (364.72)	14118.33 (313.26)	**16574.46 (334.12)**	18470.56 (304.30)	18290.21 (391.38)	**18662.76 (289.01)**
Rebalancing Costs	919.98 (29.11)	—	667.80 (21.06)	3706.65 (150.48)	—	3520.65 (58.42)	1394.40 (95.83)	—	1539.90 (102.35)
Rebalance Trips	539.80 (16.04)	—	393.10 (14.51)	1134.70 (34.26)	—	1177.80 (28.40)	290.00 (21.34)	—	321.30 (21.63)
Price	—	0.75 (0.00)	0.86 (0.00)	—	1.17 (0.00)	1.08 (0.00)	—	1.05 (0.00)	1.01 (0.00)
Wait/mins	0.26 (0.05)	0.52 (0.03)	0.51 (0.04)	0.32 (0.03)	0.45 (0.03)	0.19 (0.02)	0.64 (0.02)	0.40 (0.03)	0.58 (0.03)
Queue	1.86 (0.36)	4.60 (0.23)	5.45 (0.37)	5.20 (0.49)	5.18 (0.32)	2.49 (0.27)	14.20 (0.49)	6.95 (0.39)	12.52 (0.67)
Served Demand	971.50 (32.22)	823.60 (10.97)	1308.90 (16.82)	4254.20 (52.47)	2393.40 (51.26)	3748.70 (58.21)	3557.70 (36.60)	2971.10 (60.43)	3557.50 (35.31)
Total Demand	1092.10 (42.86)	1319.00 (34.71)	1736.90 (33.86)	4753.00 (76.79)	3174.00 (46.58)	3919.90 (71.49)	4705.70 (73.36)	3452.70 (60.75)	4511.70 (76.97)

TABLE IV
TOTAL REWARD COMPARISON ACROSS DIFFERENT CONTROL STRATEGIES IN A DUAL-OPERATOR SETUP. WE PERFORM 10 TESTS FOR EACH POLICY AND REPORT THE AVERAGE PERFORMANCE WITH STANDARD DEVIATIONS IN PARENTHESES. BOLD INDICATES THE BEST-PERFORMING POLICIES. IN THE COLUMNS, THE POLICY "NC" REPRESENTS NO CONTROL, "UD" REPRESENTS UNIFORM DISTRIBUTION, "REB." REPRESENTS REBALANCING, AND "JOINT" REPRESENTS THE JOINT PRICING AND REBALANCING POLICY.

City	NC	UD	Reb.	Pricing	Joint
San Francisco	6391.2 (232.4)	9371.4 (247.1)	**10294.4 (251.5)**	9277.4 (163.3)	10205.1 (180.3)
Washington DC	13172.3 (344.6)	15155.7 (375.8)	**15743.2 (340.6)**	14093.3 (263.8)	15669.3 (356.5)
NYC Man. South	16048.4 (416.6)	17652.8 (256.3)	18174.2 (326.2)	**18879.6 (328.2)**	17685.7 (270.9)

Remarkably, both operators achieve roughly balanced rewards across all scenarios (e.g., 5,051.2 vs. 5,243.2 in San Francisco rebalancing), suggesting that the learning dynamics converge to approximate Nash equilibria where neither operator can substantially improve by deviating unilaterally. The served demand is also nearly evenly split between operators in most configurations, with typical variations under 5%, indicating that competition produces fair market sharing despite the absence of explicit coordination mechanisms.

3) Monopolistic vs. Competitive Comparison: Comparing the best-performing policies across market structures reveals the profit losses induced by competition. In San Francisco, the monopolistic joint control achieves 12,447.47, while the best dual-operator configuration (rebalancing) yields only 10,294.4—a 17.3% profit loss. Washington DC exhibits a similar pattern with a 5.0% loss (16,574.46 vs. 15,743.2). Remarkably, NYC Manhattan South presents an anomalous case: the dual-operator pricing strategy achieves 18,879.6, representing a 1.16% gain over the monopolistic joint control (18,662.76). This counterintuitive result suggests that in stable, high-density markets, competitive pricing pressure can induce demand beyond what a monopolist would optimally choose, potentially expanding the overall pool of potential passengers through aggressive passenger acquisition strategies.

The magnitude of competition-induced profit losses correlates strongly with demand variability. High-CV environments (San Francisco, Washington DC) suffer substantial profit losses under competition, as fragmented fleet management exacerbates the challenge of matching supply to volatile demand. Conversely, the low-CV NYC environment experiences minimal profit loss or even gain, as the primary competitive dimension shifts from rebalancing efficiency to market share competition through pricing. These findings suggest that the case for regulatory intervention to consolidate AMoD markets varies systematically with urban characteristics, with high-variability cities potentially benefiting more from coordinated or regulated operations.

C. Competitive Policy Analysis in Southern Manhattan

Having established that competitive dynamics produce fundamentally different optimal strategies compared to monopolistic settings, we now conduct a detailed examination of the learned policies in the NYC Manhattan South environment across: the pricing-only policy, the rebalancing-only policy and

TABLE V

PERFORMANCE METRICS OF THE THREE POLICIES IN SAN FRANCISCO, WASHINGTON DC, AND NYC MANHATTAN SOUTH FOR DUAL-OPERATOR SETUP. THE NUMBERS IN PARENTHESES INDICATE THE STANDARD DEVIATIONS OF EACH METRIC FOR 10 TEST RUNS. "PRICE" IS THE AVERAGE PRICE SCALAR SET BY EACH OPERATOR, AND "WAIT/MINS" IS THE WAITING TIME OF THE SERVED PASSENGERS IN MINUTES. "REB." REPRESENTS REBALANCING, "PRICING" IS THE PRICING POLICY, AND "JOINT" IS THE JOINT PRICING AND REBALANCING POLICY. ALL VALUES ARE AVERAGED ACROSS ALL THE REGIONS.

	San Francisco			Washington DC			NYC Man. South		
Policy	Reb.	Pricing	Joint	Reb.	Pricing	Joint	Reb.	Pricing	Joint
Total Reward	**10294.4 (251.5)**	9277.4 (163.3)	10205.1 (180.3)	**15743.2 (340.6)**	14093.3 (263.8)	15669.3 (356.5)	18174.2 (326.2)	**18879.6 (328.2)**	17685.7 (270.9)
Reward Operator 0	5051.2 (182.6)	4684.0 (88.4)	5027.4 (121.7)	7850.0 (240.4)	6961.0 (185.0)	7598.7 (263.7)	9029.4 (228.1)	9489.0 (183.1)	8815.3 (221.2)
Reward Operator 1	5243.2 (182.6)	4593.3 (113.3)	5177.7 (137.0)	7893.2 (138.9)	7132.4 (272.6)	8070.6 (165.4)	9144.9 (295.8)	9390.6 (195.1)	8870.4 (242.2)
Total Rebalancing Costs	1040.9 (19.7)	—	571.2 (20.0)	4018.7 (121.4)	—	4105.2 (149.9)	1549.3 (113.2)	—	1379.7 (97.0)
Rebalancing Costs Operator 0	495.4 (19.8)	—	293.3 (5.7)	2022.2 (72.6)	—	2122.1 (88.5)	821.7 (83.2)	—	708.8 (81.1)
Rebalancing Costs Operator 1	545.5 (24.6)	—	277.9 (17.8)	1996.5 (98.3)	—	1983.2 (98.5)	727.6 (100.3)	—	671.0 (71.2)
Total Rebalance Trips	616.7 (9.5)	—	333.2 (14.5)	1202.9 (46.2)	—	1238.0 (43.5)	329.6 (22.2)	—	286.2 (22.3)
Rebalance Trips Operator 0	292.4 (10.0)	—	172.3 (3.4)	600.5 (37.4)	—	655.4 (32.8)	175.5 (20.7)	—	148.2 (19.2)
Rebalance Trips Operator 1	324.3 (12.0)	—	160.9 (13.2)	602.4 (14.0)	—	582.6 (18.1)	154.1 (22.1)	—	138.0 (15.8)
Total Served Demand	995.2 (23.0)	892.3 (14.5)	1385.9 (17.8)	4242.8 (51.1)	2387.4 (43.2)	4155.3 (51.6)	3532.8 (38.6)	3463.8 (43.4)	3572.5 (31.1)
Served Demand Operator 0	486.4 (16.3)	452.0 (10.1)	684.6 (11.7)	2119.1 (39.1)	1177.4 (29.2)	2022.1 (41.5)	1760.9 (26.2)	1753.4 (21.8)	1781.1 (23.0)
Served Demand Operator 1	508.8 (16.0)	440.3 (8.6)	701.3 (12.7)	2123.7 (16.2)	1210.4 (44.6)	2133.2 (20.1)	1771.9 (35.0)	1710.4 (26.6)	1791.4 (28.4)
Price Operator 0	—	0.73 (0.00)	0.67 (0.00)	—	1.16 (0.00)	1.02 (0.00)	—	0.93 (0.01)	0.97 (0.00)
Price Operator 1	—	0.71 (0.00)	0.67 (0.00)	—	1.16 (0.00)	1.01 (0.00)	—	0.93 (0.01)	0.97 (0.00)
Wait/mins Operator 0	0.52 (0.12)	0.97 (0.06)	1.35 (0.08)	0.84 (0.10)	0.74 (0.10)	0.71 (0.08)	1.29 (0.08)	1.09 (0.06)	1.43 (0.07)
Wait/mins Operator 1	0.36 (0.10)	0.82 (0.07)	1.19 (0.09)	0.79 (0.09)	0.76 (0.07)	0.85 (0.12)	1.31 (0.06)	1.02 (0.05)	1.45 (0.08)
Queue Operator 0	2.24 (0.62)	3.31 (0.77)	4.58 (0.80)	2.12 (0.70)	3.07 (0.56)	1.44 (0.60)	8.17 (0.92)	5.14 (0.72)	10.3 (1.3)
Queue Operator 1	2.10 (1.20)	3.36 (0.53)	3.98 (0.60)	1.82 (0.47)	3.17 (0.46)	2.16 (0.66)	7.73 (1.19)	4.54 (0.69)	9.98 (1.24)
Total Demand	1112.5 (43.3)	1413.5 (34.1)	2169.2 (40.8)	4746.9 (76.8)	3180.9 (47.2)	4629.0 (72.1)	4683.0 (73.9)	4425.4 (60.9)	5151.0 (61.4)
Demand Operator 0	551.9 (20.2)	731.0 (18.4)	1118.8 (30.5)	2379.4 (54.5)	1555.5 (38.0)	2210.6 (50.8)	2336.2 (56.0)	2278.4 (39.7)	2568.0 (60.4)
Demand Operator 1	560.6 (30.9)	682.5 (19.3)	1050.4 (23.5)	2367.5 (36.6)	1625.4 (29.4)	2418.4 (37.9)	2346.8 (39.4)	2147.0 (29.9)	2583.0 (40.0)

