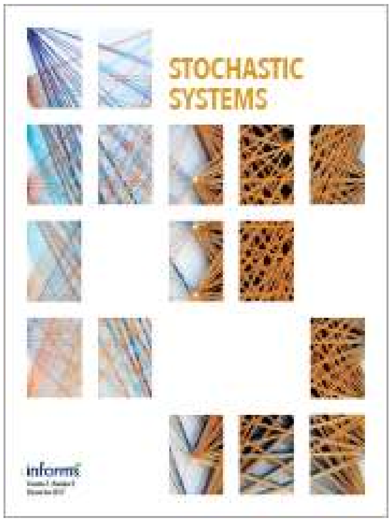


Stochastic Systems

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Dynamic Matching for Real-Time Ride Sharing

Erhun Özkan, Amy R. Ward

To cite this article:
Erhun Özkan, Amy R. Ward (2020) Dynamic Matching for Real-Time Ride Sharing. Stochastic Systems 10(1):29-70. https://doi.org/10.1287/stsy.2019.0037

http://pubsonline.informs.org/journal/stsy

STOCHASTIC SYSTEMS
Vol. 10, No. 1, March 2020, pp. 29–70
ISSN 1946-5238 (online)

Dynamic Matching for Real-Time Ride Sharing

Erhun Özkan,[a] Amy R. Ward[b]

[a] College of Administrative Sciences and Economics, Koç University, 34450 Istanbul, Turkey; [b] Booth School of Business, University of Chicago, Chicago, Illinois 60637

Contact: erhozkan@koc.edu.tr, http://orcid.org/0000-0001-6870-9495 (EÖ); amy.ward@chicagobooth.edu, http://orcid.org/0000-0003-2744-2960 (ARW)

Received: July 26, 2018
Revised: February 2, 2019
Accepted: May 10, 2019
Published Online in Articles in Advance: February 25, 2020

MSC2010 Subject Classification: 90B22, 60F99, 90C99

https://doi.org/10.1287/stsy.2019.0037

Abstract. In a ride-sharing system, arriving customers must be matched with available drivers. These decisions affect the overall number of customers matched, because they impact whether future available drivers will be close to the locations of arriving customers. A common policy used in practice is the closest driver policy, which offers an arriving customer the closest driver. This is an attractive policy because it is simple and easy to implement. However, we expect that parameter-based policies can achieve better performance. We propose matching policies based on a continuous linear program (CLP) that accounts for (i) the differing arrival rates of customers and drivers in different areas of the city, (ii) how long customers are willing to wait for driver pickup, (iii) how long drivers are willing to wait for a customer, and (iv) the time-varying nature of all the aforementioned parameters. We prove asymptotic optimality of a forward-looking CLP-based policy in a large market regime and of a myopic linear program–based matching policy when drivers are fully utilized. When pricing affects customer and driver arrival rates and parameters are time homogeneous, we show that asymptotically optimal joint pricing and matching decisions lead to fully utilized drivers under mild conditions.

History: Former designation of this paper was SSY-2018-022.

Keywords: ride sharing • dynamic matching • fluid model • asymptotic optimality

1. Introduction

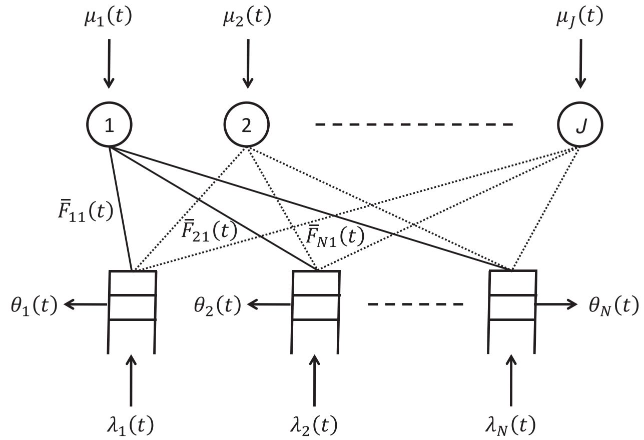
We consider the control of a two-sided matching system with multiple item types on both sides arriving randomly to the system with potentially time-varying rates; see Figure 1. The items arriving to the circles must be matched at the time of their arrival, and there is a time-varying probability that the match is acceptable. The items arriving to the queues will wait to be matched, but are impatient, and may leave their queue if left waiting for too long. The goal of the system controller is to maximize the cumulative number of (weighted) matchings in a finite time horizon.

