Iterative Network Pricing for Ridesharing Platforms

Chenkai Yu* Hongyao Ma†

August 16, 2024

Abstract

Ridesharing platforms match riders and drivers, using dynamic pricing to balance supply and demand. The origin-based "surge pricing", however, does not take into consideration market conditions at the destinations of the trips, leading to inefficient driver flows in space and incentivizes drivers to strategize. In this work, we introduce the *Iterative Network Pricing* mechanism, addressing a main challenge in the practical implementation of optimal origin-destination (OD) based prices, that the model for rider demand is hard to estimate. Assuming that the platform's surge algorithm clears the market for each origin in real-time, our mechanism updates the OD-based price adjustments week-over-week, using only information immediately observable during the same time window in the prior weeks. For stationary market conditions, we prove that our mechanism converges to an outcome that is approximately welfare-optimal. Using data made public by the City of Chicago, we illustrate (via simulation) the iterative updates under our mechanism for morning rush hours, demonstrating not only substantial welfare improvements but also remarkable robustness amid the significant fluctuations of market conditions from early 2019 through the end of 2020.

1 Introduction

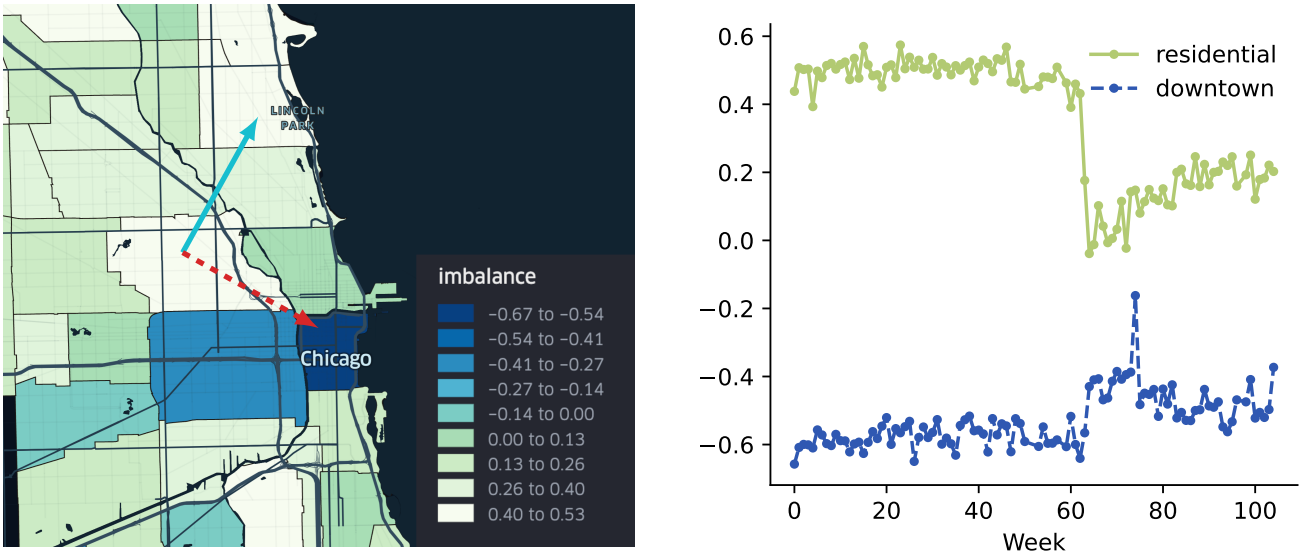
Ridesharing platforms have revolutionized the way people commute and travel in recent years. Central to their success is the emphasis placed on providing reliable transportation services to riders.[1] When rider demand surpasses available driver supply in an area, the dynamic "surge pricing" adopted by these platforms increases the prices for trips originating from the area [Castillo et al., 2017]. This maintains sufficient density

The authors would like to thank Itai Ashlagi, Francisco Castro, Yeon-Koo Che, Peter Cohen, Shuyang Du, Peter Frazier, Daniel Freund, Nikhil Garg, Sergey Gitlin, Danny Guo, Jason Hartline, Ramesh Johari, Yash Kanoria, Cinar Kilcioglu, Zhen Lian, Ilan Lobel, Elliot Lipnowski, Will Ma, Jake Marcinek, Chris Mosch, Hongseok Namkoong, Hamid Nazerzadeh, Michael Ostrovsky, David C. Parkes, Zhiwei Qin, Daniel Russo, Garrett van Ryzin, Amin Saberi, Shreyas Sekar, Peng Shi, George Sorg-Langhans, Amir Tafreshian, Carmen Wang, Jeffrey Wang, Chiwei Yan, and participants at various seminars and conferences for helpful comments and discussions.

*Decision, Risk, and Operations Division, Columbia Business School, cyu26@gsb.columbia.edu

†Decision, Risk, and Operations Division, Columbia Business School, hongyao.ma@columbia.edu

[1] Lyft, for example, advertises itself as "a ride whenever you need one" (https://www.lyft.com/, accessed May 16, 2023), and Uber aims to provide "transportation that is as reliable as running water" (https://www.uber.com/en-au/blog/melbourne/transportation-that-is-as-reliable-as-running-water/, accessed May 16, 2023).



(a) Average flow imbalance of different neighborhoods.

(b) Flow imbalance over time, 2019-2020.

Figure 1: The trip flow imbalance during weekday morning rush hours in Chicago: average of first 9 weeks of 2020 (left) and temporal trends for two neighborhoods (right). Week 60 corresponds to the COVID shutdown in March 2020, which substantially changed the market dynamics and reduces the imbalance of residential areas due to work-from-home arrangements. See Section 5 for more details.

of driver supply in space, and guarantees that rider wait times would not exceed a few minutes [Rayle et al., 2014, Castillo et al., 2017, Yan et al., 2020].

When a platform's origin-based pricing algorithm does not appropriately factor market conditions at the destination of the trips, however, the resulting spatial distribution of driver supply can be highly inefficient. This issue becomes especially pronounced when demand patterns are driven more by trip destinations than origins. For instance, during morning rush hours in major cities, a substantial number of riders seek trips that end downtown, irrespective of whether they start in the city center or in a residential area. However, relatively few riders request trips that end in the residential neighborhoods.

Figure 1a illustrates the (relative) flow imbalance of different neighborhoods in Chicago during morning rush hours. The flow imbalance of a location is defined as (outflow − inflow)/(outflow + inflow), where the *outflow* and *inflow* refer respectively to the number of trips that originate from and end at the location per hour. Downtown Chicago (usually referred to as "The Loop") is a highly popular destination. Prior to the COVID shut-down, an average of over 4000 trips ending in this area per hour while just slightly more than 1000 trips originate here. In contrast, the North Side's Lake View neighborhood sees over 1000 trips originating per hour, while only approximately 300 trips end there in the same amount of time.

Under origin-based pricing, the two trips depicted in Figure 1a will have identical prices since they start from the same location and cover the same distance. This incentivizes drivers to cherrypick, since a driver arriving in the residential area is likely to be dispatched a high surge trip immediately, while a driver who arrive in the downtown area may need to deadhead (i.e. relocate without a rider) for a substantial period of time before receiving any dispatch. This disparity in drivers' continuation earnings also suggests that origin-based pricing is inefficient in its use of drivers' time, and fails to properly distribute drivers towards areas where supply is more valuable.

In the rapidly growing literature on ridesharing platforms, some existing work already studies origin-destination (OD) based pricing to provide efficient and reliable transportation. In particular, Ma et al. [2022] show that it is not only welfare-optimal but also incentive-aligned (i.e. it is to the drivers' best interest to accept all trip dispatches) to set OD-based prices of the following form:

$$\begin{aligned}(\text{trip price}) = (\text{trip cost}) + (&\text{marginal value of a driver at trip origin})\\ &- (\text{marginal value of a driver at trip destination}).\end{aligned}$$

We refer to the difference between the marginal value of a driver at a trip's origin versus its destination as the *OD-based adjustment*. This is the network externality of completing the trip, and also the value of teleporting a driver from the destination back to the origin. Since a driver in the residential area is intuitively more "valuable" than one in downtown, the trip to downtown would command a higher price under optimal OD pricing. This not only reduces driver supply in the already saturated downtown area (by reducing rider demand for trips heading downtown), but also compensates the driver for the lower continuation earnings from downtown onward with a higher earning from the current trip.

Despite these theoretical advancements, the practical implementation of optimal OD-based pricing presents significant challenges. In theory, the marginal values of drivers at different locations are derived from the dual variables of the welfare-optimization problem. Formulating the primal and dual programs, however, necessitates a rider demand model (i.e. the expected number of riders requesting a ride at any given price for each OD pair), which proves very difficult to estimate in practice for a number of reasons:

- *Endogeneity.* Consider a trip typically priced at \$20. The platform is likely to have a solid understanding of rider demand at this price, as well as the local price sensitivity (due to natural price fluctuations resulting from fluctuating supply levels, or through local price exploration). However, estimating rider demand at \$10 or \$30 presents considerable challenges. First, the trip may never have been priced as low as \$10. Furthermore, although prices might have surged to \$30 during inclement weather or special events, such observations are not as useful when pricing trips under normal market conditions.
- *Slow market equilibration.* An abrupt drop in the price of a trip from \$20 to \$10 may induce a modest demand increase in the short run, but one could plausibly anticipate that the demand will continue to increase for many weeks, as riders react by, for instance, deciding against renewing their monthly work parking passes. Long-term price exploration is, however, not only costly but also inequitable (subjecting rider groups to different prices for extended periods of time).
- *Heterogeneity in space and time.* The shape of the demand function may vary substantially across trip origin, destination, and time of the week. These variations could be attributed to neighborhood-specific characteristics, public transit availability, fluctuating traffic conditions, etc. Consequently, assuming a universal functional form for all trips (whether in terms of prices or surge multipliers) is unlikely to yield accurate estimations of rider demand.

1.1 Our Results

In this work, we propose the *iterative network pricing* (INP) mechanism, which enables a ridesharing platform to implement optimal OD-based pricing without the need to estimate a rider demand model. Assuming that the platform's dynamic pricing algorithm finds the origin-based market-clearing multipliers in real-time, our mechanism iteratively updates the OD-based additive price adjustments week-over-week, relying only on

information immediately observable by the platform during the preceding weeks.

Take the Chicago morning rush hour as an example. Since a driver heading to the residential area has higher continuation earnings than one taking a rider downtown, it is intuitive that driver supply is less valuable in downtown, and a trip to downtown should be priced higher. thus a trip to downtown should be priced higher. Without a demand model, the platform cannot compute the optimal price difference between the two trips. The platform can, however, set the downtown trip's price to be \$1 higher than the trip to the residential area for the following week. With this additive, *relative* adjustment, the platform's surge algorithm finds the market-clearing origin-based multipliers which determine the new *absolute* levels of the trip prices. The difference in drivers' continuation earnings from trips to these two locations is likely to be reduced, but a significant disparity may still exist. At this point, the platform could increase the price difference to \$2, then review the market-clearing outcomes after another week, and so forth.

In practice, iteratively updating the OD-based adjustments can be effective, as Figure 1b suggests that the spatial imbalances of trip flows remain relatively stationary week-over-week (in the absence of turmoils such as the COVID shutdown, which took place around week 60). Moreover, it is relatively easy to adapt the platform's online-control style origin-based pricing algorithm to accommodate additive adjustments. To operationalize this idea, we address address in this work the following questions:

- Given a current origin-based market-clearing outcome, what would constitute a good "direction" for updating the OD-based adjustments for a metropolitan area with more than two neighborhoods?
- What should be an appropriate stepsize that the platform takes each week in this desired direction?
- Will this iterative process converge? If so, will it converge to the optimal OD-based prices?

Model. We study a continuous time, non-atomic setting, modeling the operation of a ridesharing platform during a particular time window of the week, e.g. morning rush hours on Wednesdays.[2] There is a fixed amount of driver supply and a set of discrete locations. Each origin-destination (OD) pair, representing a potential trip, is characterized by the duration of the trip, the cost incurred by a driver for fulfilling the trip, and a rider demand function which specifies the amount of riders requesting this trip (per unit of time) at different price levels. The platform does not know the entire rider demand functions, but can observe the demand levels and the (local) price sensitivities at the realized trip prices each week.

We assume that supply and demand are stationary during the focal time window in any given week, and that the market condition does not change week-over-week.[3] We also assume that during each week, the platforms' dynamic pricing algorithm determines origin-based multipliers that *clear the market*, meaning that (i) riders are picked up if and only if they are willing to pay the prices of their trips, (ii) all drivers are dispatched to pick up a rider or relocate to some location, and (iii) the inflow and outflow of drivers are balanced for every location. Finally, we assume that drivers always follow the platform's dispatches.

Main Results. We first establish a welfare theorem (Lemma 1), that an outcome is welfare-optimal if and only if it can be supported by OD-based trip prices in a competitive equilibrium (CE). We show that optimal CE prices consist of three parts: (i) the cost incurred by drivers to complete the trip; (ii) a multiplicative part (i.e.

[2]We focus on a particular time of the week because the demand pattern during the morning rush hours on a Wednesday is typically similar to that of the previous Wednesday (see Figure 1b), but can be very different from earlier or later times on the same day (see Figure 16 in Appendix C). A platform may adopt multiple INP mechanisms in parallel for other time windows with their own distinct demand patterns, e.g. weekday mid-day, evening rush hours, or weekend bar hours.

[3]Appendix C.1 discusses the stationarity assumption for the market conditions within our focal time window. Briefly, the duration of most trips is significantly shorter than the period over which market conditions remain relatively unchanged.

a multiplier that is uniform across all locations, times the duration of the trip) reflecting the opportunity cost of drivers' time for fulfilling the trip; and (iii) an additive adjustment dependent on the origin and destination of the trip, representing the externality imposed on the entire system by moving a driver from the trip origin to the trip destination.

A platform is not able to directly compute the optimal OD-based adjustments without the rider demand model. When the OD-based adjustments are absent or suboptimal, the platform's dynamic pricing algorithm will need different origin-based multipliers to clear the market for all locations. We show that for any given set of OD-based adjustments, there exists a unique origin-based market-clearing outcome, provided that the platform coordinates the relocation of drivers who are not assigned rider trips (Lemma 2). Furthermore, the welfare loss of such outcomes (relative to the welfare-optimal outcome) can be bounded by the differences in the multipliers across different locations (Theorem 1), despite the fact that quantifying the achieved welfare or the optimal social welfare is not possible without a demand model.

An *iterative network pricing* (INP) mechanism starts with an initial origin-based market-clearing outcome with no OD-based adjustments. At each subsequent timestep (e.g. the following week), the mechanism updates the OD-based adjustments and observes the new outcome resulting from the market-clearing process. By establishing a correspondence to Newton's method for solving systems of equations, we prove that our mechanism converges to an outcome that is approximately welfare-optimal when the market condition is stationary (Theorem 2). Using data from Chicago's morning rush hours, we illustrate (via simulation) the iterative updates under our mechanism, showcasing not only substantial welfare improvements but also remarkable robustness amid the significant fluctuations of market conditions from early 2019 through the end of 2020.

From a technical perspective, our problem is similar to various iterative mechanism design scenarios, where neither primal nor dual objective can be evaluated but a direction for improving the dual objective may be identified. Our setting, however, introduces unique challenges. In contrast to settings such as iterative combinatorial auctions [Parkes and Ungar, 2000], where the market designer can "query" the demand model at any hypothetical set of prices (by soliciting agents' preferred bundles), we are only able to observe rider demand at actual origin-based, market-clearing outcomes. This necessitates a characterization of such outcomes, which we provide in Section 3.

Furthermore, the natural idea of gradient descent specifies a direction to update not only the OD-based adjustments but also the origin-based multiplier, but this cannot be implemented since the origin-based multipliers are determined by the market-clearing process and are not directly controlled by the platform. To guarantee progress, a platform needs to anticipate how the market-clearing multipliers will change as the platform makes (small) updates to the OD-based adjustments. We show that this is possible using information immediately available from the past market-clearing outcomes, including trip throughput, driver earnings, and (local) price sensitivity (Lemma 3). The INP mechanism we propose updates OD-based adjustments towards the direction to equalize (the linear approximations of) the origin-based multipliers at all locations, and takes an appropriately sized step to ensure the multipliers don't change too drastically at any given location.

1.2 Related Work

The rapidly growing literature on ridesharing platforms covers both empirical studies of current platforms and theoretical analysis of market designs. Empirical studies have demonstrated the effectiveness of origin-based surge pricing in improving reliability and efficiency [Hall et al., 2015], increasing driver supply during

high-demand times [Chen and Sheldon, 2015], and creating incentives for drivers to relocate to high-demand areas [Lu et al., 2018]. Additional topics include consumer surplus [Cohen et al., 2016, Castillo, 2020], the labor market of Uber drivers [Hall et al., 2017], the longer-term labor market equilibration [Hall and Krueger, 2016], the value of flexible work arrangements [Chen et al., 2019, 2020, Xu et al., 2020] and the gender earnings gap associated to "learning-by-doing" [Cook et al., 2018]. To the best of our knowledge, no empirical study addresses the challenge of demand model estimation for ridesharing platforms in the presence of endogeneity, heterogeneity, and different short term and long term demand elasticity.

A variety of policy and regulation related topics have also received attention from the theoretical modeling community, including the growth of two-sided platforms [Lian and Van Ryzin, 2021, Fang et al., 2019], competition between platforms [Lian et al., 2021, Ahmadinejad et al., 2019, Fang et al., 2020], operations in the presence of autonomous vehicles [Ostrovsky and Schwarz, 2019, Lian and van Ryzin, 2020], and utilization-based minimum wage regulations [Asadpour et al., 2023].

For dynamic matching in ridesharing platforms, Ashlagi et al. [2018], Dickerson et al. [2018] and Aouad and Saritaç [2020] focus on matching between riders and drivers and the pooling of shared rides, considering the online arrival of supply and demand in space. Further, Kanoria and Qian [2020], Qin et al. [2020] and Özkan and Ward [2020] design policies that dispatch drivers from areas with relatively abundant supply, while Cai et al. [2019] and Pang et al. [2017] look at the role of information availability and transparency in platform design. Castro et al. [2022] demonstrate how to use drivers' waiting times (in airport driver queues) to align incentives and reduce inequity in earnings when some trips are more lucrative than the others. State-dependent dispatching [Banerjee et al., 2018, Castro et al., 2020], driver admission control [Afèche et al., 2023] and capacity planning [Besbes et al., 2022] are also studied using queueing-theoretical models.

On the pricing side, Castillo et al. [2017] and Yan et al. [2020] demonstrate the importance of origin-based dynamic pricing in maintaining supply density in space and reducing waiting times for riders; [Banerjee et al., 2015] demonstrate the robustness of dynamic pricing in comparison to static pricing when the platform may have imprecise information on the arrival rate of rider demand; Bimpikis et al. [2019] and Besbes et al. [2020] study revenue-optimal pricing when supply and demand are imbalanced in space; Garg and Nazerzadeh [2020] show that additive instead of multiplicative pricing is more incentive aligned for drivers when prices need to be origin-based only; Rheingans-Yoo et al. [2019] study pricing in the presence of driver location preferences; Freund and van Ryzin [2021] discuss the cycles of volatile driver supply when the platform adopts two prices when pricing trips, and Yu et al. [2023] analyzes the price cycles arising from divers collectively turning offline to drive up prices.

The closest works to ours are Ma et al. [2019] and Cashore et al. [2022], which we briefly discussed earlier in the introduction. Ma et al. [2019] study a deterministic, complete information setting to model spatial imbalance and temporal variation of supply and demand. They demonstrate that various existing market failure in today's ridesharing platforms are symptoms of inefficient pricing, and prove that optimal OD-based pricing with the structure discussed earlier is both welfare-optimal and incentive-compatible (i.e. it forms a subgame-perfect equilibrium for all drivers to accept all dispatches from the platform). Cashore et al. [2022] generalize the model to a stochastic setting, and prove that the same pricing structure is still optimal and incentive aligned in large markets. Both works, however, assume that the platform has access to the correct demand model, which has proven to be very difficult to estimate in practice. In this work, we propose mechanisms that optimizes origin-destination based pricing via iterative updates without using any model for rider demand.

2 Preliminaries

In this section, we introduce a continuous time, non-atomic model for the operation of a ridesharing platform for *a particular time period of the week*, e.g. morning rush hour on Wednesdays. During this period of time, we assume that driver supply and rider demand are stationary. We characterize welfare-optimal, competitive equilibrium (CE) outcomes for the time-window of interest, which serve as the target of the week-over-week iterative pricing process discussed in Sections 3 and 4.

Let $m > 0$ be the total units of driver supply, and $n > 1$ be the number of discrete locations, which form a set $\mathcal{L} = \{1, 2, \ldots, n\}$. Each pair of locations $(i, j) \in \mathcal{L}^2$ represents a *trip* from i to j. Let $d_{i,j} > 0$ be the *duration* of the trip, i.e. it takes a driver $d_{i,j}$ units of time to drive from i to j. Let $c_{i,j} \geq 0$ be the *cost* of the trip, including the cost of time, driving, fuel, wear-and-tear, etc., for a driver who completes a trip from i to j (with or without a rider). Trip durations and trip costs do not have to be symmetric, i.e. for $i \neq j$, it is possible that $d_{i,j} \neq d_{j,i}$ and $c_{i,j} \neq c_{j,i}$. Let $q_{i,j} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be the *demand function* for the trip, where $q_{i,j}(r)$ is the amount of riders (per unit of time) traveling from i to j and willing to pay at least r. Assume that (i) $q_{i,j}(\cdot)$ is continuously differentiable and strictly decreasing (i.e. $q'_{i,j}(r) < 0$ for all $r > 0$), and (ii) $\int_0^\infty q_{i,j}(r)\,\mathrm{d}r < \infty$, i.e. the total rider value is finite.[4]

Assume that the platform can observe the total driver supply m, trip duration $\boldsymbol{d} = (d_{i,j})_{(i,j)\in\mathcal{L}^2}$, and trip costs $\boldsymbol{c} = (c_{i,j})_{(i,j)\in\mathcal{L}^2}$. The rider demand model $\boldsymbol{q} = (q_{i,j})_{(i,j)\in\mathcal{L}^2}$ is not known to the platform, though the platform is able to observe the demand and local price sensitivity at the current trip prices. Formally, if a trip $(i, j) \in \mathcal{L}^2$ is priced at r, the platform observes $q_{i,j}(r)$ and $q'_{i,j}(r)$. The tuple $(m, \boldsymbol{d}, \boldsymbol{c}, \boldsymbol{q})$ specifies an *economy*, i.e. the market conditions during this focal time window (e.g. morning rush hour on Wednesdays).

Outcomes. An *outcome* describes the flow of drivers and riders in space, as well as payments for drivers and riders associated to each trip. Formally, an outcome is a triple $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{p})$, where for all $(i, j) \in \mathcal{L}^2$, $x_{i,j} \geq 0$ is the rider flow rate from location i to location j, $y_{i,j} \geq 0$ is the rate of driver flow from i to j (either completing a rider trip, or relocating without a rider), and $p_{i,j} \geq 0$ is the price of the (i, j) trip that is collected from all (i, j) riders who are picked up, and paid to all drivers traveling from i to j with a rider.[5] An outcome is feasible and stationary if the followings hold:

(F1) Trip prices are non-negative, i.e. $\forall i, j \in \mathcal{L},\ p_{i,j} \geq 0$.

(F2) The rider flow does not exceed the driver flow for any trip, i.e. $\forall i, j \in \mathcal{L},\ x_{i,j} \leq y_{i,j}$.

(F3) The outcome does not require more drivers than the available supply, i.e. $\sum_{i,j\in\mathcal{L}} d_{i,j} y_{i,j} \leq m$.

(F4) The driver flow into and out of every location is balanced, i.e. $\forall i \in \mathcal{L},\ \sum_{j\in\mathcal{L}} y_{i,j} = \sum_{j\in\mathcal{L}} y_{j,i}$.

Unless stated otherwise, when we mention an outcome in the rest of the paper, it is assumed to be feasible and stationary. We also assume throughout the paper that riders requesting the same trip are picked up in decreasing order of their values.[6] For each $(i, j) \in \mathcal{L}^2$, let $v_{i,j}(\cdot)$ be the inverse function of $q_{i,j}(\cdot)$, i.e. $v_{i,j}(s)$

[4] We work with demand functions in this paper, which determine riders' value distributions, and vice versa. Let $Q_{i,j} > 0$ be the number of riders (per unit of time) who would request a trip $(i, j) \in \mathcal{L}^2$ at zero price, and $F_{i,j}$ be the cumulative distribution function (CDF) of their value. Then $q_{i,j}(r) = Q_{i,j} \cdot (1 - F_{i,j}(r))$ for all $r \geq 0$. The assumption that $q_{i,j}(\cdot)$ is continuously differentiable and strictly decreasing means that the probability density function of rider values $\mathrm{d}F_{i,j}(r)/\mathrm{d}r$ is continuous and strictly positive for all $r \geq 0$. The second assumption $\int_0^\infty q_{i,j}(r)\,\mathrm{d}r < \infty$ implies $\lim_{r\to+\infty} q_{i,j}(r) = 0$.

[5] By construction, the outcomes are strictly budget-balanced. All results presented in this paper still hold if the platform takes a fixed percentage of drivers' surplus from every trip as commission.

[6] This is without loss for welfare-optimization, and always holds when rider best-response is satisfied, i.e. riders are picked up if and only if they are willing to pay the trip price (see condition (R1)). The mechanisms we propose in this paper satisfy rider

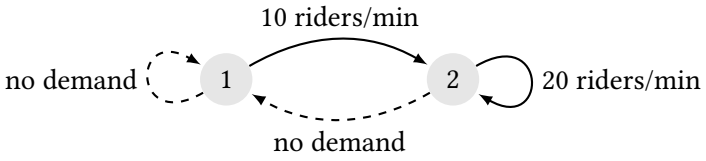


Figure 2: A stylized economy illustrating the demand pattern during morning rush hours.

the value of the s^{th} rider traveling from i to j. In this way, when $x_{i,j} \in [0, q_{i,j}(0)]$ riders are picked up for the (i,j) trip, their total value can be written as $\int_0^{x_{i,j}} v_{i,j}(s)\,\mathrm{d}s$. Since $q_{i,j}(r)$ is continuous and strictly decreasing for each $(i,j) \in \mathcal{L}^2$, its inverse $v_{i,j}(s)$ is well defined and is strictly decreasing for all $s \in (0, q_{i,j}(0))$. For convenience, we define $v_{i,j}(s) = \infty$ for $s \le 0$ and $v_{i,j}(s) = -\infty$ for $s > q_{i,j}(0)$.

Cycles and Driver Surplus. Under an outcome $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{p})$, $\boldsymbol{y}$ specifies the overall flow of drivers in space. To describe the movements of individual drivers, as well as their payoffs, it is convenient to use the cycle representation of driver flows. A *directed cycle* is a sequence of trips $((i_1, i_2), (i_2, i_3), \ldots, (i_\ell, i_1))$ visiting ℓ distinct locations before returning to the origin of the first trip. Let C be the set of all directed cycles. We say a cycle $\kappa \in C$ *covers* a trip $(i,j) \in \mathcal{L}^2$ if the cycle includes the trip from location i to location j, and we denote this as $(i,j) \in \kappa$.

By flow decomposition theorem (see Theorem 3.5 of Ahuja et al. [1988]), every balanced flow can be decomposed into a weighted sum of cycles, and vice versa. Formally, given driver flow $\boldsymbol{y} \in \mathbb{R}_{\ge 0}^{n^2}$,

$$\forall i \in \mathcal{L},\ \sum_{j\in\mathcal{L}} y_{i,j} = \sum_{j\in\mathcal{L}} y_{j,i} \iff \exists \boldsymbol{w} \in \mathbb{R}_{\ge 0}^{|C|},\ \forall i,j \in \mathcal{L},\ \sum_{\kappa\in C} w_\kappa \mathbb{1}\{(i,j)\in\kappa\} = y_{i,j}. \tag{1}$$

Here $\mathbb{1}\{\cdot\}$ is the indicator function, and we refer to $\boldsymbol{w}$ as a *(cycle) decomposition* of $\boldsymbol{y}$. Intuitively, in any balanced driver flow, every driver's movement in space forms a directed cycle, and w_κ is the amount of drivers traveling along cycle κ. Given an outcome $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{p})$, for each directed cycle $\kappa \in C$, denote

$$\mu_\kappa \triangleq \frac{\sum_{(i,j)\in\kappa}(p_{i,j} - c_{i,j})}{\sum_{(i,j)\in\kappa} d_{i,j}}. \tag{2}$$

Consider a driver traveling along the cycle κ. If for every trip on the cycle, either (i) the driver picks up a rider, or (ii) the trip price is zero when the driver relocates without a rider, then μ_κ represents the surplus rate of the driver, i.e. the total payments minus costs from all trips divided by the total trip duration.

Example 1 (The Morning Rush)**.** Consider the economy illustrated in Figure 2 with two locations: the residential neighborhood (location 1) and the downtown area (location 2). Every minute during the morning rush hours, 10 riders travel from the residential neighborhood to downtown, 20 riders travel within downtown, but no rider demands a trip ending in the residential neighborhood.

The demand functions are given by $q_{1,2}(r) = 10\exp(-r/40)$, $q_{2,2}(r) = 20\exp(-r/10)$, and $q_{1,1}(\cdot) = q_{2,1}(\cdot) = 0$. This effectively assumes that riders' values are exponentially distributed with mean 40 and 10 for trips $(1,2)$ and $(2,2)$. Trips within each location takes 10 minutes ($d_{1,1} = d_{2,2} = 10$) and trips between locations take 20 minutes ($d_{1,2} = d_{2,1} = 20$). Finally, there are $m = 240$ available drivers, and we assume for simplicity that all trip costs are zero, i.e. $c_{i,j} = 0$ for all $i,j \in \{1,2\}$.

best-response at all times.

A Feasible Outcome. We can verify that the outcome $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{p})$ specified as follows satisfies conditions (F1)–(F4), and therefore is feasible and stationary:

- the the rider flow is given by $x_{1,2} = 1$, $x_{2,2} = 20$, and $x_{1,1} = x_{2,1} = 0$,
- the driver flow is given by $y_{1,2} = y_{2,1} = 1$, $y_{2,2} = 20$, and $y_{1,1} = 0$, and
- the trip prices are $p_{1,1} = 46.05$, $p_{1,2} = 92.10$ and $p_{2,1} = p_{2,2} = 0$.

Intuitively, this outcome results from the platform's naively "clearing the market" by origin, i.e. setting a price rate for each location such that supply meets demand — location 2 has excess supply thus is priced at zero; location 1 trips are priced at \$4.605 per minute (i.e. $p_{1,1} = d_{1,1} \times 4.605 = 46.05$ and similarly $p_{1,2} = 92.10$); riders are picked up if and only if their value is (weakly) above the price. The social welfare (total rider value minus driver cost) achieved is $\int_0^{x_{1,2}} v_{1,2}(s)\,\mathrm{d}s + \int_0^{x_{2,2}} v_{2,2}(s)\,\mathrm{d}s - 0 = 332.10$ dollars/min. This outcome is inefficient, since all riders for the $(2, 2)$ trip are picked up, even those with values very close to zero, whereas many high-value $(1, 2)$ riders did not get a ride.

Cycles and Driver Surplus. There are three directed cycles in this economy: $C = \{\kappa_1, \kappa_2, \kappa_3\}$, where $\kappa_1 = ((1, 1))$ $\kappa_2 = ((2, 2))$ are "self loops" and $\kappa_3 = ((1, 2), (2, 1))$ is the cycle between the two locations (note that cycles $((1, 2), (2, 1))$ and $((2, 1), (1, 2))$ are the same). Driver flow $\boldsymbol{y}$ can be decomposed into cycles with weights $w_{\kappa_1} = 0$, $w_{\kappa_2} = 20$, and $w_{\kappa_3} = 1$. We can see that this outcome is inequitable: those drivers on κ_2 (fulfilling the $(2, 2)$ trip) gets zero surplus, whereas drivers on κ_3 (looping around the two locations) gets a surplus of $(p_{1,2} + 0)/(d_{1,2} + d_{2,1}) = 2.30$ per minute.

2.1 Optimal Outcome and CE Prices

We now formalize the optimization problem, characterize the structure of competitive equilibrium (CE) prices, and prove a welfare theorem, that an outcome is welfare-optimal if and only it is a CE.

Definition 1 (Competitive Equilibrium (CE))**.** A CE is an outcome $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{p})$ such that:

(R1) Riders are picked up if and only if their value is higher than the price: $\forall (i, j) \in \mathcal{L}^2,\ x_{i,j} = q_{i,j}(p_{i,j})$.

(D1) Trips with relocating driver flow are priced at zero: $\forall (i, j) \in \mathcal{L}^2,\ (y_{i,j} > x_{i,j} \implies p_{i,j} = 0)$.