the joint control policy. Recall from Table IV that the optimal strategy varies by policy: pricing-only achieved the highest total reward (18,879.6), followed closely by rebalancing-only (18,174.2) and joint control (17,685.7). This unexpected ordering—where specialized single-lever policies outperform joint optimization—motivates a detailed investigation into how control constraints shape competitive strategic behavior.

1) Pricing Policy: Using Prices as a Fleet Management Tool: In the pricing-only policy, operators lack direct rebalancing capability and must rely exclusively on price adjustments to influence both revenue and vehicle distribution. Figure 3 reveals the sophisticated strategy that emerges: at the initial timestep, both operators employ extreme spatial price discrimination, with scalars ranging from 0.21 to 0.95. Crucially, the lowest prices (0.21-0.38) are concentrated in the southern peripheral regions where demand is low. This reflects pricing's dual role: low prices in the south stimulate trips that relocate vehicles northward toward high-demand core regions, while higher prices in the northwest (0.82-0.95) both generate revenue and discourage trips that would pull vehicles away from where they are most needed.

Figure 4 shows the time-averaged pricing strategies, revealing how this dual-purpose pricing persists throughout the simulation. Average scalars range from 0.62 to 1.07, with southern regions maintaining substantial discounts (0.62-0.86) despite generating minimal absolute demand. The northwestern core sustains premium pricing (0.94-1.07), balancing revenue extraction against the need to avoid excessive demand that would deplete vehicle availability faster than southbound trips can replenish it. The convergence between Operator 0 and Operator 1 pricing (typical differences less than 0.05) indicates Nash equilibrium: both operators have learned similar trade-offs between using prices for revenue versus fleet management, and neither can improve by deviating.

By the final timestep (Figure 5), pricing has converged substantially but remains spatially heterogeneous, with scalars from 0.79 to 1.07. Compared to initial conditions, final prices are uniformly higher (increases of 0.2-0.6 across most regions). However, the persistence of 15-20% southern discounts even at equilibrium confirms that indirect vehicle repositioning through pricing remains essential throughout the simulation—without this sustained spatial discrimination, vehicle stocks in the northwestern core would deplete, causing service failures.

The demand allocation pattern in Figure 6 shows the outcome of this pricing strategy: Operator 0 serves 319-463 passengers in the northwestern core, with Operator 1 capturing similar volumes (303-397). The relatively balanced market shares despite the aggressive pricing strategies indicate that both operators successfully learned similar solutions to the constrained optimization problem. Total served demand in pricing-only mode (3,463.8 from Table V) is lower than in other modes, yet remarkably, this mode achieved the highest total reward (18,879.6)—a counterintuitive result that reveals the efficiency of pricing as a competitive instrument. The superior performance despite lower volumes indicates that spatially differentiated pricing enables more effective revenue extraction: premium prices in the northwestern core generate high per-trip revenue, while strategic discounts in the south cost little (due to low demand) but provide essential fleet repositioning. This demonstrates that in competitive markets, optimizing for revenue efficiency (reward per passenger) may be more valuable than maximizing served volumes.

2) Rebalancing Policy: Direct Fleet Control with Fixed Prices: In the rebalancing-only policy, operators maintain fixed reference prices and compete solely through strategic vehicle positioning. Figure 7 presents the net rebalancing flows, showing aggressive fleet redistribution patterns. Oper-
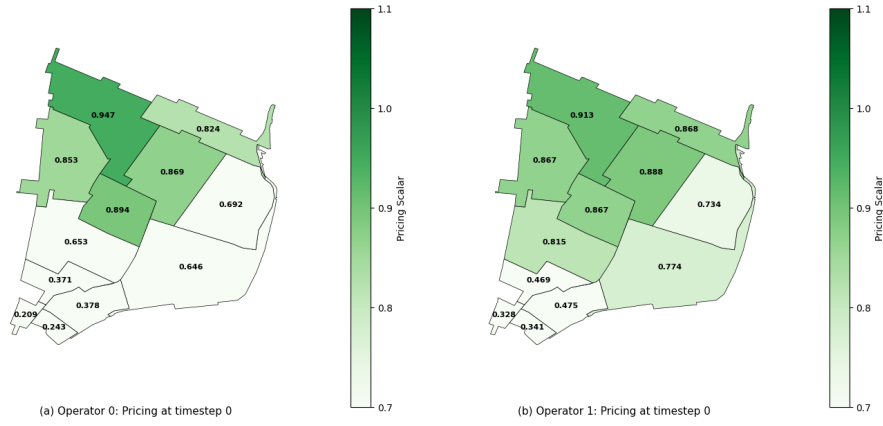
(a) Operator 0: Pricing at timestep 0

(b) Operator 1: Pricing at timestep 0

Fig. 3. Initial pricing policies at timestep t=0 for the pricing-only policy.

(a) Operator 0: Average price scalar set per region.

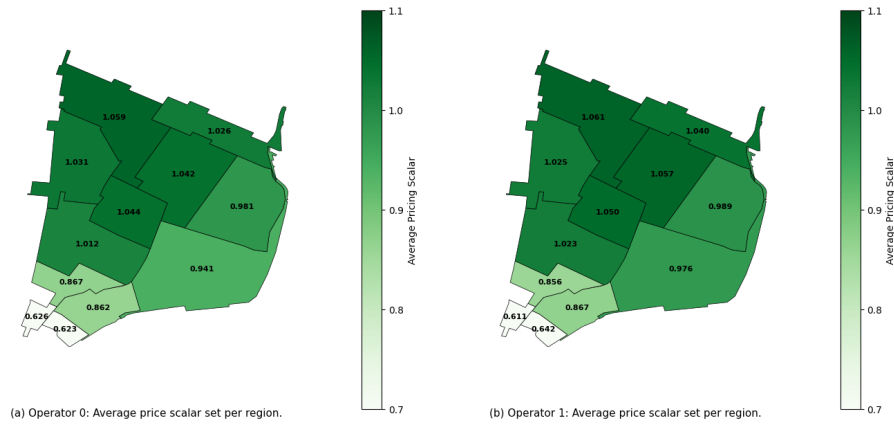(b) Operator 1: Average price scalar set per region.

Fig. 4. Time-averaged pricing scalars for the pricing-only policy, computed across all 20 timesteps.

ator 0 accumulates 33-44 net vehicles in the northwestern core regions, with particularly strong positioning in central Midtown (44 vehicles), while depleting southern and eastern regions by 13-28 vehicles. Operator 1 exhibits similar but slightly more moderate patterns (26-35 vehicles accumulated in northwest, 16-27 depleted from periphery). These flows are notably larger in absolute magnitude than those observed in joint control mode (discussed subsequently), suggesting that when rebalancing is the only competitive lever, operators rely more heavily on aggressive positioning to differentiate themselves.

The demand allocation in Figure 8 demonstrates the effectiveness of this rebalancing strategy: Operator 0 captures 375-506 passengers in the northwestern core, achieving modest but consistent market share advantages over Operator 1 (332-393 passengers in the same regions). The correlation between rebalancing intensity and demand capture reflects operational efficiency—Operator 0's more aggressive vehicle accumulation in high-demand regions enables it to serve assigned passengers more quickly, reducing abandonment due to excessive wait times. Since the discrete choice model allocates passengers based on price (which is fixed and equal across operators in this mode), travel time, and outside option utility, both operators receive similar initial demand assignments. However, Operator 0's superior vehicle positioning allows it

to actually serve a higher fraction of its assigned passengers within the maximum wait threshold, translating fleet advantage into realized demand. Total served demand (3,532.8) is slightly higher than pricing-only mode; however, the difference is modest and total reward (18,174.2) remains below pricing-only performance, suggesting that operational efficiency alone cannot compensate for the revenue optimization advantages of dynamic pricing.