Our main motivation for studying the system in Figure 1 is the customer–driver matching in ride-sharing platforms. The N queue locations in Figure 1 represent different areas in the city; the items arriving to the circles are customers, and those arriving to the queues are drivers. Customer types are categorized by factors such as their origin and destination areas and their priority status. Driver types are categorized by their current areas. Customers arrive in the system to request a ride. If the system controller, who is the corresponding ride-sharing company, offers a driver to a customer, then the customer accepts to be matched with the driver with respect to a specific probability that depends on the pickup time of the driver. If a customer must wait too long for pickup, then she may refuse the ride and use another transportation option. If a driver is not matched with a customer for a long time, he can either leave the system or travel to another region (i.e., another queue) to look for customers.

Because customers do not want to wait for driver pickup for a long time, the ride-sharing company would like to ensure there is always a nearby driver to offer to an arriving customer. However, maintaining adequate driver supply is difficult. Not only do customers choose when to request a ride, but also drivers choose when to begin work, how long to work, and where to go to search for customers. The result can be dramatic changes in customer demand and driver supply across different locations and over the course of a day, which sometimes results in significant driver shortages (see Hall et al. 2016, figures 1–3).

One common operational strategy is to match customers with the closest driver (CD) and to use pricing to incentivize drivers to move to undersupplied locations. However, surges in price can lead to negative

Figure 1. A Two-Sided Matching System with Multiple Item Types on Both Sides



publicity (see *The Economist* 2016, Michallon 2016, White 2016). This leads to the question of whether better matching decisions could reduce the need for price surges. An ideal is to solve a joint pricing and matching problem that accounts for the impact of differing customer and driver locations. We do this in a static environment, but that joint problem is very difficult in a time-varying environment. As a first step to making progress in a time-varying environment, we assume customer and driver arrival rates are given and focus on optimizing the matching decisions. The aforementioned time-varying arrival rates could be thought of as the result of a pricing policy as well as other underlying incentives that may be given to drivers or customers, such as demand information sharing or discount coupons, but we do not explicitly model that.

The CD policy is a greedy policy, and, therefore, a forward-looking policy is likely to perform better than the CD policy does. The following example illustrates this point. Figure 2 represents a region of a city partitioned into 10 disjoint hexagonal areas (for motivation that ride-sharing companies find such a representation useful see Chen and Sheldon 2015, figure 3). Suppose a customer arrives at area 1 and requests a ride. There are three drivers idle in area 4, and there is a single idle driver in area 3. Moreover, a concert recently ended in area 8, which implies a high potential customer arrival rate in that area. If the destination of the current area 1 customer is far away, the driver assigned to that customer will not return to area 1 for a long time. Then, the system controller has two options: he can offer either one of the drivers in area 4 or the driver in area 3 to the customer in area 1. Under the CD policy, the system controller offers an area 4 driver. However, offering the driver in area 3 saves all drivers in area 4 to match with the potential customers departing the concert in area 8. This prevents the potential future need to offer an area 8 customer a faraway area 3 driver, whom the customer will likely refuse, ending in no match being made. We conclude there is a nontrivial trade-off between offering the closest driver in accordance with the CD policy and offering a farther driver in order to maximize the future number of customers matched.

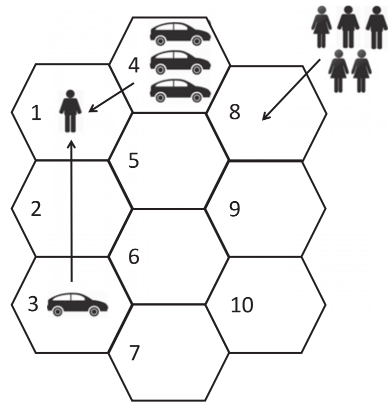
1.1. Contributions of this Paper

We analyze the two-sided matching system depicted in Figure 1 that is motivated by ride-sharing systems. The objective is to maximize the cumulative number of matchings in a finite time horizon. Our main contributions are as follows.