(D2) All drivers are used up, if there exist a cycle with a strictly positive surplus rate:

$$\max_{\kappa \in C} \mu_\kappa > 0 \implies \sum_{i,j \in \mathcal{L}} d_{i,j} y_{i,j} = m. \tag{3}$$

(D3) No driver is dispatched, if the surplus rate is strictly negative for all cycles:

$$\max_{\kappa \in C} \mu_\kappa < 0 \implies \forall (i, j) \in \mathcal{L}^2,\ y_{i,j} = 0. \tag{4}$$

(D4) The driver flow $\boldsymbol{y}$ can be decomposed into cycles with weights $\boldsymbol{w} \in \mathbb{R}_{\geq 0}^{|C|}$ such that every cycle with positive driver flow has the highest surplus rate:

$$\exists \boldsymbol{w},\ \forall \kappa \in C, \quad w_\kappa > 0 \implies \mu_\kappa = \max_{\kappa' \in C} \mu_{\kappa'}. \tag{5}$$

Intuitively, under a CE, all riders and drivers get their highest possible utility given the trip prices. (R1) can be interpreted as *rider best-response*, i.e. every rider decides whether they would like to be picked up given

their trip price. (D1)–(D4) together represent *driver best-response*, i.e. every driver is no worse off than in the scenario where they have the freedom to (i) choose any cycle and get paid for all trips on the cycle, and (ii) choose to not drive for the platform at all.[7]

When the rider demand model $\boldsymbol{q} = (q_{i,j}(\cdot))_{(i,j)\in\mathcal{L}^2}$ (and thus the inverse functions $(v_{i,j}(\cdot))_{(i,j)\in\mathcal{L}^2}$) is known, the following convex program maximizes social welfare over rider and driver flows:

$$\begin{aligned}
\underset{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^{n^2}}{\text{maximize}} \quad & \sum_{i,j\in\mathcal{L}}\left(\int_0^{x_{i,j}} v_{i,j}(s)\,\mathrm{d}s - c_{i,j}y_{i,j}\right) && \text{(6a)}\\
\text{subject to} \quad & x_{i,j} \le y_{i,j}, \qquad \forall i,j\in\mathcal{L}, && \text{(6b)}\\
& \sum_{i,j\in\mathcal{L}} d_{i,j}y_{i,j} \le m, && \text{(6c)}\\
& \sum_{j\in\mathcal{L}} y_{i,j} = \sum_{j\in\mathcal{L}} y_{j,i}, \quad \forall i\in\mathcal{L}. && \text{(6d)}
\end{aligned}$$

The objective (6a) represents the total value of riders who are picked up minus the total cost incurred by the drivers, per unit of time. Given that $v_{i,j}(\cdot)$ is strictly decreasing for all $(i,j) \in \mathcal{L}^2$, its integral is strictly concave, and thus the objective function (6a) is strictly concave. Constraints (6b)–(6d) correspond to conditions (F2)–(F4) under which the outcome is feasible and stationary.[8]

Let $p_{i,j}$, ω, and ϕ_i be the dual variables corresponding to constraints (6b), (6c), and (6d), respectively. The dual problem of the welfare-optimization problem (6) is given as follows:

$$\begin{aligned}
\underset{\boldsymbol{p}\in\mathbb{R}^{n^2},\ \omega\in\mathbb{R},\ \boldsymbol{\phi}\in\mathbb{R}^n}{\text{minimize}} \quad & m\omega + \sum_{i,j\in\mathcal{L}}\int_0^{q_{i,j}(p_{i,j})} (v_{i,j}(s) - p_{i,j})\,\mathrm{d}s && \text{(7a)}\\
\text{subject to} \quad & p_{i,j} = c_{i,j} + d_{i,j}\omega + \phi_i - \phi_j, \quad \forall i,j\in\mathcal{L}, && \text{(7b)}\\
& p_{i,j} \ge 0, \qquad \forall i,j\in\mathcal{L}, && \text{(7c)}\\
& \omega \ge 0. && \text{(7d)}
\end{aligned}$$

Lemma 1 (Welfare Theorem). *A feasible rider and driver flow $(\boldsymbol{x},\boldsymbol{y})$ is welfare-optimal if and only if there exist trip prices $\boldsymbol{p}$ such that the outcome $(\boldsymbol{x},\boldsymbol{y},\boldsymbol{p})$ is a competitive equilibrium (CE). Moreover, for any CE outcome $(\boldsymbol{x},\boldsymbol{y},\boldsymbol{p})$, there exist $\omega \ge 0$ and $\boldsymbol{\phi}\in\mathbb{R}^n$ such that*

$$p_{i,j} = c_{i,j} + d_{i,j}\omega + \phi_i - \phi_j, \quad \forall i,j\in\mathcal{L}. \tag{8}$$

This structure of CE prices are aligned with the structure of optimal and incentive compatible prices in existing literature [Ma et al., 2022, Cashore et al., 2022] (discussed in Section 1). We provide the proof of this Lemma in Appendix A.2. Briefly, we use standard observations about the connection between duality and

[7]For competitive equilibrium in two-sided markets, it is typically not necessary to require that the price be zero when supply exceeds demand, since a negative price implies a violation of the best-response on the supply side. This is the case for our setting as well, but we keep condition (D1) for simplicity of notation and presentation.

[8]Note that we did not impose non-negativity constraints on $\boldsymbol{x}$ or $\boldsymbol{y}$, or the constraint that $x_{i,j} \le q_{i,j}(0)$ for all $(i,j)\in\mathcal{L}$. This is because by construction, $v_{i,j}(s) = \infty$ for $s < 0$ and $v_{i,j}(s) = -\infty$ for $s > q_{i,j}(0)$, and thus these conditions are satisfied by any solution where the objective is not negative infinity.

market equilibrium [Shapley and Shubik, 1971, Bertsekas, 1990, Parkes and Ungar, 2000, Ma et al., 2022]. The dual variables $\boldsymbol{p}$ and ω can be interpreted as trip prices and drivers' surplus rate, and the difference between the dual and the primal objectives corresponds to the total violation of the driver and rider best-response conditions (R1)–(D4). An outcome is therefore welfare-optimal if and only if it is a competitive equilibrium.

Note that we work with alternative primal and dual formulations (see Appendix A.1) that provide sufficient flexibility to map any set of prices $\boldsymbol{p}$ to the dual variables. Driver best-response, however, requires that drivers be indifferent towards all feasible paths between each pair of locations. This allows us to prove that CE prices must have the structure in (8), consisting of three parts:

- $c_{i,j}$: the cost incurred by the driver for completing each trip $(i, j) \in \mathcal{L}^2$,
- $d_{i,j}\omega$: the opportunity cost of the trip, which is multiplicative and scales with trip duration. ω represents the marginal value of driver supply, and is referred to as the *multiplier*,
- $\phi_i - \phi_j$: the additive, origin-destination (OD) based adjustment, corresponding to the marginal value of a driver at the trip origin i minus that of a driver at the trip destination j. Intuitively, this can be interpreted as the value of "teleporting" a driver from j back to i.

There are $n - 1$ instead of n degrees of freedom in the OD-based adjustments, since shifting all ϕ_i's by the same constant has no impact on the dual problem (7). Technically, this is because one of the flow-balance constraints (6d) is redundant (the flow-balance of any $n - 1$ locations imply that the remaining location is also balanced). We prove in Proposition 1 in Appendix A.3 that the optimal dual solution is unique up to a constant shift in $\boldsymbol{\phi}$, and in the rest of this paper, we use ω^* and $\boldsymbol{\phi}^*$ to denote the optimal multiplier and the optimal OD-based adjustments with $\phi_n^* = 0$.

Example 1 (Continued)**.** Under the optimal outcome, the platform picks up (per minute) $x_{1,2} = 4$ units of riders going from the residential area to downtown, and $x_{2,2} = 8$ units of riders traveling within downtown. The optimal welfare is approximately 459.91 dollars per minute, and the corresponding driver flow is $y_{1,1} = 0$, $y_{1,2} = y_{2,1} = 4$, and $y_{2,2} = 8$. Despite the fact that there is no demand from downtown to the residential area ($x_{2,1} = 0$), some drivers relocate without a rider ($y_{2,1} = 4$) in order to satisfy the high-value demand on the opposite direction.

CE Prices. The CE prices are $p_{1,1} = p_{2,2} = 9.16$, $p_{1,2} = 36.65$ and $p_{2,1} = 0$, corresponding to the optimal dual solution $\omega^* = 0.916$ (dollars per minute) and $\phi_1^* - \phi_2^* = 18.33$ (dollars). It is intuitive that $\phi_1^* > \phi_2^*$, i.e. the marginal value of a driver in the residential area is higher than that in downtown. Adjusting trip prices using $\boldsymbol{\phi}^*$, the platform charges $p_{2,2} = c_{2,2} + d_{2,2}\omega^* + \phi_2^* - \phi_2^* = 10\omega^* > 0$ (so that the low-value downtown riders are not picked up), while setting $p_{2,1} = c_{2,1} + d_{2,1}\omega^* + \phi_2^* - \phi_1^* = 0$ and relocate drivers from downtown back to the residential area.

Equitable Driver Earnings. For drivers cycling between locations 1 and 2, each cycle takes $d_{1,2} + d_{2,1} = 40$ minutes, from which the drivers get a surplus of $(p_{1,2} + 0)/40 = 0.916 = \omega^*$ per minute. The drivers who loop around location 2 get the same surplus rate: each $(2, 2)$ trip takes $d_{2,2} = 10$ minutes, and the surplus rate is also $p_{2,2}/10 = 0.916 = \omega^*$.

In practice, without access to the rider demand model, a platform is not able to directly compute $\boldsymbol{\phi}^*$, the marginal value of drivers by location. Instead, under the outcome (discussed in Example 1 earlier in this section) that results from the platforms' naively clearing the market by origin without any OD-based adjustments, the platform achieves only $332.1/459.81 = 72\%$ of the optimal social welfare. Moreover, drivers earn substantially more from trips originating from the residential area than from downtown. The large gap

of earnings suggests that drivers in the residential area are more valuable than those in downtown (which is indeed the case, as $\phi_1^* > \phi_2^*$). It is therefore natural to consider making some small adjustment, for example $\phi_1 - \phi_2 = 1$, and then see what happens the next week. In the next section, we characterize the outcome of such "surge pricing" algorithms that clear the market using origin-based multipliers in the presence of additive OD-based adjustments.

3 Origin-Based Market Clearing

In this section, we characterize the outcome of origin-based market-clearing with OD-based adjustments, i.e. when the platform determines origin-based "surge" multipliers to balance supply and demand for each location. We prove that the market-clearing outcome is unique when driver relocation is coordinated by the platform, and that the welfare-loss (in comparison to the optimal outcome discussed in Section 2) is bounded by the difference in the multipliers at different locations. This characterization allows us to discuss in Section 4 how a platform updates the OD-based adjustments week-over-week, based on the observed market-clearing outcomes in prior weeks.

As we have shown in Lemma 1, given the optimal OD-based adjustments $\boldsymbol{\phi}^*$, a platform can clear the market and optimize welfare using a uniform multiplier ω^* for all locations. In the absence of OD-based adjustment or when the adjustments are suboptimal, it is impossible to satisfy the best-response conditions for both drivers and riders. Instead, as in practice, the platforms adopt *origin-based multipliers* $\boldsymbol{\pi} = (\pi_i)_{i\in\mathcal{L}}$ to balance supply and demand for each location, using prices of the form:

$$p_{i,j} = c_{i,j} + d_{i,j}\pi_i + \phi_i - \phi_j, \quad \forall i, j \in \mathcal{L}. \tag{9}$$

For instance, Example 1 illustrates such an outcome for the morning rush hours with $\boldsymbol{\phi} = \mathbf{0}$. With origin-based multipliers $\pi_1 = 4.605$ and $\pi_2 = 0$, (i) riders are picked up if and only if their values are above the price, and (ii) prices are non-negative but trips with excess supply are priced at zero. When a platform's dynamic pricing algorithm sets $\boldsymbol{\pi}$ to achieve the desired properties (i) and (ii), however, the resulting outcome may not be unique and will depend the actions of drivers who are relocating without a rider.[9] This lack of uniqueness in the potential market-clearing outcome imposes a challenge on the week-over-week update of the OD-based adjustments.

Coordinated Driver Relocation. In this work, we assume that the platform coordinates the relocation of drivers who are not dispatched rider trips. This can be modeled using *phantom demand functions* $\tilde{\boldsymbol{q}} = (\tilde{q}_{i,j}(\cdot))_{(i,j)\in\mathcal{L}^2}$. Specifically, if the price of a trip $(i, j) \in \mathcal{L}$ is $p_{i,j} \geq 0$, the platform dispatches $\tilde{q}_{i,j}(p_{i,j})$ units of drivers to relocate from i to j (per unit of time) without a rider, in addition to dispatching $q_{i,j}(p_{i,j})$ drivers to rider trips. The platform has flexibility in choosing any phantom demand function $\tilde{\boldsymbol{q}}$ such that:

(I) For all $i, j \in \mathcal{L}$, $\tilde{q}_{i,j}(\cdot)$ is non-negative, continuously differentiable, and decreasing.

(II) For all $i, j \in \mathcal{L}$, $d_{i,j}\tilde{q}_{i,j}(0) > m$.

Condition (II) guarantees that the relocation flow at zero price is sufficiently high for each trip, so that the platform will provide a relocation recommendation for every driver who is not dispatched to pick up a

[9] For instance, when a half of the drivers who are undispatched at location 2 drives to location 1 while the other half stay in location 2, the market-clearing multipliers are given by $\pi_1 = 5.99$ and $\pi_2 = 0$, and corresponding rider and driver flows are $x_{1,1} = x_{2,1} = 0$, $x_{1,2} = 0.5$, $x_{2,2} = 20$, and $y_{1,1} = 0$, $y_{1,2} = y_{2,1} = 0.5$, $y_{2,2} = 22$.

rider. When prices are strictly positive, however, it is inefficient to have drivers relocate without a rider (and as a result not getting paid for the trip). Define this "slackness" of $\tilde{\boldsymbol{q}}$ as

$$e_{i,j} \triangleq \sup_{r \geq 0} r \cdot \tilde{q}_{i,j}(r), \quad \forall i, j \in \mathcal{L}. \tag{10}$$

Then $\sum_{i,j \in \mathcal{L}} e_{i,j}$ is an upper bound for the welfare loss due to driver relocation at positive prices under $\tilde{\boldsymbol{q}}$. Unless stated otherwise, we assume (I) and (II) in the rest of the paper.

Definition 2 (Origin-based market clearing)**.** For any economy $(m, \boldsymbol{d}, \boldsymbol{c}, \boldsymbol{q})$, any phantom demand $\tilde{\boldsymbol{q}}$, and any OD-based additive adjustment $\boldsymbol{\phi} \in \mathbb{R}^n$, a set of origin-based multipliers $\boldsymbol{\pi} \in \mathbb{R}^n$ is *market clearing* if there exists a feasible outcome $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{p})$ with prices given by (9) such that:

(MC1) Riders best-respond to the trip prices (i.e. $x_{i,j} = q_{i,j}(p_{i,j})$ for all $(i, j) \in \mathcal{L}^2$), and the driver relocation is coordinated by the phantom demand (i.e. $y_{i,j} - x_{i,j} = \tilde{q}_{i,j}(p_{i,j})$ for all $(i, j) \in \mathcal{L}^2$).

(MC2) All drivers are dispatched, i.e. $\sum_{i,j \in \mathcal{L}} d_{i,j} y_{i,j} = m$.

By defining the origin-based market-clearing outcome, we effectively postulate capabilities and behaviors of the platform's origin-based dynamic pricing algorithm. Condition (MC2), for instance, implies that the platform dispatches all drivers who are on the road to either a rider or a relocation trip, and does not recommend any driver to stop driving and go offline. Among the conditions for competitive equilibria, rider best-response (R1) is still satisfied since it is imperative for the ridesharing platforms to guarantee reliable service for riders (see Footnote 1). In contrast, the driver best-response conditions are relaxed, mirroring the current state of platforms, as without optimal OD-based adjustments a platform cannot ensure best-response for both sides of the market.

Lemma 2 (Existence and Uniqueness)**.** *For any economy $(m, \boldsymbol{d}, \boldsymbol{c}, \boldsymbol{q})$, any phantom demand function $\tilde{\boldsymbol{q}}$, and any OD-based price adjustment $\boldsymbol{\phi} \in \mathbb{R}^n$, there exists a unique vector of origin-based multipliers $\boldsymbol{\pi} \in \mathbb{R}^n$ that clears the market.*

We prove this lemma in Appendix A.4. Briefly, fixing the economy and the OD-based adjustments, the flow balance constraints (only $n - 1$ of them are "independent", as discussed in Section 2) together with the condition on total supply (MC2) impose n non-linear constraints on the origin-based multipliers $\boldsymbol{\pi} \in \mathbb{R}^n$. Exploiting the structure of the flow-balance constraints, we show that the set of multipliers that induce balanced driver flows form a one-dimensional manifold, within which exactly one set of multipliers lead to an outcome that dispatch exactly m units of drivers.

By Definition 2 and Lemma 2, we know that the origin-based multipliers $\boldsymbol{\pi}$ and the corresponding market-clearing outcome $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{p})$ are fully determined by the OD-based adjustments $\boldsymbol{\phi}$. We refer to $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\pi})$ as the *origin-based market-clearing outcome* associated with OD-based adjustment $\boldsymbol{\phi} \in \mathbb{R}^n$.

One challenge for analyzing the social welfare achieved by an orogin-based market-clearing outcome using primal and dual formulations (6)–(7) is that the dual variables satisfying constraint (7b) does not provide sufficient flexibility to represent the trip prices (9) (which now have more than n degrees of freedom). As a result, we use the following equivalent primal formulation with additional constraints.

$$\underset{\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{n^2},\ \boldsymbol{z} \in \mathbb{R}^n}{\text{maximize}} \quad \sum_{i,j \in \mathcal{L}} \left(\int_0^{x_{i,j}} v_{i,j}(s)\, \mathrm{d}s - c_{i,j} y_{i,j} \right) \tag{11a}$$

$$\text{subject to} \quad x_{i,j} \le y_{i,j}, \qquad \forall i, j \in \mathcal{L}, \tag{11b}$$

$$\sum_{j \in \mathcal{L}} d_{i,j} y_{i,j} = z_i, \qquad \forall i \in \mathcal{L}, \tag{11c}$$

$$\sum_{i \in \mathcal{L}} z_i \le m, \tag{11d}$$

$$\sum_{j \in \mathcal{L}} y_{k,j} = \sum_{i \in \mathcal{L}} y_{i,k}, \quad \forall k \in \mathcal{L}. \tag{11e}$$

Intuitively, replacing the supply constraint (6c) with a *local* supply constraint (11c) in addition to the *global* supply constraint (11d), the resulting dual problem below now includes additional variables π_i (corresponding to (11c)) that can represent origin-based price multipliers:

$$\underset{\boldsymbol{p} \in \mathbb{R}^{n^2},\, \omega \in \mathbb{R},\, \boldsymbol{\pi} \in \mathbb{R}^n,\, \boldsymbol{\phi} \in \mathbb{R}^n}{\text{minimize}} \quad m\omega + \sum_{i,j \in \mathcal{L}} \int_0^{q_{i,j}(p_{i,j})} (v_{i,j}(s) - p_{i,j})\, \mathrm{d}s \tag{12a}$$

$$\text{subject to} \quad p_{i,j} = c_{i,j} + d_{i,j}\pi_i + \phi_i - \phi_j, \quad \forall i, j \in \mathcal{L}, \tag{12b}$$

$$p_{i,j} \ge 0, \qquad \forall i, j \in \mathcal{L}, \tag{12c}$$

$$\omega \ge \pi_i, \qquad \forall i \in \mathcal{L}, \tag{12d}$$

$$\omega \ge 0. \tag{12e}$$

For each location $i \in \mathcal{L}$, π_i can be interpreted as the marginal value of a unit of drivers' time for drivers who are on trip originating from location i. Intuitively, when the origin-based multipliers are higher for some locations than the others, or when some of the multipliers are negative, we are making inefficient use of drivers' time. In the following result, we establish an upper bound on the welfare suboptimality of any origin-based market-clearing outcome (i.e. the difference between the achieved and the highest possible social welfare), which consists of this inefficiency as well as the slack introduced by driver relocation.

Theorem 1. *Given any origin-based market-clearing outcome* $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\pi})$*, the welfare suboptimality is upper bounded by*

$$\sum_{i,j \in \mathcal{L}} d_{i,j} y_{i,j} \left(\max\left\{ \max_{k \in \mathcal{L}} \pi_k,\, 0 \right\} - \pi_i \right) + \sum_{i,j \in \mathcal{L}} p_{i,j}(y_{i,j} - x_{i,j}), \tag{13}$$

which is further upper bounded by

$$m \left(\max\left\{ \max_{k \in \mathcal{L}} \pi_k,\, 0 \right\} - \min_{i \in \mathcal{L}} \pi_i \right) + \sum_{i,j \in \mathcal{L}} e_{i,j}. \tag{14}$$

Importantly, the bounds (13) and (14) are computable using only information immediately available from a current market-clearing outcome: origin-based multipliers, trip prices, and driver and rider flows. This is of practical relevance as it allows a platform to assess the welfare suboptimality even when the overall social welfare cannot be determined due to the absence of a demand model.

Proof. Given any origin-based market-clearing outcome $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{p}, \boldsymbol{\pi}, \boldsymbol{\phi})$, let $z_i = \sum_{j \in \mathcal{L}} d_{i,j} y_{i,j}$ for all $i \in \mathcal{L}$, and denote $\omega = \max\{\max_{i \in \mathcal{L}}\{\pi_i\},\, 0\}$. Intuitively, z_i is the amount of drivers (at any moment of time) who are on trip from each location $i \in \mathcal{L}$, and ω represents the "best-response" driver surplus rate (i.e. when the driver

has the freedom to choose any route or to not drive for the platform). It is straightforward to verify that $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ and $(\boldsymbol{p}, \omega, \boldsymbol{\pi}, \boldsymbol{\phi})$ form feasible solutions of the primal (11) and the dual (12), respectively.

The primal objective (11a) represents welfare, which is equal to the total driver and rider surplus under the market-clearing outcome. The dual objective (12a), on the other hand, can be interpreted as the total "best response" surplus of drives and riders, i.e. every driver gets a surplus rate of ω per unit of time, and all riders who are willing to pay the price of their trips are picked up. The difference

$$\sum_{i,j\in\mathcal{L}} \int_{x_{i,j}}^{q_{i,j}(p_{i,j})} (v_{i,j}(s) - p_{i,j})\,\mathrm{d}s + \sum_{i,j\in\mathcal{L}} p_{i,j}(y_{i,j} - x_{i,j}) + \sum_{i\in\mathcal{L}} z_i(\omega - \pi_i) + \omega\Big(m - \sum_{i\in\mathcal{L}} z_i\Big), \tag{15}$$

and can is therefore the total violation of driver and rider best-response conditions:

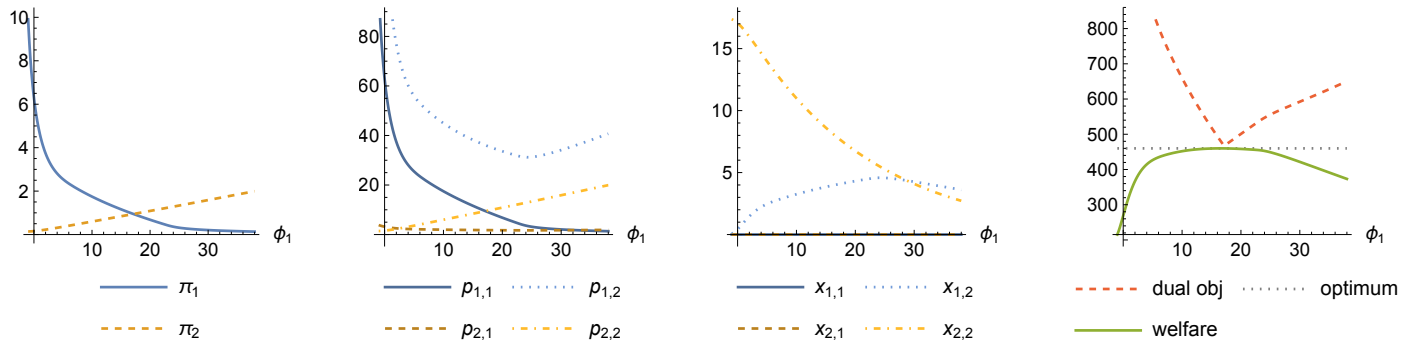
- The first term $\sum_{i,j\in\mathcal{L}} \int_{x_{i,j}}^{q_{i,j}(p_{i,j})} (v_{i,j}(s) - p_{i,j})\,\mathrm{d}s$ represents the violation of rider best-response: $q_{i,j}(p_{i,j})$ riders who like to be picked up for trip $(i,j) \in \mathcal{L}^2$ when the price is given by $p_{i,j}$, but only $x_{i,j}$ of them are picked up and the rest each looses $v_{i,j}(s) - p_{i,j}$ in comparison to their "best-response" outcome. Given (MC1), we know that this term is zero under any market-clearing outcome.
- The second term $\sum_{i,j\in\mathcal{L}} p_{i,j}(y_{i,j} - x_{i,j})$ is the violation of driver best-response due to drivers being dispatched to relocate without getting paid when prices are positive. This can be rewritten as $\sum_{i,j\in\mathcal{L}} p_{i,j}\tilde{q}_{i,j}(p_{i,j})$, which is bounded by $\sum_{i,j\in\mathcal{L}} e_{i,j}$ by definition of the slack (10).
- The third term $\sum_{i\in\mathcal{L}} z_i(\omega - \pi_i)$ is the violation of driver best-response due to the fact that at any moment in time z_i drivers are *on-trip* originating from location $i \in \mathcal{L}$ but the the surplus rate π_i these drivers get may be lower than the highest possible surplus rate ω.
- The last term $\omega\Big(m - \sum_{i\in\mathcal{L}} z_i\Big)$ represents the violation of driver best-response due to the fact that some drivers may not be dispatched even when some other drivers are getting a positive surplus of ω per unit of time. This terms is always zero under market-clearing outcomes due to (MC2).

The welfare-suboptimality of the current market-clearing outcome (i.e. the current primal objective) is bounded by the difference between the primal and dual objectives (15), since the dual objective is always weakly above the optimal primal objective. Combining the non-zero (i.e. the 2nd and 3rd) terms of (15) gives us (13), which we can see is upper bounded by (14). □

We now revisit the morning rush-hour example discussed in Section 2, and illustrate the origin-based market-clearing outcomes at different OD-based adjustments.

Example 1 (Continued)**.** Recall that in the morning rush hour example, all riders request trips ending in the downtown area (location 2). To coordinate driver relocation, the platform adopts phantom demand function $\tilde{q}_{i,j}(r) = 24(\max\{0, 1 - r/5\})^4$ for all $i, j \in \{1, 2\}$, i.e. relocation occurs only when the trip price falls below \$5. Fixing $\phi_2 = 0$, Figure 3 illustrates the market-clearing outcomes as ϕ_1 varies. When $\phi_1 = 0$, which is the case of the naïve origin-based pricing, $\pi_1 \gg \pi_2$, meaning that the origin-based multiplier for location 1 has to be substantially higher than that at location 2. The resulting outcome is very inefficient, since drivers spend time serving many $(2, 2)$ riders with very low values, while many very high value $(1, 2)$ riders were not picked up.

As ϕ_1 increases, $p_{2,1}$ decreases relative to $p_{2,2}$. This leads to higher $x_{1,2}$ and lower $x_{2,2}$, since more drivers at location 2 relocate to location 1 to pick up the $(1, 2)$ riders instead of picking up the lower value riders at location 2. This reduces the gap in the multipliers and improves welfare. Starting from the market-clearing



(a) Multipliers. (b) Trip prices. (c) Rider flow rates. (d) Objectives.

Figure 3: The origin-based market clearing outcome for different ϕ_1 in Example 1 (fixing $\phi_2 = 0$).

outcome with $\phi_1 = 0$, $\pi_1 \gg \pi_2$ indicates that the marginal value of drivers is higher at location 1 than that at location 2. It is intuitive that the platform should increase ϕ_1 until the $\pi_1 - \pi_2$ is sufficiently small, in which case welfare is approximately optimal (Theorem 1).

With more than two locations, however, it is unclear what would be a good direction for updating the OD-based adjustment $\boldsymbol{\phi}$, how big the stepsizes should be, and whether the process will converge to the optimal outcome (or at all). We answer these questions in the next section.

4 Iterative Network Pricing

In this section, we introduce the Iterative Network Pricing (INP) mechanism, which observes the market-clearing outcomes during a specific time window (e.g. Wed. morning rush hour) in preceding timesteps, and updates the OD-based additive price adjustments to be adopted in this time window for the upcoming timestep.[10] We prove the main result of this paper, that when the market condition is stationary over time, the INP mechanism converges to an outcome that is approximately welfare-optimal.

We first define a general iterative pricing mechanism. Let $t = 0, 1, \ldots$ be the index of timesteps. Time $t = 0$ represents the status quo with $\boldsymbol{\phi} = \mathbf{0}$, i.e. the platform clears the market by origin without any OD-based adjustments. Time $t = 1$ and $t = 2$ represent the first and second timesteps (e.g. weeks), respectively, for the mechanism to start updating the OD-based adjustments, and so forth. We make the following assumptions:

(A1) The economy is stationary over time, i.e. $(m, \boldsymbol{d}, \boldsymbol{c}, \boldsymbol{q})$ remains the same for all $t \geq 0$. The phantom demand $\tilde{\boldsymbol{q}}$ used to coordiante relocation also stays the same for all $t \geq 0$.[11]

(A2) During each timestep $t \geq 0$, the platform is able to observe the origin-based market-clearing outcome $(\boldsymbol{x}^{(t)}, \boldsymbol{y}^{(t)}, \boldsymbol{p}^{(t)}, \boldsymbol{\phi}^{(t)}, \boldsymbol{\pi}^{(t)})$ corresponding to OD-based adjustments $\boldsymbol{\phi}^{(t)}$, as well as the (local) price sensitivity $(q'_{i,j}(p^{(t)}_{i,j}))_{(i,j)\in\mathcal{L}^2}$ at the current prices.

$h_t \triangleq ((\boldsymbol{x}^{(t')}, \boldsymbol{y}^{(t')}, \boldsymbol{p}^{(t')}, \boldsymbol{\phi}^{(t')}, \boldsymbol{\pi}^{(t')}, (q'_{i,j}(p^{(t')}_{i,j}))_{(i,j)\in\mathcal{L}^2}))_{t'\leq t-1}$ denotes the *history* up to time t, and we denote the set of all possible histories under an iterative mechanism up to time t as $\mathcal{H}_t$.

[10] The timesteps represent the frequency at which the platform updates the OD-based adjustments. When a timestep corresponds to a week, for example, the platform determines the OD-based adjustments for Wednesday morning rush hour each week using market-clearing outcomes from the prior weeks during the same time window.

[11] Assumption (A1) is used to establish the optimality of the INP mechanism in theory. The mechanism is well defined when the market condition is not stationary (see Section 5.2 for simulation results for Chicago morning rush hours, 2019–2020).

Definition 3 (Iterative Pricing Mechanisms)**.** An iterative pricing mechanism starts with no OD-based adjustment $\boldsymbol{\phi}^{(0)} = \mathbf{0}$ at time $t = 0$. At each $t \geq 1$, the mechanism determines $\boldsymbol{\phi}^{(t)}$ based on the observable market-condition $(m, \boldsymbol{d}, \boldsymbol{c})$ and the history of past market-clearing outcomes h_t.

Given history h_t, a seemingly straightforward method for updating the OD-based adjustment $\boldsymbol{\phi}^{(t)}$ is to use the gradient of the dual objective. Recall from Section 3 that given any $\boldsymbol{\phi}$ and the corresponding market-clearing $\boldsymbol{\pi}$, we can construct a feasible solution to the dual problem (12): $\boldsymbol{p}$ is determined by constraint (12b), and $\omega = \max\{\max_{i\in\mathcal{L}} \pi_i, 0\}$ achieves the smallest dual objective once $(\boldsymbol{p}, \boldsymbol{\pi}, \boldsymbol{\phi})$ is fixed. In this way, we can rewrite the dual objective (12a) as

$$\text{Dual}(\boldsymbol{\phi}, \boldsymbol{\pi}) \triangleq m \max\left\{\max_{i\in\mathcal{L}} \pi_i,\ 0\right\} + \sum_{i,j\in\mathcal{L}} \int_{c_{i,j}+d_{i,j}\pi_i+\phi_i-\phi_j}^{\infty} q_{i,j}(r)\,\mathrm{d}r. \tag{16}$$

It is straightforward to verify that the gradient (or a subgradient, at non-differentiable points) of Dual with respect to $\boldsymbol{\pi}$ and $\boldsymbol{\phi}$ can be computed using the trip prices, distances, and rider flow at current prices, quantities that are immediately observable by the platform at the current market-clearing outcome.[12] However, this natural idea of gradient descent with respect to $(\boldsymbol{\phi}, \boldsymbol{\pi})$ does not solve our problem for two reasons:

- First, $\nabla_{\boldsymbol{\phi}}\text{Dual}(\boldsymbol{\phi}, \boldsymbol{\pi}) = \mathbf{0}$ at any outcome where the rider flow is balanced, in which case the gradient does not provide a direction for updating $\boldsymbol{\phi}$, even when the outcome is far from optimal.
- Moreover, while the gradient with respect to the origin-based multipliers $\boldsymbol{\pi}$ is non-zero, $\boldsymbol{\pi}$ is the result of the market-clearing process, and is not determined by the platform ahead of time.

Since the OD-based adjustments $\boldsymbol{\phi}$ are the only dual variable directly controlled by the platform, a first challenge a mechanism needs to address is to "anticipate" how the market-clearing $\boldsymbol{\pi}$ will change as the platform updates $\boldsymbol{\phi}$. The following result shows that this is possible using only information immediately available to the platform at any market-clearing outcome. Recall Lemma 2, that there exists a unique market-clearing $\boldsymbol{\pi} \in \mathbb{R}^n$ given any $\boldsymbol{\phi} \in \mathbb{R}^n$. Since shifting all ϕ_i's by the same constant has no impact (see Section 2), we fix $\phi_n = 0$ for the rest of the paper, and treat $\boldsymbol{\phi} \in \mathbb{R}^{n-1}$ as a vector of $n-1$ dimensions. In this way, we denote $\boldsymbol{\Pi} : \mathbb{R}^{n-1} \to \mathbb{R}^n$ as the mapping from the OD-based adjustments to the corresponding market-clearing multipliers.

Lemma 3. *The mapping $\boldsymbol{\Pi} : \mathbb{R}^{n-1} \to \mathbb{R}^n$ is continuously differentiable. The Jacobian matrix at the time $t-1$ market-clearing outcome $\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{(t-1)}) \in \mathbb{R}^{n\times(n-1)}$ can be computed using history h_t.*

The proof of this lemma is provided in Appendix A.5. Briefly, the market-clearing conditions (Definition 2) provide n equality constraints on the OD based adjustments $\boldsymbol{\phi}$ and the corresponding market-clearing $\boldsymbol{\pi}$ (which implicitly define the mapping $\boldsymbol{\Pi}$ from $\boldsymbol{\phi}$ to $\boldsymbol{\pi}$). We show that the violation of these n constraints (which is a n-dimensional function) is continuously differentiable, and that its Jacobian matrix with respect to $\boldsymbol{\pi}$ is invertable. The implicit function theorem then implies that $\boldsymbol{\Pi}$ is continuously differentiable. It is straightforward to verify that its Jacobian matrix depends only on the trip durations $(d_{i,j})_{(i,j)\in\mathcal{L}^2}$ and the local slope of the demand function $(q'_{i,j}(p_{i,j}))_{(i,j)\in\mathcal{L}^2}$, both of which can be observed by the platform.

With Lemma 3, we can consider the dual objective (16) as a function of only the OD-based adjustments $\text{Dual}(\boldsymbol{\phi}, \boldsymbol{\Pi}(\boldsymbol{\phi}))$. The function is, however, not convex in $\boldsymbol{\phi}$. Figure 3d illustrates this for an economy with

[12] This is similar in spirit to many other settings such as that of iterative combinatorial auctions [Parkes and Ungar, 2000], where the market designer can is able to identity a direction to update the dual variables that is guaranteed to improve the dual objective, even though neither the primal nor the dual objective can be evaluated.

two locations. Moreover, Example 2 in Appendix B shows that quasi-convexity does not hold either. Consequently, we need to address the challenge of finding the global optimum using only local information, and in the absence of convexity.

4.1 The Iterative Network Pricing Mechanism

We define the INP mechanism by providing the direction towards which it updates the OD-based adjustments, the stepsize it takes along that direction, as well as a policy for "backtracking".