3) Joint Control Policy: The joint control policy permits simultaneous optimization of both pricing and rebalancing, theoretically enabling superior coordination. However, as Table IV revealed, joint control achieved the lowest total reward (17,685.7) among the three modes in competitive settings—a surprising result that demands explanation.

Figure 9 reveals the first mechanism of failure: joint-mode pricing exhibits remarkable spatial uniformity, with average scalars tightly clustered around 0.967 (range 0.962-0.968). Notably, these prices are higher on average than the pricing-only mode's range of 0.62-1.07. This combination—higher average prices but compressed spatial variation—indicates a coordination failure. In the pricing-only mode, operators used spatial price differentiation strategically: low prices in peripheral areas attracted demand where vehicles naturally accumulated, while higher prices in congested areas both maximized revenue and managed fleet distribution. With joint
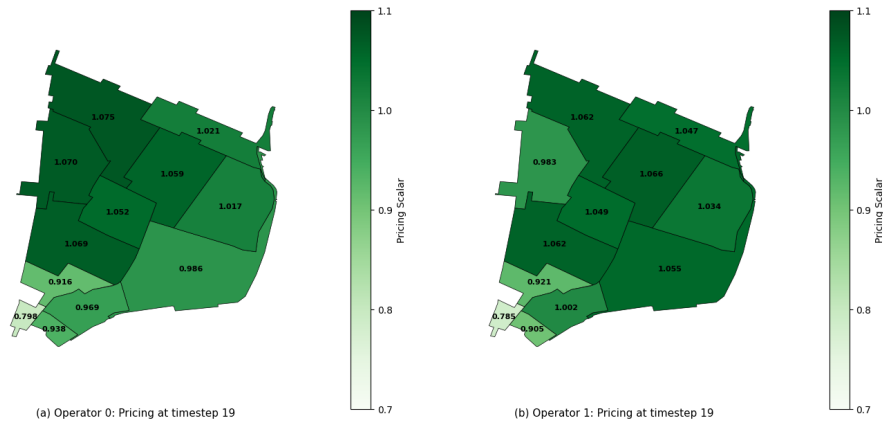
(a) Operator 0: Pricing at timestep 19

(b) Operator 1: Pricing at timestep 19

Fig. 5. Final pricing policies at timestep 19 for the pricing-only policy.

(a) Operator 0: Total demand originating from each region.

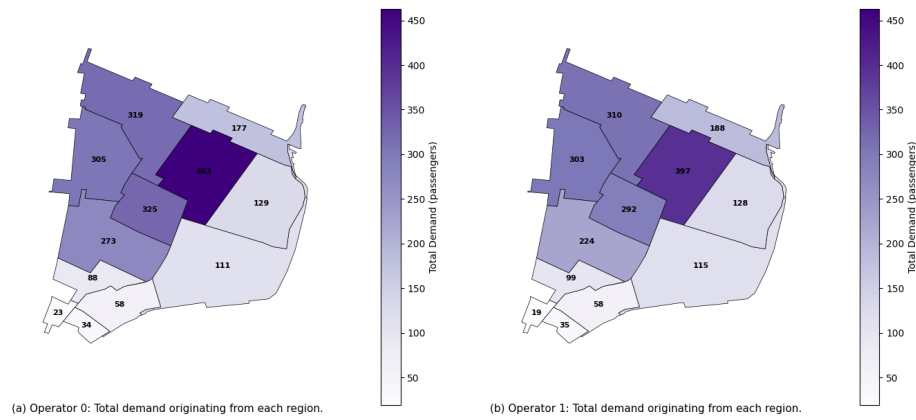(b) Operator 1: Total demand originating from each region.

Fig. 6. Total passenger demand allocation for the pricing-only policy, summed over all timesteps.

control, operators set uniformly high prices across space while relying on rebalancing for fleet management. However, this separation of functions proves counterproductive in competitive markets: high uniform pricing reduces the total market size (passengers choosing outside options), while rebalancing costs accumulate without the revenue efficiency that strategic price differentiation provided with the pricing-only policy.

The second failure mechanism appears in the rebalancing flows (Figure 10): both operators engage in parallel repositioning toward the same high-demand areas, creating wasteful competition for spatial positioning. Operator 0 accumulates 29-33 vehicles in the northwestern core, while Operator 1 shows 17-25 in overlapping regions. This represents wasted rebalancing effort—both operators incur repositioning costs to move vehicles to the same locations, where they then compete for the same passengers. Unlike the rebalancing-only mode where spatial strategies could differentiate through pure positioning, or the pricing-only mode where natural vehicle distribution guided by demand patterns avoided explicit repositioning costs, the joint mode combines the costs of active rebalancing with the competitive interference of simultaneous spatial competition. Both operators rebalancing to identical targets creates redundancy that destroys value: the benefits of better positioning are negated by the competitor's parallel actions, while the costs are fully realized by both parties.

The demand allocation in Figure 11 shows the consequences of these dual failures: Operator 0 serves 360-550 passengers in the northwest (highest absolute volumes across all modes), with Operator 1 capturing 373-550. However, total served demand (3,572.5) remains modest. This paradox—higher absolute demand per operator but lower total reward—directly reflects the combination of high uniform pricing and wasted rebalancing. The uniformly high pricing (0.967 average) shrinks the total addressable market as price-sensitive passengers opt out, while both operators simultaneously incur substantial repositioning costs to reach the same high-demand areas. The strategic error is clear: operators maintain high prices expecting rebalancing to ensure service quality, but the redundant repositioning costs consume the revenue gains from premium pricing, while the lack of spatial price differentiation prevents the demand-shaping efficiency that made the pricing-only mode successful.

This analysis reveals a fundamental insight for competitive AMoD markets: strategic flexibility without strategic clarity breeds inefficiency. The joint control mode's poor performance stems not from inadequate learning, but from the inherent structure of two-dimensional competition. When operators compete simultaneously on price and space, they fail to achieve the synergy available in monopolistic settings. Instead, high uniform pricing reduces market size while redundant
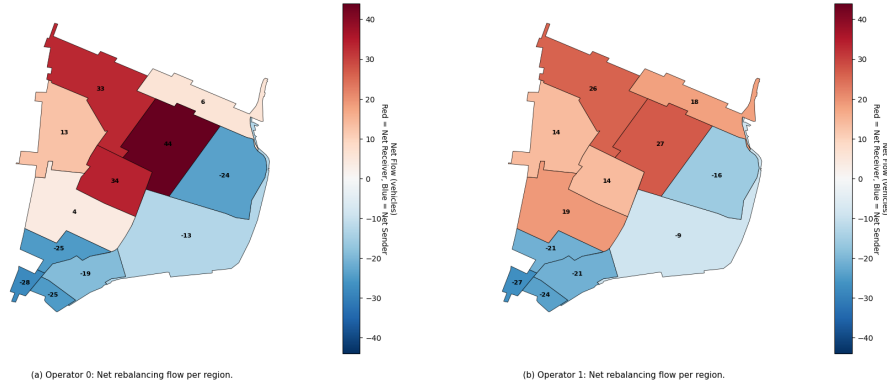
(a) Operator 0: Net rebalancing flow per region.

(b) Operator 1: Net rebalancing flow per region.

Fig. 7. Net rebalancing flows for the rebalancing-only policy, showing cumulative vehicle movements across all timesteps.

(a) Operator 0: Total demand originating from each region.
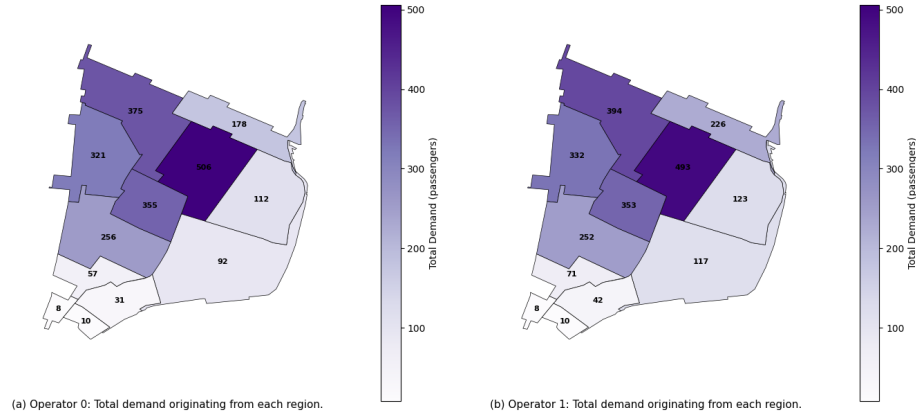

(b) Operator 1: Total demand originating from each region.

Fig. 8. Total passenger demand allocation for the rebalancing-only policy.

rebalancing wastes resources through competitive interference. The superior performance of specialized single-mode policies demonstrates that in competitive settings, constrained strategy spaces—which force differentiation along a single dimension—can outperform unconstrained optimization that enables mutually destructive behavior.

D. Impact of Pricing Information

We investigate the impact of information sharing between competing operators by comparing scenarios where operators can observe competitor prices versus when they cannot. Table VI presents the performance metrics for both information sharing configurations across the three operational modes. The results reveal nuanced differences in operator behavior and system performance depending on whether competitor prices are visible.