Proposing Asymptotically Optimal Matching Policies Based on a Continuous Linear Program and a Linear Program. We consider a large matching market, in which the arrival rates of the customers and drivers grow without bound, so as to approximate the case of a large city. For any matching policy that does not know the future with certainty, we establish that the solution to a continuous linear program (CLP) is an asymptotic upper bound on the cumulative number of matchings done in a finite time horizon under fluid scaling (see Theorem 1). That upper bound is very strong in the sense that it is valid on almost every sample path. Then, we propose a matching policy based on an optimal solution of the CLP, which is asymptotically optimal (see Corollary 1). The CLP leads to a linear program (LP) when drivers are fully utilized or the CLP parameters are time homogeneous, which motivates an asymptotically optimal LP-based matching policy in each case (see Theorems 3 and 4, respectively). When pricing affects customer and driver arrival rates and parameters are time homogeneous, we provide an asymptotically optimal pricing and matching policy (see Corollary 3) and show drivers are fully utilized under that policy under very mild conditions (see Example 1).

Published in Stochastic Systems on February 25, 2020 as DOI: 10.1287/stsy.2019.0037.
This article has not been copyedited or formatted. The final version may differ from this version.

Figure 2. An Intuitive Explanation of Why the CD Policy May Not Assign the Right Driver to a Customer



Providing Simulation Experiments Illustrating the Superior Performance of the Proposed Policies Against the CD Policy. Our simulation experiments in Section 5 show that both the CLP-based and the LP-based proposed matching policies can significantly outperform the CD policy. Consistent with intuition, we see that demand spikes coupled with low nearby driver availability, such as in Figure 2, lead to the poor performance of the CD policy. This is exactly when we recommend spending the extra effort of estimating parameters, such as customer and driver arrival rates, in order to be able to implement a CLP- or an LP-based matching policy. (In comparison, the CD policy requires no network information to implement.)

The remainder of this paper is organized as follows. We conclude this section with a literature review (see Section 1.2) and a summary of our mathematical notation (see Section 1.3). Section 2 presents our model. We formalize our large matching market regime, an asymptotic upper bound on the cumulative number of matchings, and prove that a CLP-based matching policy achieves that upper bound in Section 3. We provide conditions under which an LP-based matching policy achieves the asymptotic upper bound and consider a joint pricing and matching problem in Section 4. Section 5 presents some simulation experiments. Finally, we make concluding remarks in Section 6. The proofs of all results can be found in the appendices.

1.2. Literature Review

Dynamic matching control has been studied in the literature in the context of kidney exchanges (Ünver 2010), housing markets (Leshno 2016), online matching platforms such as Upwork or Airbnb (Arnosti, Johari and Kanoria 2016), assemble-to-order manufacturing systems (Plambeck and Ward 2006, Reiman and Wang 2015), and more abstract queueing models (Gurvich and Ward 2014). However, the ride-sharing model is different enough that it is not clear whether any of the results of these studies carry over.

The spirit of our methodology is drawn from the queueing control literature, where (i) an asymptotic regime is defined, (ii) a control policy is derived from the solution of an optimization problem, and (iii) asymptotic optimality of the control policy is proven (see Harrison 2000). We consider a large market asymptotic regime in which number of drivers and customers grow without a bound. A similar large market regime is considered by Plambeck and Ward (2006), Gurvich and Ward (2014), Arnosti, Johari and Kanoria (2016), and Leshno (2016).

In recent years, academic research related to ride-sharing platforms has grown rapidly, alongside the use of these platforms. An overview of research problems on ride-sharing platforms can be seen in the work of Azevedo and Weyl (2016). Although the effects of pricing have been well studied—see Chen et al. (2015), Chen and Sheldon (2015), Riquelme et al. (2015), Bimpikis et al. (2019), Banerjee et al. (2016), Castillo et al. (2016), Hall et al. (2016), Cachon et al. (2017), Guda and Subramanian (2019), and Besbes et al. (2018)—none of those papers optimize the matching decisions.