The Direction. Recall from Theorem 1 that the inefficiency of a market-clearing outcome is bounded by the differences in the origin-based multipliers at different locations. We prove that at any market-clearing outcome, there exists a unique direction for equalizing the (linear approximations of the) origin-based multipliers $\{\Pi_i(\boldsymbol{\phi})\}_{i\in\mathcal{L}}$, and we choose this direction for updating the OD-based adjustments $\boldsymbol{\phi}$ under the INP mechanism. Formally, let $\mathbf{1}_n$ be the all-one vector of dimension n, and denote $\boldsymbol{\chi}_{1:n-1} \in \mathbb{R}^{n-1}$ as the first $n-1$ entries of any vector $\boldsymbol{\chi} \in \mathbb{R}^n$. Given any OD-based adjustments $\boldsymbol{\phi} \in \mathbb{R}^{n-1}$, the direction for updating $\boldsymbol{\phi}$ is given by

$$\boldsymbol{\delta} = \left(\begin{bmatrix} -\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}) & \mathbf{1}_n \end{bmatrix}^{-1} \boldsymbol{\pi}\right)_{1:n-1}, \tag{17}$$

where $\boldsymbol{\pi} = \boldsymbol{\Pi}(\boldsymbol{\phi})$ is the corresponding market-clearing multipliers, $\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}) \in \mathbb{R}^{n\times(n-1)}$ is the Jacobian matrix of $\boldsymbol{\Pi}$ evaluated at the market-clearing outcome, and the matrix $\begin{bmatrix} -\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}) & \mathbf{1}_n \end{bmatrix}$ is guaranteed to be full-rank (see Lemma 4 in Appendix A.6). In this way, there exists $\xi \in \mathbb{R}$ s.t. $\boldsymbol{\pi} + \mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi})\boldsymbol{\delta} = \xi\mathbf{1}_n$. [13] Note that $\boldsymbol{\delta}$ corresponds to the update direction under Newton's method for solving the following system of $n-1$ nonlinear equations:

$$\Pi_1(\boldsymbol{\phi}) = \Pi_2(\boldsymbol{\phi}) = \cdots = \Pi_n(\boldsymbol{\phi}). \tag{18}$$

The Stepsize. It is desirable in practice to avoid volatile changes in trip prices and driver earning rates, since as discussed in Section 1, both drivers and riders learn and adapt slowly to changes in the marketplace. Given any $\boldsymbol{\phi} \in \mathbb{R}^{n-1}$ and the corresponding update direction $\boldsymbol{\delta}$, we choose stepsize

$$\alpha = \min\{1,\ \tau/\|\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi})\boldsymbol{\delta}\|_\infty\}, \tag{19}$$

where $\tau > 0$ is a constant and $\|\cdot\|_\infty$ is the ℓ^∞ norm of vectors. In this way, based on the current linear approximation of $\boldsymbol{\Pi}$, the origin-based multiplier π_i for each location $i \in \mathcal{L}$ will change by at most τ under the new market-clearing outcome at $\boldsymbol{\phi} + \alpha\boldsymbol{\delta}$.

Moreover, define the following auxiliary function (also known as a Lyapunov function) representing the sum of the squared differences between the origin-based multipliers and their mean:

$$f(\boldsymbol{\phi}) \triangleq \sum_{i\in\mathcal{L}} \left(\Pi_i(\boldsymbol{\phi}) - \frac{1}{n}\sum_{j\in\mathcal{L}} \Pi_j(\boldsymbol{\phi})\right)^2. \tag{20}$$

To guarantee convergence, our mechanism uses f for *backtracking line search*, i.e. repeatedly shrinks the step size by a constant factor until a sufficient decrease of f is observed, relative to the expected decrease based on

[13] To see this, observe that $-\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi})\boldsymbol{\delta} + \xi\mathbf{1}_n$ can be written as $\begin{bmatrix} -\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}) & \mathbf{1}_n \end{bmatrix}\begin{bmatrix} \boldsymbol{\delta} \\ \xi \end{bmatrix}$.

the step size and the local gradient of f. The gradient of f at any market-clearing outcome can be evaluated using the origin-based multipliers and the Jacobian of $\boldsymbol{\Pi}$.

We now define the INP mechanism, which updates the OD-based adjustments in the direction for equalizing the origin-based multipliers, chooses the stepsize to target a maximum change of the multipliers, and reduces the stepsize if necessary via backtracking line search using the Lyapunov function f.

Definition 4 (Iterative Network Pricing)**.** An *iterative network pricing* (INP) mechanism is an iterative mechanism parameterized by $\tau > 0$, $\beta \in (0,1)$, $\sigma \in (0,1)$. At each time $t > 0$, given history h_t:

- The mechanism determines $\boldsymbol{\phi}^{(t)} = \boldsymbol{\phi}^{(t-1)} + \alpha^{(t)}\boldsymbol{\delta}^{(t)}$ according to

$$\boldsymbol{\delta}^{(t)} = \left(\begin{bmatrix} -\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{(t-1)}) & \mathbf{1}_n \end{bmatrix}^{-1} \boldsymbol{\pi}^{(t-1)}\right)_{1:n-1}, \tag{21}$$

$$\alpha^{(t)} = \min\{1,\ \tau/\|\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{(t-1)})\boldsymbol{\delta}^{(t)}\|_\infty\}, \tag{22}$$

  if $t = 1$ or sufficient progress in f was made (i.e. $f(\boldsymbol{\phi}^{(t-1)}) < f(\boldsymbol{\phi}^{(t')}) + \sigma\nabla f(\boldsymbol{\phi}^{(t')})^\mathsf{T}\alpha^{(t-1)}\boldsymbol{\delta}^{(t-1)}$ where t' is the timestep prior to the most-recent update of direction according to (21)).
- If $t > 1$ and not enough progress was made (i.e. $f(\boldsymbol{\phi}^{(t-1)}) \geq f(\boldsymbol{\phi}^{(t')}) + \sigma\nabla f(\boldsymbol{\phi}^{(t')})^\mathsf{T}\alpha^{(t-1)}\boldsymbol{\delta}^{(t-1)}$), the mechanism backtracks, setting $\boldsymbol{\phi}^{(t)} = \boldsymbol{\phi}^{(t')} + \alpha^{(t)}\boldsymbol{\delta}^{(t)}$ with $\boldsymbol{\delta}^{(t)} = \boldsymbol{\delta}^{(t-1)}$ and $\alpha^{(t)} = \beta\alpha^{(t-1)}$.

We now state the main result of the present paper, that the INP mechanism converges to an outcome that is approximately welfare-optimal.

Theorem 2. *Assuming (A1) and (A2), given any economy, the iterative network pricing mechanism converges to an origin-based market-clearing outcome where all multipliers are equal, i.e.* $\exists\omega^\star \in \mathbb{R}$ *s.t.* $\lim_{t\to\infty}\pi_i^{(t)} = \omega^\star$ *for all* $i \in \mathcal{L}$*. The welfare suboptimality of the limit outcome is upper bounded by*

$$m\max\{0, -\omega^\star\} + \sum_{i,j\in\mathcal{L}} e_{i,j}. \tag{23}$$

Intuitively, we prove that under assumptions (A1) and (A2), the INP mechanism can be interpreted as an iterative algorithm that is guaranteed to converge to the unique $\boldsymbol{\phi}^\star \in \mathbb{R}^{n-1}$ for which (18) holds. When the platform does not have an over-supply of drivers, we have $\omega^\star \geq 0$. In this case, the only inefficiency in the limit outcome arises from drivers relocating when prices are strictly positive (see (10)). Conversely, when the platform is over-supplied, $\omega^\star$ can be strictly negative if all m units of drivers continue to show up, even when facing a negative surplus. If we introduce the assumption that drivers will gradually exit the platform when faced with a consistently negative surplus, the suboptimality will converge to $\sum_{i,j\in\mathcal{L}} e_{i,j}$.

Proof. First, observe that fixing any economy $(m, \boldsymbol{d}, \boldsymbol{c}, \boldsymbol{q})$, the phantom demand function $\tilde{\boldsymbol{q}}$ satisfying conditions (I) and (II), and assuming (A1) and (A2), the INP mechanism can be interpreted as an iterative algorithm for solving (18), which we formally describe in Algorithm 1.

Lines 2 and 13 in Algorithm 1 correspond to the platform observing the market-clearing outcome given the OD-based adjustments for each timestep. The observed information is then used to evaluate ∇f and $\mathrm{D}\boldsymbol{\Pi}$. Line 7 marks the timesteps at which a sufficiently small stepsize for the current update direction is found, and the market-clearing outcomes at these timesteps are used to compute the new direction in the following timestep.

Algorithm 1 Iterative Network Pricing

Input: $\tau > 0$, $\beta \in (0,1)$, $\sigma \in (0,1)$.

1: $\boldsymbol{\phi}^{(0)} \leftarrow \mathbf{0}$
2: $\boldsymbol{\pi}^{(0)} \leftarrow \boldsymbol{\Pi}(\boldsymbol{\phi}^{(0)})$
3: **for** $t = 1, 2, \ldots$ **do**
4: **if** $t = 1$ **or** $(f(\boldsymbol{\phi}^{(t-1)}) < f(\boldsymbol{\phi}^{(t')}) + \sigma \nabla f(\boldsymbol{\phi}^{(t')})^{\mathsf{T}} \alpha^{(t-1)} \boldsymbol{\delta}^{(t-1)})$ **then**
5: $\boldsymbol{\delta}^{(t)} \leftarrow \left(\left[-\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{(t-1)}) \quad \mathbf{1}\right]^{-1} \boldsymbol{\pi}^{(t-1)}\right)_{1:n-1}$
6: $\alpha^{(t)} \leftarrow \min\{1,\ \tau / \|\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{(t-1)}) \boldsymbol{\delta}^{(t)}\|_\infty\}$
7: $t' \leftarrow t - 1$
8: **else**
9: $\boldsymbol{\delta}^{(t)} \leftarrow \boldsymbol{\delta}^{(t-1)}$
10: $\alpha^{(t)} \leftarrow \beta \alpha^{(t-1)}$
11: **end if**
12: $\boldsymbol{\phi}^{(t)} \leftarrow \boldsymbol{\phi}^{(t')} + \alpha^{(t)} \boldsymbol{\delta}^{(t)}$
13: $\boldsymbol{\pi}^{(t)} \leftarrow \boldsymbol{\Pi}(\boldsymbol{\phi}^{(t)})$
14: **end for**

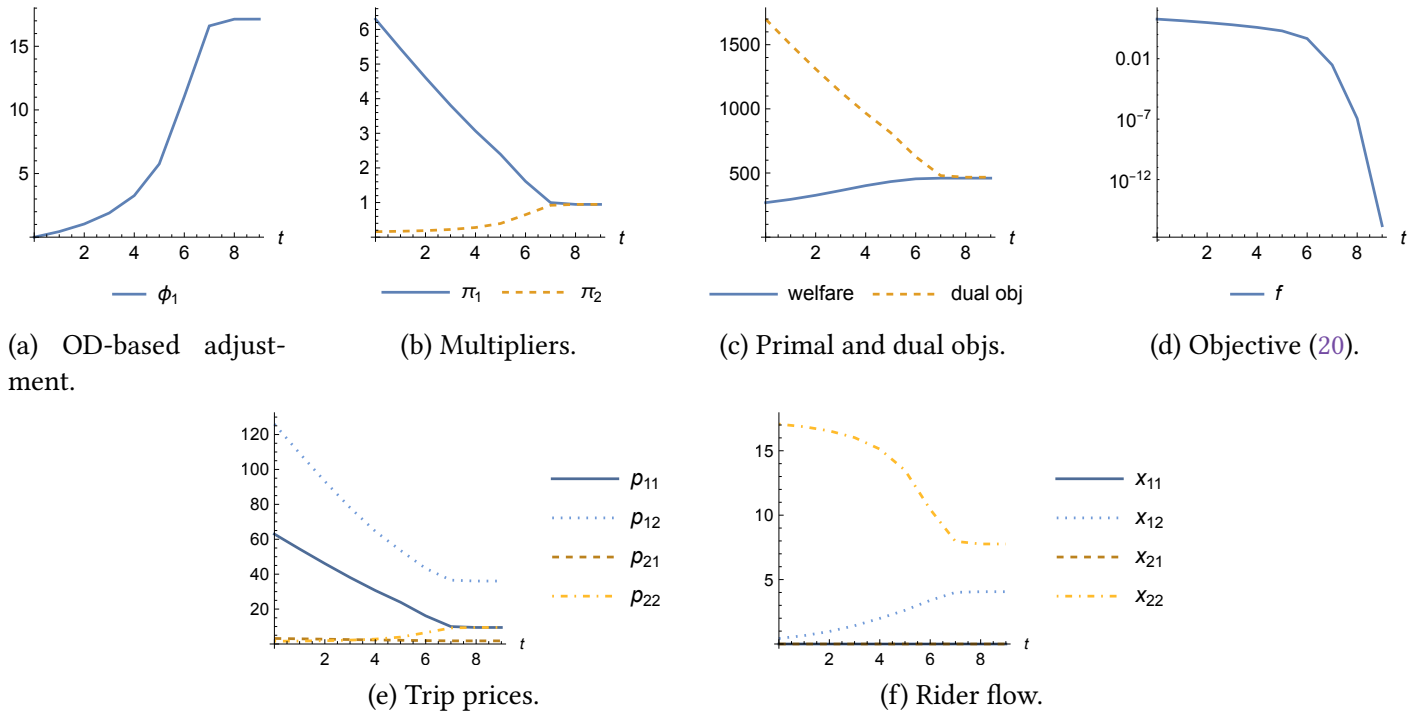
By constructing another auxiliary optimization problem and its dual problem, we prove in Lemma 5 in Appendix A.7 that there exists a unique set of OD-based adjustments $\boldsymbol{\phi}^\star \in \mathbb{R}^{n-1}$ solving (18), at which $\Pi_i(\boldsymbol{\phi}^\star) = \omega^\star$ for all $i \in \mathcal{L}$. Once we establish that the algorithm converges to $\boldsymbol{\phi}^\star$, the welfare suboptimality is a straightforward application of Theorem 1. What is left to prove is that $\boldsymbol{\phi}^{(t)}$ converges to $\boldsymbol{\phi}^\star$ as $t \to \infty$.

First, observe that $\boldsymbol{\phi}^\star \in \mathbb{R}^{n-1}$ is also the unique minimizer of the Lyapunov function f, because f is non-negative and achieves the global minimum of 0 only when (18) is satisfied. f is not necessary convex (or quasi-convex; see Figure 14b), but we prove in Appendix A.7 that f does not have any saddle points or spurious local minima (Proposition 2), and that f is coercive, meaning that every sublevel set of f is bounded (Proposition 3).

Let $t_0, t_1, t_2, \ldots$ be the sequence of values that t' in Line 7 of Algorithm 1 ever takes, i.e. the timesteps at which a new update direction is determined using the current market-clearing outcome. We prove that the sequence of update directions is *gradient related* (see Proposition 1.2.1 of Bertsekas [2016]), which, in conjunction with Propositions 2 and 3, implies that $\boldsymbol{\phi}^{(t_k)}$ converges to $\boldsymbol{\phi}^\star$ as $k \to \infty$ (Proposition 5). Finally, we show in Proposition 6 that $\boldsymbol{\delta}^{(t_k)}$ converges to $\mathbf{0}$ as $t \to \infty$, which implies that $\boldsymbol{\delta}^{(t)}$ also converges to $\mathbf{0}$. For any integers $k \geq 0$ and $t \in (t_k, t_{k+1}]$, we know $\|\boldsymbol{\phi}^{(t)} - \boldsymbol{\phi}^\star\| = \|\boldsymbol{\phi}^{(t_k)} + \alpha^{(t)} \boldsymbol{\delta}^{(t)} - \boldsymbol{\phi}^\star\| \leq \|\boldsymbol{\phi}^{(t_k)} - \boldsymbol{\phi}^\star\| + \|\alpha^{(t)} \boldsymbol{\delta}^{(t)}\| \leq \|\boldsymbol{\phi}^{(t_k)} - \boldsymbol{\phi}^\star\| + \|\boldsymbol{\delta}^{(t)}\|$. Therefore, $\limsup_{t\to\infty} \|\boldsymbol{\phi}^{(t)} - \boldsymbol{\phi}^\star\| \leq \limsup_{k\to\infty} \|\boldsymbol{\phi}^{(t_k)} - \boldsymbol{\phi}^\star\| + \limsup_{t\to\infty} \|\boldsymbol{\delta}^{(t)}\| = 0$. This completes the proof of this theorem. □

Note that the convergence rate is superlinear if $\sigma < 1/2$, according to Proposition 1.3.2 of Bertsekas [2016]. To verify the assumption in that proposition, we show that $\alpha^{(t_k+1)} = 1$ for large enough k, i.e. the first stepsize for any new direction is eventually always 1. By the definition of $\alpha^{(t)}$, it suffices to show that $\|\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{(t_k)}) \boldsymbol{\delta}^{(t_k+1)}\|_\infty \to 0$ as $k \to \infty$. Since we have proven that $\boldsymbol{\delta}^{(t)} \to \mathbf{0}$ as $t \to \infty$, it suffices to show that $\|\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{(t_k)})\|_\infty$ is bounded uniformly for all $k \geq 0$. This follows from Proposition 5, continuity of $\mathrm{D}\boldsymbol{\Pi}$, and the fact that continuous functions on a compact set are bounded.

We now revisit our running example and illustrate the iterative updates under the INP mechanism.



(a) OD-based adjustment. (b) Multipliers. (c) Primal and dual objs. (d) Objective (20).

(e) Trip prices. (f) Rider flow.

Figure 4: Outcomes under the INP mechanism over iterations, for the economy in Example 1.

Example 1 (Continued)**.** Recall from Section 2.1 that ixing $\phi_2 = 0$, the optimal OD-based adjustment for the morning rush hour example has $\phi_1 = 18.33$ (i.e. driver supply at location 1 (the residential area) is more valuable than that at location 2 (downtown). Starting from from $\phi_1^{(0)} = 0$ with $\pi_1^{(0)} \gg \pi_2^{(0)}$, Figure 4 illustrates the iterative updates under the INP mechanism, when the maximum change of each π_i between timesteps is limited to $\tau = 1$.

We can see that the difference between the origin-based multipliers reduces over time (Figure 4b), and that both the primal and the dual objectives converge to the optimal welfare (Figure 4c). Moreover, the prices of the $(1, 1)$ and $(2, 2)$ trips converge to the same level since the two trips take the same amount of time (Figure 4e). We also observe superlinear convergence of f when ϕ_1 is sufficiently close to the optimum.

4.2 Discussions

Alternative Update Directions. Under the INP mechanism, the OD-based adjustments $\boldsymbol{\phi}$ are iteratively updated towards the direction (17) for equalizing the origin-based multipliers, which corresponds to the update direction under Newton's method for solving the system of equations (18). Note that this is different from Newton's method for optimization, which will determine a different update direction and requires additional higher-order information (e.g. the second order derivatives of the demand functions).

Since the Lyapunov function f is coercive (Proposition 3) and has a single stationary point corresponding to the unique global optimum (Proposition 2), gradient descent with respect to f (i.e. taking $-\nabla f$ as the update direction) in conjunction with backtracking line search has the same convergence guarantees. This method, however, converges slowly (see simulation results in Appendix C.4) and does not require less information in comparison to the INP mechanism since the Jacobian $\mathrm{D}\boldsymbol{\Pi}$ is needed for evaluating ∇f.

A simple mechanism that requires very little information is to directly take last iteration's market-clearing

multipliers $\boldsymbol{\pi}^{(t-1)}$ as the update direction. In Appendix C.5, we show that simply setting $\boldsymbol{\phi}^{(t)} = \boldsymbol{\phi}^{(t-1)} + \alpha \boldsymbol{\pi}^{(t-1)}$ (where α is some small constant) performs very well in simulation. Intuitively, moving $\boldsymbol{\phi}$ in the direction of $\boldsymbol{\pi}$ may reduce f because increasing ϕ_i for some location $i \in \mathcal{L}$ with a high multiplier π_i will likely lead to lower prices for trips ending at location i for the next iteration, thereby increasing driver supply for this location and decreasing π_i. However, $\boldsymbol{\pi}$ is not necessarily a descent direction of f, and we are not yet able to prove or disprove the convergence of this simple mechanism.

Non-Stationary Market Conditions. Backtracking line search with respect to f is integrated in the design of the INP mechanism in order to guarantee convergence in theory. In simulations, for appropriately chosen τ and σ, our mechanism converges without ever backtracking for a stationary setting (see Section 5.1). The practical implementation of INP should not include the backtracking mechanism since the market conditions fluctuate week-over-week (see Section 5.2 for simulation results under a non-stationary setting).

By design, the INP mechanism does not include a termination criteria. In theory, with stationary market conditions, we may terminate the algorithm once the origin-based multipliers are sufficiently close. In practice, however, the INP mechanism should be "always on" even after substantially improving the outcomes in order to track potential shifts in the overall market conditions. We illustrate this in simulations in the next section.

5 Simulation Results

In this section, we illustrate the iterative updates under the INP mechanism for scenarios representing the morning rush hours in Chicago. We fit the market conditions using trip-level data made public by the City of Chicago,[14] and take the 77 community areas in Chicago as the set of locations (i.e. $n = |\mathcal{L}| = 77$).[15] Detailed descriptions of the simulation setup and the overall market dynamics in Chicago are provided in Appendix C, together with additional simulation results.

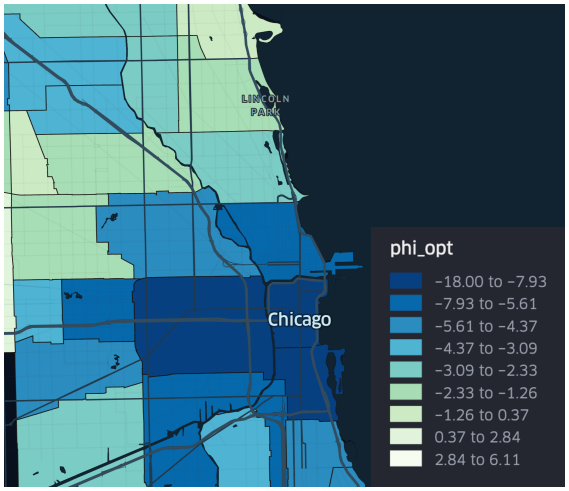
5.1 Stationary Market Condition

We first illustrate the iterative updates under the INP mechanism when the market condition is stationary and corresponds to 7–8 a.m. on Wednesdays during the first 9 weeks of 2020. For each OD pair $(i, j) \in \mathcal{L}^2$, we set the distance $d_{i,j}$ to be the average duration of all recorded trips from i to j in terms of *hours*, and assume that the trip cost is \$20/hour, i.e. $c_{i,j} = 20 d_{i,j}$.[16] For rider demand, we assume riders' values are exponentially distributed with mean \$60/hour for all OD pairs, and we determine the total demand for each OD to match

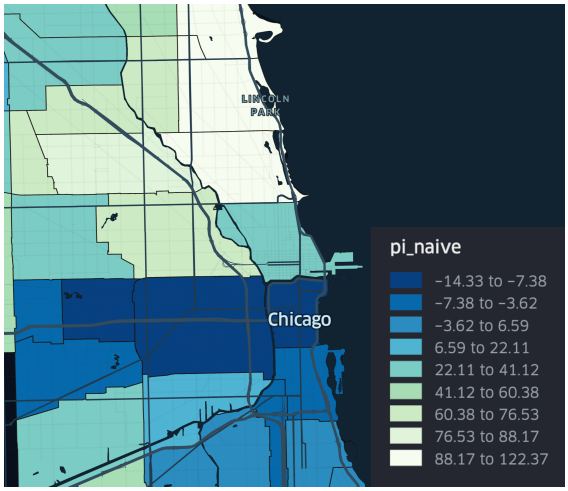
[14] https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips-2018-2022-/m6dm-c72p, accessed May 16, 2023. The dataset contains trips from transportation network companies including both Uber and Lyft, and provides for each trip the starting and ending timestamps (rounded to the nearest 15 minutes), price (rounded to the nearest \$2.5), duration (seconds), distance (miles), and the pickup and dropoff locations.

[15] https://en.wikipedia.org/wiki/Community_areas_in_Chicago (accessed May 16, 2023) provides a map of the community areas broken down by region. The pick-up (or drop-off) location of a trip is not provided for trips starting (or ending) outside the city boundary. The locations may also be omitted due to privacy considerations. Our simulations use only trips with both the pick-up and drop-off community areas, which consist of 86% of all trips provided in the dataset. We do not expect any qualitative change in the simulation results if all trips in the greater Chicago area are taken into consideration.

[16] This is roughly aligned with a proposal from Uber in 2019 (see https://p2a.co/H9gttWA, accessed September 14, 2020) for ensuring drivers are paid an average of \$21 per hour while on trip. The earnings per hour online could be slightly lower, depending on the average utilization level.



(a) Optimal OD-based additive adjustments $\boldsymbol{\phi}^*$.



(b) Origin-based multipliers $\boldsymbol{\pi}$ corresponding to $\boldsymbol{\phi} = \mathbf{0}$.

Figure 5: Optimal OD-based adjustments and naive origin-based multipliers, Chicago morning rush hours.

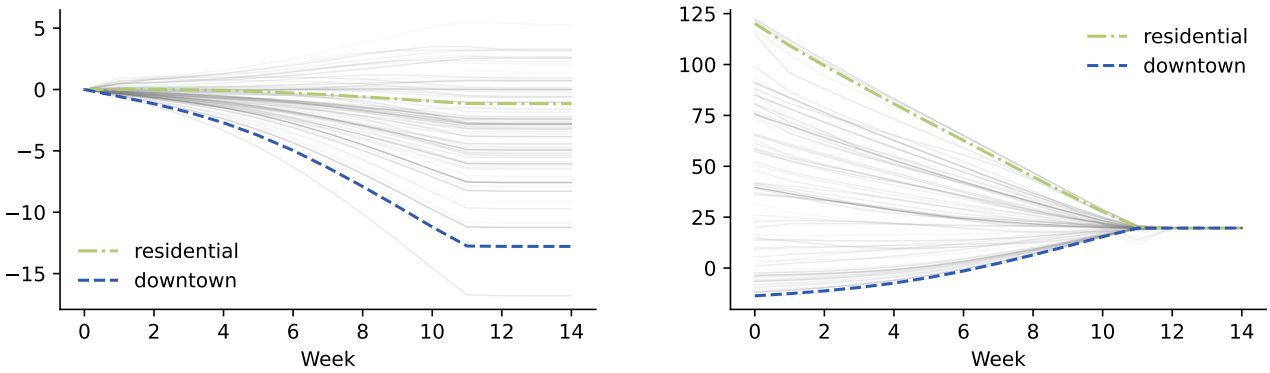
the observed average trip volume given the observed average prices. For total driver supply m, we compute the minimum amount of drivers that can fulfill all observed trips $(x^{\text{obs}}_{i,j})_{(i,j)\in\mathcal{L}^2}$ and maintain flow balance. See Appendix C.2.1 for a more detailed description of this simulation setup.

Naive Surge and Optimal OD-Based Adjustments. In Figure 5, we illustrate (a) the optimal OD-based adjustments and (b) the origin-based multipliers when the platform naively clears the market without any OD-based adjustments. The naive origin-based multipliers vary substantially in space: drivers can earn more than \$120/hour from trips originating from the North Side residential neighborhoods, but as low as negative \$15/hour in Downtown (meaning that drivers are paid only \$5/hour since the trip cost is \$20/hour). In contrast, the optimal OD-based adjustments $\boldsymbol{\phi}^*$ as shown in Figure 5a has a substantially smaller spread, and is also appropriately smooth in space, reflecting the fact that the marginal values for divers in neighboring locations do not differ substantially.

Iterative Updates Under INP. In Figure 6, we present the OD-based additive adjustments $\boldsymbol{\phi}^{(t)}$ and the corresponding market-clearing multipliers $\boldsymbol{\pi}^{(t)}$ over time under the INP mechanism. Each curve in the figures represents a location (i.e. a community area), and we highlight the residential area (Lake View) and the downtown area (the Loop), the destinations of the two trips that we discussed in Figure 1 in Section 1. The shade of the remaining curves correspond to the volume of trips originating from each location.

We can see that the OD-based adjustments start from $\boldsymbol{\phi}^{(0)} = \mathbf{0}$ and smoothly "spread out" over time. On the other hand, the origin-based multipliers start from $\boldsymbol{\pi}^{(0)}$ as shown in Figure 5b, and each converges to \$19.65/hour after 13 iterations. With $\tau = 10$, the mechanism aims to limit the week-over-week change of π_i for each location $i \in \mathcal{L}$ within \$10/hour. This is based on the linear approximation of the mapping from $\boldsymbol{\phi}$ to $\boldsymbol{\pi}$, and we do observe in Figure 6b larger changes for some locations at certain times.

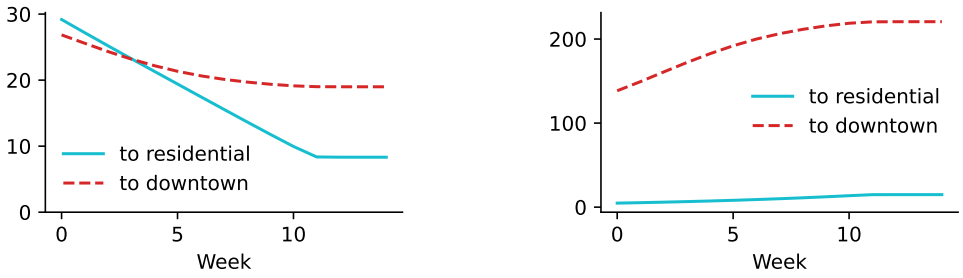
Figure 7 plots the trip prices and rider flow rates for the two trips discussed in Figure 1 in Section 1. We can see that the prices for both trips decrease over time, and as a result, more riders are picked up for both trips. Initially, both prices are very high because the origin of the two trips is very under-supplied during morning rush hours, and therefore has a very high multiplier when market is naively cleared by origin. Proper OD-based price adjustments increase the supply to the location, thereby reducing the multiplier and



(a) The additive adjustments $\boldsymbol{\phi}^{(t)}$.

(b) The multipliers $\boldsymbol{\pi}^{(t)}$.

Figure 6: $\boldsymbol{\phi}^{(t)}$ and $\boldsymbol{\pi}^{(t)}$ over iterations t, with the residential area and downtown highlighted.



(a) Trip prices $\boldsymbol{p}^{(t)}$.

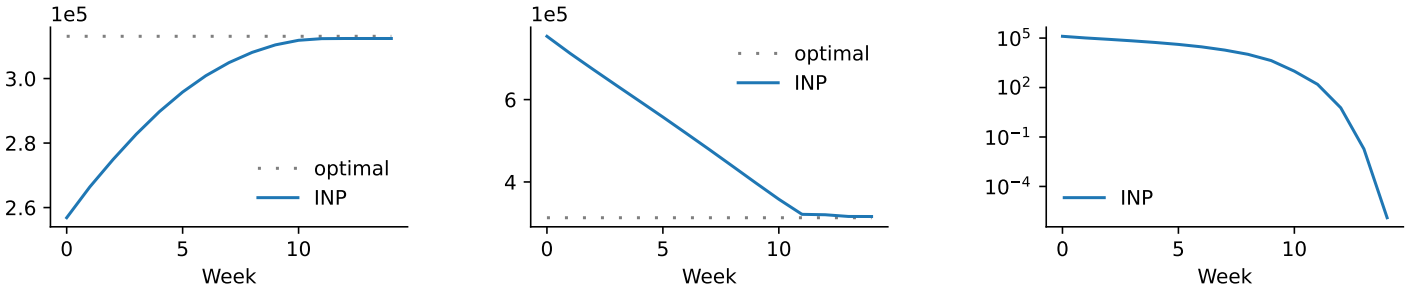
(b) Rider flow rate $\boldsymbol{x}^{(t)}$.

Figure 7: Trip prices and rider flow rates for the two trips as shown in Figure 1, ending in the residential area and in downtown, respectively.

the prices of the two trips. Moreover, the price for the trip ending in the residential area drops faster than that to downtown. This is because the OD-based adjustment is higher for the residential area in comparison, and this difference increases over time (see Figure 6a).

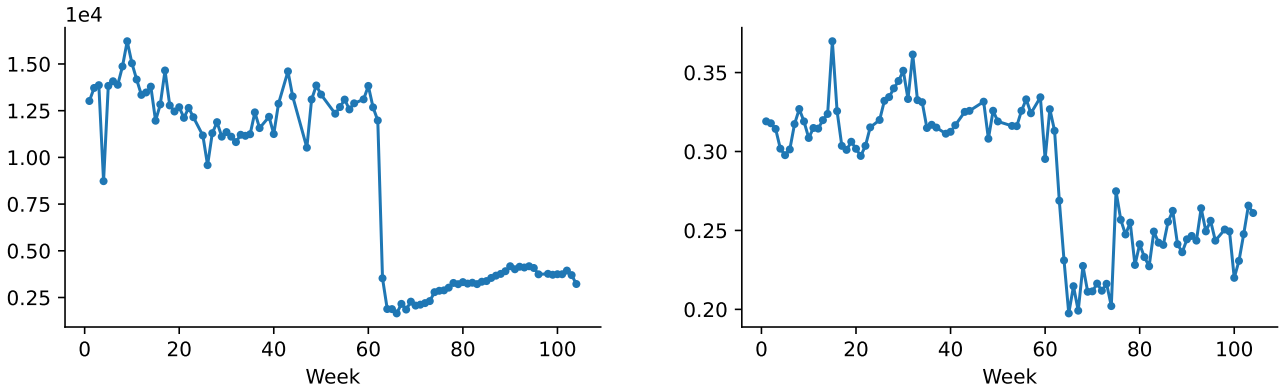
In Figure 8, we plot the social welfare (i.e. the primal objective (11a)), the dual objective (12a), and the Lyapunov function f (defined in (20)) corresponding to the market-clearing outcome at each time t. In this setting, the optimal social welfare is \$313 102 per hour. The initial market clearing outcome without OD-based adjustments achieves a welfare of \$256 846 per hour, which is 82.0% of optimum. Within 13 iterations, the welfare increases to \$312 422 per hour (99.8% of optimum), and the gap of 0.2% is a result of drivers relocating without a rider when the trip prices are positive.

The Lyapunov function f defined in (20) achieves a value of 1×10^{-6} at time $t = 14$, and we observe superlinear local convergence in Figure 8c. With $\tau = 10$, $\beta = 1/2$ and $\sigma = 10^{-3}$, sufficient progress in f was made at every timestep and the mechanism never has to backtrack. In Appendix C.3, we illustrate the iterative updates when the mechanism does not limit week-over-week changes in multipliers (i.e. when $\tau = +\infty$). We can see that in addition to limiting the volatility of prices for riders and earnings rates for drivers (which are desirable in practice), limiting the stepsize also leads to faster convergence, even though in theory this is not necessary for convergence due to backtracking using f.



(a) Social welfare ($/hour). (b) Dual objective. (c) Function f.

Figure 8: Social welfare, dual objective, and function f over iterations t.



(a) Trip volume. (b) Overall flow imbalance.

Figure 9: Trip volume and overall flow imbalance during 7–8 a.m. on Wednesdays, 2019-2020.