The overall system performance remains relatively stable across information sharing configurations. With the rebalancing policy only, total rewards are nearly identical: 18190.5 without information sharing versus 18174.2 with information sharing. Similarly, the pricing policy only shows comparable performance with 18813.1 and 18879.6, respectively. The joint control exhibits similar total rewards with 17491.9 without information sharing and 17685.7 with information sharing, representing a modest 1.1% increase.

The most notable differences emerge with the joint-control policy, where price visibility influences strategic behavior. With information sharing, both operators converge to identical pricing strategies (0.97 for both operators) compared to the slightly differentiated pricing without visibility (0.94 and 0.95). Operator 0 experiences increased rebalancing costs (708.8 versus 610.8) and rebalancing trips (148.2 versus 125.1) with information sharing, indicating that price visibility may lead to more aggressive spatial competition. Figure 12 illustrates the training convergence dynamics for both information sharing scenarios across the three operational policies. The results demonstrate that information sharing accelerates convergence, particularly with the pricing policy or the rebalancing policy. This faster convergence suggests that observing competitor prices provides valuable learning signals that help operators discover effective strategies more quickly.

(a) Operator 0: Average price scalar set per region.

(b) Operator 1: Average price scalar set per region.

Fig. 9. Time-averaged pricing scalars for the joint control policy.

(a) Operator 0: Net rebalancing flow per region.

(b) Operator 1: Net rebalancing flow per region.
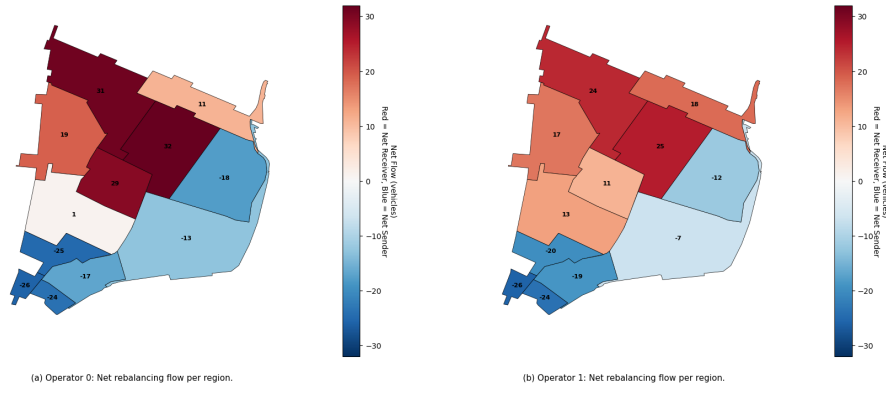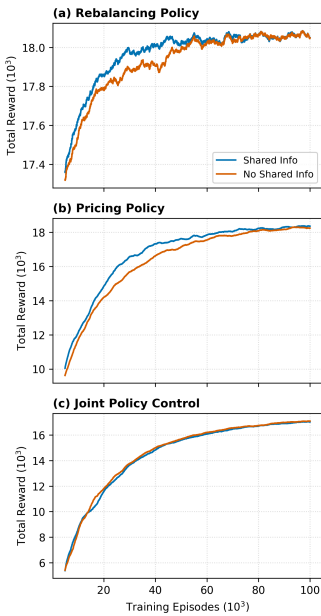
Fig. 10. Net rebalancing flows for the joint control policy.

Fig. 12. Convergence dynamics comparing scenarios with and without competitor price visibility across operational policies: (a) Rebalancing policy, (b) Pricing policy, (c) Joint policy. Curves are smoothed over 30 episodes and the first 5,000 episodes are excluded for clarity. Note rewards are training rewards, based on drawing from the policy distributions, and may therefore be lower than test rewards.

Interestingly, the convergence advantage is less pronounced with the joint policy. The joint policy presents a more complex learning problem where both pricing and rebalancing must be coordinated, potentially diluting the relative advantage of price visibility.

E. Fleet Size Sensitivity Analysis

Table VII compares system performance across three operational policies under varying fleet sizes, with vehicles evenly distributed between the two operators. The results highlight the advantages of joint pricing and rebalancing optimization.

As fleet size increases from 450 to 1250 vehicles (225 to 625 per operator), the joint control policy exhibits adaptive pricing that responds to fleet availability. Prices decrease from 1.01/1.03 to 0.76/0.72 for Operators 0 and 1 respectively, stimulating demand to maintain high utilization. At fleet size 1250, this dynamic pricing enables the system to serve 6291.7 passengers versus 3934.1 for NC and 4633.2 for UD—increases of 60% and 36% respectively. However, the increase in passengers due to the lower prices results in increased wait times (1.09 and 1.34 minutes) and queue lengths (12.8 and 16.8 passengers) compared to the baseline policies (0.60 and 0.17 minutes for NC and UD).

The Joint Policy achieves competitive total rewards across all configurations. At smaller fleet sizes (450 and 650), it

(a) Operator 0: Total demand originating from each region.   (b) Operator 1: Total demand originating from each region.
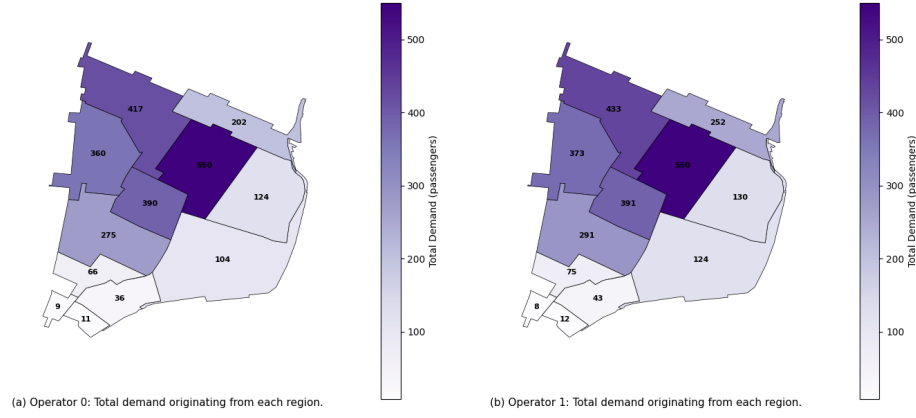
Fig. 11. Total passenger demand allocation for the joint control policy.

TABLE VI

PERFORMANCE COMPARISON WITH AND WITHOUT COMPETITOR PRICE VISIBILITY IN NYC MANHATTAN SOUTH IN DUAL-OPERATOR SETUP. THE
NUMBERS IN PARENTHESES INDICATE THE STANDARD DEVIATIONS OF EACH METRIC FOR 10 TEST RUNS. "PRICE" IS THE AVERAGE PRICE SCALAR SET
BY EACH OPERATOR, AND "WAIT/MINS" IS THE WAITING TIME OF THE SERVED PASSENGERS IN MINUTES. "REB." REPRESENTS REBALANCING,
"PRICING" IS THE PRICING POLICY, AND "JOINT" IS THE JOINT PRICING AND REBALANCING POLICY.

| Policy | No Information Sharing | | | Information Sharing | | |
	Reb.	Pricing	Joint	Reb.	Pricing	Joint
Total Reward	18190.5 (272.0)	18813.1 (277.6)	17491.9 (203.1)	18174.2 (326.2)	18879.6 (328.2)	17685.7 (270.9)
Reward Operator 0	9035.4 (206.1)	9236.9 (172.1)	8833.7 (150.9)	9029.4 (228.1)	9489.0 (183.1)	8815.3 (221.2)
Reward Operator 1	9155.0 (289.1)	9576.2 (139.8)	8658.2 (187.5)	9144.9 (295.8)	9390.6 (195.1)	8870.4 (242.2)
Total Rebalancing Costs	1481.0 (94.3)	—	1256.7 (62.7)	1549.3 (113.2)	—	1379.7 (97.0)
Rebalancing Costs Operator 0	776.5 (69.5)	—	610.8 (40.7)	821.7 (83.2)	—	708.8 (81.1)
Rebalancing Costs Operator 1	704.4 (90.2)	—	645.9 (58.7)	727.6 (100.3)	—	671.0 (71.2)
Total Rebalance Trips	316.1 (20.0)	—	257.3 (14.4)	329.6 (22.2)	—	286.2 (22.3)
Rebalance Trips Operator 0	166.4 (18.4)	—	125.1 (9.3)	175.5 (20.7)	—	148.2 (19.2)
Rebalance Trips Operator 1	149.7 (19.9)	—	132.2 (12.4)	154.1 (22.1)	—	138.0 (15.8)
Total Served Demand	3523.6 (33.1)	3467.1 (40.3)	3583.3 (24.7)	3532.8 (38.6)	3463.8 (43.4)	3572.5 (31.1)
Served Demand Operator 0	1755.6 (24.8)	1676.7 (23.9)	1799.7 (13.9)	1760.9 (26.2)	1753.4 (21.8)	1781.1 (23.0)
Served Demand Operator 1	1768.0 (34.8)	1790.4 (18.4)	1783.6 (22.9)	1771.9 (35.0)	1710.4 (26.6)	1791.4 (28.4)
Price Operator 0	—	0.94 (0.01)	0.94 (0.00)	—	0.93 (0.01)	0.97 (0.00)
Price Operator 1	—	0.93 (0.00)	0.95 (0.00)	—	0.93 (0.01)	0.97 (0.00)
Wait/mins Operator 0	1.31 (0.07)	1.05 (0.04)	1.48 (0.07)	1.29 (0.08)	1.09 (0.06)	1.43 (0.07)
Wait/mins Operator 1	1.31 (0.05)	1.07 (0.06)	1.51 (0.06)	1.31 (0.06)	1.02 (0.05)	1.45 (0.08)
Queue Operator 0	8.14 (0.99)	4.69 (0.81)	10.7 (1.0)	8.17 (0.92)	5.14 (0.72)	10.3 (1.3)
Queue Operator 1	7.62 (1.32)	5.67 (0.78)	11.2 (1.1)	7.73 (1.19)	4.54 (0.69)	9.98 (1.24)
Total Demand	4683.0 (73.9)	4423.6 (58.2)	5358.0 (54.1)	4683.0 (73.9)	4425.4 (60.9)	5151.0 (61.4)
Demand Operator 0	2336.2 (56.0)	2121.9 (35.0)	2651.7 (53.0)	2336.2 (56.0)	2278.4 (39.7)	2568.0 (60.4)
Demand Operator 1	2346.8 (39.4)	2301.7 (37.7)	2706.3 (34.5)	2346.8 (39.4)	2147.0 (29.9)	2583.0 (40.0)