Özkan (2018) studies joint optimization of the pricing and the matching decisions in a setting with time-homogeneous parameters. He shows that optimizing the pricing decisions while fixing the matching decisions to same area matchings (i.e., customers can be matched only with the drivers from the same area) can be suboptimal. Therefore, matching decisions have first-order importance for the ride-sharing firms. Özkan (2018) considers a steady-state fluid model, whereas we rigorously prove the convergence of a prelimit model to a fluid model in a large market regime.

There are only a few papers that study the effect of matching decisions to the ride-sharing firms (see Hu and Zhou 2015, Banerjee et al. 2018). Hu and Zhou (2015) consider dynamic matching control of a two-sided,

Stochastic Systems, 2020, vol. 10, no. 1, pp. 29–70, © 2020 The Author(s)

discrete-time matching system where the objective is to maximize the expected total discounted profit. They derive conditions under which the CD policy is optimal. Banerjee et al. (2018) consider dynamic matching control of a two-sided, continuous-time matching system where the objective is to maximize the number of matchings. They propose a state-dependent matching policy that achieves the asymptotically optimal system performance with the fastest possible rate as the market size increases. A main modeling difference is that in the papers by both Hu and Zhou (2015) and Banerjee et al. (2018), matchings between different types occur with either probability zero or one, whereas we assume probabilistic matching. Another difference is that Banerjee et al. (2018) consider a closed network in the sense that number of drivers in the network is constant at all times, whereas Hu and Zhou (2015) and our study consider an open network such that drivers can enter and leave the network.

There are also ride-sharing studies focusing on fairness of the carpooling decisions to the customers (see Gopalakrishnan et al. 2016), competition between ride-sharing firms (see Nikzad 2017), optimal driver decisions (see Chaudhari et al. 2018), and empty car routing (see Braverman et al. 2016). Braverman et al. (2016) consider a closed network where the drivers are centrally controlled (e.g., driverless vehicles) and propose asymptotically optimal routing policies for empty vehicles in a large market regime. However, customers can be matched only with the vehicles from the same area in their setting; that is, there is no matching optimization.

There are many ride-sharing papers that consider a model with time-homogeneous parameters in steady state (see Riquelme et al. 2015, Banerjee et al. 2016, Braverman et al. 2016, Özkan 2018, Bimpikis et al. 2019). However, customer demand and driver supply can change dramatically within minutes in practice (see Hall et al. 2016, figures 1–3), which implies that the system may never reach steady state. Consequently, the time-varying nature of the problem is important. Among all of the aforementioned ride-sharing studies, the only ones that allow time-varying parameters are that by Hu and Zhou (2015) and our study.

1.3. Notation and Terminology

The set of nonnegative integers and strictly positive integers are denoted by $\mathbb{N}$ and $\mathbb{N}_+$, respectively. The k dimensional ($k \in \mathbb{N}_+$) Euclidean space is denoted by $\mathbb{R}^k$, and $\mathbb{R}_+$ denotes $[0, +\infty)$. For $x, y \in \mathbb{R}$, $x \vee y := \max\{x, y\}$, $x \wedge y := \min\{x, y\}$, and $(x)^+ := x \vee 0$. We let $\mathcal{B}(\mathbb{R}^k)$ denote the Borel σ-algebra on $\mathbb{R}^k$ for all $k \in \mathbb{N}_+$, and let $\mathcal{L}(\mathbb{R})$ denote the collection of Lebesgue-measurable subsets of $\mathbb{R}$, which is a σ-algebra on $\mathbb{R}$. For all $T \in \mathbb{R}_+$, $\mathcal{B}([0,T])$ and $\mathcal{L}([0,T])$ denote the Borel and Lebesgue σ-algebras on the interval $[0,T]$, respectively. A function $f : X \to \mathbb{R}$ defined in measure space $(X, \mathcal{X})$ is called $\mathcal{X}$-measurable (denoted by $f \in \mathcal{X}$) if it is $(\mathcal{X}, \mathcal{B}(\mathbb{R}))$-measurable. If the measure space $(X, \mathcal{X})$ is $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ $((\mathbb{R}, \mathcal{L}(\mathbb{R})))$, we say that f is Borel (Lebesgue) measurable.