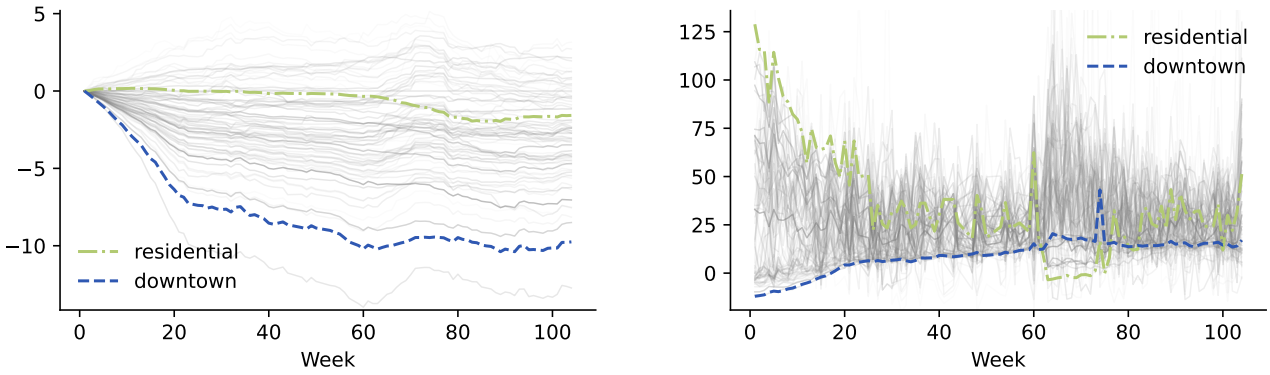
5.2 Non-Stationary Market Condition

We now demonstrate the robustness of the INP mechanism under non-stationary market conditions focusing on the Chicago morning rush hours from the beginning of 2019 to the end of 2020. Figure 9 presents the total trip volume and the overall flow imbalance for 7–8 a.m. on Wednesdays over the two years.[17] We can observe seasonality (lower trip volume during the summer, weeks 20–40), week-over-week fluctuations, and the substantial shift in market dynamics due to the COVID-19 pandemic. The shutdown starts at week $t = 60$, after which the trip volume drops from over 12k to 2k per hour. The overall imbalance decreases from 0.3 to 0.2, since the demand from residential areas drops substantially as people start to work from home.

Appendix C.2.2 describes in detail the construction of the economies for each week. Briefly, the main differences (compared to Section 5.1) are (i) using the observed trip duration and prices from each week (instead of the average over all weeks) to fit the rider demand model, and (ii) taking the minimum number of drivers needed to fulfill the realized trip flow of each week as the driver supply for that week.[18] Moreover,

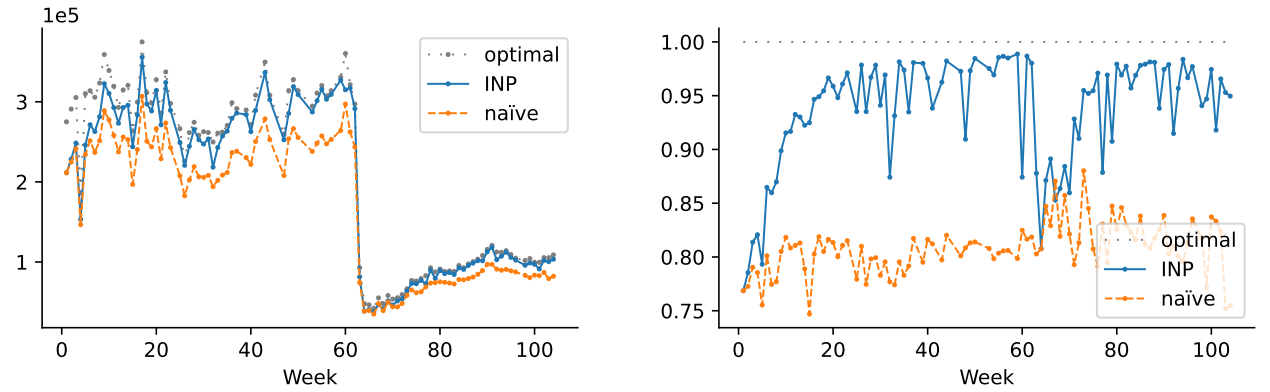
[17] The *overall imbalance* is defined as the average of the absolute value of the flow imbalance (defined in Section 1) of all locations, i.e. $\frac{1}{n}\sum_{i\in\mathcal{L}}\left|\sum_{j\in\mathcal{L}} x^{\text{obs}}_{i,j} - \sum_{j\in\mathcal{L}} x^{\text{obs}}_{j,i}\right| / \left(\sum_{j\in\mathcal{L}} x^{\text{obs}}_{i,j} + \sum_{j\in\mathcal{L}} x^{\text{obs}}_{j,i}\right)$, where $x^{\text{obs}}_{i,j}$ is the observed trip volume from i to j.

[18] With (i), we attribute all week-over-week variability in the data to the changes in underlying market conditions. This is likely an overestimation, since even with stationary market conditions, the realized trip flows will be different due to stochasticity. (ii) means that driver supply responds instantaneously to changes in market conditions. Our mechanism does not explicit model drivers' decisions on whether to driver for a platform. With an over 80% drop in trip volume, however, it is not meaningful to assume a fixed level of total supply. In practice, the total driver supply equilibrates in a few weeks after a substantial change in the earnings level [Hall et al., 2017].



(a) The OD-based additive adjustments $\boldsymbol{\phi}^{(t)}$. (b) The multipliers $\boldsymbol{\pi}^{(t)}$.

Figure 10: $\boldsymbol{\phi}^{(t)}$ and $\boldsymbol{\pi}^{(t)}$ over weeks under the INP mechanism.



(a) Social welfare. (b) The welfare ratio.

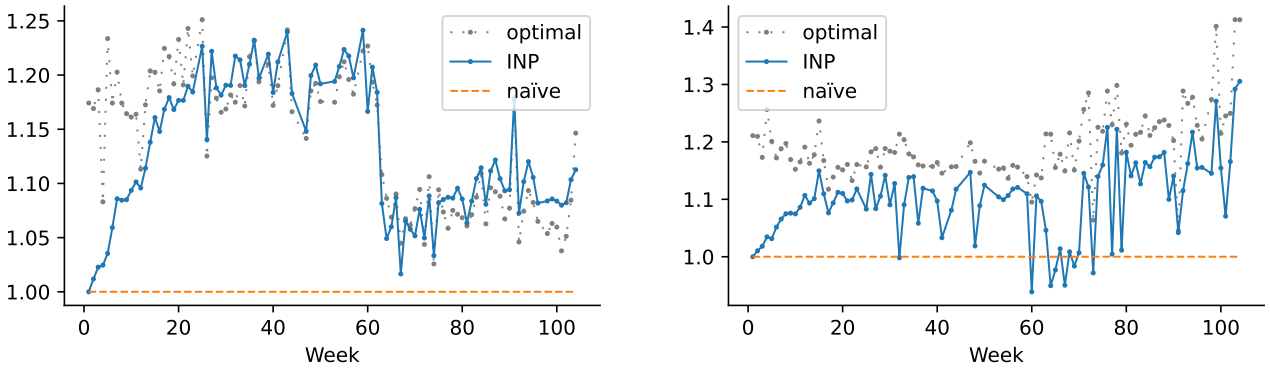
Figure 11: Social welfare and the welfare ratio (relative to the optimal welfare of each week) over weeks.

we excluded a total of 5 "event days", during which the McCormick Place (Chicago's main convention center) held large events that draw substantial rider demand into the neighborhood immediately south of the Chicago Loop during the morning rush hours.[19]

Simulation Results. In Figure 10, we present the OD-based additive adjustments $\boldsymbol{\phi}^{(t)}$ and the corresponding market-clearing multipliers $\boldsymbol{\pi}^{(t)}$ under the INP mechanism with $\tau = 5$, without backtracking line search. Figure 11 compares the social welfare achieved by INP mechanism, the naïve origin-based surge (i.e. the market clearing outcome in the absence of OD-based adjustments), and the welfare-optimal outcome for each week. Note that the optimal outcome represents the highest possible welfare *in hindsight* (i.e. using the optimal OD-based adjustments for the economy constructed for each particular week) and cannot be achieved in practice. Overall, the INP mechanism consistently outperforms the naïve origin-based pricing, achieving over 95% of the optimal welfare for most of the weeks.

During the first 25 weeks, the welfare ratio achieved by the INP mechanism increases from 80% to over 95%, as the OD-based adjustments approaches the levels of the pre-COVID "steady state". The difference between the OD-based adjustments of the residential area and downtown smoothly increases to around $10,

[19]Major events are typically arranged months ahead of time, thus in practice, the platform can maintain a separate set of OD-based adjustments corresponding to the scenario where substantially more riders demand trips ending in south of loop.



(a) Gross booking ratio.

(b) Throughput ratio.

Figure 12: The gross booking and throughput ratios (relative to those under the naive outcome of each week).

and the difference between the multipliers decreases from over \$100/hour to close to 0. The naïve origin-based pricing without OD-based adjustments achieves only around 80–85% of the optimum welfare. There are a few weeks where the INP mechanism achieves around 90% of the optimal welfare. These are days during which the south of loop location (where convention center resides) observes larger than normal trip inflows, but did not reach the threshold to be excluded from the simulations.

After the COVID shutdown, we observe an abrupt drop of the multiplier π_i and a gradual decrease of the OD-based adjustment ϕ_i for the residential area. This is a result of the reduced imbalance for the residential area (most likely due to work-from-home arrangements) as discussed in Figure 1b. Despite the substantial impact of the shutdown, the INP mechanism maintains over 80% of the optimal welfare, and recovers to over 95% within only a few weeks. The performance of the naïve origin-based pricing increases slightly after the shutdown. This improvement is a result of an overall less imbalanced demand pattern (see Figure 9b), and diminishes as the economy returns to normal in 2021-2022.

Since social welfare cannot be directly measured by the ridesharing platforms in practice, we compare in Figure 12 the gross bookings (GB, i.e. the total payments made by all riders) and trip throughput achieved under our mechanism and benchmarks. The "optimal" benchmark optimizes welfare and may achieve worse GB and/or throughput in comparison to the other policies, thus we present here the ratios relative to the outcomes under naive origin-based surge pricing (see Figures 28 and 29 for the actual GB and throughput achieved each week). We see that our mechanism also achieves substantial improvements in both metrics.

6 Conclusion

We introduce the *iterative network pricing* (INP) mechanism, addressing a main challenge in the practical implementation of optimal origin-destination (OD) based prices, that the model for rider demand is hard to estimate. Our mechanism, assuming real-time market clearance through dynamic "surge" pricing by origin, updates OD-based additive adjustments week-over-week, using only the data immediately available from previous weeks' market outcomes. By establishing a correspondence to Newton's method, we prove that the INP mechanism converges to an approximately optimal outcome when the market condition is stationary. Using data from Chicago's morning rush hours from early 2019 to late 2020, we demonstrate via simulations not only substantial welfare improvements but also remarkable robustness amid significant fluctuations of market conditions.

References

Philipp Afèche, Zhe Liu, and Costis Maglaras. Ride-hailing networks with strategic drivers: The impact of platform control capabilities on performance. *Manufacturing & Service Operations Management*, 25(5): 1890–1908, 2023.

AmirMahdi Ahmadinejad, Hamid Nazerzadeh, and Amin Saberi. Competition in ride-hailing markets. *Web and Internet Economics LNCS 11920*, page 333, 2019.

Ravindra K Ahuja, Thomas L Magnanti, and James B Orlin. Network flows. 1988.

Ali Aouad and Ömer Saritaç. Dynamic stochastic matching under limited time. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 789–790, 2020.

Arash Asadpour, Ilan Lobel, and Garrett van Ryzin. Minimum earnings regulation and the stability of marketplaces. *Manufacturing & Service Operations Management*, 25(1):254–265, 2023.

Itai Ashlagi, Maximilien Burq, Chinmoy Dutta, Patrick Jaillet, Amin Saberi, and Chris Sholley. Maximum weight online matching with deadlines. *arXiv preprint arXiv:1808.03526*, 2018.

Siddhartha Banerjee, Ramesh Johari, and Carlos Riquelme. Pricing in ride-sharing platforms: A queueing-theoretic approach. In *ACM Conference on Economics and Computation (EC)*, 2015.

Siddhartha Banerjee, Yash Kanoria, and Pengyu Qian. State dependent control of closed queueing networks. In *Abstracts of the 2018 ACM International Conference on Measurement and Modeling of Computer Systems*, 2018.

Dimitri Bertsekas. *Nonlinear Programming*. Athena Scientific, 2016.

Dimitri P Bertsekas. The auction algorithm for assignment and other network flow problems: A tutorial. *Interfaces*, 20(4):133–149, 1990.

Omar Besbes, Francisco Castro, and Ilan Lobel. Surge pricing and its spatial supply response. *Management Science*, 2020.

Omar Besbes, Francisco Castro, and Ilan Lobel. Spatial capacity planning. *Operations Research*, 70(2):1271–1291, 2022.

Kostas Bimpikis, Ozan Candogan, and Daniela Saban. Spatial pricing in ride-sharing networks. *Operations Research*, 67(3):744–769, 2019.

Desmond Cai, Subhonmesh Bose, and Adam Wierman. On the role of a market maker in networked cournot competition. *Mathematics of Operations Research*, 44(3):1122–1144, 2019.

J Massey Cashore, Peter I Frazier, and Eva Tardos. Dynamic pricing provides robust equilibria in stochastic ride-sharing networks. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 301–302, 2022.

Juan Camilo Castillo. Who benefits from surge pricing? *Available at SSRN 3245533*, 2020.

Juan Camilo Castillo, Dan Knoepfle, and Glen Weyl. Surge pricing solves the wild goose chase. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 241–242, 2017.

Francisco Castro, Peter Frazier, Hongyao Ma, Hamid Nazerzadeh, and Chiwei Yan. Matching queues, flexibility and incentives. *arXiv preprint arXiv:2006.08863*, 2020.

Francisco Castro, Hongyao Ma, Hamid Nazerzadeh, and Chiwei Yan. Randomized fifo mechanisms. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 60–60, 2022.

Kuan-Ming Chen, Claire Ding, John A List, and Magne Mogstad. Reservation wages and workers' valuation of job flexibility: Evidence from a natural field experiment. Technical report, National Bureau of Economic Research, 2020.

M Keith Chen and Michael Sheldon. Dynamic pricing in a labor market: Surge pricing and flexible work on the uber platform. 2015. URL https://www.anderson.ucla.edu/faculty_pages/keith.chen/papers/SurgeAndFlexibleWork_WorkingPaper.pdf.

M Keith Chen, Peter E Rossi, Judith A Chevalier, and Emily Oehlsen. The value of flexible work: Evidence from uber drivers. *Journal of Political Economy*, 127(6):2735–2794, 2019.

Peter Cohen, Robert Hahn, Jonathan Hall, Steven Levitt, and Robert Metcalfe. Using big data to estimate consumer surplus: The case of uber. *NBER Working Paper No. 22627*, 2016.

Cody Cook, Rebecca Diamond, Jonathan Hall, John A List, and Paul Oyer. The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. Technical report, National Bureau of Economic Research, 2018.

John Dickerson, Karthik Sankararaman, Aravind Srinivasan, and Pan Xu. Allocation problems in ride-sharing platforms: Online matching with offline reusable resources. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Zhixuan Fang, Longbo Huang, and Adam Wierman. Prices and subsidies in the sharing economy. *Performance Evaluation*, 2019.

Zhixuan Fang, Longbo Huang, and Adam Wierman. Loyalty programs in the sharing economy: Optimality and competition. *Performance Evaluation*, 2020.

Daniel Freund and Garrett van Ryzin. Pricing fast and slow: Limitations of dynamic pricing mechanisms in ride-hailing. *Available at SSRN 3931844*, 2021.

Nikhil Garg and Hamid Nazerzadeh. Driver surge pricing. *Proceedings of the 21st ACM Conference on Economics and Computation*, 2020.

Jonathan Hall, Cory Kendrick, and Chris Nosko. The effects of uber's surge pricing: A case study. Technical report, The University of Chicago Booth School of Business, 2015.

Jonathan V Hall and Alan B Krueger. An analysis of the labor market for uber's driver-partners in the united states. *NBER Working Paper No. 22843*, 2016.

Jonathan V Hall, John J Horton, and Daniel T Knoepfle. Labor market equilibration: Evidence from uber. Technical report, New York University Stern School of Business, 2017.

Yash Kanoria and Pengyu Qian. Blind dynamic resource allocation in closed networks via mirror backpressure. In *EC'20: Proceedings of the 21st ACM Conference on Economics and Computation*, 2020.

Zhen Lian and Garrett van Ryzin. Autonomous vehicle market design. *Available at SSRN*, 2020.

Zhen Lian and Garrett Van Ryzin. Optimal growth in two-sided markets. *Management Science*, 2021.

Zhen Lian, Sebastien Martin, and Garrett van Ryzin. Larger firms pay more in the gig economy. *Available at SSRN*, 2021.

Alice Lu, Peter I Frazier, and Oren Kislev. Surge pricing moves uber's driver-partners. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 3–3, 2018.

Hongyao Ma, Fei Fang, and David C Parkes. Spatio-temporal pricing for ridesharing platforms. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 583–583, 2019.

Hongyao Ma, Fei Fang, and David C Parkes. Spatio-temporal pricing for ridesharing platforms. *Operations Research*, 70(2):1025–1041, 2022.

Michael Ostrovsky and Michael Schwarz. Carpooling and the economics of self-driving cars. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 581–582, 2019.

Erhun Özkan and Amy R Ward. Dynamic matching for real-time ride sharing. *Stochastic Systems*, 10(1): 29–70, 2020.

John ZF Pang, Hu Fu, Won I Lee, and Adam Wierman. The efficiency of open access in platforms for networked cournot markets. In *IEEE INFOCOM 2017*, pages 1–9, 2017.

David C Parkes and Lyle H Ungar. Iterative combinatorial auctions: Theory and practice. *Aaai/iaai*, 7481:53, 2000.

Zhiwei Qin, Xiaocheng Tang, Yan Jiao, Fan Zhang, Zhe Xu, Hongtu Zhu, and Jieping Ye. Ride-hailing order dispatching at DiDi via reinforcement learning. *INFORMS Journal on Applied Analytics*, 50(5):272–286, 2020.

Lisa Rayle, Susan Shaheen, Nelson Chan, Danielle Dai, and Robert Cervero. App-based, on-demand ride services: Comparing taxi and ridesourcing trips and user characteristics in san francisco. Technical report, University of California Transportation Center (UCTC), 2014.

Duncan Rheingans-Yoo, Scott Duke Kominers, Hongyao Ma, and David C Parkes. Ridesharing with driver location preferences. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 2019.

Lloyd S Shapley and Martin Shubik. The assignment game i: The core. *International Journal of game theory*, 1(1):111–130, 1971.

Zhengtian Xu, Daniel AMC Vignon, Yafeng Yin, and Jieping Ye. An empirical study of the labor supply of ride-sourcing drivers. *Transportation Letters*, pages 1–4, 2020.

Chiwei Yan, Helin Zhu, Nikita Korolko, and Dawn Woodard. Dynamic pricing and matching in ride-hailing platforms. *Naval Research Logistics (NRL)*, 67(8):705–724, 2020.

Chenkai Yu, Hongyao Ma, and Adam Wierman. Price cycles in ridesharing platforms. In *International Conference on Web and Internet Economics*, pages 618–636. Springer, 2023.

Appendices

We provide in Appendix A the proofs omitted from the body of the paper. Appendix B provides additional discussions and examples. We include in Appendix C detailed description of the data from the City of Chicago, the market dynamics, the simulation setup as well as additional simulation results.

A Proofs

We provide in this appendix proofs that are omitted from the body of the paper.

Contents

A.1 A Cycle Formulation . . . 32
A.2 Proof of Lemma 1 . . . 34
A.3 Uniqueness of Optimal Dual Solution . . . 37
A.4 Proof of Lemma 2 . . . 37
A.5 Proof of Lemma 3 . . . 42
A.6 The Update Direction . . . 44
A.7 Properties of the Lyapunov Function . . . 46
A.7.1 Proof of Proposition 2 . . . 47
A.7.2 Proof of Proposition 3 . . . 48
A.8 Proof of Theorem 2 . . . 50

A.1 A Cycle Formulation

Before proceeding with the proofs, we first provide an alternative formulation of the welfare-optimization problem and its dual, which we use to prove Lemma 1.

$$\begin{aligned}
\underset{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^{n^2},\,\boldsymbol{w}\in\mathbb{R}^{|C|}}{\text{maximize}} \quad & \sum_{i,j\in\mathcal{L}}\left(\int_0^{x_{i,j}} v_{i,j}(s)\,\mathrm{d}s - c_{i,j}y_{i,j}\right) && \text{(24a)}\\
\text{subject to} \quad & x_{i,j} \le y_{i,j}, \quad \forall i,j\in\mathcal{L}, && \text{(24b)}\\
& \sum_{i,j\in\mathcal{L}} d_{i,j}y_{i,j} \le m, && \text{(24c)}\\
& \sum_{\kappa\in C} \mathbb{1}\{(i,j)\in\kappa\}w_\kappa = y_{i,j}, \quad \forall i,j\in\mathcal{L}, && \text{(24d)}\\
& w_\kappa \ge 0, \quad \forall \kappa\in C. && \text{(24e)}
\end{aligned}$$

Compared to the flow formulation (6), this *cycle* formulation introduces additional decision variables $\boldsymbol{w}\in\mathbb{R}^{|C|}$ representing the amount of driver flow on each directed cycle. It is straightforward to see that the two formulations are equivalent. The objective function (24a) and the constraints on demand and supply

(24b)–(24c) are the same as those in the flow formulation (6). Constraints (24d)–(24e) are equivalent to constraint (6d) because, by the flow decomposition theorem, the driver flow is balanced if and only if it can be decomposed into cycles with non-negative weights, as formally stated in (1).

Claim 1. *The primal problem of the cycle formulation* (24) *is equivalent to the primal problem of the flow formulation* (6) *in the sense that for any* feasible *solution to one problem, there exists a* feasible *solution to the other problem with the same objective value.*

Let $p_{i,j}$, ω and $\eta_{i,j}$ be the dual variables corresponding to constraints (24b)–(24d), respectively. The dual problem can be written as:

$$\begin{aligned}
\underset{\boldsymbol{p} \in \mathbb{R}^{n^2},\, \omega \in \mathbb{R},\, \boldsymbol{\eta} \in \mathbb{R}^{n^2}}{\text{minimize}} \quad & m\omega + \sum_{i,j \in \mathcal{L}} \int_0^{q_{i,j}(p_{i,j})} (v_{i,j}(s) - p_{i,j})\, \mathrm{d}s && && \text{(25a)}\\
\text{subject to} \quad & p_{i,j} = c_{i,j} + d_{i,j}\omega - \eta_{i,j}, && \forall i, j \in \mathcal{L}, && \text{(25b)}\\
& \sum_{(i,j) \in \kappa} \eta_{i,j} \geq 0, && \forall \kappa \in C, && \text{(25c)}\\
& p_{i,j} \geq 0, && \forall i, j \in \mathcal{L}, && \text{(25d)}\\
& \omega \geq 0. && && \text{(25e)}
\end{aligned}$$

Strong duality holds between the primal problem (24) and the dual problem (25) because the primal objective (24a) is concave and Slater's condition can be verified by the following solution. Let $w_\kappa = \varepsilon > 0$ for all $\kappa \in C$ (strictly satisfying (24e)). Let $y_{i,j} = \sum_{\kappa \in C} \mathbb{1}\{(i,j) \in \kappa\} w_\kappa$ for all $(i,j) \in \mathcal{L}^2$ (satisfying (24d)). Let $x_{i,j} = y_{i,j}/2$ for all $(i,j) \in \mathcal{L}^2$ (strictly satisfying (24b)). Finally, let ε be sufficiently small such that $\sum_{i,j \in \mathcal{L}} d_{i,j} y_{i,j} \leq m$ (strictly satisfying (24c)).

Difference Between Primal and Dual Objectives. Given any feasible primal solution $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w})$ and any feasible dual solution $(\boldsymbol{p}, \omega, \boldsymbol{\eta})$, the difference between the primal and dual objectives can be written as:

$$\sum_{i,j \in \mathcal{L}} \int_{x_{i,j}}^{q_{i,j}(p_{i,j})} (v_{i,j}(s) - p_{i,j})\, \mathrm{d}s + \sum_{\kappa \in C} w_\kappa \sum_{(i,j) \in \kappa} \eta_{i,j} + \sum_{i,j \in \mathcal{L}} p_{i,j}(y_{i,j} - x_{i,j}) + \omega\Big(m - \sum_{i,j \in \mathcal{L}} d_{i,j} y_{i,j}\Big). \tag{26}$$

The first term of (26) is weakly positive because for every trip $(i,j) \in \mathcal{L}^2$, one of the following three cases must hold:

- When $x_{i,j} = q_{i,j}(p_{i,j})$, the integral is zero.
- When $x_{i,j} < q_{i,j}(p_{i,j})$, the fact that $v_{i,j}(\cdot)$ is strictly decreasing implies that for all $s \in [x_{i,j}, q_{i,j}(p_{i,j}))$, $v_{i,j}(s) > p_{i,j}$. The integral is therefore strictly positive.
- If $x_{i,j} > q_{i,j}(p_{i,j})$, $v_{i,j}(s) < p_{i,j}$ for all $s \in (q_{i,j}(p_{i,j}), x_{i,j}]$ holds again due to the strict monotonicity of $v_{i,j}(\cdot)$. The integral is therefore also strictly positive.

The remaining three terms are also always weakly positive due to the primal and dual constraints. As a result, (26) is zero (meaning that both the primal and the dual solutions are optimal) if and only if each of the four terms is zero.

We now show an equivalence between the dual problems of the two formulations. Note that the equivalence between the two primal formulations and the two dual formulations imply that strong duality also

holds between (6) and (7).

Claim 2. *The dual problem of the cycle formulation* (25) *is equivalent to the dual problem of the flow formulation* (7) *in the sense that for any* optimal *solution to one problem, there exists an* optimal *solution to the other problem with the same objective value. Moreover, variables* $\boldsymbol{p}$ *and* ω *are the same in the two dual problems, and the correspondence between* $\boldsymbol{\eta}$ *and* $\boldsymbol{\phi}$ *is given by* $\eta_{i,j} = \phi_j - \phi_i$ *for all* $(i,j) \in \mathcal{L}^2$.

Proof. Let $(\boldsymbol{p}, \omega, \boldsymbol{\phi})$ be an optimal solution to the flow dual problem (7). We can construct a solution $(\boldsymbol{p}, \omega, \boldsymbol{\eta})$ to the cycle dual problem (25) with $\boldsymbol{p}$ and ω being the same as in the flow dual problem (7), and with $\eta_{i,j} = \phi_j - \phi_i$ for all $(i,j) \in \mathcal{L}^2$. We can verify that constraint (25b) holds by construction; constraint (25c) holds because $\sum_{(i,j)\in\kappa} \eta_{i,j} = \sum_{(i,j)\in\kappa}(\phi_j - \phi_i) = 0$ for all $\kappa \in C$; constraints (25d) and (25e) hold since $(\boldsymbol{p}, \omega, \boldsymbol{\phi})$ is feasible solution to (7). The objective functions are the same as in the flow dual problem (7), therefore, $(\boldsymbol{p}, \omega, \boldsymbol{\eta})$ is a feasible solution to the cycle dual problem (25) with the same (hence optimal) objective value.

For the other direction, let $(\boldsymbol{p}, \omega, \boldsymbol{\eta})$ be an optimal solution to the cycle dual problem (25). We construct a solution $(\boldsymbol{p}, \omega, \boldsymbol{\phi})$ to the flow dual problem (7) with $\boldsymbol{p}$ and ω being the same as in the cycle dual problem (25), and with ϕ_i for each $i \in \mathcal{L}$ being the length of the shortest path from node n to node i in the complete directed graph where each edge $(i,j) \in \mathcal{L}^2$ has weight $\eta_{i,j}$. The shortest paths are well-defined because, by (25c), there is no negative cycle in this directed graph. The rest of this shows that $\{\eta_{i,j}\}_{i,j\in\mathcal{L}}$ has the structure such that it can be written as $\eta_{i,j} = \phi_j - \phi_i$ for all $(i,j) \in \mathcal{L}^2$. Then, (7b) holds by construction, and as a result, $(\boldsymbol{p}, \omega, \boldsymbol{\phi})$ is a feasible solution to the flow dual problem (7) with the same (hence optimal) objective value.

First, observe that for any pair of nodes i and j, the length of the shortest path from n to j is at most the length of the shortest path from n to i plus the length of edge (i,j), which is $\eta_{i,j}$. Therefore we have $\phi_j \le \phi_i + \eta_{i,j}$, which implies $\eta_{i,j} - \phi_j + \phi_i \ge 0$.

Since $(\boldsymbol{p}, \omega, \boldsymbol{\eta})$ is an optimal solution to the cycle dual problem (25), we know from strong duality that for any optimal solution $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w})$ to the cycle primal problem (24), the difference between the primal and dual objectives (26) is zero. This implies that every term in (26) is zero. In particular, we have (i) $x_{i,j} = q_{i,j}(p_{i,j})$ for all $(i,j) \in \mathcal{L}^2$ and (ii) $\sum_{(i,j)\in\kappa} \eta_{i,j} = 0$ for all $\kappa \in C$ with $w_\kappa > 0$. (i) implies that $y_{i,j} \ge x_{i,j} > 0$ for all $(i,j) \in \mathcal{L}^2$, i.e. the driver flow on every edge is strictly positive (recall that $q_{i,j}(r) > 0$ for all $r \ge 0$ since $q_{i,j}$ is assumed to be strictly decreasing for all $r \ge 0$). Since $\boldsymbol{w}$ is a cycle decomposition of $\boldsymbol{y}$, for all $(i,j) \in \mathcal{L}^2$, we have $\sum_{\kappa:(i,j)\in\kappa\in C} w_\kappa = y_{i,j} > 0$. Therefore, there exists a cycle $\kappa \in C$ such that $(i,j) \in \kappa$ and $w_\kappa > 0$. Applying (ii), we know that for all $(i,j) \in \mathcal{L}^2$, there exists a cycle κ covering (i,j) such that $\sum_{(i',j')\in\kappa} \eta_{i',j'} = 0$. For this cycle κ, we know that

$$\sum_{(i',j')\in\kappa} (\eta_{i',j'} - \phi_{j'} + \phi_{i'}) = \sum_{(i',j')\in\kappa} \eta_{i',j'} + \sum_{(i',j')\in\kappa} (\phi_{i'} - \phi_{j'}) = 0,$$

and that $\eta_{i',j'} - \phi_{j'} + \phi_{i'} \ge 0$ for all $(i',j') \in \kappa$. Therefore, it must be the case that $\eta_{i',j'} = \phi_{j'} - \phi_{i'}$ for all $(i',j') \in \kappa$, and this completes the proof that $\eta_{i,j} = \phi_j - \phi_i$ for all $(i,j) \in \mathcal{L}^2$. □

A.2 Proof of Lemma 1

Lemma 1 (Welfare Theorem). *A feasible rider and driver flow* $(\boldsymbol{x}, \boldsymbol{y})$ *is welfare-optimal if and only if there exist trip prices* $\boldsymbol{p}$ *such that the outcome* $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{p})$ *is a competitive equilibrium (CE). Moreover, for any CE outcome* $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{p})$, *there exist* $\omega \ge 0$ *and* $\boldsymbol{\phi} \in \mathbb{R}^n$ *such that*

$$p_{i,j} = c_{i,j} + d_{i,j}\omega + \phi_i - \phi_j, \quad \forall i,j \in \mathcal{L}. \tag{8}$$

Proof. The following two claims prove the equivalence between welfare-optimal and competitive equilibrium outcomes.

Claim 3. *Given any optimal solution $(\boldsymbol{x}, \boldsymbol{y})$ to the primal problem (6), there exists $\boldsymbol{p}$ such that $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{p})$ forms a competitive equilibrium.*

Claim 4. *Given any competitive equilibrium $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{p})$, there exist $\boldsymbol{w}$, ω and $\boldsymbol{\eta}$ such that $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w})$ is an optimal solution to the primal problem (24) and $(\boldsymbol{p}, \omega, \boldsymbol{\eta})$ is an optimal solution to the dual problem (25).*

To establish the structure of the CE prices, note that for any CE outcome $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{p})$, by Claim 4, there exist $\boldsymbol{w}$, ω and $\boldsymbol{\eta}$ such that $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w})$ is an optimal solution to the primal problem (24) and $(\boldsymbol{p}, \omega, \boldsymbol{\eta})$ is an optimal solution to the dual problem (25), with $p_{i,j} = c_{i,j} + d_{i,j}\omega - \eta_{i,j}$ for all $(i, j) \in \mathcal{L}^2$. Moreover, the equivalence between the dual problems (7) and (25) (Claim 2) imply that there exists $\boldsymbol{\phi} \in \mathbb{R}^n$ such that $(\boldsymbol{p}, \omega, \boldsymbol{\phi})$ forms an optimal solution to the dual problem (7), and that $\eta_{i,j} = \phi_j - \phi_i$ holds for all $(i, j) \in \mathcal{L}^2$. As a result, the CE prices must be of the form $p_{i,j} = c_{i,j} + d_{i,j}\omega - \phi_j + \phi_i$ for all $(i, j) \in \mathcal{L}^2$. □

We now provide the proof for the two claims.

Proof of Claim 3. Intuitively, given any set of optimal driver and rider flow $(\boldsymbol{x}, \boldsymbol{y})$, we derive trip prices from the optimal dual solution, and argue that the equality between primal and dual objectives implies all rider best-response and driver best-response conditions in Definition 1.

Formally, for any decomposition $\boldsymbol{w}$ of the driver flow $\boldsymbol{y}$, $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w})$ is an optimal solution to the primal problem (24). Let $(\boldsymbol{p}, \omega, \boldsymbol{\eta})$ be any optimal solution to the dual problem (25). To prove that $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{p})$ forms a competitive equilibrium, we first derive a useful property of the optimal ω. For all $\kappa \in C$, constraints (25b) and (25c) imply that

$$\omega \sum_{(i,j)\in\kappa} d_{i,j} - \sum_{(i,j)\in\kappa} (p_{i,j} - c_{i,j}) = \sum_{(i,j)\in\kappa} \eta_{i,j} \geq 0.$$

Rearranging the terms, we get $\omega \geq \sum_{(i,j)\in\kappa}(p_{i,j} - c_{i,j}) \big/ \sum_{(i,j)\in\kappa} d_{i,j}$ for all $\kappa \in C$. Moreover, constraint (25e) ensures that $\omega \geq 0$. Since ω appears in the dual objective that shall be minimized, in every optimal solution, we must have

$$\omega = \max\left\{\max_{\kappa\in C} \frac{\sum_{(i,j)\in\kappa}(p_{i,j} - c_{i,j})}{\sum_{(i,j)\in\kappa} d_{i,j}},\ 0\right\} = \max\left\{\max_{\kappa\in C} \mu_\kappa,\ 0\right\}, \tag{27}$$

where $\mu_\kappa = \sum_{(i,j)\in\kappa}(p_{i,j} - c_{i,j}) \big/ \sum_{(i,j)\in\kappa} d_{i,j}$ is the surplus rate of cycle κ as defined in (2).