outperforms both baselines, though at 650 vehicles the advantage over UD is minimal (17685.7 versus 17652.8). At larger fleet sizes (1050 and 1250), NC achieves slightly higher rewards (20667.8 and 21955.0) due to zero rebalancing costs but serves substantially less demand, indicating a trade-off between operational efficiency and service coverage.

Rebalancing efficiency varies considerably across policies. The Joint Policy demonstrates strategically deployed rebalancing costs that scale with fleet size, reaching 2971.9 at fleet size 1250 with 649.8 trips. In contrast, UD exhibits substantially higher costs (6200.6) with 1491.3 trips, suggesting inefficient vehicle repositioning. The NC policy's zero rebalancing costs come at the expense of served demand, confirming that

strategic rebalancing remains essential for maximizing service coverage with dynamic pricing.

F. Asymmetric Fleet Distribution

We evaluate the joint policy approach across different fleet split configurations between the two competing operators, with results presented in Table VIII. The results reveal several key patterns in the competitive dynamics between operators. First, the total system reward exhibits a non-monotonic relationship with fleet allocation, peaking at the 3:7 split (17689.3) before declining as the imbalance increases further. This suggests that moderate asymmetry in fleet sizes can benefit overall system

TABLE VII
SYSTEM PERFORMANCE COMPARISON ACROSS THREE POLICIES (JOINT POLICY, NC, UD) UNDER VARYING FLEET SIZES FOR NYC MANHATTAN SOUTH. FLEET VEHICLES ARE EVENLY DISTRIBUTED BETWEEN THE TWO OPERATORS. WE PERFORM 10 TESTS FOR EACH CONFIGURATION AND REPORT THE AVERAGE PERFORMANCE WITH STANDARD DEVIATIONS IN PARENTHESES.

Fleet Size	Joint Policy						NC			UD				
	Reward	Served	Rebal. Costs	Rebal. Trips	Price A0	Price A1	Reward	Served	Price Both	Reward	Served	Rebal. Costs	Rebal. Trips	Price Both
450	14199.7 (210.8)	2578.3 (24.6)	821.0 (64.2)	167.6 (16.0)	1.01 (0.00)	1.03 (0.00)	12270.7 (251.8)	2210.0 (43.8)	1.00 (0.00)	13228.7 (136.7)	2550.7 (14.3)	996.8 (69.7)	202.3 (15.9)	1.00 (0.00)
650	17685.7 (270.9)	3572.5 (31.1)	1379.7 (97.0)	286.2 (22.3)	0.97 (0.00)	0.97 (0.00)	16048.4 (416.6)	2894.2 (73.9)	1.00 (0.00)	17652.8 (256.3)	3497.1 (31.4)	1865.5 (83.2)	389.8 (18.1)	1.00 (0.00)
850	20096.1 (273.2)	4612.0 (36.5)	1620.0 (88.7)	331.9 (19.6)	0.86 (0.00)	0.86 (0.00)	18683.9 (461.8)	3367.4 (83.3)	1.00 (0.00)	19916.4 (339.7)	4168.0 (42.7)	3411.9 (132.2)	767.0 (29.3)	1.00 (0.00)
1050	20461.6 (276.2)	5542.5 (44.4)	2059.8 (94.3)	424.0 (21.1)	0.78 (0.00)	0.80 (0.00)	20667.8 (467.6)	3714.9 (83.3)	1.00 (0.00)	20433.6 (558.6)	4533.2 (71.6)	5000.1 (161.6)	1184.5 (37.5)	1.00 (0.00)
1250	20909.3 (294.3)	6291.7 (30.2)	2971.9 (101.7)	649.8 (27.5)	0.76 (0.00)	0.72 (0.00)	21955.0 (397.2)	3934.1 (71.2)	1.00 (0.00)	19831.4 (534.9)	4633.2 (78.1)	6200.6 (219.1)	1491.3 (57.1)	1.00 (0.00)

performance, likely due to reduced direct competition for the same demand.

The reward distribution between operators naturally reflects their fleet sizes, with the larger fleet consistently capturing higher rewards. Notably, Operator 1, with the larger fleet in asymmetric configurations, achieves substantially higher rewards than Operator 0, indicating effective market capture. The pricing strategies adapt accordingly: Operator 0 progressively increases prices from 0.95 to 1.05 as its fleet shrinks, while Operator 1 maintains lower prices (0.92-0.94) to leverage its capacity advantage.

Rebalancing behavior demonstrates notable patterns. The total number of rebalancing trips remains relatively stable across configurations (249.9-300.9), but the distribution shifts dramatically. Operator 0's rebalancing activity drops from 145.1 trips in the balanced 5:5 split to just 8.10 trips in the 1:9 configuration, while Operator 1's rebalancing increases correspondingly. This reflects the reduced need for proactive repositioning when fleet capacity is severely limited. Demand service patterns align with fleet allocations, though not perfectly proportionally. In the 1:9 split, Operator 1 serves 3107.3 requests (88.8% of total served demand) with 90% of the fleet, suggesting diminishing returns to fleet size. The total served demand remains relatively stable (3500.9-3612.3) across configurations, indicating that system capacity rather than competitive dynamics primarily constrains service levels.

G. Regional Wage Heterogeneity

To evaluate the impact of spatially varying passenger income structures on the joint pricing and rebalancing policy, we conduct experiments in the NYC Manhattan South network where passenger wages differ substantially across regions. This reflects real-world scenarios where regional income levels vary based on factors such as local economic conditions, employment patterns, and socioeconomic characteristics. Understanding how the learned policy adapts to this income heterogeneity is critical for practical deployment in markets with diverse economic conditions.

The wage distribution is calibrated based on regional income characteristics, with average passenger wages ranging from approximately $10 per hour in southeastern regions to over $35 per hour in northwestern regions, as shown in Figure 13. This heterogeneity creates spatially varying demand patterns where passenger behavior and willingness to pay differ across regions. The two-operator system has a total fleet size of 650 vehicles, and both operators learn their policies jointly using the A2C algorithm over the same training horizon as previous experiments.
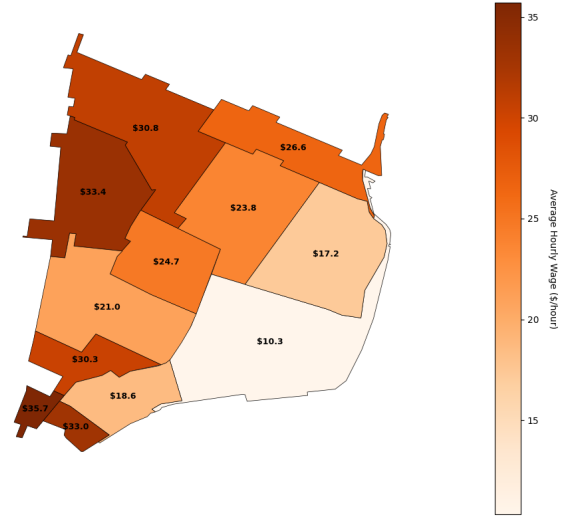

Fig. 13. Average hourly passenger wage distribution across regions in NYC Manhattan South under regional income heterogeneity, showing substantial spatial variation with passenger wages ranging from approximately $10/hour in southeastern regions to over $35/hour in northwestern regions.

Table IX presents the performance metrics for the joint pricing and rebalancing policy under regional wage heterogeneity. The system achieves a total reward of 24624.2 across both operators, with Operator 0 earning slightly higher rewards (12630.5) compared to Operator 1 (11993.7). This asymmetry is reflected in the served demand, where Operator 0 serves 1549.6 passengers compared to Operator 1's 1483.0 passen-

TABLE VIII

SYSTEM PERFORMANCE UNDER VARYING FLEET SPLITS IN NYC MANHATTAN SOUTH WITH DUAL POLICY. WE PERFORM 10 TESTS FOR EACH FLEET CONFIGURATION AND REPORT THE AVERAGE PERFORMANCE WITH STANDARD DEVIATIONS IN PARENTHESES. O0 AND O1 REPRESENT OPERATOR 0 AND OPERATOR 1 RESPECTIVELY. EACH MODEL HAS BEEN TRAINED FOR 100,000 EPISODES AND THE 5:5 SPLIT THEREFORE DIVERGES SLIGHTLY FROM OTHER RESULTS PRESENTED.