For each $k \in \mathbb{N}_+$, $\mathbb{D}^k$ denotes the space of all $\omega : \mathbb{R}_+ \to \mathbb{R}^k$ that are right continuous with left limits. Let $\mathbf{0}, e \in \mathbb{D}$ be such that $\mathbf{0}(t) = 0$ and $e(t) = t$ for all $t \in \mathbb{R}_+$. We abbreviate the phrase "independent and identically distributed" as "i.i.d.," "almost surely" as "a.s.," and "uniformly on compact intervals" as "u.o.c." The notation $\xrightarrow{a.s.}$ denotes almost sure convergence. For $f \in \mathbb{D}$ and $t \in \mathbb{R}_+$, we let $\|f\|_t := \sup_{0 \le s \le t} |f(s)|$. Let $\{X_n, n \in \mathbb{N}\}$ be a sequence in $\mathbb{D}$ and $X \in \mathbb{D}$. Then $X_n \to X$ u.o.c. as $n \to \infty$, if $\|X_n - X\|_t \to 0$ as $n \to \infty$ for all $t \in \mathbb{R}_+$. We assume that all of the random variables and stochastic processes are defined in the same complete probability space $(\Omega, \mathcal{F}, \mathbf{P})$, $\mathbf{E}$ denotes the expectation under $\mathbf{P}$, and $\mathbf{P}(A, B) := \mathbf{P}(A \cap B)$. We let $\sigma\{X\}$ denote the σ-field generated by the random variable X, let $\mathbb{I}$ denote the indicator function, and let $\perp$ denote independence.

2. The Ride-Sharing Model

We partition the considered region of the city into $N \in \mathbb{N}_+$ disjoint areas, as illustrated in Figure 2, and assume that in the aggregate the individual driver decisions (regarding when to begin and end working, where to relocate if left waiting too long to be matched with a customer, and where to go after dropping off a customer) result in Poisson process arrivals to each area, independent of the matching decisions made by the system controller (see Remark 2). Specifically, type i drivers arrive at area $i \in \mathcal{N} := \{1, 2, \ldots, N\}$ according to a nonhomogeneous Poisson process having instantaneous rate $\lambda_i(t)$ at time $t \in \mathbb{R}_+$, and cumulative rate function $\Lambda_i(t) := \int_0^t \lambda_i(s)\mathrm{d}s$. We further assume that the amount of time a type i driver will wait in an area to be matched with an arriving customer is exponentially distributed with time dependent rate $\theta_i(t)$ for all $i \in \mathcal{N}$ and $t \in \mathbb{R}_+$.

Customers arrive in the system and request to be matched. The $J \in \mathbb{N}_+$ different customer types are categorized by factors such as their origin and destination areas and their priority status. Type $j \in \mathcal{J} := \{1, 2, \ldots, J\}$ customers arrive in accordance with a nonhomogeneous Poisson process having instantaneous rate $\mu_j(t)$ at time $t \in \mathbb{R}_+$, and cumulative rate function $\Gamma_j(t) := \int_0^t \mu_j(s)\mathrm{d}s$. A matching policy $\pi = (\pi_1, \ldots, \pi_J)$ determines which driver type to offer an arriving customer. Each component π_j tracks the sequence of driver types offered to type j customers; that is, when $\pi_j(k) = i$, then the system controller attempts to match the kth arriving type j

customer with a type i driver, for $j \in \mathcal{J}, k \in \mathbb{N}_+, i \in \mathcal{N} \cup \{0\}$. The notation $\pi_j(k) = 0$ implies no driver is offered to the customer, in which case the customer is lost. The customer accepts the offered driver if the waiting time required for pickup is not too large (in a sense specified precisely below), and otherwise the customer is lost and the offered driver stays in his current area. The implication is that customers are classified as matched or unmatched (i.e., lost) at the time of their arrival, even though a matched customer must still wait to be picked up by a driver. The process $D^\pi_{ij}(t)$ tracks the cumulative number of type i drivers matched to type j customers under policy π in $[0,t]$.