Strong duality implies that given optimal primal and solutions, the difference between the primal and dual objectives (26) is zero, and thus all four terms must be zero since each term is non-negative. This allows us to verify all conditions for competitive equilibria:

- To establish rider best-response (R1), observe that $\sum_{i,j\in\mathcal{L}} \int_{x_{i,j}}^{q_{i,j}(p_{i,j})} (v_{i,j}(s) - p_{i,j})\, \mathrm{d}s = 0$ implies that $x_{i,j} = q_{i,j}(p_{i,j})$ must hold for all $i, j \in \mathcal{L}$. As we've discussed in Appendix A.1 above, the integral $\int_{x_{i,j}}^{q_{i,j}(p_{i,j})} (v_{i,j}(s) - p_{i,j})\, \mathrm{d}s$ is strictly positive when $x_{i,j} \neq q_{i,j}(p_{i,j})$.
- The third term $\sum_{i,j\in\mathcal{L}} p_{i,j}(y_{i,j} - x_{i,j}) = 0$ implies that either $p_{i,j} = 0$ or $y_{i,j} = x_{i,j}$, since both $p_{i,j} \geq 0$ and $y_{i,j} - x_{i,j} \geq 0$ hold for all $(i, j) \in \mathcal{L}^2$. This gives us (D1), i.e. trips with relocation flows have zero prices.
- We now show (D2), that all drivers are used up when some cycle has a positive surplus. This is because $\omega\big(m - \sum_{i,j\in\mathcal{L}} d_{i,j} y_{i,j}\big) = 0$ together with (27) and primal feasibility imply that whenever $\max_{\kappa\in C} \mu_\kappa > 0$, we must have $\omega > 0$ and thus $m - \sum_{i,j\in\mathcal{L}} d_{i,j} y_{i,j} = 0$.

- Finally, the second term $\sum_{\kappa\in C} w_\kappa \sum_{(i,j)\in\kappa} \eta_{i,j} = 0$, together with constraint (25c), imply that for each cycle $\kappa \in C$, either $w_\kappa = 0$ and $\sum_{(i,j)\in\kappa} \eta_{i,j} = 0$. For each cycle κ with $w_\kappa > 0$, it must be the case that $\sum_{(i,j)\in\kappa} \eta_{i,j} = 0$. Given (25b), we know that $\omega = \sum_{(i,j)\in\kappa}(p_{i,j} - c_{i,j}) \big/ \sum_{(i,j)\in\kappa} d_{i,j} = \mu_\kappa$ must hold.
  - When $\max_{\kappa\in C} \mu_\kappa \geq 0$, $\omega = \max_{\kappa\in C} \mu_\kappa$ given (27). As a result, for each cycle κ with $w_\kappa > 0$, we have $\mu_\kappa = \omega = \max_{\kappa'\in C} \mu_{\kappa'}$. This gives us (D4), that all drivers get the highest possible surplus rate among all cycles.
  - When $\max_{\kappa\in C} \mu_\kappa < 0$, we need to show that $y_{i,j} = 0$ for all $i, j \in \mathcal{L}$ so that (D3) holds. Note that (27) implies $\omega = 0$ and therefore $\omega > \mu_\kappa$ for all $\kappa \in C$. Since $\omega = \mu_\kappa$ for all $\kappa \in C$ with $w_\kappa > 0$, it must be the case that $w_\kappa = 0$ for all $\kappa \in C$, implying that all driver flows are zero.[20]

This completes the proof that the outcome $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{p})$ constructed above is a competitive equilibrium. Additionally, we emphasize that since the difference of the primal and dual objective is zero for *any* optimal primal and dual solution, this above argument holds for any non-negative decomposition of $\boldsymbol{y}$. As a result, we have a (seemingly) stronger version of condition (D4) with "$\forall \boldsymbol{w}$" replacing "$\exists \boldsymbol{w}$":

$$\forall \boldsymbol{w},\ \forall \kappa \in C, \quad w_\kappa > 0 \implies \mu_\kappa = \max_{\kappa'\in C} \mu_{\kappa'}. \tag{28}$$

□

Proof of Claim 4. Given any CE outcome $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{p})$, for any cycle decomposition of the driver flow $\boldsymbol{w}$, we know $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w})$ is a feasible solution to the primal problem (24). We will prove the optimality of $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w})$ and $(\boldsymbol{p}, \omega, \boldsymbol{\eta})$ by constructing ω and $\boldsymbol{\eta}$ such that $(\boldsymbol{p}, \omega, \boldsymbol{\eta})$ forms a feasible dual solution, and showing that the difference between the primal and dual objectives (26) is zero.

Let μ_κ be the surplus rate of each cycle $\kappa \in C$ as defined in (2). We construct ω and $\boldsymbol{\eta}$ as follows:

$$\omega \triangleq \max\left\{\max_{\kappa\in C} \frac{\sum_{(i,j)\in\kappa}(p_{i,j} - c_{i,j})}{\sum_{(i,j)\in\kappa} d_{i,j}},\ 0\right\} = \max\left\{\max_{\kappa\in C} \mu_\kappa,\ 0\right\}, \tag{29}$$

$$\eta_{i,j} \triangleq c_{i,j} + d_{i,j}\omega - p_{i,j}, \qquad \forall (i,j) \in \mathcal{L}^2. \tag{30}$$

We first argue that $(\boldsymbol{p}, \omega, \boldsymbol{\eta})$ forms a feasible solution to the dual problem (25). The feasibility constraints (25b) and (25e) hold by construction. Constraint (25d) holds because trip prices are non-negative (as in condition (F1)). To see (25c), observe that by construction of ω, given any cycle $\kappa \in C$, we have $\omega \geq \sum_{(i,j)\in\kappa}(p_{i,j} - c_{i,j}) \big/ \sum_{(i,j)\in\kappa} d_{i,j}$. This implies

$$\sum_{(i,j)\in\kappa} \eta_{i,j} = \sum_{(i,j)\in\kappa} (c_{i,j} + d_{i,j}\omega - p_{i,j}) = \omega \sum_{(i,j)\in\kappa} d_{i,j} - \sum_{(i,j)\in\kappa} (p_{i,j} - c_{i,j}) \geq 0. \tag{31}$$

We now prove that the difference between the primal and dual objectives is zero. Consider each of the four terms in (26):

- The rider best response condition (R1) implies that $x_{i,j} = q_{i,j}(p_{i,j})$ for all $(i,j) \in \mathcal{L}^2$, thus the first term $\sum_{i,j\in\mathcal{L}} \int_{x_{i,j}}^{q_{i,j}(p_{i,j})} (v_{i,j}(s) - p_{i,j})\, \mathrm{d}s$ is zero.

[20] Note that this also implies that given any pair of optimal solutions, $\max_{\kappa\in C} \mu_\kappa < 0$ is not possible under the assumption that $q_{i,j}(\cdot)$ is strictly decreasing for all $r \geq 0$, for all $(i,j) \in \mathcal{L}^2$. This is because these conditions imply that $q_{i,j}(r) > 0$ for all $r \geq 0$, meaning that when rider best-response is satisfied (as we have established above), $x_{i,j} > 0$ must hold for all $(i,j) \in \mathcal{L}^2$, implying $y_{i,j} > 0$ as well due to primal feasibility.

- Condition (D1), i.e. trip price must be zero when the driver flow exceeds the rider flow, implies that the third term $\sum_{i,j\in\mathcal{L}} p_{i,j}(y_{i,j} - x_{i,j})$ is zero.
- For the fourth term $\omega\big(m - \sum_{i,j\in\mathcal{L}} d_{i,j}y_{i,j}\big)$, note that by construction (29), ω is strictly positive only if $\mu_\kappa > 0$ for some $\kappa \in C$, i.e. there exists some cycle with strictly positive surplus. In this case, condition (D2) implies that all drivers are used up, i.e. $m - \sum_{i,j\in\mathcal{L}} d_{i,j}y_{i,j} = 0$.
- What is left to prove that the second term of (26), i.e. $\sum_{\kappa\in C} w_\kappa \sum_{(i,j)\in\kappa} \eta_{i,j}$, is also zero. Intuitively, $\sum_{(i,j)\in\kappa} \eta_{i,j}$ is the total earnings a driver on cycle κ loses per cycle, relative to making a surplus of ω per unit of time. Driver best response therefore requires that either $\sum_{(i,j)\in\kappa} \eta_{i,j} = 0$ (i.e. drivers on this cycle is making the highest possible surplus rate) or $w_\kappa = 0$ (i.e. no driver is on the cycle). Formally, we need to discuss the following two cases:
  - If $\max_{\kappa\in C} \mu_\kappa < 0$, then by condition (D3), $y_{i,j} = 0$ for all trips $(i,j) \in \mathcal{L}$. In this case, $w_\kappa = 0$ for all cycles $\kappa \in \mathcal{L}$ thus $\sum_{\kappa\in C} w_\kappa \sum_{(i,j)\in\kappa} \eta_{i,j} = 0$ trivially holds.[21]
  - If $\max_{\kappa\in C} \mu_\kappa \geq 0$, $\omega = \max_{\kappa\in\mathcal{L}} \mu_\kappa$ given (29). Condition (D4) then implies that for all cycles $\kappa \in C$ with $w_\kappa > 0$, we must have $\mu_\kappa = \omega$. This implies that for all $\kappa \in C$ s.t. $w_\kappa > 0$, $\omega \sum_{(i,j)\in\kappa} d_{i,j} = \sum_{(i,j)\in\kappa}(p_{i,j} - c_{i,j})$, which further implies $\sum_{(i,j)\in\kappa} \eta_{i,j} = \sum_{(i,j)\in\kappa} c_{i,j} + d_{i,j}\omega - p_{i,j} = 0$.

This completes the proof that $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w})$ and $(\boldsymbol{p}, \omega, \boldsymbol{\eta})$ are optimal solutions to the primal problem (24) and the dual problem (25), respectively. □

A.3 Uniqueness of Optimal Dual Solution

Proposition 1. *There exists $\omega^* \geq 0$ and $\boldsymbol{\phi}^* \in \mathbb{R}^n$ with $\phi_n^* = 0$ such that any optimal solution $(\boldsymbol{p}, \omega, \boldsymbol{\phi})$ of the dual problem (7) satisfies*

(i) $\omega = \omega^*$,

(ii) $\phi_i - \phi_n = \phi_i^*$ *for all* $i \in \mathcal{L}$.

Proof. Let $(\boldsymbol{x}, \boldsymbol{y})$ be an optimal solution to the primal problem (6), and $(\boldsymbol{p}, \omega, \boldsymbol{\phi})$ and optimal solution to the dual problem (7). Similar to the proof of Lemma 1, we can decompose the difference of the primal objective (6a) and dual objective (7a) in way similar to (26), and prove that in order for the difference in the objectives to be zero, we must have $x_{i,j} = q_{i,j}(p_{i,j})$ for all $i, j \in \mathcal{L}$.

Note that the difference in the objectives is zero between any pair of optimal primal and dual solutions. The strict monotonicity of $q_{i,j}$ for all $i, j \in \mathcal{L}$ therefore implies that all optimal dual solutions must have the same prices $\boldsymbol{q}$. This immediately pin down the unique ω^* and gives us (i), since $p_{n,n} = c_{n,n} + \omega d_{n,n}$ must be the same for all optimal dual solutions thus it is impossible for two optimal dual solutions to have different ω's.

Now consider the prices $p_{i,n}$ for all $i \neq n$. The fact that $p_{i,n} = c_{i,n} + \omega d_{i,n} + \phi_i - \phi_n$ must be the same across all optimal dual solutions, together with the uniqueness of optimal ω, imply that $\phi_i - \phi_n$ must be the same for all optimal ϕ's. This finishes the proof of part (ii) and concludes the proof of this proposition. □

A.4 Proof of Lemma 2

We first state and prove the following result, which we use in the proofs of Lemma 2 and Lemma 3.

[21] Similar to the discussion in the proof of Claim 3, we note here that this case is not possible for any competitive equilibrium under the assumption that $q_{i,j}(r)$ is strictly decreasing for all $r \geq 0$ for all $(i,j) \in \mathcal{L}^2$.

Claim 5. *Suppose $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ is a strictly diagonally dominant matrix (i.e. $|M_{i,i}| > \sum_{j\neq i} |M_{i,j}|$ for all $i = 1, 2, \ldots, n$) with strictly positive diagonal entries and strictly negative non-diagonal entries. Then $\boldsymbol{M}$ is invertible and all entries of $\boldsymbol{M}^{-1}$ are strictly positive.*

Proof. Assume without loss of generality that $\max_i M_{i,i} < 1$ (otherwise scale $\boldsymbol{M}$ by a positive factor). Let $\boldsymbol{W} = \boldsymbol{I} - \boldsymbol{M}$, where $\boldsymbol{I}$ is the identity matrix. We know $\boldsymbol{W}$ is a matrix with positive entries, and for each i,

$$\sum_j |W_{i,j}| = W_{i,i} + \sum_{j\neq i} W_{i,j} = 1 - M_{i,i} + \sum_{j\neq i} |M_{i,j}| < 1.$$

This implies $\|\boldsymbol{W}\|_\infty < 1$. Therefore, as k goes to infinity, $\boldsymbol{W}^k$ converges to the zero matrix, and $\boldsymbol{I} + \boldsymbol{W} + \boldsymbol{W}^2 + \cdots + \boldsymbol{W}^k$ converges to $(\boldsymbol{I} - \boldsymbol{W})^{-1} = \boldsymbol{M}^{-1}$. Since all entries of $\boldsymbol{W}$ are positive, entries of $\boldsymbol{M}^{-1}$ are also positive. □

Define $\hat{q}_{i,j}(r) \triangleq q_{i,j}(r) + \tilde{q}_{i,j}(r)$ for all $i, j \in \mathcal{L}$ and $r \in \mathbb{R}$. This is the need of drivers on trip (i, j) to satisfy both demand and relocation at price r. The feasibility constraints (F1)–(F4), the price structure, and the market clearing conditions (MC1) and (MC2) together impose the following constraints on $\boldsymbol{\pi}$ and $\boldsymbol{\phi}$:

$$\sum_{j\in\mathcal{L}} \hat{q}_{i,j}(c_{i,j} + d_{i,j}\pi_i + \phi_i - \phi_j) = \sum_{j\in\mathcal{L}} \hat{q}_{j,i}(c_{j,i} + d_{j,i}\pi_j + \phi_j - \phi_i), \quad \forall i \in \mathcal{L},$$
$$\sum_{i,j\in\mathcal{L}} d_{i,j}\hat{q}_{i,j}(c_{i,j} + d_{i,j}\pi_i + \phi_i - \phi_j) = m,$$
$$c_{i,j} + d_{i,j}\pi_i + \phi_i - \phi_j \geq 0, \quad \forall i, j \in \mathcal{L}.$$

We prove Lemma 2, which states that the origin-based market-clearing outcome exists and is unique.

Lemma 2 (Existence and Uniqueness)**.** *For any economy $(m, \boldsymbol{d}, \boldsymbol{c}, \boldsymbol{q})$, any phantom demand function $\tilde{\boldsymbol{q}}$, and any OD-based price adjustment $\boldsymbol{\phi} \in \mathbb{R}^n$, there exists a unique vector of origin-based multipliers $\boldsymbol{\pi} \in \mathbb{R}^n$ that clears the market.*

Proof of Lemma 2. The fact that $\boldsymbol{\pi}$ clears the market means that trip prices are non-negative, the driver flow is balanced for each location, and exactly m units of drivers are dispatched.

We first prove the uniqueness by showing that the set of surge multipliers that achieves flow-balance is "ordered", i.e. if both $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ lead to flow-balanced outcomes, $\pi'_i < \pi_i$ for some $i \in \mathcal{L}$ implies that $\pi'_j < \pi_j$ for all $j \in \mathcal{L}$. Then, we prove that there exists a vector $\boldsymbol{\pi} \in \mathbb{R}^n$ that satisfies all three conditions above.

(i) Orderedness/Uniqueness. We first prove that for some fixed $\boldsymbol{\phi} \in \mathbb{R}^n$, if both $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ lead to flow balanced outcomes, then $\exists i \in \mathcal{L}$ s.t. $\pi_i > \pi'_i \implies \pi_j > \pi'_j$, $\forall j \in \mathcal{L}$. Equivalently, we show that $\mathcal{J} \neq \varnothing \implies \mathcal{J} = \mathcal{L}$, where $\mathcal{J} \triangleq \{i \in \mathcal{L} \mid \pi'_i < \pi_i\}$. Assume towards a contradiction, that $\mathcal{J} \neq \varnothing$ and $\mathcal{J} \neq \mathcal{L}$. The total flow from $\mathcal{J}$ to $\mathcal{L} \setminus \mathcal{J}$ is strictly higher under $\boldsymbol{\pi}'$ in comparison to that under $\boldsymbol{\pi}$, i.e.

$$\sum_{i\in\mathcal{J},\, j\in\mathcal{L}\setminus\mathcal{J}} \hat{q}_{i,j}(c_{i,j} + d_{i,j}\pi'_i + \phi_i - \phi_j) > \sum_{i\in\mathcal{J},\, j\in\mathcal{L}\setminus\mathcal{J}} \hat{q}_{i,j}(c_{i,j} + d_{i,j}\pi_i + \phi_i - \phi_j). \tag{32}$$

At the same time, the total flow from $\mathcal{L} \setminus \mathcal{J}$ to $\mathcal{J}$ is weakly lower under $\boldsymbol{\pi}'$, i.e.

$$\sum_{i\in\mathcal{L}\setminus\mathcal{J},\, j\in\mathcal{J}} \hat{q}_{i,j}(c_{i,j} + d_{i,j}\pi'_i + \phi_i - \phi_j) \leq \sum_{i\in\mathcal{L}\setminus\mathcal{J},\, j\in\mathcal{J}} \hat{q}_{i,j}(c_{i,j} + d_{i,j}\pi_i + \phi_i - \phi_j). \tag{33}$$

This implies that the flow-balance constraints for locations in $\mathcal{U}$ cannot be satisfied under both $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$, since under flow balance, the left-hand-sides of (32) and (33) must be equal, and so do the right-hand-sides of the two inequalities. This completes the proof of part (i).

Note that this also implies the *uniqueness* of origin-based market-clearing outcomes, i.e. there can be at most one market-clearing $\boldsymbol{\pi}$, if any exists at all. This is because if $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ both satisfy flow-balance and $\boldsymbol{\pi} \neq \boldsymbol{\pi}'$, either $\pi_i > \pi_i'$ for all $i \in \mathcal{L}$ or $\pi_i < \pi_i'$ for all $i \in \mathcal{L}$. The resulting outcomes therefore cannot use the same number of drivers, thus (MC2) cannot hold under both $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$.

(ii) Existence. Before proving that a market clearing $\boldsymbol{\pi} \in \mathbb{R}^n$ exists for any $\boldsymbol{\phi} \in \mathbb{R}^n$, we first introduce some notations.

For simplicity, define for all $i, j \in \mathcal{L}$ and all $\pi_i \in \mathbb{R}$ s.t. $c_{i,j} + d_{i,j}\pi_i + \phi_i - \phi_j \geq 0$,

$$\rho_{i,j}(\pi_i) \triangleq \hat{q}_{i,j}(c_{i,j} + d_{i,j}\pi_i + \phi_i - \phi_j).$$

Note that $\rho_{i,j}(\cdot)$ is defined for fixed $\boldsymbol{\phi} \in \mathbb{R}^n$, and represents the rate of driver flow on the (i, j) trip as a function of the surge multiplier π_i. Moreover, for all $\pi_i \geq (\phi_j - \phi_i - c_{i,j})/d_{i,j}$, $\rho_{i,j}(\cdot)$ is strictly positive, continuously differentiable and strictly decreasing in π_i.

We also denote the number of drivers needed to satisfy the demand originating from each location $i \in \mathcal{L}$ (as a function of π_i) as

$$S_i(\pi_i) \triangleq \sum_{j \in \mathcal{L}} d_{i,j}\rho_{i,j}(\pi_i).$$

Since trip prices must be non-negative, $S_i(\pi_i)$ is defined only for $\pi_i \geq \underline{\pi}_i$, where $\underline{\pi}_i$ is the lowest value of π_i such that $p_{i,j} \geq 0$ for all $j \in \mathcal{L}$:

$$\underline{\pi}_i \triangleq \max_{j \in \mathcal{L}} \frac{\phi_j - \phi_i - c_{i,j}}{d_{i,j}}, \quad \forall i \in \mathcal{L}. \tag{34}$$

Given properties of $\{\rho_{i,j}(\cdot)\}_{j \in \mathcal{L}}$, $S_i(\pi_i)$ is also strictly positive, continuously differentiable, and strictly decreasing in π_i for all $\pi_i \geq \underline{\pi}_i$. Finally, denote total the number of drivers needed to satisfy the driver flow as:

$$S(\boldsymbol{\pi}) \triangleq \sum_{i \in \mathcal{L}} S_i(\pi_i).$$

S is well-defined for $\boldsymbol{\pi} \in \mathbb{R}^n$ s.t. $\pi_i \geq \underline{\pi}_i$ for all $i \in \mathcal{L}$. We denote the domain of S as

$$\mathcal{D} \triangleq \{\boldsymbol{\pi} \in \mathbb{R}^n | \pi_i \geq \underline{\pi}_i, \ \forall i \in \mathcal{L}\} \tag{35}$$

To establish that a market-clearing outcome exists, what we need to prove is that there exists a $\boldsymbol{\pi} \in \mathcal{D}$ such that the resulting outcome is flow balanced and uses up all drivers:

$$\sum_{j \in \mathcal{L}} \rho_{k,j}(\pi_k) = \sum_{i \in \mathcal{L}} \rho_{i,k}(\pi_i), \ \forall k \in \mathcal{L}, \tag{36}$$

$$S(\boldsymbol{\pi}) = m. \tag{37}$$

We prove this by showing that the set of multipliers satisfying the flow balance constraints (36) forms a one-dimensional manifold, and then proving that there exist a $\boldsymbol{\pi}$ in this manifold s.t. (37) is satisfied.

A first challenge is that it is not straightforward to prove that there exists any $\boldsymbol{\pi} \in \mathcal{D}$ s.t. (36) is satisfied.

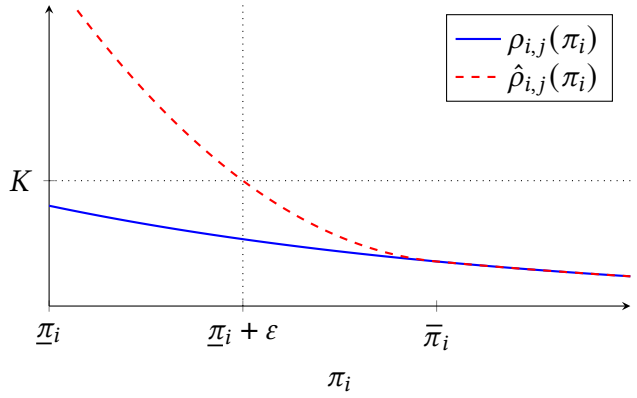


Figure 13: Sketch of functions $\rho_{i,j}(\pi_i)$ and $\hat{\rho}_{i,j}(\pi_i)$.

To overcome this issue, we construct an alternative $\hat{\boldsymbol{\rho}} = (\rho_{i,j}(\cdot))_{(i,j)\in\mathcal{L}^2}$ s.t.

(I) flow-balance (36) (stated in terms of $\hat{\boldsymbol{\rho}}$ instead of $\boldsymbol{\rho}$) is satisfied by construction for $\underline{\boldsymbol{\pi}}$, and

(II) $\hat{\boldsymbol{\rho}}$ coincides with $\boldsymbol{\rho}$ for all outcomes that use m units of drivers supply (i.e. $\hat{\boldsymbol{\rho}}(\boldsymbol{\pi}) = \boldsymbol{\rho}(\boldsymbol{\pi})$ for all $\boldsymbol{\pi} \in \mathcal{D}$ s.t. (37) holds under $\hat{\boldsymbol{\rho}}$).

First, observe that $S_i(\underline{\pi}_i) > m$ for all $i \in \mathcal{L}$ since given (34), $p_{i,j} = 0$ for at least one destination $j \in \mathcal{L}$, in which case the (i,j) trip would use $d_{i,j}\rho_{i,j}(\underline{\pi}_i) = d_{i,j}\hat{q}_{i,j}(0) \geq m$ units of drivers. Given that $S_i(\pi_i)$ is continuous and strictly decreasing in π_i, there must exist a unique $\overline{\pi}_i > \underline{\pi}_i$ s.t. $S_i(\overline{\pi}_i) = m$. Let $\varepsilon > 0$ be a small enough constant such that $\underline{\pi}_i + \varepsilon < \overline{\pi}_i$ for all $i \in \mathcal{L}$. Denote

$$K \triangleq \max_{i,j\in\mathcal{L}} \rho_{i,j}\big(\underline{\pi}_i + \varepsilon\big),$$

and define $\hat{\rho}_{i,j}(\cdot)$ for all $(i,j) \in \mathcal{L}^2$ as

$$\hat{\rho}_{i,j}(\pi_i) \triangleq \rho_{i,j}(\pi_i) + \big(K - \rho_{i,j}\big(\underline{\pi}_i + \varepsilon\big)\big)\left(\frac{\max\{\overline{\pi}_i - \pi_i, 0\}}{\overline{\pi}_i - (\underline{\pi}_i + \varepsilon)}\right)^2.$$

We know that $\hat{\rho}_{i,j}\big(\underline{\pi}_i + \varepsilon\big) = K$ for all $(i,j) \in \mathcal{L}$, and that $\hat{\rho}_{i,j}(\pi_i) = \rho_{i,j}(\pi_i)$ for all $\pi_i \geq \overline{\pi}_i$. See Figure 13 for an illustration.

It is straightforward to verify that (I) holds— the flow balance constraints are trivially satisfied since $\hat{\rho}_{i,j}\big(\underline{\pi}_i\big) = K$ for all $(i,j) \in \mathcal{L}^2$. We now argue that (II) also holds, by showing that for any $\boldsymbol{\pi} \in \mathcal{D}$ s.t. (37), we must have $\pi_i > \overline{\pi}_i$ for all $i \in \mathcal{L}$ since by definition of $\overline{\pi}_i$. $S_i(\overline{\pi}_i) = m$ for all $i \in \mathcal{L}$ and for all other $j \in \mathcal{L}$, $S_j(\pi_j)$ is strictly positive for all $\pi_j \geq \underline{\pi}_j$. As a result, if there exists any $i \in \mathcal{L}$ s.t. $\pi_i \leq \overline{\pi}_i$, $S(\boldsymbol{\pi}) > m$ must hold and (37) cannot be satisfied.

Denote $\hat{S}(\boldsymbol{\pi}) \triangleq \sum_{(i,j\in\mathcal{L})} d_{i,j}\hat{\rho}_{i,j}(\pi_i)$, we now define the following two subsets of $\mathcal{D}$ where the flow-balance constraint and the total supply constraint are satisfied, respectively:

$$\mathcal{F} = \left\{\boldsymbol{\pi} \in \mathcal{D} \;\middle|\; \sum_{j\in\mathcal{L}} \hat{\rho}_{k,j}(\pi_k) = \sum_{i\in\mathcal{L}} \hat{\rho}_{i,k}(\pi_i),\ \forall k \in \mathcal{L}\right\}, \tag{38}$$

$$\mathcal{T} = \big\{\boldsymbol{\pi} \in \mathcal{D} \;\big|\; \hat{S}(\boldsymbol{\pi}) = m\big\}. \tag{39}$$

What is left to prove is that $\mathcal{T} \cap \mathcal{F} \neq \emptyset$. The fact that $\hat{\boldsymbol{\rho}} = \boldsymbol{\rho}$ for all $\boldsymbol{\pi} \in \mathcal{T}$ then implies that (36) and (37) are both satisfied for any $\boldsymbol{\rho} \in \mathcal{T} \cap \mathcal{F}$, completing the proof of of the existence of an origin-based market-clearing outcome.

To prove $\mathcal{T} \cap \mathcal{F} \neq \emptyset$, we first define the following auxiliary function $\boldsymbol{b} : \mathcal{D} \to \mathbb{R}^{n-1}$ such that

$$b_k(\boldsymbol{\pi}) = \sum_{i \in \mathcal{L}} \hat{\rho}_{i,k}(\pi_i) - \sum_{j \in \mathcal{L}} \hat{\rho}_{k,j}(\pi_k), \quad k = 1, \ldots, n-1.$$

Intuitively, b_k represents the total amount of driver flow into location k minus the total driver flow out of location k. As a result, $\boldsymbol{b}(\boldsymbol{\pi}) = \mathbf{0}$ holds if and only if $\boldsymbol{\pi} \in \mathcal{F}$, meaning that $\mathcal{F}$ is the preimage of $\mathbf{0}_{n-1}$ under function $\boldsymbol{b}$.

Given that $\hat{\rho}_{i,j}(\cdot)$ is strictly positive, strictly decreasing, and continuously differentiable for each $(i, j) \in \mathcal{L}^2$, we know that for each $k = 1, \ldots, n-1$, $b_k(\boldsymbol{\pi})$ is also continuously differentiable in $\boldsymbol{\pi}$. Now, consider the set

$$\mathcal{S} \triangleq \mathcal{F} \cap \{\boldsymbol{\pi} \in \mathcal{D} \mid \hat{S}(\boldsymbol{\pi}) \geq m\},$$

we know $\underline{\boldsymbol{\pi}} = \left(\underline{\pi}_i\right)_{i \in \mathcal{L}} \in \mathcal{S}$ since as we've discussed earlier, $\boldsymbol{b}\left(\underline{\boldsymbol{\pi}}\right) = \mathbf{0}$, and the corresponding outcome uses strictly more than m units of drivers. We also know that $\mathcal{S}$ is closed, since

- $\mathcal{D}$ is closed,
- $\mathcal{F}$ as the preimage of $\mathbf{0}$ under the continuous function $\boldsymbol{b}$ is closed, and
- set $\{\boldsymbol{\pi} \in \mathcal{D} \mid \hat{S}(\boldsymbol{\pi}) \geq m\}$ as the preimage of the closed set $[m, +\infty)$ under the continuous function $\hat{S}$ is also closed.

The set $\mathcal{S}$ is also bounded above. Otherwise, if $\pi_i \to \infty$ for any $i \in \mathcal{L}$, then the prices $p_{i,j} \to \infty$ for all $j \in \mathcal{L}$ and the flow out of location i must go to zero. To satisfy the flow balance constraint, the flow into location i must also go to zero. Since $\boldsymbol{\phi}$ is fixed, this implies that $\pi_j \to \infty$ for all $j \in \mathcal{L}$. This further implies that $\hat{S}(\boldsymbol{\pi}) \to 0$, contradicting the fact that $\hat{S}(\boldsymbol{\pi}) \geq m$.

Let $\boldsymbol{\pi}^* = (\pi_1^*, \ldots, \pi_n^*)$ be the $\boldsymbol{\pi} \in \mathcal{S}$ with the largest π_n coordinate, i.e. $\pi_n^* = \sup_{\boldsymbol{\pi} \in \mathcal{S}} \pi_n$. We argue in the rest of the proof that $\hat{S}(\boldsymbol{\pi}^*) = m$ must hold, i.e. $\boldsymbol{\pi}^* \in \mathcal{F} \cap \mathcal{T}$.

We first show that the mapping $\boldsymbol{b}$ has no critical point. To see, this, divide the Jacobian matrix of function $\boldsymbol{b}$ into two blocks, with $\boldsymbol{A} \in \mathbb{R}^{(n-1)\times(n-1)}$, $\boldsymbol{\beta} \in \mathbb{R}^{n-1}$:

$$\mathrm{D}\boldsymbol{b}(\boldsymbol{\pi}) = \left[\begin{array}{ccc|c} \frac{\partial b_1}{\partial \pi_1}(\boldsymbol{\pi}) & \cdots & \frac{\partial b_1}{\partial \pi_{n-1}}(\boldsymbol{\pi}) & \frac{\partial b_1}{\partial \pi_n}(\boldsymbol{\pi}) \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\partial b_{n-1}}{\partial \pi_1}(\boldsymbol{\pi}) & \cdots & \frac{\partial b_{n-1}}{\partial \pi_{n-1}}(\boldsymbol{\pi}) & \frac{\partial b_{n-1}}{\partial \pi_n}(\boldsymbol{\pi}) \end{array}\right] \triangleq \left[\boldsymbol{A} \;\middle|\; \boldsymbol{\beta}\right].$$

For $k, j = 1, 2, \ldots, n-1$, $j \neq k$, we have:

$$\begin{aligned} A_{k,k} =& \frac{\partial b_k}{\partial \pi_k}(\boldsymbol{\pi}) = -\sum_{j \in \mathcal{L} \setminus \{k\}} \frac{\mathrm{d}}{\mathrm{d}\pi_k} \hat{\rho}_{k,j}(\pi_k) > 0, \\ A_{k,j} =& \frac{\partial b_k}{\partial \pi_j}(\boldsymbol{\pi}) = \frac{\mathrm{d}}{\mathrm{d}\pi_j} \hat{\rho}_{j,k}(\pi_j) < 0. \end{aligned}$$

Moreover, observe that

$$\left|A_{k,k}\right| = \sum_{j\in\mathcal{L}\setminus\{k\}}\left|\frac{\mathrm{d}}{\mathrm{d}\pi_k}\hat{\rho}_{k,j}(\pi_k)\right| > \sum_{j\in\mathcal{L}\setminus\{k,n\}}\left|\frac{\mathrm{d}}{\mathrm{d}\pi_k}\hat{\rho}_{k,j}(\pi_k)\right| = \sum_{j\in\mathcal{L}\setminus\{k,n\}}\left|A_{j,k}\right|.$$

By Claim 5, $\boldsymbol{A}$ is invertible, and this holds at any $\boldsymbol{\pi}$, i.e. $\boldsymbol{b}$ does not have any critical point. By the implicit function theorem, there exists an open set $\mathcal{U}\in\mathbb{R}$ containing π_n^* such that there exists a unique continuously differentiable function $u:\mathcal{U}\to\mathbb{R}^{n-1}$ with $u(\pi_n^*)=(\pi_1^*,\ldots,\pi_{n-1}^*)$ and $\boldsymbol{b}(u(\pi_n),\pi_n)=0$ for all $\pi_n\in\mathcal{U}$.