Fleet Split (O0:O1)	Total Reward	O0 Reward	O1 Reward	Total Rebal. Trips	O0 Rebal. Trips	O1 Rebal. Trips	O0 Price	O1 Price	Total Served	O0 Served	O1 Served
5:5	17402.5 (191.7)	8510.6 (155.5)	8891.9 (184.5)	263.7 (20.6)	145.1 (13.9)	118.6 (12.8)	0.95 (0.00)	0.94 (0.00)	3587.0 (24.6)	1769.0 (19.4)	1818.0 (21.6)
4:6	17562.2 (236.1)	7497.2 (137.0)	10065.1 (192.0)	268.1 (20.7)	96.0 (10.8)	172.1 (14.8)	0.97 (0.00)	0.94 (0.00)	3586.4 (27.4)	1464.7 (14.6)	2121.7 (24.9)
3:7	17689.3 (209.5)	6010.7 (127.4)	11678.6 (168.8)	249.9 (20.9)	65.6 (7.2)	184.3 (16.4)	0.99 (0.00)	0.92 (0.00)	3612.3 (26.1)	1115.5 (15.0)	2496.8 (21.9)
2:8	17112.8 (238.7)	4309.5 (117.8)	12803.4 (194.0)	278.7 (16.4)	34.7 (4.3)	244.0 (15.1)	1.01 (0.00)	0.93 (0.00)	3553.3 (31.8)	762.8 (12.7)	2790.5 (26.4)
1:9	16626.8 (218.0)	2423.9 (84.6)	14202.9 (226.7)	300.9 (17.6)	8.10 (1.70)	292.8 (18.0)	1.05 (0.00)	0.92 (0.00)	3500.9 (29.1)	393.6 (9.1)	3107.3 (29.9)

gers, despite similar demand for both operators.

TABLE IX

PERFORMANCE METRICS FOR JOINT PRICING AND REBALANCING POLICY WITH REGIONAL WAGE HETEROGENEITY IN NYC MANHATTAN SOUTH. THE NUMBERS IN PARENTHESES INDICATE THE STANDARD DEVIATIONS ACROSS 10 TEST RUNS. ALL VALUES ARE AVERAGED ACROSS ALL REGIONS.

Metric	Joint Policy
Total Reward	24624.2 (490.1)
Reward Operator 0	12630.5 (264.3)
Reward Operator 1	11993.7 (303.2)
Total Rebalancing Costs	2486.7 (99.8)
Rebalancing Costs Operator 0	1161.0 (61.8)
Rebalancing Costs Operator 1	1325.7 (49.8)
Total Rebalance Trips	572.7 (25.2)
Rebalance Trips Operator 0	266.8 (13.4)
Rebalance Trips Operator 1	305.9 (15.1)
Total Served Demand	3032.6 (38.0)
Served Demand Operator 0	1549.6 (20.3)
Served Demand Operator 1	1483.0 (26.1)
Price Operator 0	1.27 (0.00)
Price Operator 1	1.31 (0.00)
Wait/mins Operator 0	0.93 (0.06)
Wait/mins Operator 1	0.95 (0.09)
Queue Operator 0	4.67 (0.98)
Queue Operator 1	5.93 (1.14)
Total Demand	3510.9 (47.3)
Demand Operator 0	1775.4 (20.4)
Demand Operator 1	1735.5 (30.9)
Average Wage	25.8 (0.1)

The rebalancing behavior under wage heterogeneity, shown in Figure 14, reveals a clear pattern: both operators exhibit net positive vehicle flows toward northwestern regions and net negative flows from southeastern regions. Operator 1 demonstrates more aggressive rebalancing, with net flows reaching up to 57 vehicles in northwestern regions and -54 vehicles in southeastern regions, compared to Operator 0's maximum flows of 41 and -71 vehicles respectively. This corresponds to Operator 1's higher total rebalancing costs (1325.7) and more rebalancing trips (305.9) compared to Operator 0 (1161.0 costs

and 266.8 trips).

Figure 15 shows the spatial distribution of demand for both operators. Demand is heavily concentrated in northwestern regions, with Operator 0 serving 300-324 passengers and Operator 1 serving 338-392 passengers in these areas. In contrast, southeastern regions exhibit minimal demand, with fewer than 31 passengers for Operator 0 and fewer than 19 passengers for Operator 1. This demand pattern aligns closely with the income distribution, where northwestern regions have passenger wages exceeding $30 per hour while southeastern regions have passenger wages below $20 per hour. The correlation between higher passenger incomes and higher demand suggests that regions with greater economic prosperity naturally generate more ride requests.

The pricing strategies in Figure 16 demonstrate that both operators set higher price scalars in high-income, high-demand northwestern regions (1.40-1.45) compared to low-income, low-demand southeastern regions (1.15-1.28). Operator 1 maintains a slightly higher average price scalar (1.31) compared to Operator 0 (1.27), consistent with its higher rebalancing costs. The spatial variation in pricing indicates that both operators have learned to charge premium prices in regions where demand is high and passenger incomes are elevated, while moderating prices in areas with lower income levels and demand.

These results demonstrate that the joint pricing and rebalancing policy successfully adapts to regional income heterogeneity by coordinating vehicle allocation and pricing decisions across spatially varying demand patterns. The rebalancing patterns clearly show that both operators concentrate their fleets in high-income, high-demand northwestern regions, accepting the higher rebalancing costs to position vehicles where revenue potential is greatest. This behavior is economically rational: despite northwestern regions having passenger incomes exceeding $30 per hour compared to $10-15 per hour in southeastern regions, the substantially higher demand (300+ passengers versus fewer than 30) justifies the increased rebalancing costs.

(a) Operator 0: Net rebalancing flow per region.   (b) Operator 1: Net rebalancing flow per region.
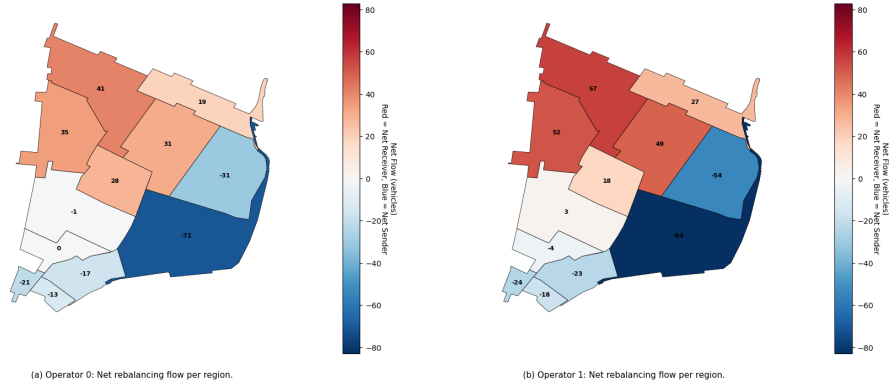
Fig. 14. Net rebalancing flows per region for (a) Operator 0 and (b) Operator 1 under regional income heterogeneity. Positive values (red) indicate net inflow of vehicles, while negative values (blue) indicate net outflow. Both operators rebalance vehicles from low-income southeastern regions toward high-income northwestern regions.
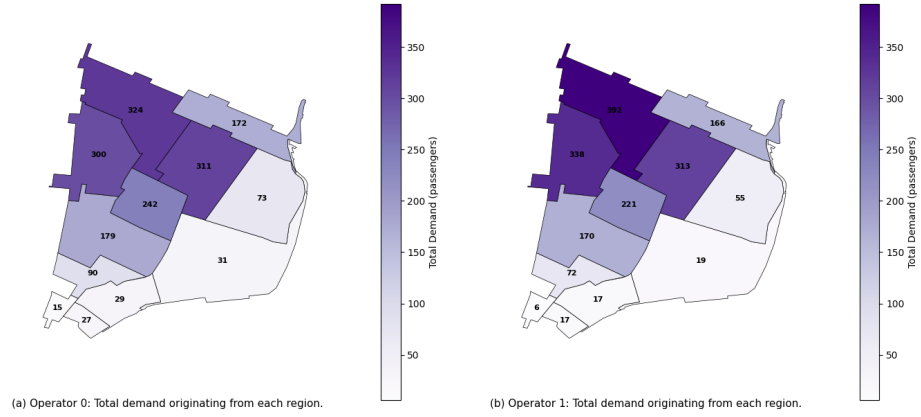

(a) Operator 0: Total demand originating from each region.   (b) Operator 1: Total demand originating from each region.

Fig. 15. Total demand originating from each region for (a) Operator 0 and (b) Operator 1 under regional income heterogeneity. Demand is concentrated in northwestern regions with higher passenger incomes, while southeastern regions with lower passenger incomes exhibit substantially lower demand.

VI. CONCLUSION

This thesis has presented a comprehensive multi-operator reinforcement learning framework for the joint control of pricing and fleet rebalancing in Autonomous Mobility-on-Demand systems. By integrating a discrete choice model directly into the learning loop, this work has formulated a dynamic environment where passenger demand is endogenous and sensitive to the strategic decisions of interacting operators. A central contribution of this research is the demonstration of the framework's capacity to facilitate effective policy learning and analysis across diverse urban environments and economic conditions. The results indicate that the framework is highly responsive to the distinct topological and economic characteristics of the cities in which it is deployed, providing a robust tool for analyzing competitive mobility markets.