We now assume towards a contradiction, that $S(\boldsymbol{\pi}^*)\neq m$. By definition of $\mathcal{S}$, we must have $S(\boldsymbol{\pi}^*)>m$. Let $\{\pi_n^{(a)}\}_{a=1}^{\infty}$ be a sequence in $\mathcal{U}$ that converges to π_n^* s.t. $\pi_n^{(a)}>\pi_n^*$ for all $a=1,2,\ldots$, and denote $\boldsymbol{\pi}^{(a)}\triangleq(u(\pi_n^{(a)}),\pi_n^{(a)})$. Given the continuity of u and $\hat{S}$, we know that $\hat{S}(\boldsymbol{\pi}^{(a)})$ converges to $\hat{S}(\pi^*)$ as $a\to+\infty$. This implies for a that is sufficiently large, $\hat{S}(\boldsymbol{\pi}^{(a)})>m$ holds, which implies that $\boldsymbol{\pi}^{(a)}\in\{\boldsymbol{\pi}\in\mathcal{D}|\hat{S}(\boldsymbol{\pi})\geq m\}$. Since $\pi_n^{(a)}\in\mathcal{U}$, we also know that $\boldsymbol{\pi}^{(a)}\in\mathcal{F}$. This implies that $\boldsymbol{\pi}^{(a)}\in\mathcal{S}$, and this contradicts the fact that $\boldsymbol{\pi}^*$ is the point in $\mathcal{S}$ with the largest π_n. This completes the proof of this lemma. □

A.5 Proof of Lemma 3

Lemma 3. *The mapping $\boldsymbol{\Pi}:\mathbb{R}^{n-1}\to\mathbb{R}^n$ is continuously differentiable. The Jacobian matrix at the time $t-1$ market-clearing outcome $\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{(t-1)})\in\mathbb{R}^{n\times(n-1)}$ can be computed using history h_t.*

Proof. We first define an auxiliary function $\boldsymbol{g}:\mathbb{R}^{2n-1}\to\mathbb{R}^n$, where

$$g_k(\boldsymbol{\pi},\boldsymbol{\phi})\triangleq\sum_{i\in\mathcal{L}}\hat{q}_{i,k}(c_{i,k}+d_{i,k}\pi_i+\phi_i-\phi_k)-\sum_{j\in\mathcal{L}}\hat{q}_{k,j}(c_{k,j}+d_{k,j}\pi_k+\phi_k-\phi_j),\ k=1,2,\ldots,n-1, \tag{40}$$

$$g_n(\boldsymbol{\pi},\boldsymbol{\phi})\triangleq m-\sum_{i,j\in\mathcal{L}}d_{i,j}\hat{q}_{i,j}(c_{i,j}+d_{i,j}\pi_i+\phi_i-\phi_j). \tag{41}$$

Intuitively, given (MC1), the first $n-1$ components of $\boldsymbol{g}(\boldsymbol{\pi},\boldsymbol{\phi})$ correspond to the total flow of drivers into each location $k=1,\ldots,n-1$ minus the flow of drivers out of location k per unit of time, when trip prices are given by $p_{i,j}=c_{i,j}+d_{i,j}\pi_i+\phi_i-\phi_j$. The last component of $\boldsymbol{g}(\boldsymbol{\pi},\boldsymbol{\phi})$ is the difference between the amount of available drivers m and the amount of drivers needed to fulfill the flow. $\boldsymbol{g}(\boldsymbol{\pi},\boldsymbol{\phi})=\mathbf{0}$ thereby implies that the flow balance constraints and the total supply constraint (MC2) both hold, i.e. $\boldsymbol{\pi}$ clears the market given destination-based adjustment $\boldsymbol{\phi}$ (recall that the first $n-1$ flow balance constraints imply the flow-balance at location n). Fixing any $\boldsymbol{\phi}\in\mathbb{R}^{n-1}$, Lemma 2 implies that there exists a unique $\boldsymbol{\pi}\in\mathbb{R}^n$ s.t. $\boldsymbol{g}(\boldsymbol{\pi},\boldsymbol{\phi})=\mathbf{0}$, which is denoted as $\boldsymbol{\Pi}(\boldsymbol{\phi})$.

The function $\boldsymbol{g}$ is continuously differentiable since $\hat{\boldsymbol{q}}$ is continuously differentiable. The Jacobian matrix of $\boldsymbol{g}$ can be partitioned into two blocks, representing the Jacobian of $\boldsymbol{g}$ with respect to $\boldsymbol{\pi}$ and $\boldsymbol{\phi}$, respectively:

$$(\mathrm{D}\boldsymbol{g})(\boldsymbol{\pi},\boldsymbol{\phi})=\left[\begin{array}{ccc|ccc}\frac{\partial g_1}{\partial\pi_1}(\boldsymbol{\pi},\boldsymbol{\phi})&\cdots&\frac{\partial g_1}{\partial\pi_n}(\boldsymbol{\pi},\boldsymbol{\phi})&\frac{\partial g_1}{\partial\phi_1}(\boldsymbol{\pi},\boldsymbol{\phi})&\cdots&\frac{\partial g_1}{\partial\phi_{n-1}}(\boldsymbol{\pi},\boldsymbol{\phi})\\ \vdots&\ddots&\vdots&\vdots&\ddots&\vdots\\ \frac{\partial g_n}{\partial\pi_1}(\boldsymbol{\pi},\boldsymbol{\phi})&\cdots&\frac{\partial g_n}{\partial\pi_n}(\boldsymbol{\pi},\boldsymbol{\phi})&\frac{\partial g_n}{\partial\phi_1}(\boldsymbol{\pi},\boldsymbol{\phi})&\cdots&\frac{\partial g_n}{\partial\phi_{n-1}}(\boldsymbol{\pi},\boldsymbol{\phi})\end{array}\right]\triangleq\left[\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g}\ \middle|\ \mathrm{D}_{\boldsymbol{\phi}}\boldsymbol{g}\right]. \tag{42}$$

When $\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g}$ is invertable, the implicit function theoremimplies that $\boldsymbol{\Pi}(\boldsymbol{\phi})$ is continuously differentiable, with

$$\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi})=-(\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g})^{-1}\mathrm{D}_{\boldsymbol{\phi}}\boldsymbol{g}. \tag{43}$$

The rest of this proof establishes that $\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi})$ is invertable, and that the entries of $\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g}$ and $\mathrm{D}_{\boldsymbol{\phi}}\boldsymbol{g}$ depend on the trip durations $\{d_{i,j}\}_{(i,j)\in\mathcal{L}^2}$ and the local slope of the augmented demand function $\{\hat{q}'_{i,j}(p_{i,j})\}_{(i,j)\in\mathcal{L}^2}$, both of which can be observed by the platform.

We start from proving that $\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g}$ is invertable. First, partition $\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g}$ into four blocks, with $\boldsymbol{A}\in\mathbb{R}^{(n-1)\times(n-1)}$, $\boldsymbol{\beta}\in\mathbb{R}^{n-1}$, $\boldsymbol{\gamma}\in\mathbb{R}^{n-1}$, and $\lambda\in\mathbb{R}$:

$$\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g} \triangleq \begin{bmatrix} \boldsymbol{A} & \boldsymbol{\beta} \\ \boldsymbol{\gamma}^{\mathsf{T}} & \lambda \end{bmatrix}. \tag{44}$$

When A is invertable, the determinant of $\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g}$ can be written as:

$$\det(\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g}) = \det(\boldsymbol{A})\det(\lambda - \boldsymbol{\gamma}^{\mathsf{T}}\boldsymbol{A}^{-1}\boldsymbol{\beta}) \neq 0.$$

We will prove that $\det(\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g}) \neq 0$ by proving that (i) $\boldsymbol{A}$ is indeed invertable, and (ii) $\lambda - \boldsymbol{\gamma}^{\mathsf{T}}\boldsymbol{A}^{-1}\boldsymbol{\beta} > 0$.

To prove that (i) $\boldsymbol{A}$ is indeed invertable, observe that the diagonal entries of $\boldsymbol{A}$ is given by

$$A_{k,k} = \frac{\partial g_k}{\partial \pi_k}(\boldsymbol{\pi},\boldsymbol{\phi}) = -\sum_{j\in\mathcal{L}\setminus\{k\}} d_{k,j}\hat{q}'_{k,j}(c_{k,j} + d_{k,j}\pi_k + \phi_k - \phi_j),\ \forall k = 1,\ldots,n-1, \tag{45}$$

and that the off-diagonal entries are of the form

$$A_{k,\ell} = \frac{\partial g_k}{\partial \pi_\ell}(\boldsymbol{\pi},\boldsymbol{\phi}) = d_{\ell,k}\hat{q}'_{\ell,k}(c_{\ell,k} + d_{\ell,k}\pi_\ell + \phi_\ell - \phi_k),\ \forall k = 1,\ldots,n-1,\ \forall \ell = 1,\ldots,n-1,\ \ell \neq k \tag{46}$$

Since $\hat{q}_{i,j}(\cdot)$ is strictly decreasing for all $(i,j)\in\mathcal{L}^2$, we know that $A_{k,k} > 0$ for all $k \leq n-1$, and that $A_{k,\ell} < 0$ for all $k \leq n-1$ and all $\ell \neq k$. Moreover, note that

$$|A_{k,k}| - \sum_{\ell\neq k} |A_{\ell,k}| = -d_{k,n}\hat{q}'_{k,n}(c_{k,n} + d_{k,n}\pi_k + \phi_k - \phi_n) > 0.$$

It follows from Claim 5 that $\boldsymbol{A}$ is invertible and all entries of $\boldsymbol{A}^{-1}$ are strictly positive.

To prove that (ii) $\lambda - \boldsymbol{\gamma}^{\mathsf{T}}\boldsymbol{A}^{-1}\boldsymbol{\beta} > 0$, observe that

$$\beta_k = \frac{\partial g_k}{\partial \pi_n}(\boldsymbol{\pi},\boldsymbol{\phi}) = d_{\ell,k}\hat{q}'_{n,k}(c_{n,k} + d_{n,k}\pi_n + \phi_n - \phi_k) < 0,\ \forall k = 1,\ldots,n-1, \tag{47}$$

$$\gamma_\ell = \frac{\partial g_n}{\partial \pi_\ell}(\boldsymbol{\pi},\boldsymbol{\phi}) = -\sum_{j\in\mathcal{L}} d^2_{\ell,j}q'_{\ell,j}(c_{\ell,j} + d_{\ell,j}\pi_\ell + \phi_\ell - \phi_j) > 0,\ \forall \ell = 1,\ldots,n-1, \tag{48}$$

$$\lambda = \frac{\partial g_n}{\partial \pi_n}(\boldsymbol{\pi},\boldsymbol{\phi}) = -\sum_{j\in\mathcal{L}} d^2_{n,j}q'_{n,j}(c_{n,j} + d_{n,j}\pi_h + \phi_n - \phi_j) > 0. \tag{49}$$

As a result, $\boldsymbol{\gamma}^{\mathsf{T}}\boldsymbol{A}^{-1}\boldsymbol{\beta} < 0$, and thus $\lambda - \boldsymbol{\gamma}^{\mathsf{T}}\boldsymbol{A}^{-1}\boldsymbol{\beta} > 0$.

What is left to show is that the entries of $\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g}$ and $\mathrm{D}_{\boldsymbol{\phi}}\boldsymbol{g}$ depend on the trip durations and the local slope of the augmented demand function. For $\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g}$, this is shown in Equations (45)–(49). For $\mathrm{D}_{\boldsymbol{\phi}}\boldsymbol{g}$, it is convenient

to partition the Jacobian matrix as

$$\mathrm{D}_{\boldsymbol{\phi}}\boldsymbol{g} \triangleq \begin{bmatrix} \boldsymbol{B} \\ \boldsymbol{\theta}^{\mathsf{T}} \end{bmatrix}, \tag{50}$$

where $\boldsymbol{B} \in \mathbb{R}^{(n-1)\times(n-1)}$ and $\boldsymbol{\theta} \in \mathbb{R}^{(n-1)}$. For $k = 1, 2, \ldots, n-1$, the diagonal entries of $\boldsymbol{B}$ are given by

$$B_{k,k} = \frac{\partial g_k}{\partial \phi_k}(\boldsymbol{\pi}, \boldsymbol{\phi}) = -\sum_{i \in \mathcal{L}\setminus\{k\}} \hat{q}'_{i,k}(c_{i,k} + d_{i,k}\pi_i + \phi_i - \phi_k) - \sum_{j \in \mathcal{L}\setminus\{k\}} \hat{q}'_{k,j}(c_{k,j} + d_{k,j}\pi_k + \phi_k - \phi_j). \tag{51}$$

For $k = 1, 2, \ldots, n-1$ and $\ell = 1, 2, \ldots, n-1$ with $\ell \neq k$, the off-diagonal entry

$$B_{k,\ell} = \frac{\partial g_k}{\partial \phi_\ell}(\boldsymbol{\pi}, \boldsymbol{\phi}) = \hat{q}'_{\ell,k}(c_{\ell,k} + d_{\ell,k}\pi_\ell + \phi_\ell - \phi_k) + \hat{q}'_{k,\ell}(c_{k,\ell} + d_{k,\ell}\pi_k + \phi_k - \phi_\ell). \tag{52}$$

Finally, for $\ell = 1, 2, \ldots, n-1$,

$$\theta_\ell = \frac{\partial g_n}{\partial \phi_\ell}(\boldsymbol{\pi}, \boldsymbol{\phi}) = -\sum_{j \in \mathcal{L}\setminus\{\ell\}} d_{\ell,j}\hat{q}'_{\ell,j}(c_{\ell,j} + d_{\ell,j}\pi_\ell + \phi_\ell - \phi_j) + \sum_{i \in \mathcal{L}\setminus\{\ell\}} d_{i,\ell}\hat{q}'_{i,\ell}(c_{i,\ell} + d_{i,\ell}\pi_i + \phi_i - \phi_\ell). \tag{53}$$

This completes the proof of this lemma. □

A.6 The Update Direction

In this section, we prove the following result, which shows that any the market-clearing outcome corresponding to any $\boldsymbol{\phi} \in \mathbb{R}^{n-1}$, there exists a unique $\boldsymbol{\phi}' \in \mathbb{R}^{n-1}$ at which all π_i's are equalized in the linear approximation of $\boldsymbol{\Pi}$, i.e. there exists $\xi \in \mathbb{R}$ s.t. $\Pi(\boldsymbol{\phi}) + \mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi})(\boldsymbol{\phi}' - \boldsymbol{\phi}) = \xi \mathbf{1}_n$. This is the direction towards which the INP mechanism updates the OD-based price adjustments week-over-week.

Lemma 4. *The matrix* $\begin{bmatrix} \mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}) & -\mathbf{1} \end{bmatrix} \in \mathbb{R}^{n\times n}$ *has full rank at any* $\boldsymbol{\phi} \in \mathbb{R}^{n-1}$.

Proof. Recall from the proof of Lemma 3 that the Jacobian of $\boldsymbol{\Pi}$ is of the form $\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}) = -(\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g})^{-1}\mathrm{D}_{\boldsymbol{\phi}}\boldsymbol{g}$, where $\boldsymbol{g}$ is given by (40) and (41) and $\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g}$ is full rank. The matrix $\begin{bmatrix} \mathrm{D}\boldsymbol{\Pi} & -\mathbf{1} \end{bmatrix}$ can therefore be written as

$$\begin{bmatrix} \mathrm{D}\boldsymbol{\Pi} & -\mathbf{1} \end{bmatrix} = \begin{bmatrix} -(\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g})^{-1}\mathrm{D}_{\boldsymbol{\phi}}\boldsymbol{g} & -\mathbf{1} \end{bmatrix} = -(\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g})^{-1}\begin{bmatrix} \mathrm{D}_{\boldsymbol{\phi}}\boldsymbol{g} & (\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g})\mathbf{1} \end{bmatrix}.$$

As a result, it is sufficient to show that $\begin{bmatrix} \mathrm{D}_{\boldsymbol{\phi}}\boldsymbol{g} & (\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g})\mathbf{1} \end{bmatrix}$ also has full rank. To prove this, we partition $\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g}$ and $\mathrm{D}_{\boldsymbol{\phi}}\boldsymbol{g}$ as:

$$\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g} = \begin{bmatrix} \bar{\boldsymbol{A}} \\ \bar{\boldsymbol{\gamma}}^{\mathsf{T}} \end{bmatrix},$$
$$\mathrm{D}_{\boldsymbol{\phi}}\boldsymbol{g} = \begin{bmatrix} \boldsymbol{B} \\ \boldsymbol{\theta}^{\mathsf{T}} \end{bmatrix}.$$

Here, $\boldsymbol{B} \in \mathbb{R}^{(n-1)\times(n-1)}$ and $\boldsymbol{\theta} \in \mathbb{R}^{n-1}$ are exactly the same as in the partition (50), given by (51)–(53). On the other hand, $\bar{\boldsymbol{A}} \in \mathbb{R}^{(n-1)\times n}$ and $\bar{\boldsymbol{\gamma}} \in \mathbb{R}^n$ are given by

$$\bar{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{\beta} \end{bmatrix},$$

$$\bar{\boldsymbol{\gamma}}^\top = \begin{bmatrix}\boldsymbol{\gamma}^\top & \lambda\end{bmatrix},$$

where $\boldsymbol{A}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and λ are components of the partition (44) of $\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g}$ (see (45)–(49)).

It is straightforward to verify that

- the row sum of $\bar{\boldsymbol{A}}$ is equal to $\boldsymbol{\theta}$, i.e. $\bar{\boldsymbol{A}}\mathbf{1} = \boldsymbol{\theta}$, following from (45)–(47) and (53),
- $\bar{\boldsymbol{\gamma}}^\top\mathbf{1} > 0$ because $\bar{\gamma}_i > 0$ for all $i \in \mathcal{L}$ given (48) and (49), and
- $\boldsymbol{B}$ is symmetric, given (52).

The matrix $\begin{bmatrix}\mathrm{D}_{\boldsymbol{\phi}}\boldsymbol{g} & (\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g})\mathbf{1}\end{bmatrix}$ can therefore be written as:

$$\begin{bmatrix}\mathrm{D}_{\boldsymbol{\phi}}\boldsymbol{g} & (\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g})\mathbf{1}\end{bmatrix} = \begin{bmatrix}\boldsymbol{B} & \bar{\boldsymbol{A}}\mathbf{1} \\ \boldsymbol{\theta}^\top & \bar{\boldsymbol{\gamma}}^\top\mathbf{1}\end{bmatrix} = \begin{bmatrix}\boldsymbol{B} & \boldsymbol{\theta} \\ \boldsymbol{\theta}^\top & \bar{\boldsymbol{\gamma}}^\top\mathbf{1}\end{bmatrix} = \begin{bmatrix}\boldsymbol{I} & \boldsymbol{\theta}/\bar{\boldsymbol{\gamma}}^\top\mathbf{1} \\ \mathbf{0} & 1\end{bmatrix}\begin{bmatrix}\boldsymbol{B} - \boldsymbol{\theta}\boldsymbol{\theta}^\top/\bar{\boldsymbol{\gamma}}^\top\mathbf{1} & \mathbf{0} \\ \mathbf{0}^\top & \bar{\boldsymbol{\gamma}}^\top\mathbf{1}\end{bmatrix}\begin{bmatrix}\boldsymbol{I} & \mathbf{0} \\ \boldsymbol{\theta}^\top/\bar{\boldsymbol{\gamma}}^\top\mathbf{1} & 1\end{bmatrix}.$$

It suffices to show that $\boldsymbol{B} - \boldsymbol{\theta}\boldsymbol{\theta}^\top/\bar{\boldsymbol{\gamma}}^\top\mathbf{1}$ is positive definite. To this end, we show that for any $n-1$ dimensional vector $\boldsymbol{\zeta} \in \mathbb{R}^{n-1}$ such that $\boldsymbol{\zeta} \neq \mathbf{0}_{n-1}$, $\boldsymbol{\zeta}^\top(\bar{\boldsymbol{\gamma}}^\top\mathbf{1}\boldsymbol{B} - \boldsymbol{\theta}\boldsymbol{\theta}^\top)\boldsymbol{\zeta} > 0$.

In the rest of this proof, we make use of the structure of $\bar{\boldsymbol{A}}$ and $\boldsymbol{B}$ to show that

$$\boldsymbol{\zeta}^\top(\bar{\boldsymbol{\gamma}}^\top\mathbf{1}\boldsymbol{B} - \boldsymbol{\theta}\boldsymbol{\theta}^\top)\boldsymbol{\zeta} = \boldsymbol{\zeta}^\top\bar{\boldsymbol{\gamma}}^\top\mathbf{1}\boldsymbol{B}\boldsymbol{\zeta} - \boldsymbol{\zeta}^\top\boldsymbol{\theta}\boldsymbol{\theta}^\top\boldsymbol{\zeta} = \bar{\boldsymbol{\gamma}}^\top\mathbf{1}(\boldsymbol{\zeta}^\top\boldsymbol{B}\boldsymbol{\zeta}) - (\boldsymbol{\zeta}^\top\boldsymbol{\theta})^2 > 0,\ \forall\boldsymbol{\zeta} \in \mathbb{R}^{n-1},\ \boldsymbol{\zeta} \neq \mathbf{0}_{n-1}, \tag{54}$$

or we have a counter-example of the Cauchy–Schwarz inequality. For simplicity of notation, we define $\zeta_n \triangleq 0$, despite the fact that $\boldsymbol{\zeta}$ is a vector with dimension $n-1$. First, given (48) and (49) (recall that $\bar{\gamma}_n = \lambda$), the element-wise sum of $\bar{\boldsymbol{\gamma}}$ can be written as

$$\bar{\boldsymbol{\gamma}}^\top\mathbf{1} = -\sum_{i=1}^{n}\sum_{j=1}^{n} d_{i,j}^2\hat{q}'_{i,j}.$$

Combining (51) and (52), $\boldsymbol{\zeta}^\top\boldsymbol{B}\boldsymbol{\zeta}$ can be rewritten as:

$$\begin{aligned}
\boldsymbol{\zeta}^\top\boldsymbol{B}\boldsymbol{\zeta} &= \sum_{i=1}^{n-1}\sum_{j=1}^{n-1}\zeta_i\zeta_j\left(\hat{q}'_{j,i} + \hat{q}'_{i,j} - \mathbb{1}[j=i]\left(\sum_{k=1}^{n}\hat{q}'_{k,i} + \sum_{\ell=1}^{n}\hat{q}'_{i,\ell}\right)\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\zeta_i\zeta_j\left(\hat{q}'_{j,i} + \hat{q}'_{i,j} - \mathbb{1}[j=i]\left(\sum_{k=1}^{n}\hat{q}'_{k,i} + \sum_{\ell=1}^{n}\hat{q}'_{i,\ell}\right)\right) \\
&= -\sum_{i=1}^{n}\zeta_i^2\left(\sum_{k=1}^{n}\hat{q}'_{k,i} + \sum_{\ell=1}^{n}\hat{q}'_{i,\ell}\right) + \sum_{i=1}^{n}\sum_{j=1}^{n}\zeta_i\zeta_j(\hat{q}'_{i,j} + \hat{q}'_{j,i}) \\
&= -\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\zeta_j^2\hat{q}'_{i,j} + \sum_{i=1}^{n}\sum_{j=1}^{n}\zeta_i^2\hat{q}'_{i,j}\right) + 2\sum_{i=1}^{n}\sum_{j=1}^{n}\zeta_i\zeta_j\hat{q}'_{i,j} \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{n}(\zeta_i - \zeta_j)^2\hat{q}'_{i,j},
\end{aligned}$$

and $\boldsymbol{\zeta}^\mathsf{T}\boldsymbol{\theta}$ can be written as

$$\begin{aligned}
\boldsymbol{\zeta}^\mathsf{T}\boldsymbol{\theta} &= \sum_{i=1}^{n-1} \zeta_i \left(-\sum_{j=1}^{n} d_{i,j}\hat{q}'_{i,j} + \sum_{k=1}^{n} d_{k,i}\hat{q}'_{k,i} \right) \\
&= \sum_{i=1}^{n} \zeta_i \left(-\sum_{j=1}^{n} d_{i,j}\hat{q}'_{i,j} + \sum_{k=1}^{n} d_{k,i}\hat{q}'_{k,i} \right) \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{n} \zeta_i d_{i,j}\hat{q}'_{i,j} + \sum_{i=1}^{n}\sum_{j=1}^{n} \zeta_j d_{i,j}\hat{q}'_{i,j} \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{n} (\zeta_i - \zeta_j) d_{i,j}\hat{q}'_{i,j}.
\end{aligned}$$

By Cauchy–Schwarz inequality, $(\bar{\boldsymbol{y}}^\mathsf{T}\mathbf{1})(\boldsymbol{\zeta}^\mathsf{T}\boldsymbol{B}\boldsymbol{\zeta}) \geq (\boldsymbol{\zeta}^\mathsf{T}\boldsymbol{\theta})^2$. Moreover, the equality is achieved if and only if the ratios between the $d_{i,j}$'s and $\zeta_i - \zeta_j$'s are the same for all $i, j = 1, \ldots, n$. This is impossible because while $d_{i,j}$ is always positive, $\zeta_i - \zeta_j$ is zero when $i = j$ and $\zeta_i - \zeta_n = \zeta_i \neq 0$ for some $i \leq n-1$. Therefore, $\boldsymbol{\zeta}^\mathsf{T}(\bar{\boldsymbol{y}}^\mathsf{T}\mathbf{1}\boldsymbol{B} - \boldsymbol{\theta}\boldsymbol{\theta}^\mathsf{T})\boldsymbol{\zeta} > 0$ must hold. This completes the proof that $\begin{bmatrix} \mathrm{D}_{\boldsymbol{\phi}}\boldsymbol{g} & (\mathrm{D}_{\boldsymbol{\pi}}\boldsymbol{g})\mathbf{1} \end{bmatrix}$ is positive definite, and also the proof of this lemma. □

A.7 Properties of the Lyapunov Function

We establish in this section useful properties of the alternative objective function f as defined in (20). We use $\mathbf{0}_\ell$ and $\mathbf{1}_\ell$ denote ℓ-dimensional vectors of all zeros and all ones, respectively, and drop the subscript when the dimension is clear from the context.

Lemma 5. *The function f as defined in (20) is continuously differentiable, has a single critical point corresponding to the unique global minimum, and is coercive (i.e. every sublevel set is bounded).*

Proof of Lemma 5. We first prove that f has a unique minimizer $\boldsymbol{\phi}^\star \in \mathbb{R}^{n-1}$, achieving $f(\boldsymbol{\phi}^\star) = 0$. Given any augmented economy $(\boldsymbol{d}, \boldsymbol{c}, \boldsymbol{q}, m, \tilde{\boldsymbol{q}})$, consider the following optimization problem, where $\hat{v}(\cdot) = \hat{q}^{-1}(\cdot)$ is the inverse function of $\hat{q}(\cdot)$ (recall that $\hat{q}_{i,j}(\cdot) = q_{i,j}(\cdot) + q_{i,j}(\cdot)$ is strictly decreasing for all $(i,j) \in \mathcal{L}^2$):

$$\begin{aligned}
\underset{\boldsymbol{y} \in \mathbb{R}^{n^2}}{\text{maximize}} \quad & \sum_{i,j\in\mathcal{L}} \left(\int_0^{y_{i,j}} \hat{v}_{i,j}(s)\,\mathrm{d}s - c_{i,j}y_{i,j} \right) && \text{(55a)} \\
\text{subject to} \quad & \sum_{j\in\mathcal{L}} y_{k,j} = \sum_{i\in\mathcal{L}} y_{i,k}, \quad \forall k = 1, 2, \ldots, n-1, && \text{(55b)} \\
& \sum_{i,j\in\mathcal{L}} d_{i,j}y_{i,j} = m && \text{(55c)}
\end{aligned}$$

Let $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_{n-1})$ denote the dual variables corresponding to (55b), and let ω be the dual variable

corresponding to the total supply constraint (55c), and fix $\phi_n \triangleq 0$.[22] The dual of (55) can be written as

$$\underset{\omega \in \mathbb{R},\, \boldsymbol{\phi} \in \mathbb{R}^{n-1}}{\text{minimize}} \quad m\omega + \sum_{i,j \in \mathcal{L}} \int_{c_{i,j}+d_{i,j}\omega+\phi_i-\phi_j}^{\infty} \hat{q}_{i,j}(r)\, \mathrm{d}r, \tag{56a}$$

and difference between primal and dual objectives is of the form

$$\sum_{i,j \in \mathcal{L}} \int_{y_{i,j}}^{\hat{q}_{i,j}(c_{i,j}+d_{i,j}\omega+\phi_i-\phi_j)} (\hat{v}_{i,j}(s) - (c_{i,j}+d_{i,j}\omega+\phi_i-\phi_j))\, \mathrm{d}s + \omega\Big(m - \sum_{i,j \in \mathcal{L}} d_{i,j} y_{i,j}\Big) + \sum_{k \in \mathcal{L}} \phi_k \Big(\sum_{i \in \mathcal{L}} y_{i,k} - \sum_{j \in \mathcal{L}} y_{k,j}\Big). \tag{57}$$

Let $\boldsymbol{y}^\star \in \mathbb{R}^{n^2}$ and $(\boldsymbol{\phi}^\star, \omega^\star)$ be optimal solutions of the primal (55c) and dual (56a), respectively. The primal-dual difference (57) is zero for $\boldsymbol{y}^\star \in \mathbb{R}^{n^2}$ and $(\boldsymbol{\phi}^\star, \omega^\star)$ since the primal problem is convex. It is straightforward to verify that (57) being zero implies that the surge multiplier $\boldsymbol{\pi} = \omega^\star \mathbf{1}_n$ (i.e. $\pi_1 = \cdots = \pi_n = \omega^\star$) clears the market when the OD-based adjustments are given by $\boldsymbol{\phi}^\star$. In other words, $\boldsymbol{\Pi}(\boldsymbol{\phi}^\star) = \omega^\star \mathbf{1}$, implying $f(\boldsymbol{\phi}^\star) = 0$. This is a global minimum, since f is non-negative by construction.

We now prove that the global minimum is unique. Suppose $\boldsymbol{\phi}' \in \mathbb{R}^{n-1}$ also minimize (20), i.e. $f(\boldsymbol{\phi}') = 0$. We know that the market-clearing surge multipliers $\boldsymbol{\pi}' = \boldsymbol{\Pi}(\boldsymbol{\phi}')$ must also satisfy $\pi'_1 = \cdots = \pi'_n$. Let $\omega' = \pi'_1$, and let $\boldsymbol{y}'$ be the resulting driver flow under this market-clearing outcome. $\boldsymbol{y}'$ is a feasible primal solution. We can verify that the market-clearing conditions in Definition 2 imply that the difference between the primal objective achieved by $\boldsymbol{y}'$ and the dual objective achieved by $(\boldsymbol{\phi}', \omega')$ is zero, meaning that both solutions are optimal. With an argument similar to the proof of Proposition 1, we can prove that the optimal dual solution is unique. This implies that $\boldsymbol{\phi}^\star = \boldsymbol{\phi}'$, i.e. (20) has a unique minimum.

To complete the proof of this theorem, we establish the following results.

Proposition 2. *Every stationary point of f is a global minimum.*

Proposition 3. *f as defined in (20) is coercive, i.e. every sublevel set of f is bounded. Formally, $\forall U \in \mathbb{R}, \exists M > 0$, $\forall \boldsymbol{\phi} \in \mathbb{R}^{n-1}, f(\boldsymbol{\phi}) \leq U \implies \|\boldsymbol{\phi}\|_\infty \leq M$.*

□

A.7.1 Proof of Proposition 2

Proposition 2. *Every stationary point of f is a global minimum.*

Proof. First note that the alternative objective function f as defined in (20) can be written as:

$$f(\boldsymbol{\phi}) = \sum_{i \in \mathcal{L}} \Big(\Pi_i(\boldsymbol{\phi}) - \frac{1}{n} \sum_{j \in \mathcal{L}} \Pi_j(\boldsymbol{\phi})\Big)^2 = \frac{1}{2n} \sum_{i,j \in \mathcal{L}} (\Pi_i(\boldsymbol{\phi}) - \Pi_j(\boldsymbol{\phi}))^2.$$

Define $\boldsymbol{\xi} : \mathbb{R}^{n-1} \to \mathbb{R}^n$ such that for each $i = 1, 2, \ldots, n$,

$$\xi_i(\boldsymbol{\phi}) \triangleq 2\Pi_i(\boldsymbol{\phi}) - \frac{2}{n} \sum_{j \in \mathcal{L}} \Pi_j(\boldsymbol{\phi}).$$

[22] Note that the first $n-1$ flow balance constraints (55b) imply that location n is also flow balanced. For convenience of notation, we drop the redundant flow balance constraint, which is equivalent to fixing the dual variable at $\phi_n = 0$.

The gradient of f can be written as:

$$\begin{aligned}\nabla f(\boldsymbol{\phi}) &= \frac{1}{n}\sum_{i,j\in\mathcal{L}}(\Pi_i(\boldsymbol{\phi})-\Pi_j(\boldsymbol{\phi}))(\nabla\Pi_i(\boldsymbol{\phi})-\nabla\Pi_j(\boldsymbol{\phi}))\\ &= \frac{2}{n}\sum_{i,j\in\mathcal{L}}(\Pi_i(\boldsymbol{\phi})-\Pi_j(\boldsymbol{\phi}))\nabla\Pi_i(\boldsymbol{\phi})\\ &= \sum_{i\in\mathcal{L}}\Big(2\Pi_i(\boldsymbol{\phi})-\frac{2}{n}\sum_{j\in\mathcal{L}}\Pi_j(\boldsymbol{\phi})\Big)\nabla\Pi_i(\boldsymbol{\phi})\\ &= \mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi})^{\mathsf{T}}\boldsymbol{\xi}(\boldsymbol{\phi}).\end{aligned}$$

Observe that the entries of $\boldsymbol{\xi}(\boldsymbol{\phi})$ always sum to zero, i.e. $\mathbf{1}^{\mathsf{T}}\boldsymbol{\xi}(\boldsymbol{\phi})=0$, $\forall\boldsymbol{\phi}\in\mathbb{R}^{n-1}$. As a result,

$$\nabla f(\boldsymbol{\phi})=\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi})^{\mathsf{T}}\boldsymbol{\xi}(\boldsymbol{\phi})=\mathbf{0}_{n-1}\iff\begin{bmatrix}\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}) & -\mathbf{1}_n\end{bmatrix}^{\mathsf{T}}\boldsymbol{\xi}(\boldsymbol{\phi})=\mathbf{0}_n.$$

By Lemma 4, the matrix $\begin{bmatrix}\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}) & -\mathbf{1}_n\end{bmatrix}$ has full rank at all $\boldsymbol{\phi}\in\mathbb{R}^{n-1}$. $\nabla f(\boldsymbol{\phi})=\mathbf{0}$ then implies that $\boldsymbol{\xi}(\boldsymbol{\phi})=\mathbf{0}$, meaning that for all $i\in\mathcal{L}$,

$$\Pi_i(\boldsymbol{\phi})=\frac{1}{n}\sum_{j\in\mathcal{L}}\Pi_j(\boldsymbol{\phi}),$$

In other words, $\Pi_1(\boldsymbol{\phi})=\cdots=\Pi_n(\boldsymbol{\phi})$, in which case $f(\boldsymbol{\phi})=0$. Since f is non-negative, this completes the proof that every stationary point of f is a global minimum. □

A.7.2 Proof of Proposition 3

Proposition 3. *f as defined in (20) is coercive, i.e. every sublevel set of f is bounded. Formally, $\forall U\in\mathbb{R}$, $\exists M>0$, $\forall\boldsymbol{\phi}\in\mathbb{R}^{n-1}$, $f(\boldsymbol{\phi})\le U\implies\|\boldsymbol{\phi}\|_\infty\le M$.*

Proof. Assume toward contradiction that f has an unbounded sublevel set, i.e.

$$\exists U^*\in\mathbb{R},\ \forall M>0,\ \exists\boldsymbol{\phi},\ (f(\boldsymbol{\phi})\le U^*)\wedge(\|\boldsymbol{\phi}\|_\infty>M).\tag{58}$$

We know it must be the case that $U^*>0$, since (as we have established earlier in this section) f is non-negative and achieves the global minimum $f=0$ at a unique $\boldsymbol{\phi}^\star\in\mathbb{R}^{n-1}$.

The first term in (58), $f(\boldsymbol{\phi})\le U^*$, implies that $\big(\Pi_i(\boldsymbol{\phi})-\frac{1}{n}\sum_{j\in\mathcal{L}}\Pi_j(\boldsymbol{\phi})\big)^2\le U^*$ holds for all $i\in\mathcal{L}$. As a result, at any $\boldsymbol{\phi}\in\mathbb{R}^{n-1}$ s.t. $f(\boldsymbol{\phi})\le U^*$, we must have

$$\max_{i\in\mathcal{L}}\Pi_i(\boldsymbol{\phi})-\min_{i\in\mathcal{L}}\Pi_i(\boldsymbol{\phi})=\Big(\max_{i\in\mathcal{L}}\Pi_i(\boldsymbol{\phi})-\frac{1}{n}\sum_{i\in\mathcal{L}}\Pi_i(\boldsymbol{\phi})\Big)+\Big(\frac{1}{n}\sum_{i\in\mathcal{L}}\Pi_i(\boldsymbol{\phi})-\min_{i\in\mathcal{L}}\Pi_i(\boldsymbol{\phi})\Big)\le 2\sqrt{U^*}.\tag{59}$$

The second term in (58), $\|\boldsymbol{\phi}\|_\infty>M$, implies that there exists some $i\le n-1$ s.t. $|\phi_i|=|\phi_i-\phi_n|>M$. Assume w.l.o.g. that $|\phi_1-\phi_n|>M$ (otherwise re-label the locations). Consider the prices of the two trips between location 1 and location n:

$$p_{1,n}=c_{1,n}+d_{1,n}\Pi_1(\boldsymbol{\phi})+\phi_1-\phi_n,$$

$$p_{n,1} = c_{n,1} + d_{n,1}\Pi_n(\boldsymbol{\phi}) + \phi_n - \phi_1.$$

Trip prices being non-negative implies that

$$\Pi_1(\boldsymbol{\phi}) \geq \frac{\phi_n - \phi_1 - c_{1,n}}{d_{1,n}},$$
$$\Pi_n(\boldsymbol{\phi}) \geq \frac{\phi_1 - \phi_n - c_{n,1}}{d_{n,1}}.$$

Since $|\phi_1 - \phi_n| > M$, we have

$$\max_i \Pi_i(\boldsymbol{\phi}) \geq \max\{\Pi_1(\boldsymbol{\phi}),\ \Pi_n(\boldsymbol{\phi})\} \geq \min\left\{\frac{M - c_{1,n}}{d_{1,n}},\ \frac{M - c_{n,1}}{d_{n,1}}\right\}.$$

Furthermore, applying (59), we get

$$\min_i \Pi_i(\boldsymbol{\phi}) \geq \max_i \Pi_i(\boldsymbol{\phi}) - 2\sqrt{U^*} \geq \min\left\{\frac{M - c_{1,n}}{d_{1,n}},\ \frac{M - c_{n,1}}{d_{n,1}}\right\} - 2\sqrt{U^*}. \tag{60}$$

The rest of the proof applies (60) to establish that there exists a sufficiently large $\bar{M} > 0$ s.t. for all $\boldsymbol{\phi} \in \mathbb{R}^{n-1}$ such that $\|\boldsymbol{\phi}\|_\infty > \bar{M}$, the corresponding market-clearing outcome cannot satisfy $f(\boldsymbol{\phi}) \leq U^*$. We start from constructing one such $\bar{M}$.

First, for each $(i, j) \in \mathcal{L}$, we define $\bar{p}_{i,j} \geq 0$ as the smallest non-negative price for the (i, j) trip s.t. the driver flow for (i, j) trip is no greater than $m/(2n^2 \max_{i,j\in\mathcal{L}} d_{i,j})$:

$$\bar{p}_{i,j} = \inf\left\{r \in \mathbb{R}_{\geq 0} \,\middle|\, \hat{q}_{i,j}(r) \leq \varepsilon\right\},$$

where

$$\varepsilon \triangleq \frac{m}{(n^3 + n^2 + n)\max_{i,j\in\mathcal{L}}\{d_{i,j}\}}.$$

If $\hat{q}_{i,j}(0) \leq \varepsilon$, we know $\bar{p}_{i,j} = 0$. If $\hat{q}_{i,j}(0) > \varepsilon$, since $\hat{q}_{i,j}$ is continuous, strictly decreasing, and converges to 0 as the price approaches infinity, $\bar{p}_{i,j}$ is the unique prices at which $\hat{q}_{i,j}(\bar{p}_{i,j}) = \varepsilon$.

For each origin, let $\overline{\pi}_i$ be the smallest π_i such that $c_{i,j} + d_{i,j}\pi_i$ is at least $\bar{p}_{i,j}$ for all destinations $i \in \mathcal{L}$:

$$\overline{\pi}_i \triangleq \max_{j\in\mathcal{L}}\left\{\left(\bar{p}_{i,j} - c_{i,j}\right)/d_{i,j}\right\}, \tag{61}$$

and let $\bar{M}$ be sufficiently large such that (60) guarantees that $\min_i \Pi_i(\boldsymbol{\phi})$ is no smaller than the maximum $\overline{\pi}_i$ among all $i \in \mathcal{L}$:

$$\bar{M} \triangleq \max\left\{\left(\max_{i\in\mathcal{L}} \overline{\pi}_i + 2\sqrt{U^*}\right)d_{1,n} + c_{1,n},\ \left(\max_{i\in\mathcal{L}} \overline{\pi}_i + 2\sqrt{U^*}\right)d_{n,1} + c_{n,1}\right\}. \tag{62}$$

Now consider a $\boldsymbol{\phi} \in \mathbb{R}^{n-1}$ s.t. $\|\boldsymbol{\phi}\|_\infty > \bar{M}$, under which $f(\boldsymbol{\phi}) \leq U^*$ is satisfied ((58) guarantees that such $\boldsymbol{\phi}$ exists). Applying (60), (61) and (62), we know that for all $i, j \in \mathcal{L}$ s.t. $\phi_i \geq \phi_j$, the trip price under the

market-clearing outcome must satisfy

$$p_{i,j} = c_{i,j} + d_{i,j}\Pi_i(\boldsymbol{\phi}) + \phi_i - \phi_j \geq c_{i,j} + d_{i,k} \min_i \Pi_i(\boldsymbol{\phi}) \geq \bar{p}_{i,j}.$$

This implies that for all $i, j \in \mathcal{L}$ such that $\phi_i \geq \phi_j$, the driver flow rate

$$y_{i,j} = \hat{q}(p_{i,j}) \leq \hat{q}(\bar{p}_{i,j}) \leq \varepsilon.$$

Relabel the locations s.t. $\phi_1 \geq \phi_2 \geq \cdots \geq \phi_n$ and apply Claim 6 below, we know that the total amount of drivers dispatched under the market-clearing outcome is bounded by:

$$\sum_{i,j\in\mathcal{L}} d_{i,j} y_{i,j} \leq \max_{i,j\in\mathcal{L}}\{d_{i,j}\} \sum_{i,j\in\mathcal{L}} y_{i,j} \leq \max_{i,j\in\mathcal{L}}\{d_{i,j}\}(n + n^2 + n^3)\varepsilon = m.$$

This is a contradiction, since given (MC2), any market-clearing outcome must use exactly m units of drivers. This completes the proof that f is coercive. □

Claim 6. *Suppose driver flow $\boldsymbol{y} \in \mathbb{R}^{n^2}$ is balanced (i.e. $\sum_{i\in\mathcal{L}} y_{i,k} = \sum_{j\in\mathcal{L}} y_{k,j}$ for all $k \in \mathcal{L}$) and satisfies*

$$y_{i,j} \leq \varepsilon, \ \forall i, j \in \mathcal{L} \text{ s.t. } i \leq j,$$

the sum of driver flow over all trips is bounded by

$$\sum_{i,j\in\mathcal{L}} y_{i,j} < (n + n^2 + n^3)\varepsilon.$$

Proof. First, we decompose the sum $\sum_{i,j\in\mathcal{L}} y_{i,j}$ into two parts:

$$\sum_{i,j\in\mathcal{L}} y_{i,j} = \sum_{i,j\in\mathcal{L}, i\leq j} y_{i,j} + \sum_{i,j\in\mathcal{L}, i>j} y_{i,j}.$$

Since $y_{i,j} \leq \varepsilon$ if $i \leq j$, the first term is bounded by

$$\sum_{i,j\in\mathcal{L}, i\leq j} y_{i,j} \leq \varepsilon \frac{n(n+1)}{2} < (n + n^2)\varepsilon.$$

Due to the flow-balance constraint, for each $k \geq 2$ we must have $\sum_{i\geq k, j<k} y_{i,j} = \sum_{i\geq k, j<k} y_{j,i}$. As a result, the second term can be written as

$$\sum_{i,j\in\mathcal{L}, i>j} y_{i,j} = \sum_{k=2}^{n} \sum_{i\geq k, j<k} y_{i,j} \leq \sum_{k=2}^{n} (k-1)(n-k+1)\varepsilon < n^3\varepsilon.$$

This completes the proof of this claim. □

A.8 Proof of Theorem 2

Theorem 2. *Assuming (A1) and (A2), given any economy, the iterative network pricing mechanism converges to an origin-based market-clearing outcome where all multipliers are equal, i.e. $\exists \omega^\star \in \mathbb{R}$ s.t. $\lim_{t\to\infty} \pi_i^{(t)} = \omega^\star$*

for all $i \in \mathcal{L}$. The welfare suboptimality of the limit outcome is upper bounded by

$$m \max\{0, -\omega^\star\} + \sum_{i,j \in \mathcal{L}} e_{i,j}. \tag{23}$$

The following proposition is a direct corollary of Proposition 1.2.1 of Bertsekas [2016].

Proposition 4. *Let f be the objective function defined in (20), and $(\boldsymbol{\phi}^{[k]})_{k\in\mathbb{N}}$ be a sequence generated by:*

$$\boldsymbol{\phi}^{[k+1]} = \boldsymbol{\phi}^{[k]} + \nu^{[k]} \alpha^{[k]} \boldsymbol{\delta}^{[k]}.$$

Assume the followings:

(i) $(\alpha^{[k]}\boldsymbol{\delta}^{[k]})_{k\in\mathbb{N}}$ is gradient related, i.e. for any subsequence $(\boldsymbol{\phi}^{[k]})_{k\in\mathcal{K}}$ (for some $\mathcal{K} \subseteq \mathbb{N}$, $|\mathcal{K}| = \infty$) that converges to a non-stationary point of f, the corresponding subsequence $(\alpha^{[k]}\boldsymbol{\delta}^{[k]})_{k\in\mathcal{K}}$ is bounded and satisfies

$$\limsup_{k\to\infty,\, k\in\mathcal{K}} \nabla f(\boldsymbol{\phi}^{[k]})^\mathsf{T} \alpha^{[k]} \boldsymbol{\delta}^{[k]} < 0.$$

(ii) $\nu^{[k]}$ is chosen by the Armijo rule, i.e. $\nu^{[k]} = \beta^{N_k}$, where $\beta \in (0,1)$, $\sigma \in (0,1)$, and N_k is the smallest non-negative integer N for which

$$f(\boldsymbol{\phi}^{[k]}) - f(\boldsymbol{\phi}^{[k]} + \beta^N \alpha^{[k]} \boldsymbol{\delta}^{[k]}) > -\sigma\beta^N \nabla f(\boldsymbol{\phi}^{[k]})^\mathsf{T} \alpha^{[k]} \boldsymbol{\delta}^{[k]}.$$

Then, every limit point of $(\boldsymbol{\phi}^{[k]})_{k\in\mathbb{N}}$ is a stationary point of f.

Recall that $t_0, t_1, t_2, \ldots$ are the values that t' in Line 7 of Algorithm 1 ever takes. For each integer $k \geq 0$, define

$$\begin{aligned}
\boldsymbol{\phi}^{[k]} &\triangleq \boldsymbol{\phi}^{(t_k)}, \\
\boldsymbol{\pi}^{[k]} &\triangleq \boldsymbol{\pi}^{(t_k)}, \\
\boldsymbol{\delta}^{[k]} &\triangleq \boldsymbol{\delta}^{(t_k+1)}, \\
\alpha^{[k]} &\triangleq \alpha^{(t_k+1)}.
\end{aligned}$$

We know that

$$\boldsymbol{\phi}^{[k+1]} = \boldsymbol{\phi}^{[k]} + \beta^{N_k} \alpha^{[k]} \boldsymbol{\delta}^{[k]},$$

where $N_k = t_{k+1} - t_k - 1$ is the smallest integer $N \geq 0$ such that

$$f(\boldsymbol{\phi}^{[k]} + \beta^N \alpha^{[k]} \boldsymbol{\delta}^{[k]}) < f(\boldsymbol{\phi}^{[k]}) + \sigma \nabla f(\boldsymbol{\phi}^{[k]})^\mathsf{T} \beta^N \alpha^{[k]} \boldsymbol{\delta}^{[k]}. \tag{63}$$

Proposition 5. *The sublevel set $\{\boldsymbol{\phi} \in \mathbb{R}^{n-1} \mid f(\boldsymbol{\phi}) \leq f(\boldsymbol{\phi}^{(0)})\}$ is compact, and it contains $\boldsymbol{\phi}^{[k]}$ for all $k \in \mathbb{N}$. $\boldsymbol{\phi}^{[k]}$ converges to $\boldsymbol{\phi}^\star$.*

Proof. We first apply Proposition 4 and show that every limit point of $(\boldsymbol{\phi}^{[k]})_{k\in\mathbb{N}}$ is a stationary point of f. By construction, the backtracking factor is chosen by the Armijo rule thus condition (ii) of Proposition 4 is satisfied. What is left to verify is condition (i), that $(\alpha^{[k]}\boldsymbol{\delta}^{[k]})_{k\in\mathbb{N}}$ is gradient related. By definition of the INP

mechanism (see Algorithm 1), we know that

$$\boldsymbol{\delta}^{[k]} = \left(\begin{bmatrix} -\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{[k]}) & \mathbf{1}_n \end{bmatrix}^{-1} \boldsymbol{\pi}^{[k]}\right)_{1:n-1},$$
$$\alpha^{[k]} = \min\left\{1,\ \frac{\tau}{\|\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{[k]})\boldsymbol{\delta}^{[k]}\|_\infty}\right\},$$

Moreover, the gradient of f with respect to $\boldsymbol{\phi}$ at $\boldsymbol{\phi}^{[k]}$ is of the form:

$$\nabla f(\boldsymbol{\phi}^{[k]}) = 2\big(\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{[k]})\big)^\mathsf{T}\left(\boldsymbol{\pi}^{[k]} - \left(\frac{1}{n}\mathbf{1}_n^\mathsf{T}\boldsymbol{\pi}^{[k]}\right)\mathbf{1}_n\right) \tag{64}$$
$$= 2\big(\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{[k]})\big)^\mathsf{T}\left(\boldsymbol{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\mathsf{T}\right)\boldsymbol{\pi}^{[k]}. \tag{65}$$

Since $\mathbf{1}_n^\mathsf{T}\big(\boldsymbol{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\mathsf{T}\big)\boldsymbol{\pi}^{[k]} = 0$, we can rewrite ∇f with a zero appended at the end as

$$\begin{bmatrix} \nabla f(\boldsymbol{\phi}^{[k]}) \\ 0 \end{bmatrix} = 2\begin{bmatrix} \mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{[k]}) & -\mathbf{1}_n \end{bmatrix}^\mathsf{T}\left(\boldsymbol{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\mathsf{T}\right)\boldsymbol{\pi}^{[k]}.$$

Therefore,

$$\begin{aligned}
\nabla f(\boldsymbol{\phi}^{[k]})^\mathsf{T}\boldsymbol{\delta}^{[k]} &= \begin{bmatrix} \nabla f(\boldsymbol{\phi}^{[k]}) \\ 0 \end{bmatrix}^\mathsf{T} \begin{bmatrix} -\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{[k]}) & \mathbf{1}_n \end{bmatrix}^{-1} \boldsymbol{\pi}^{[k]} \\
&= 2\big(\boldsymbol{\pi}^{[k]}\big)^\mathsf{T}\left(\boldsymbol{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\mathsf{T}\right)^\mathsf{T} \begin{bmatrix} \mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{[k]}) & -\mathbf{1}_n \end{bmatrix} \begin{bmatrix} -\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{[k]}) & \mathbf{1}_n \end{bmatrix}^{-1} \boldsymbol{\pi}^{[k]} \\
&= -2\big(\boldsymbol{\pi}^{[k]}\big)^\mathsf{T}\left(\boldsymbol{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\mathsf{T}\right)\boldsymbol{\pi}^{[k]} \\
&\leq 0.
\end{aligned} \tag{66}$$

The last inequality holds with equality if and only if $\pi_1^{[k]} = \pi_2^{[k]} = \cdots = \pi_n^{[k]}$, at which point $f = 0$ achieves the global minimum. As a result, $\nabla f(\boldsymbol{\phi}^{[k]})^\mathsf{T}\boldsymbol{\delta}^{[k]} < 0$ (i.e. $\boldsymbol{\delta}^{[k]}$ is a descent direction) for f whenever $\boldsymbol{\phi}^{[k]}$ is not optimal. What this implies is that the last term of (63), $\sigma\nabla f(\boldsymbol{\phi}^{[k]})^\mathsf{T}\beta^N\alpha^{[k]}\boldsymbol{\delta}^{[k]}$, is weakly negative, and that the objective value $f(\boldsymbol{\phi}^{[k]})$ is weakly decreasing in k. As a result, every point in the sequence $(\boldsymbol{\phi}^{[k]})_{k\in\mathbb{N}}$ must be contained in the sublevel set $\{\boldsymbol{\phi} \in \mathbb{R}^{n-1} \mid f(\boldsymbol{\phi}) \leq f(\boldsymbol{\phi}^{(0)})\}$, which is bounded (and compact) given Proposition 3.

Observe that both $\boldsymbol{\delta}^{[k]}$ and $\alpha^{[k]}$ can be written as functions of $\boldsymbol{\phi}^{[k]}$. The functions are continuous because $\boldsymbol{\Pi}$ is continuously differentiable. Since every continuous function over a compact space must be bounded, $(\alpha^{[k]}\boldsymbol{\delta}^{[k]})_{k\in\mathbb{N}}$ is bounded. For the same reason, $(\|\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{[k]})\boldsymbol{\delta}^{[k]}\|_\infty)_{k\in\mathbb{N}}$ is bounded, and thus $(\alpha^{[k]})_{k\in\mathbb{N}}$ is bounded away from zero.

Now consider a subsequence $(\boldsymbol{\phi}^{[k]})_{k\in\mathcal{K}}$ that converges to a non-stationary point. We know that as $k \to \infty$, the limits of both $\alpha^{[k]}$ and $\nabla f(\boldsymbol{\phi}^{[k]})^\mathsf{T}\boldsymbol{\delta}^{[k]}$ exist due to continuity. $\lim_{k\to\infty,\,k\in\mathcal{K}} \alpha^{[k]} > 0$ since for each $k \in \mathbb{N}$, $\alpha^{[k]}$ is positive and bounded away from zero. Moreover, $\lim_{k\to\infty,\,k\in\mathcal{K}} \nabla f(\boldsymbol{\phi}^{[k]})^\mathsf{T}\boldsymbol{\delta}^{[k]} < 0$ since $\lim_{k\to\infty,\,k\in\mathcal{K}} \boldsymbol{\phi}^{[k]}$ is not a stationary point of f. This implies that

$$\limsup_{k\to\infty,\,k\in\mathcal{K}} \nabla f(\boldsymbol{\phi}^{[k]})^\mathsf{T}\alpha^{[k]}\boldsymbol{\delta}^{[k]} < 0,$$

and completes the proof of condition (i), that $(\alpha^{[k]}\boldsymbol{\delta}^{[k]})_{k\in\mathbb{N}}$ is gradient related.

It follows from Proposition 4 that every limit point of $(\boldsymbol{\phi}^{[k]})_{k\in\mathbb{N}}$ is a stationary point of f. By Lemma 5, there is a single stationary point corresponding to the unique global minimum of f. It then follows that every limit point of $(\boldsymbol{\phi}^{[k]})_{k\in\mathbb{N}}$ is the unique global minimum, and therefore $(\boldsymbol{\phi}^{[k]})_{k\in\mathbb{N}}$ converges to the global minimum, completing the proof of this proposition.

To see the last statement, assume toward contradiction that the sequence $(\boldsymbol{\phi}^{[k]})_{k\in\mathbb{N}}$ does not converge to the global minimum $\boldsymbol{\phi}^\star$. In this case, there exists $\varepsilon > 0$ and $\mathcal{K} \subseteq \mathbb{N}$ with $|\mathcal{K}| = \infty$ such that $\boldsymbol{\phi}^{[k]}$ resides outside of the open ball $\mathcal{B}_\varepsilon(\boldsymbol{\phi}^\star) \triangleq \{\boldsymbol{\phi} \in \mathbb{R}^{n-1} \mid \|\boldsymbol{\phi} - \boldsymbol{\phi}^\star\|_2 < \varepsilon\}$ for all $k \in \mathcal{K}$. (Here $\|\cdot\|_2$ denotes the 2-norm for vectors). Observe that the infinite subsequence $(\boldsymbol{\phi}^{[k]})_{k\in\mathcal{K}}$ is contained in the compact set $\{\boldsymbol{\phi} \in \mathbb{R}^{n-1} \mid f(\boldsymbol{\phi}) \le f(\boldsymbol{\phi}^{(0)})\} \setminus \mathcal{B}_\varepsilon(\boldsymbol{\phi}^\star)$, thus must have a limit point in that set, which does not contain the global minimum. This contradicts with the fact that every limit point of $(\boldsymbol{\phi}^{[k]})_{k\in\mathbb{N}}$ is the global minimum. □

Proposition 6. *$(\boldsymbol{\delta}^{[k]})_{k\in\mathbb{N}}$ converges to $\mathbf{0}_{n-1}$.*

Proof. We prove this proposition by showing that the 2-norm of $\boldsymbol{\delta}^{[k]}$ converges to zero. By definition, $\boldsymbol{\delta}^{[k]}$ is the first $n-1$ entries of $\left(\begin{bmatrix}-\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{[k]}) & \mathbf{1}_n\end{bmatrix}^{-1}\boldsymbol{\pi}^{[k]}\right)$. As a result, for each $k \ge 0$, there exist $\xi^{[k]} \in \mathbb{R}$ s.t.

$$\begin{bmatrix}\boldsymbol{\delta}^{[k]} \\ \xi^{[k]}\end{bmatrix} = \begin{bmatrix}-\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{[k]}) & \mathbf{1}_n\end{bmatrix}^{-1}\boldsymbol{\pi}^{[k]},$$

which gives us

$$\begin{bmatrix}-\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{[k]}) & \mathbf{1}_n\end{bmatrix}\begin{bmatrix}\boldsymbol{\delta}^{[k]} \\ \xi^{[k]}\end{bmatrix} = \boldsymbol{\pi}^{[k]}.$$

Subtracting $\frac{1}{n}\mathbf{1}_n^\mathsf{T}\boldsymbol{\pi}^{[k]}\mathbf{1}_n$ from both sides of the equation, we have

$$\begin{bmatrix}-\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{[k]}) & \mathbf{1}_n\end{bmatrix}\begin{bmatrix}\boldsymbol{\delta}^{[k]} \\ \xi^{[k]} - \frac{1}{n}\mathbf{1}_n^\mathsf{T}\boldsymbol{\pi}^{[k]}\end{bmatrix} = \boldsymbol{\pi}^{[k]} - \frac{1}{n}\mathbf{1}_n^\mathsf{T}\boldsymbol{\pi}^{[k]}\mathbf{1}_n,$$

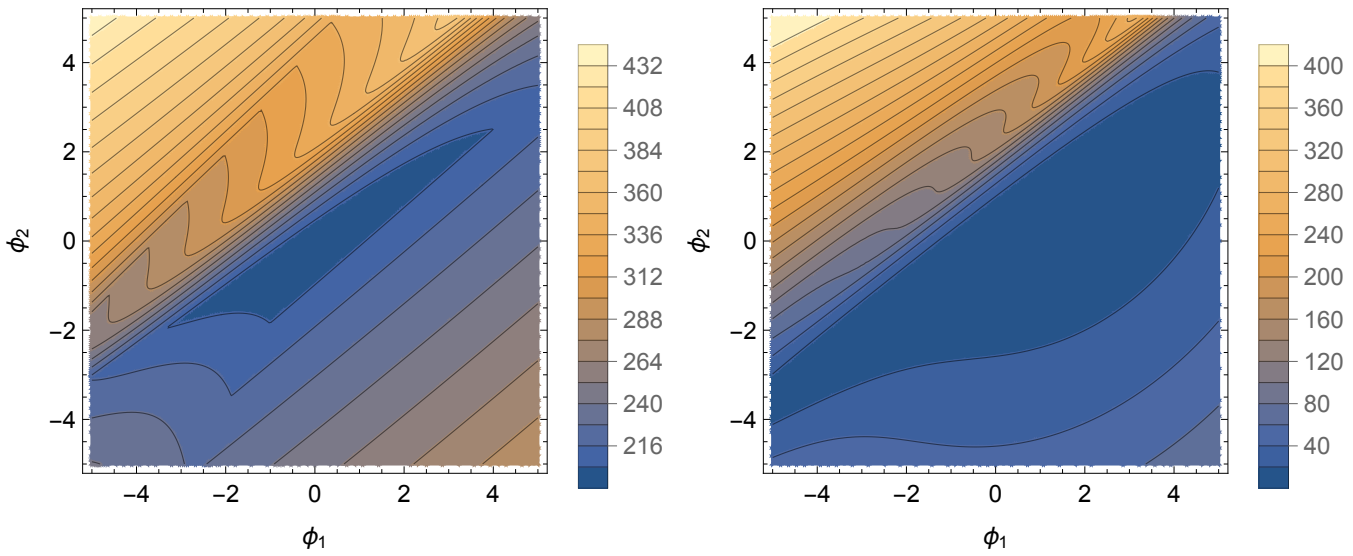
and equivalently,

$$\begin{bmatrix}\boldsymbol{\delta}^{[k]} \\ \xi^{[k]} - \frac{1}{n}\mathbf{1}_n^\mathsf{T}\boldsymbol{\pi}^{[k]}\end{bmatrix} = \begin{bmatrix}-\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{[k]}) & \mathbf{1}_n\end{bmatrix}^{-1}\left(\boldsymbol{\pi}^{[k]} - \frac{1}{n}\mathbf{1}_n^\mathsf{T}\boldsymbol{\pi}^{[k]}\mathbf{1}_n\right).$$

Let $\|\cdot\|_2$ denote the 2-norm for vectors and the induced 2-norm for matrices. We have

$$\begin{aligned}\|\boldsymbol{\delta}^{[k]}\|_2^2 \le \left\|\begin{bmatrix}\boldsymbol{\delta}^{[k]} \\ \xi^{[k]} - \frac{1}{n}\mathbf{1}_n^\mathsf{T}\boldsymbol{\pi}^{[k]}\end{bmatrix}\right\|_2^2 &\le \left\|\begin{bmatrix}-\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{[k]}) & \mathbf{1}_n\end{bmatrix}^{-1}\right\|_2^2\left\|\boldsymbol{\pi}^{[k]} - \frac{1}{n}\mathbf{1}^\mathsf{T}\boldsymbol{\pi}^{[k]}\mathbf{1}_n\right\|_2^2 \\ &= \left\|\begin{bmatrix}-\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{[k]}) & \mathbf{1}_n\end{bmatrix}^{-1}\right\|_2^2 f(\boldsymbol{\phi}^{[k]}).\end{aligned}$$

Given Proposition 5 and the continuity of $\mathrm{D}\boldsymbol{\Pi}$, we know that as $k \to \infty$, $\left\|\begin{bmatrix}-\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^{[k]}) & \mathbf{1}_n\end{bmatrix}^{-1}\right\|_2^2$ converges to $\left\|\begin{bmatrix}-\mathrm{D}\boldsymbol{\Pi}(\boldsymbol{\phi}^\star) & \mathbf{1}_n\end{bmatrix}^{-1}\right\|_2^2$, and $f(\boldsymbol{\phi}^{[k]})$ converges to 0. Therefore, $\|\boldsymbol{\delta}^{[k]}\|_2^2$ converges to 0, completing the proof of this proposition. □



(a) Dual objective.

(b) Lyanupov function f.

Figure 14: Contour plots of the non-convex dual objective and the Lyapunov function f for Example 2.

B Additional Examples

The following example demonstrates that the dual objective (as a function of the OD-based adjustments $\boldsymbol{\phi}$) and the Lyanupov function as defined in (20) are not necessarily quasi-convex.

Example 2. Consider an economy with three-locations, a total of $m = 8.6$ units of drivers, and rider demand functions

$$q_{1,1}(r) = q_{1,2}(r) = q_{2,3}(r) = q_{3,1}(r) = q_{3,2}(r) = q_{3,3}(r) = \exp(-r/30),$$
$$q_{1,3}(r) = q_{2,1}(r) = q_{2,2}(r) = 4\exp(-r).$$

The trip costs are zero, i.e. $c_{i,j} = 0$ for all $i, j \in \mathcal{L}$, and the durations are given by $d_{i,j} = 1$ for all i, j. Figure 14 shows the dual objective and the Lyapunov function f as a function of ϕ_1 and ϕ_2, fixing $\phi_3 = 0$. We can see that the contours are not convex, meaning that the dual objective and the function f are not quasi-convex.

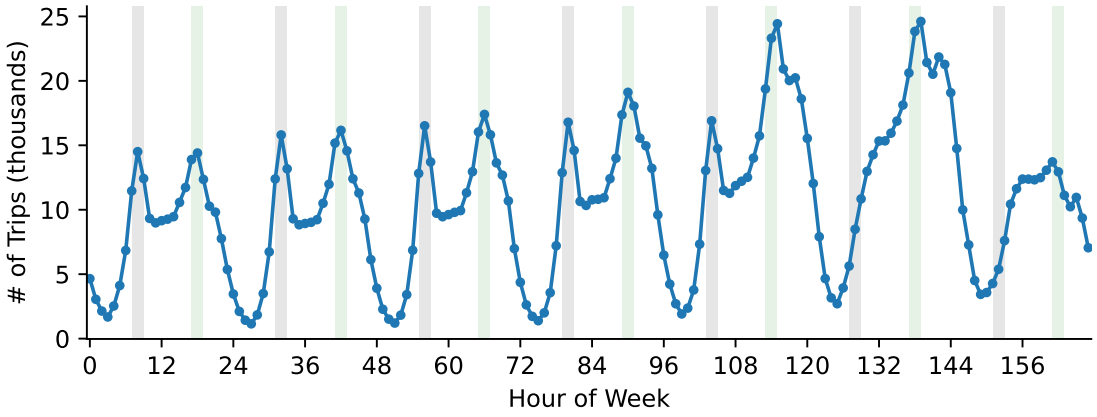


Figure 15: Average number of trips by hour-of-week.

C Simulation Details and Additional Results

In this section, we provide detailed descriptions of the simulation settings that are omitted from Section 5 of the paper (Appendix C.2). Moreover, we describe the overall market dynamics in Chicago (Appendix C.1), and include additional simulation results on:

(i) the *pure Newton* (as oppose to *damped Newton*) method that does not limit the maximum allowed changes in the origin-based multipliers (Appendix C.3),

(ii) an iterative mechanism based on gradient descent with respect to the f (Appendix C.4), and

(iii) a simple method that directly uses the multipliers as the direction for updating the OD-based adjustments (Appendix C.5).

C.1 Chicago Market Dynamics

In this section, we offer a high-level overview of the market dynamics in the City of Chicago, supporting the assumption that market conditions remain stationary during the focal time window, such as the morning rush hours.

First, Figure 15 presents the average (over the first 9 weeks of 2020) number of trips originating during each *hour-of-week*. The 0^{th} hour-of-week corresponds to midnight–1 a.m. on Mondays, and the 1^{st} hour-of-week corresponds to 1–2 a.m. on Mondays, and so on. The gray stripes indicate the morning rush hours (7–10 a.m.) and the green stripes indicate the evening rush hours (5–8 p.m.). We can observe clear weekly patterns: during weekdays, trip volume peaks during morning and evening rush hours; the market is the most busy during Friday and Saturday evenings (when people take rides to commute from work as well as to and from restaurants and bars).

In Figure 16, we illustrate the average flow imbalance of the two locations discussed in Section 1, the residential area (Lake View) and the downtown area (The Loop). The residential area sees more trips originating than ending in the area during the weekday morning rush hours, and the opposite during the evening rush hours. Downtown, on the other hand, sees more trips ending in the area in the morning, and leaving the area in the evening. Importantly, observe that the trip flow imbalance of both locations remain relatively constant during the weekday morning rush hours (7–10 a.m.) and the evening rush hours (5–8 p.m.).

We present the distribution of trip duration (in minutes) in Figure 17, for all trips during the first 10

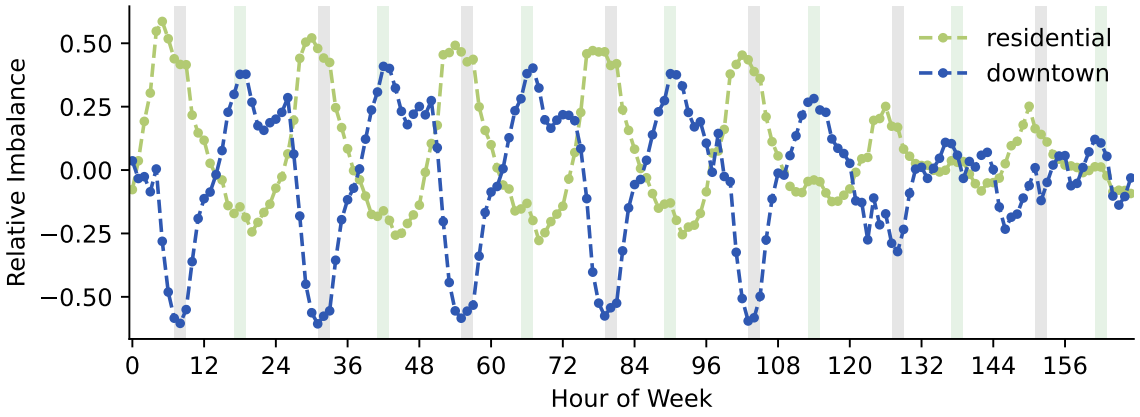
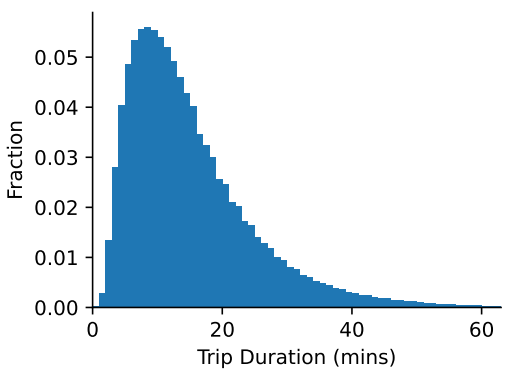
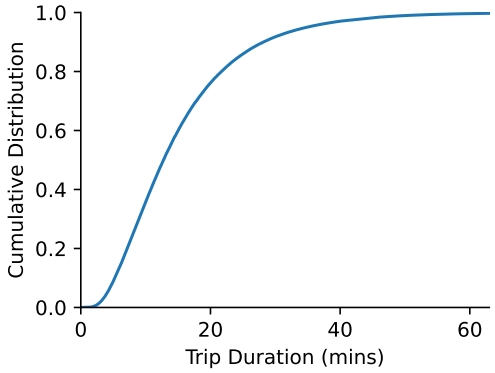


Figure 16: Average trip flow imbalance of the residential area and downtown, by hour-of-week.