A primary finding of this research is the significant impact of environmental context on the performance of different control configurations. The experimental results demonstrated that the transition from a monopolistic to a competitive setting fundamentally alters the efficacy of operational strategies. While joint pricing and rebalancing policies proved most effective for a single operator, the results in competitive scenarios, particularly in high-density environments like Southern Manhattan, showed that specialized policies such as exclusive pricing control often yielded superior performance. These outcomes suggest that in highly adversarial settings, the increased complexity of joint control can lead to mutually destructive interference between operators, such as parallel vehicle repositioning toward identical high-demand areas and the maintenance of uniformly high prices that reduce total market size. This highlights the importance of matching the control dimensionality to the specific competitive landscape to avoid strategic redundancies.

The framework further demonstrated its adaptability through a comparative analysis across San Francisco, Washington DC, and New York City. The results showed that the system successfully navigated varying demand patterns and city scales, producing policies that reflected the unique requirements of each location. For instance, in environments with high demand variance, the operators' learned behaviors emphasized the necessity of proactive fleet management, whereas in stable, high-density areas, pricing emerged as a more critical tool for demand regulation. Furthermore, the experiments on regional wage heterogeneity underscored the framework's sensitivity to socio-economic factors. When exposed to spatially varying income levels, the resulting policies exhibited clear adaptations to local economic conditions. Operators autonomously learned to charge premium prices in high-income regions while si-
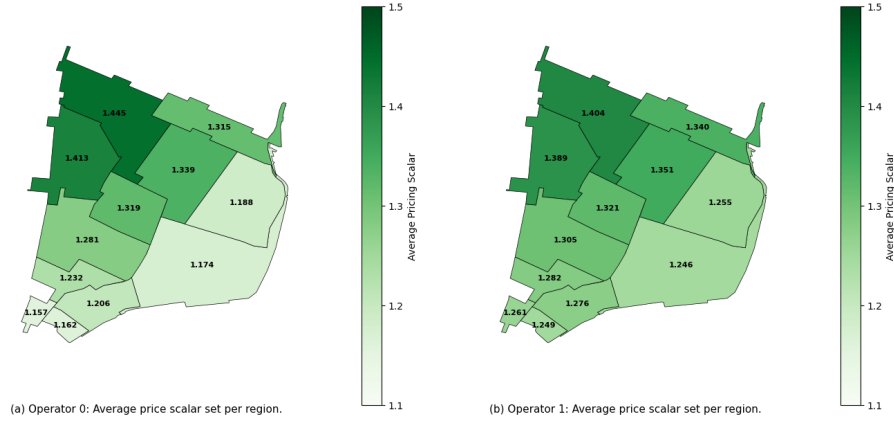
Fig. 16. Average pricing scalars set by (a) Operator 0 and (b) Operator 1 per region under regional income heterogeneity. Both operators set higher prices in northwestern regions with higher passenger incomes and demand, and lower prices in southeastern regions with lower passenger incomes and demand.

multaneously rebalancing their fleets away from lower-income areas to capture higher-value demand. While economically rational from a profit-maximization perspective, these findings reveal a significant tension between private operator incentives and equitable service distribution.

The insights gained from this study suggest several promising avenues for future research to further enhance the realism and social utility of the framework. A critical next step is the refinement of the passenger choice model to explicitly incorporate expected waiting times. Currently, the system responds primarily to price and immediate availability; integrating service latency into the utility function would compel operators to compete on a multidimensional basis, balancing fare optimization with service quality constraints. Finally, the observation that operators naturally neglect low-income regions in favor of high-value demand centers points to the critical need for research into regulatory mechanism design. Future work should investigate how external interventions, such as congestion taxes, rebalancing subsidies, or minimum service level requirements for underserved zones, can be integrated into the multi-operator reward structure. By modeling these regulatory constraints, researchers can identify mechanism designs that align the profit-driven objectives of autonomous fleets with broader social welfare and equity goals. Such advancements would facilitate the development of autonomous mobility systems that are not only operationally efficient and competitively robust but also socially responsible and inclusive.

Acknowledgments

Appendix A
Vehicle Rebalancing Model

The rebalancing model follows the approach used in [3]. At each time step t, the model determines the rebalancing flows $\{y_{i,j,o}^t\}$ for operator o that minimize the total rebalancing cost while satisfying the desired vehicle distribution specified by the actor network. The optimization problem is formulated as:

$$\min \sum_{(i,j)\in\mathcal{E}} c_{i,j,o}^t y_{i,j,o}^t \tag{13}$$

subject to:

$$\sum_{j\neq i}(y_{j,i,o}^t - y_{i,j,o}^t) + m_{i,o}^t \geq \tilde{m}_{i,o}^t, \quad i \in \mathcal{V} \tag{14}$$

$$\sum_{j\neq i} y_{i,j,o}^t \leq m_{i,o}^t, \quad i \in \mathcal{V} \tag{15}$$

$$y_{i,j,o}^t \geq 0, \quad (i,j) \in \mathcal{E} \tag{16}$$

where $\tilde{m}_{i,o}^t$ denotes the number of desired vehicles at region i for operator o at time t. Objective (13) minimizes the rebalancing cost. Constraint (14) ensures that the desired vehicle number is satisfied, accounting for the current idle vehicles $m_{i,o}^t$ and the net inflow of rebalanced vehicles. Constraint (15) limits the rebalancing flow from each region by the number of available idle vehicles. The desired vehicle distribution is calculated by $\tilde{m}_{i,o}^t = \lfloor w_{i,o}^t \sum_{i\in\mathcal{V}} m_{i,o}^t \rfloor$, where $w_{i,o}^t$ is the rebalancing weight output by the actor network for region i.

APPENDIX B
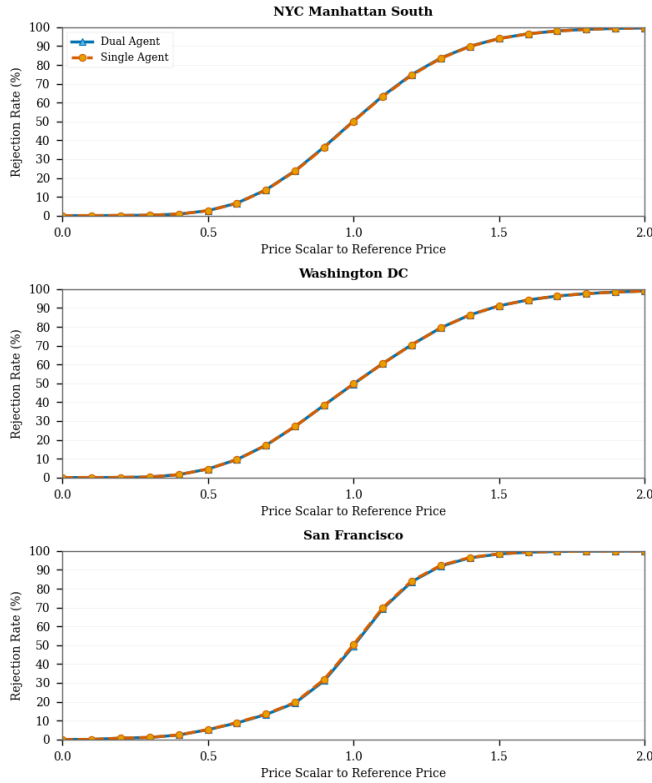## CHOICE MODEL CALIBRATION

Fig. 17. Rejection rate versus price scalar relative to the historical reference price across studied datasets for both single and dual-operator setups, with the model calibrated to a 50% rejection rate at the historical reference price.

APPENDIX C
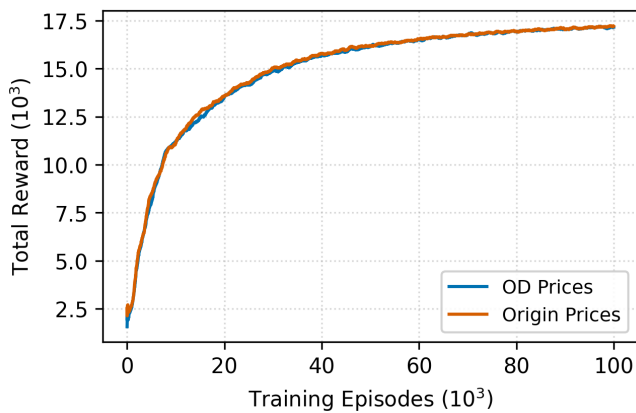## OD VERSUS ORIGIN BASED PRICING

Fig. 18. Episode rewards for OD based versus origin based pricing for the dual-operator setup with joint policy control. The line is smoothed by using a 100 episode rolling mean.

REFERENCES

[1] A. A. Ceder, "Urban mobility and public transport: future perspectives and review," *International Journal of Urban Sciences*, vol. 25, no. 4, pp. 455–479, 2021. [Online]. Available: https://doi.org/10.1080/12265934.2020.1799846

[2] Z. Wang and S. Li, "Competition between autonomous and traditional ride-hailing platforms: Market equilibrium and technology transfer," *Transportation Research Part C: Emerging Technologies*, vol. 165, p. 104728, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0968090X24002493

[3] D. Gammelli, K. Yang, J. Harrison, F. Rodrigues, F. C. Pereira, and M. Pavone, "Graph neural network reinforcement learning for autonomous mobility-on-demand systems," in *2021 60th IEEE Conference on Decision and Control (CDC)*, 2021, pp. 2996–3003.