(a) Histogram of trip duration.



(b) Cumulative distribution of trip duration.

Figure 17: Distribution of trip duration (in minutes).

weeks of 2020 (the distribution is not qualitatively different if we focus only on trips that take place during the morning rush hours). We can see that the vast majority of trips take fewer than 60 minutes, and roughly 80% of all trips are shorter than 20 minutes. Given that the flow imbalance of the two locations remain relatively constant for three hours during the morning rush (as shown in Figure 16), the market condition should not change substantially during the time a driver completes an average trip.

C.2 Simulation Details

We conduct two types of simulations in this paper: a stationary setting where the rider demand and driver supply are stationary over time and correspond to the average market condition during the first 10 weeks of 2020, and a dynamic setting where the rider demand and driver supply change over time according to the realized trip flows from the beginning of 2019 to the end of 2020.

C.2.1 Stationary Setting

This section describes the way we construct the economy for the stationary setting, using data from the first 10 weeks of 2020. This economy is used in Section 5.1, Appendix C.3, Appendix C.4, and Appendix C.5.1.

Trip Duration and Cost. We consider the $n = 77$ community areas in the City of Chicago as our set of locations. Among all $n^2 = 5929$ origin-destination (OD) pairs, 5826 of them have at least one observed trip. For each of these OD pairs (i, j), we calculate the average duration of the trips and use it as $d_{i,j}$. For the remaining 16 edges for which we did not observe any trip, we impute the duration with the length of the shortest path using the edges with observed trips.[23] We assume that all trips cost \$20/hour, i.e. $c_{i,j} = 20d_{i,j}$ for all $i, j \in \mathcal{L}$.

Rider Demand. Among the approximately 18 million recorded trips, 148 thousands started during 7–8 a.m. on Wednesdays. For each OD pair $(i, j) \in \mathcal{L}^2$, we calculate the average number of observed trips per hour, and denote it as $x^{\text{obs}}_{i,j}$. Among all $n^2 = 5929$ OD pairs, 3331 are traversed by at least one trip during the time window that we focus on. For each $(i, j) \in \mathcal{L}^2$ such that we observed no trips in the data, we assume $q_{i,j}(r) = 0$ for all $r \geq 0$. A constant zero demand function does not satisfy our assumption that $q_{i,j}(\cdot)$ is strictly decreasing for all $i, j \in \mathcal{L}$. This assumption is, however, sufficient but not necessary for establishing our theoretical results, and having zero demand for these OD pairs does not lead to any issues in our simulations.

For each $(i, j) \in \mathcal{L}^2$ for which we observe at least one trip, we calculate the average price (the sum of the *Fare* and *Additional Charges* columns) of the recorded trips, and denote it as $p^{\text{obs}}_{i,j}$. To construct the rider demand functions, we assume that (i) the riders' average willingness to pay for a trip is \$1 per minute (\$60/hour), and (ii) the willingness to pay follows an exponential distribution. Technically, this is assuming that the demand function is of the form

$$q_{i,j}(r) = Q_{i,j} \exp(-r/(60d_{i,j}))$$

for some constants $(Q_{i,j})_{(i,j)\in\mathcal{L}^2}$. Intuitively, $Q_{i,j}$ can be interpreted as the total amount of riders interested in traveling from i to j. For each $(i, j) \in \mathcal{L}^2$ s.t. at least one trip is recoreded, we determine $Q_{i,j}$ by solving $x^{\text{obs}}_{i,j} = Q_{i,j} \exp(-p^{\text{obs}}_{i,j}/(60d_{i,j}))$.

Driver Supply. For the total number of drivers m, we use the minimum number that can fulfill the observed rider flow $(x^{\text{obs}}_{i,j})_{(i,j)\in\mathcal{L}^2}$ and maintain flow balance. Formally, we assume

$$\begin{aligned} m = \ \text{minimize}_{\boldsymbol{y} \in \mathbb{R}^{n^2}} \quad & \sum_{i,j\in\mathcal{L}} d_{i,j} y_{i,j} \\ \text{subject to} \quad & y_{i,j} \geq x^{\text{obs}}_{i,j}, \qquad \forall i, j \in \mathcal{L}, \\ & \sum_{j\in\mathcal{L}} y_{i,j} = \sum_{j\in\mathcal{L}} y_{j,i}, \quad \forall i \in \mathcal{L}. \end{aligned} \tag{67}$$

This gives us $m = 4475$, which is almost certainly an underestimation of the number of online drivers since in practice it's very unlikely for drivers to relocate in the most efficient manner. In comparison, the observed total supply hour that's spent "on trip" is $\sum_{i,j\in\mathcal{L}} d_{i,j} x^{\text{obs}}_{i,j} = 3368$. This implies that in practice the efficiency (i.e. fraction of drivers' time spent on trip) is most likely lower than $3368/4475 = 75.3\%$. The platforms may observe seemingly a higher level of utilization, which indicates that some drivers might be relocating offline and not being counted towards the total supply hours.

[23] There is little or no rider demand between these extremely sparse origin-destination pairs. We nevertheless estimate the duration of these trips such that we can properly account for the time it takes a driver to relocate without a rider.

Driver Relocation. To coordinate driver relocation, we use phantom demand functions given by $\tilde{q}_{i,j}(r) = 500(\max\{0, 1 - r/3\})^4$ for all $i, j \in \mathcal{L}$.[24] As a result, relocation occurs only when the price drops below \$3.

C.2.2 Non-Stationary Setting

This section describes the non-stationary setting used in Section 5.2 and Appendix C.5.2, representing Chicago's market conditions for 7–8 a.m. on Wednesdays, from the beginning of 2019 through the end of 2020. There are a total of 139 million trips during the two years. We removed a total of 43 trips with distances that are more than 30 times the median distance of trips with the same origin and destination— these are most likely due to errors in the data recording/processing proces.

Trip Duration and Cost. Among all $n^2 = 5929$ OD pairs, 5925 of them have at least one trip in 2019 or 2020. For these OD pairs, we compute the average trip duration over the two years and denote it as $d_{i,j}$ for OD pair (i, j). The remaining 4 OD pairs have no trip in 2019 and 2020, and we set their trip duration to be the shortest path distance using the duration of the observed trips. Similar to the stationary setting, we set the trip cost to be the average trip duration multiplied by the driver cost rate \$20 per hour.

Rider Demand. The rider demand functions are constructed in the same way we construct the rider demand function for the stationary setting, except that we use only recorded trips from each week to determine the demand function for each week $\boldsymbol{q}^{(t)}$. More specifically, for each OD pair $(i, j) \in \mathcal{L}^2$ and each week t, let $x^{\text{obs}}_{i,j,t}$ be the number of recorded trips from i to j that started during 7–8 a.m. on Wednesday. If at least one trip is recorded, i.e. if $x^{\text{obs}}_{i,j,t} > 0$, we use the average price of these trips as $p^{\text{obs}}_{i,j,t}$. The rider demand functions are defined as

$$q^{(t)}_{i,j}(r) = Q^{(t)}_{i,j} \exp(-r/(60 d_{i,j})),$$

where $Q^{(t)}_{i,j}$ is the solution of $x^{\text{obs}}_{i,j,t} = Q^{(t)}_{i,j} \exp(-p^{\text{obs}}_{i,j,t}/(60 d_{i,j}))$.

Driver Supply. Similar to the stationary setting, for each week t, we compute the minimum number of drivers needed to serve all observed trips while maintaining flow balance, and use it as the number of drivers $m^{(t)}$ in the simulation. Technically, we solve (67) for each week t, using $x^{\text{obs}}_{i,j,t}$ instead of $x^{\text{obs}}_{i,j}$.

Driver Relocation. To coordinate the flow of driver relocation, we use phantom demand functions $\tilde{q}_{i,j}(r) = 1000(\max\{0, 1 - r/4\})^4$ for all OD pairs (i, j). Relocation occurs only when the price is below \$4.

Events at McCormic Place. There are a total of 101 Wednesdays that are working days during 2019 and 2020. As we have briefly explained in Section 5.2, we excluded a total of 5 of them from our simulations, during which it is very likely that major events were held at McCormick Place, which is located in the South Side, the community area just south of The Loop. In particular, we count the trips from North Side and Downtown to South Side during 7–8 a.m. (see Figure 18), and excluded days for which this number greater than 300.

During a typical "non-event" day, the South Side community area has an imbalance of around 0.4, meaning that more trips originate from the area than end in the area (i.e. the area is a popular origin). During the event days, however, the imbalance can drop to as low as −0.4, when substantially more riders request trips ending in the area. This leads to a very different flow of driver supply in space, thus a ridesharing platform should

[24] This phantom demand function does not satisfy our assumption (II) that $d_{i,j}\tilde{q}_{i,j}(0) > m$. The assumption is used to prove existence of the market clearing surge multiplier $\boldsymbol{\pi}$ for any $\boldsymbol{\phi}$. In practice, we are only interested in a limited range of reasonable values of $\boldsymbol{\phi}$, so the assumption is only sufficient but not necessary.

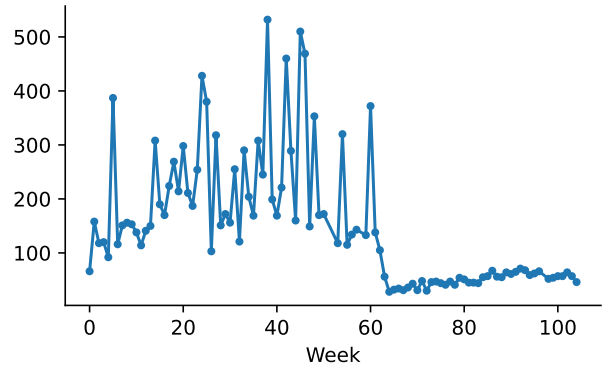
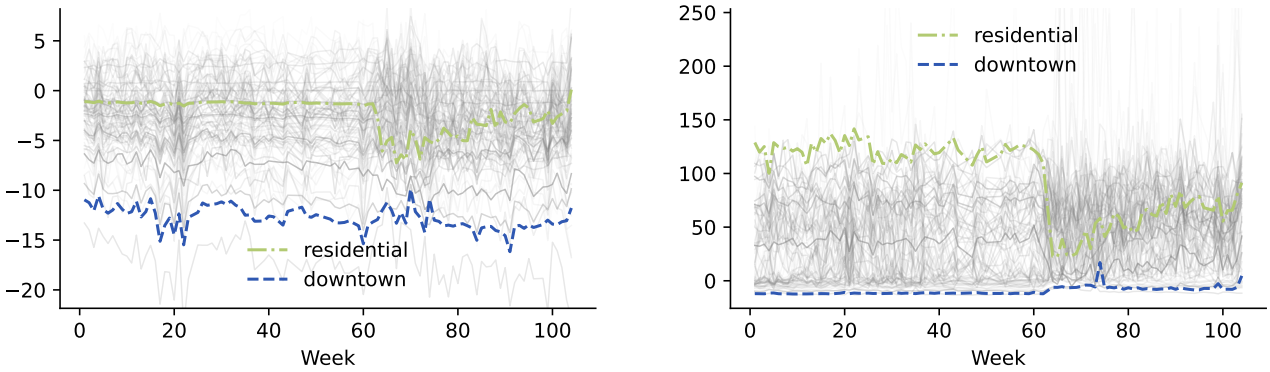


Figure 18: Total trip flow into South Side over weeks.



(a) Optimal OD-based additive adjustments $\boldsymbol{\phi}^{(t)}$.

(b) Origin-based multipliers $\boldsymbol{\pi}^{(t)}$.

Figure 19: Optimal OD-based adjustments and naïve origin-based multipliers $\boldsymbol{\phi}^{(t)}$ and $\boldsymbol{\pi}^{(t)}$ for each week.

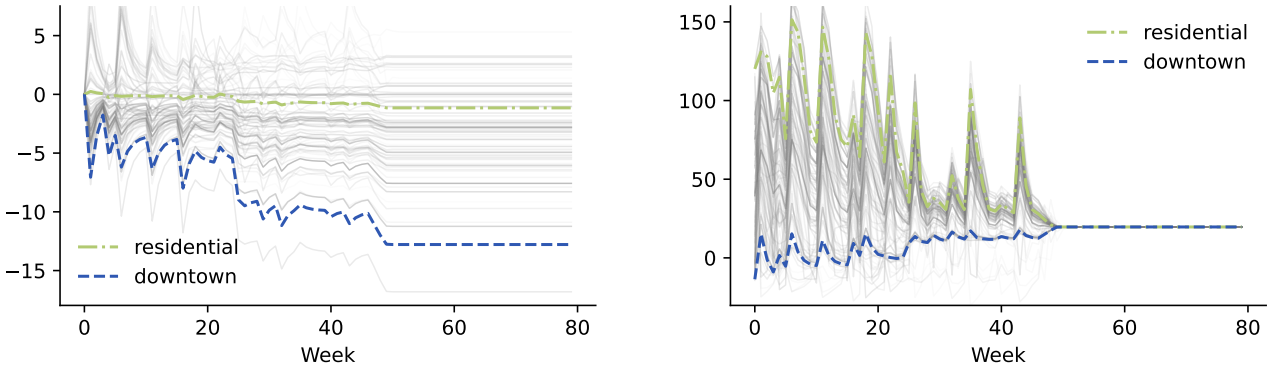
adjust the marginal values of drivers for this area as well as the neighboring areas. This is practically feasible since major events are typically planned months ahead of time.

Optimal OD-Based Adjustments and the Naïve Origin-Based Multipliers. Figure 19 shows the optimal OD-based adjustments $\boldsymbol{\phi}^{(t)}$ and the naïve origin-based multipliers $\boldsymbol{\pi}^{(t)}$ for each week t. In Figure 19a, the optimal adjustment of the residential area stays relatively constant over time, until the pandemic hits in March 2020, after which the adjustment drops substantially. Overall, both the optimal adjustments and the naïve origin-based multipliers become more volatile after the pandemic hits.

C.3 The Pure Newton Method

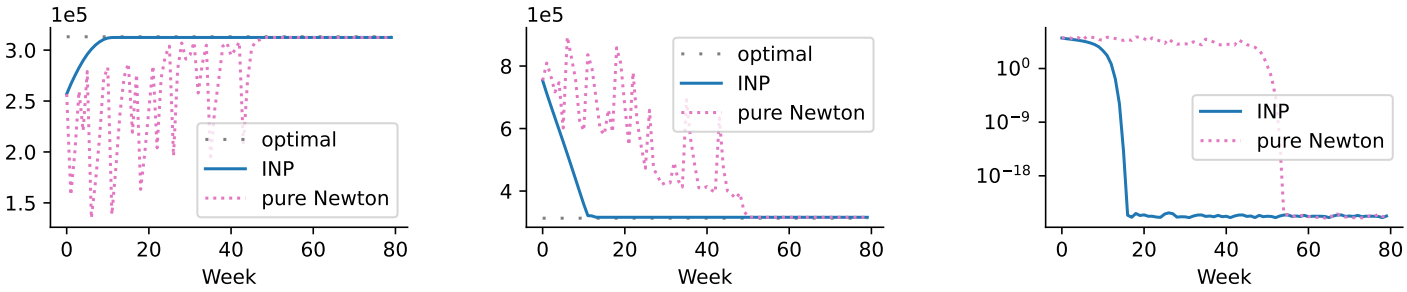
Under the INP mechanism defined in Section 4.1, the step size (22) is determined such that the (linear approximation of the) origin-based multipliers will not change week-over-week by more than some constant $\tau > 0$ for any location. Since the update direction (21) corresponds to the one under the Newton method for solving systems of equations, the INP mechanism can be considered as a "damped Newton" algorithm. If we remove the stepsize limit (i.e. setting $\alpha^{(t)} = 1$ for all t) and run the *pure Newton* method, the results for the stationary setting (same as the setting studied in Section 5.1) are shown in Figure 20 and Figure 21.

We can see that despite the fact that the pure Newton algorithm is still able to converge, the very large updates and frequent backtracking leads to very volatile market-clearing multipliers and social welfare (which



(a) The price adjustments $\boldsymbol{\phi}^{(t)}$. (b) The multipliers $\boldsymbol{\pi}^{(t)}$.

Figure 20: $\boldsymbol{\phi}^{(t)}$ and $\boldsymbol{\pi}^{(t)}$ over iterations t for the pure Newton method in the stationary setting.



(a) Social welfare ($/hour). (b) Dual objective. (c) Function f.

Figure 21: Social welfare, dual objective, and the Lyanupov function f over iterations t for the pure Newton method in the stationary setting.

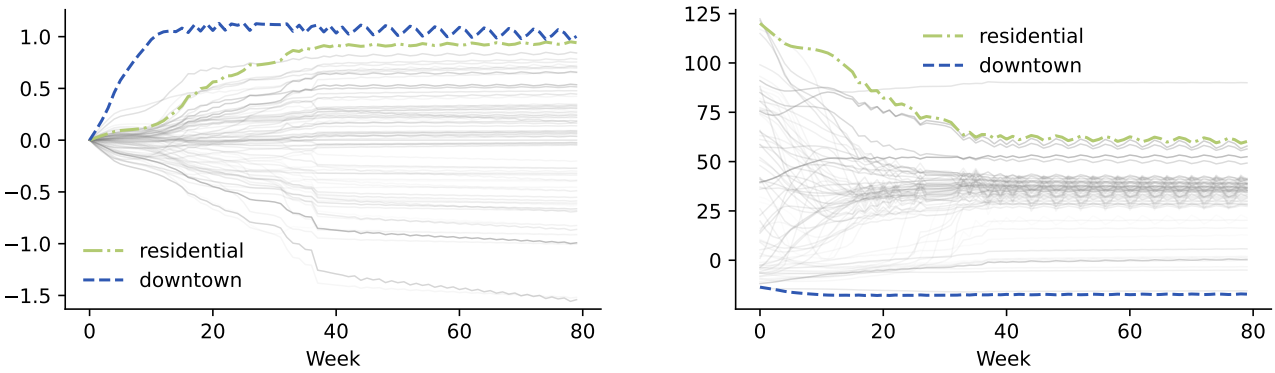
are highly undesirable in practice). The convergence also takes more iterations in comparison to the results presented in Section 5.1, where the properly "paced" INP mechanism never has to backtrack.

C.4 Gradient Descent on the Lyapunov Function

Given that the gradient of the Lyanupov function f at a market clearing outcome can be computed using information available in the market, a natural alternative (to INP) to consider is the gradient descent algorithm with respect to f. More specifically, instead of using the Newton direction (21), we can use $-\nabla f$ at the previous market clearing outcome as the update direction. In conjunction with backtracking line search (recall that f as a function of $\boldsymbol{\phi}$ is not necessarily convex), the algorithm is guaranteed to converge to the unique global minimum of f where all origin-based multipliers are equal.

In Figure 22 and Figure 23, we present the outcome under this algorithm (labeled GD) for the same stationary setting as in Section 5.1. Similar to the INP mechanism, we limit the stepsize such that the expected week-over-week change in the origin-based multipliers is at most $\tau = 10$, and use the same parameters for backtracking line search. We can see that despite the theoretical convergence guarantees, the convergence is very slow— after 80 iterations, the outcome is still far from optimal (and not much progress was made even after we run the algorithm for a total of 500 iterations).

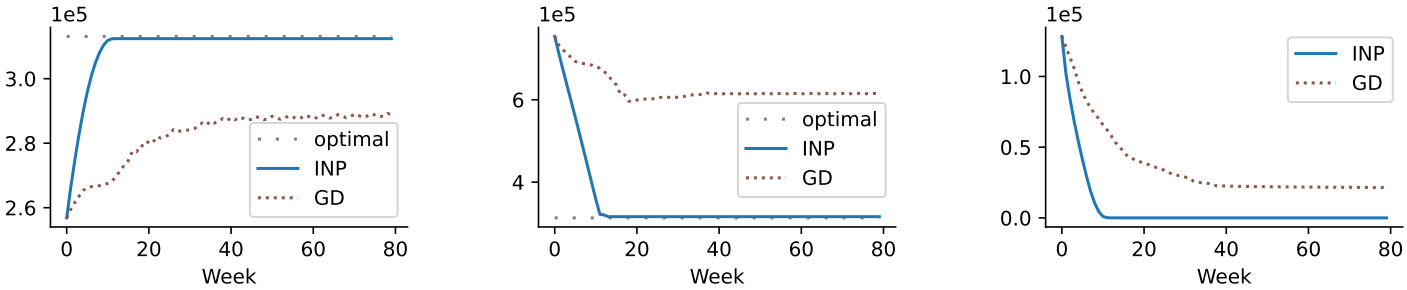
The cyclic oscillations in $\boldsymbol{\phi}^{(t)}$ and $\boldsymbol{\pi}^{(t)}$ starting around the 20th iteration is not a result of backtracking line search, and a smaller stepsize limit τ or different backtracking parameters does not help. This suggests



(a) The price adjustments $\boldsymbol{\phi}^{(t)}$.

(b) The multipliers $\boldsymbol{\pi}^{(t)}$.

Figure 22: $\boldsymbol{\phi}^{(t)}$ and $\boldsymbol{\pi}^{(t)}$ over iterations t for the gradient descent method in the static setting.



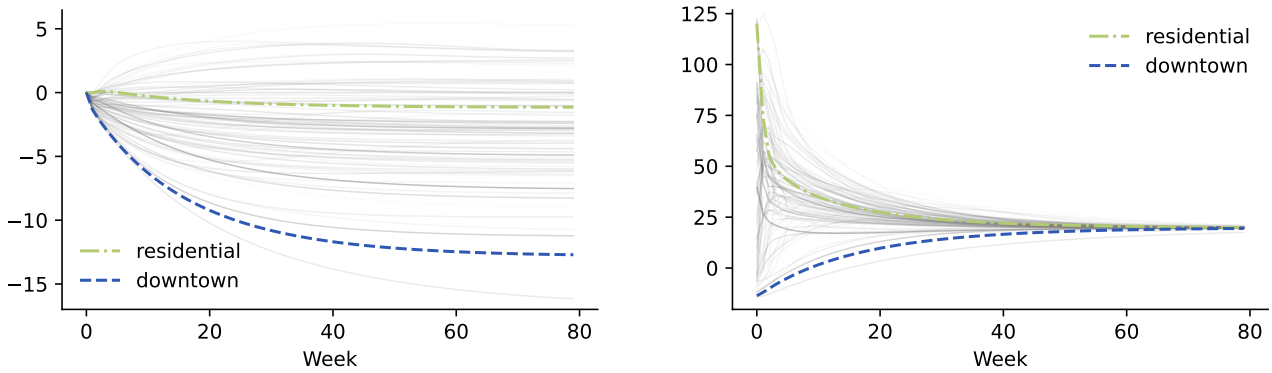
(a) Social welfare ($/hour).

(b) Dual objective.

(c) Function f in linear scale.

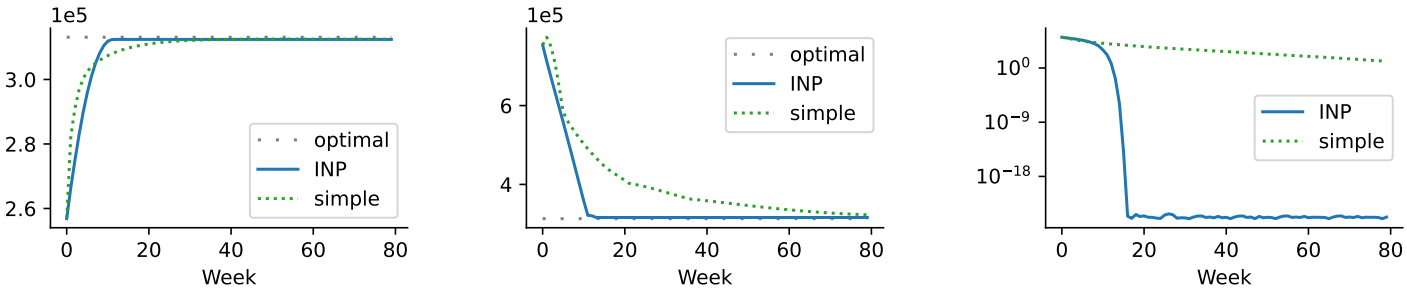
Figure 23: Social welfare, dual objective, and f over iterations t for gradient descent in the static setting.

that the gradient of f is not an effective direction for updating the OD-based adjustments. Also note that this method does not require less information than INP, since the local price sensitivity of rider demand is also necessary for computing ∇f.



(a) The price adjustments $\boldsymbol{\phi}^{(t)}$. (b) The multipliers $\boldsymbol{\pi}^{(t)}$.

Figure 24: $\boldsymbol{\phi}^{(t)}$ and $\boldsymbol{\pi}^{(t)}$ over iterations t for the simple direction in the static setting.



(a) Social welfare ($/hour). (b) Dual objective. (c) Function f.

Figure 25: Social welfare, dual objective, and f over iterations t for the simple direction in the static setting.

C.5 A Simple Update Direction

In this section, we present simulation results for a very simple mechanism, where we update the OD-based adjustments using only the origin-based multipliers. Specifically, we adjust ϕ_i for each $i \in \mathcal{L}$ by $\pi_i^{(t-1)} - \pi_n^{(t-1)}$ (multiplied by some constant), and without using the observed trip volume or the price sensitivity.[25]

We are unable to prove any theoretical guarantee for this method so far, since the update direction may not be a descent direction of the Lyanupov function f as defined in (20). As we will see, however, the method works well and is quite robust. Intuitively, increasing the ϕ_i for a location with high multiplier π_i reduces the prices of trips ending in this location (relative to other locations). As a result, this is likely to increase the driver supply for this location, thereby reducing the multiplier for this location in the subsequent iterations.

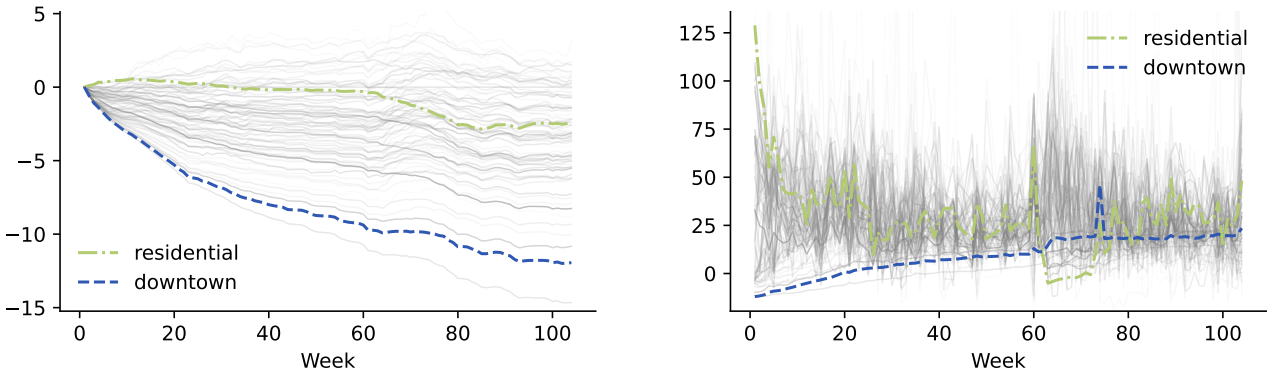
C.5.1 Stationary Market Condition

We first consider the simple direction in the static setting studied in Section 5.1, with

$$\boldsymbol{\phi}^{(t)} = \boldsymbol{\phi}^{(t-1)} + 0.01\boldsymbol{\pi}^{(t-1)}.$$

Figure 24 plots the OD-based additive adjustments $\boldsymbol{\phi}^{(t)}$ and the origin-based multipliers $\boldsymbol{\pi}^{(t)}$ over iterations t, and Figure 25 compares the social welfare, the dual objective, and the function f over iterations t with those

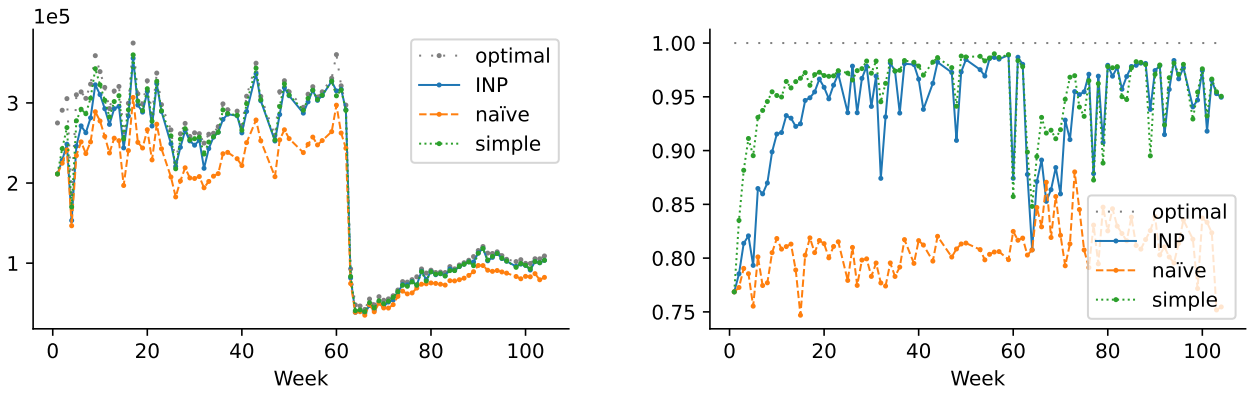
[25] Whether to subtract $\pi_n^{(t-1)}$ has no impact since the trip prices depend only on the differences in the OD-based adjustments. The current forms, however, keeps $\phi_n^{(t)}$ at zero for all t, making it easier to observe the changes of $\boldsymbol{\phi}^{(t)}$ over time.



(a) The OD-based adjustments $\boldsymbol{\phi}^{(t)}$.

(b) The origin-based multipliers $\boldsymbol{\pi}^{(t)}$.

Figure 26: $\boldsymbol{\phi}^{(t)}$ and $\boldsymbol{\pi}^{(t)}$ over weeks t, for the simple update direction in the dynamic setting.



(a) Social welfare.

(b) Welfare ratio to the optimum.

Figure 27: Social welfare and welfare ratio over iterations t for the simple direction in the dynamic setting.

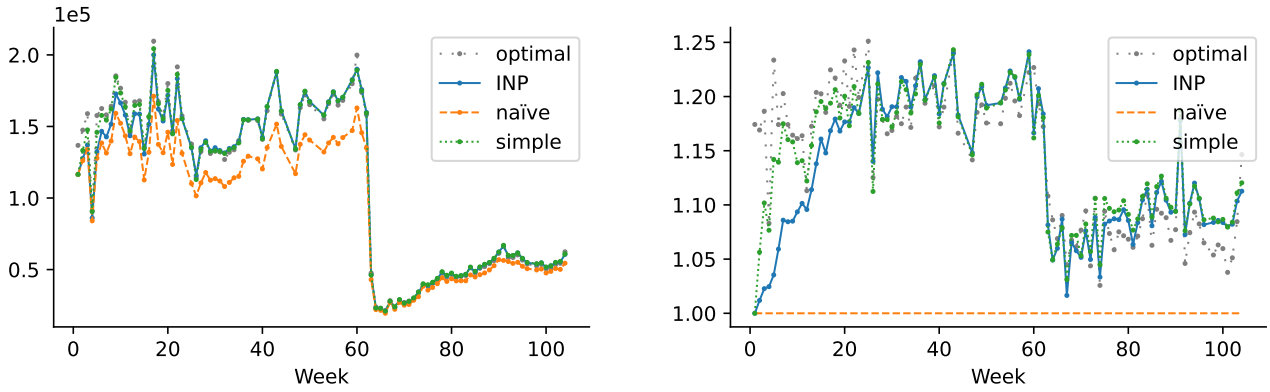
under the INP mechanism. The multipliers $\boldsymbol{\pi}^{(t)}$ at different locations eventually converge to the same value, and the social welfare converges to 99.8% of the optimal social welfare. The local convergence rate appears to be linear (in contrast to the superlinear local convergence achieved by INP), as shown in Figure 25c.

C.5.2 Non-Stationary Market Conditions

We also test the simple update direction in the non-stationary setting studied in Section 5.2, setting

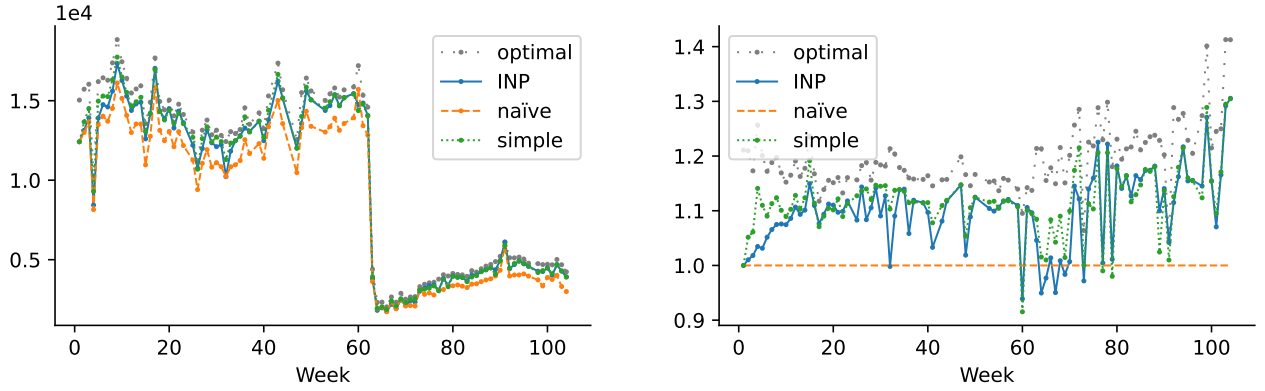
$$\boldsymbol{\phi}^{(t)} = \boldsymbol{\phi}^{(t-1)} + 0.005\boldsymbol{\pi}^{(t-1)}.$$

The results are presented in Figures 26 through 29. Despite having a worse empirical local convergence rate compared to the INP mechanism (linear vs. superlinear, as suggested by Figure 25c), the simple direction performs quite well in the dynamic setting. In particular, using the origin-based multipliers as the update direction appears to be more robust to the non-stationarity of the demand than the INP mechanism, resulting in smaller week-over-week fluctuations in welfare as shown in Figure 27b. Also note that the smaller fluctuations did not come at a cost of its ability to "track" changes in the market conditions— in comparison to the INP mechanism, social welfare in fact bounces back faster after COVID under this simple policy.



(a) Gross booking. (b) Gross booking ratio.

Figure 28: Gross booking (GB) and GB ratio over iterations t for the simple direction in the dynamic setting.



(a) Trip throughput. (b) Throughput ratio.

Figure 29: Trip throughput and throughput ratio over iterations t for the simple direction in dynamic setting.