[4] J. Holler, R. Vuorio, Z. Qin, X. Tang, Y. Jiao, T. Jin, S. Singh, C. Wang, and J. Ye, "Deep reinforcement learning for multi-driver vehicle dispatching and repositioning problem," in *2019 IEEE International Conference on Data Mining (ICDM)*, 2019, pp. 1090–1095.

[5] C. Fluri, C. Ruch, J. Zilly, J. Hakenberg, and E. Frazzoli, "Learning to operate a fleet of cars," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019, pp. 2292–2298.

[6] J. Wen, J. Zhao, and P. Jaillet, "Rebalancing shared mobility-on-demand systems: A reinforcement learning approach," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 2017, pp. 220–225.

[7] X. Tang, M. Li, X. Lin, and F. He, "Online operations of automated electric taxi fleets: An advisor-student reinforcement learning framework," *Transportation Research Part C: Emerging Technologies*, vol. 121, p. 102844, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0968090X20307464

[8] Z. Lei and S. V. Ukkusuri, "Scalable reinforcement learning approaches for dynamic pricing in ride-hailing systems," *Transportation Research Part B: Methodological*, vol. 178, p. 102848, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S019126152300173X

[9] C. Chen, F. Yao, D. Mo, J. Zhu, and X. M. Chen, "Spatial-temporal pricing for ride-sourcing platform with reinforcement learning," *Transportation Research Part C: Emerging Technologies*, vol. 130, p. 103272, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0968090X21002849

[10] J. Huang, L. Huang, M. Liu, H. Li, Q. Tan, X. Ma, J. Cui, and D.-S. Huang, "Deep reinforcement learning-based trajectory pricing on ride-hailing platforms," *ACM Trans. Intell. Syst. Technol.*, vol. 13, no. 3, Mar. 2022. [Online]. Available: https://doi.org/10.1145/3474841

[11] X. Li, C. Schmidt, D. Gammelli, and F. Rodrigues, "Learning joint rebalancing and dynamic pricing policies for autonomous mobility-on-demand," *IEEE Transactions on Intelligent Transportation Systems*, vol. 26, no. 10, pp. 16 619–16 634, 2025.

[12] R. Zhang and M. Pavone, "Control of robotic mobility-on-demand systems: A queueing-theoretical perspective," *The International Journal of Robotics Research*, vol. 35, no. 1-3, pp. 186–203, 2016. [Online]. Available: https://doi.org/10.1177/0278364915581863

[13] M. Pavone, S. L. Smith, E. Frazzoli, and D. Rus, "Robotic load balancing for mobility-on-demand systems," *The International Journal of Robotics Research*, vol. 31, no. 7, pp. 839–854, 2012. [Online]. Available: https://doi.org/10.1177/0278364912444766

[14] G. C. Calafiore, C. Bongiorno, and A. Rizzo, "A robust mpc approach for the rebalancing of mobility on demand systems," *Control Engineering Practice*, vol. 90, pp. 169–181, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0967066119300887

[15] J. Warrington and D. Ruchti, "Two-stage stochastic approximation for dynamic rebalancing of shared mobility systems," *Transportation Research Part C: Emerging Technologies*, vol. 104, pp. 110–134, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0968090X18314104

[16] M. Tsao, R. Iglesias, and M. Pavone, "Stochastic model predictive control for autonomous mobility on demand," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 3941–3948.

[17] R. Iglesias, F. Rossi, K. Wang, D. Hallac, J. Leskovec, and M. Pavone, "Data-driven model predictive control of autonomous mobility-on-demand systems," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 6019–6025.

[18] X. Guo, N. S. Caros, and J. Zhao, "Robust matching-integrated vehicle rebalancing in ride-hailing system with uncertain demand," *Transportation Research Part B: Methodological*, vol. 150, pp. 161–189, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0191261521001004

[19] X. Guo, Q. Wang, and J. Zhao, "Data-driven vehicle rebalancing with predictive prescriptions in the ride-hailing system," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 3, pp. 251–266, 2022.

[20] J. Dai, Q. Zhu, N. Jiang, and W. Wang, "Rebalancing autonomous vehicles using deep reinforcement learning," *International Journal of Circuits, Systems and Signal Processing*, vol. 16, pp. 646–652, 01 2022.

[21] S. Banerjee, C. Riquelme, and R. Johari, "Pricing in ride-share platforms: A queueing-theoretic approach," *SSRN Electronic Journal*, 01 2015.

[22] K. Bimpikis, "Spatial pricing in ride-sharing networks," *SSRN Electronic Journal*, 01 2016.

[23] G. P. Cachon, K. M. Daniels, and R. Lobel, "The role of surge pricing on a service platform with self-scheduling capacity," *Operations Strategy eJournal*, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:4533020

[24] M. Chen, "Dynamic pricing in a labor market: Surge pricing and flexible work on the uber platform," 07 2016, pp. 455–455.

[25] H. Qiu, R. li, and J. Zhao, "Dynamic pricing in shared mobility on demand service," 02 2018.

[26] Y. Guan, "Design and optimization of shared mobility on demand: Dynamic routing and dynamic pricing," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, 2021. [Online]. Available: https://dspace.mit.edu/handle/1721.1/130843

[27] S. Wollenstein-Betech, I. C. Paschalidis, and C. G. Cassandras, "Joint pricing and rebalancing of autonomous mobility-on-demand systems," in *2020 59th IEEE Conference on Decision and Control (CDC)*, 2020, pp. 2573–2578.

[28] W. Tang, H. Wang, Y. Wang, C. Chen, and X. Chen, "A bi-level optimization model for ride-sourcing platform's spatial pricing strategy," *Journal of Advanced Transportation*, vol. 2022, pp. 1–22, 02 2022.

[29] Y. Yang and M. Ramezani, "The intraday competition in a duopoly ride-hailing market," in *Extended abstract submitted for presentation at the 12th Triennial Symposium on Transportation Analysis conference (TRISTAN XII)*, Okinawa, Japan, June 2025, june 22-27, 2025. [Online]. Available: https://tristan2025.org/proceedings/TRISTAN2025_ExtendedAbstract_289.pdf

[30] R. Engelhardt, P. Malcolm, F. Dandl, and K. Bogenberger, "Competition and cooperation of autonomous ridepooling services: Game-based simulation of a broker concept," *Frontiers in Future Transportation*, vol. 3, p. 915219, 06 2022.

[31] S. Wang, G. Homem de Almeida Correia, and H. Lin, "Modeling the competition between multiple automated mobility on-demand operators: An agent-based approach," *SSRN Electronic Journal*, 01 2022.

[32] T. S. Walunj, S. Singhal, V. Kavitha, and J. Nair, "Pricing, competition and market segmentation in ride hailing," in *2022 58th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2022, pp. 1–8.

[33] G. Zardini, N. Lanzetti, G. Belgioioso, C. Hartnik, S. Bolognani, F. Dörfler, and E. Frazzoli, "Strategic interactions in multi-modal mobility systems: A game-theoretic perspective," 09 2023, pp. 5452–5459.

[34] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[35] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 61–80, 2009.

[36] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *International Conference on Learning Representations (ICLR)*, 2017.

[37] D. Khulbe, R. Seshadri, and M. Ben-Akiva, "A probabilistic simulation framework to assess the impacts of ridesharing and congestion charging in new york city," *Data Science for Transportation*, vol. 5, no. 1, pp. 1–25, 2023.

[38] U.S. Census Bureau, "Income in the past 12 months (subject table s1901), 2013 american community survey," data.census.gov, 2013, accessed: 2026-01-13. [Online]. Available: https://data.census.gov/

[39] ——, "Mean income in the past 12 months (subject table s1902), 2013 american community survey," data.census.gov, 2013, accessed: 2026-01-13. [Online]. Available: https://data.census.gov/

[40] ——, "Income in the past 12 months (subject table s1901), 2019 american community survey," data.census.gov, 2019, accessed: 2026-01-13. [Online]. Available: https://data.census.gov/

[41] ——, "Mean income in the past 12 months (subject table s1902), 2019 american community survey," data.census.gov, 2019, accessed: 2026-01-13. [Online]. Available: https://data.census.gov/

[42] ——, "Income in the past 12 months (subject table s1901), 2011 american community survey," data.census.gov, 2011, accessed: 2026-01-13. [Online]. Available: https://data.census.gov/

[43] ——, "Mean income in the past 12 months (subject table s1902), 2011 american community survey," data.census.gov, 2011, accessed: 2026-01-13. [Online]. Available: https://data.census.gov/

[44] D. Gammelli, K. Yang, J. Harrison, F. Rodrigues, F. C. Pereira, and M. Pavone, "Graph meta-reinforcement learning for transferable autonomous mobility-on-demand," 2022. [Online]. Available: https://arxiv.org/abs/2202.07147

[45] B. Donovan and D. Work, "New york city taxi trip data (2010-2013)," 2016. [Online]. Available: https://doi.org/10.13012/J8PN93H8