The number of drivers in area i at time $t \in \mathbb{R}_+$ depends on the matching policy π. Then, for A_i and R_i unit rate Poisson processes, the number of unmatched type i drivers at time t is

$$Q^\pi_i(t) = Q_i(0) + A_i(\Lambda_i(t)) - R_i\left(\int_0^t \theta_i(s)Q^\pi_i(s)\mathrm{d}s\right) - \sum_{j\in\mathcal{J}} D^\pi_{ij}(t) \geq 0, \tag{1}$$

where $\{Q_i(0), i \in \mathcal{N}\}$ are random variables independent of all other stochastic primitives. The second term on the right-hand side of (1) represents the cumulative number of driver arrivals to area i in $[0,t]$, whereas the third and fourth terms represent the cumulative number of driver departures from area i in $[0,t]$ by unmatched and matched drivers, respectively.

We would like to match as many customers with drivers as possible. This is straightforward if there are many matching policies under which $Q^\pi_i(t) > 0$ for all $i \in \mathcal{N}$ and $t \in \mathbb{R}_+$, because then there is always a driver near to an arriving customer. The difficulty is that in general not every area will have an available driver—and which areas have available drivers depends on earlier matching decisions. In this case, the arriving customer is matched only if the time required for the driver to pick up the customer is less than the amount of time that customer is willing to wait for a driver.

The time a customer is willing to wait for a driver can depend on the time of day. For example, during working hours, a customer may be more time-constrained than during nonworking hours. We represent this using a step (piecewise constant) function that depends on the customer arrival time. First, we partition the time horizon into countably many disjoint intervals by defining the deterministic sequence $\{\tau_m, m \in \mathbb{N}\}$ such that $\tau_m \in \mathbb{R}_+$ and $\tau_m < \tau_{m+1}$ for all $m \in \mathbb{N}$, and $\tau_m \to \infty$ as $m \to \infty$. Second, to allow for potentially changing traffic conditions, we define the pickup time of a type i driver for a type j customer who arrived in the system at time $t \in [\tau_m, \tau_{m+1})$ by $t_{ij}(m) \in \mathbb{R}_+$ for all $i \in \mathcal{N}$, $j \in \mathcal{J}$, and $m \in \mathbb{N}$. Third, we denote the time the kth type j customer arrival is willing to wait for pickup given the arrival occurred at time $t \in [\tau_m, \tau_{m+1})$ by the random variable $a^k_j(m)$ for all $k \in \mathbb{N}_+$, $j \in \mathcal{J}$, and $m \in \mathbb{N}$. The sequence $\{a^k_j(m), k \in \mathbb{N}_+\}$ is i.i.d. and independent of all other stochastic primitives for all $m \in \mathbb{N}$ and $j \in \mathcal{J}$. Then, the probability that the kth type j customer accepts a type i driver, given the arrival time is t, is

$$\bar{F}_{ij}(t) := \sum_{m=0}^{\infty} \mathbf{P}\left(a^k_j(m) \geq t_{ij}(m)\right)\mathbb{I}(t \in [\tau_m, \tau_{m+1})) \tag{2}$$

for all $i \in \mathcal{N}$, $j \in \mathcal{J}$, $k \in \mathbb{N}_+$, and $t \in \mathbb{R}_+$.

The closest driver policy, denoted by π_{CD}, offers a type j customer that arrives at time $t \in \mathbb{R}_+$ a driver type from the set

$$\underset{\{i\in\mathcal{N}:\ Q^{\pi_{CD}}_i(t-)>0\}}{\operatorname{argmin}} \sum_{m\in\mathbb{N}} t_{ij}(m)\mathbb{I}(t \in [\tau_m, \tau_{m+1})). \tag{3}$$

If the set in (3) is not a singleton, the offered driver is chosen randomly from the closest drivers. The question is, Does the CD policy match as many customers as possible?

The total cumulative number of matchings under any policy π depends on the number of customers willing to wait for their offered driver. Specifically, if E_j is a unit rate Poisson process and $\nu_j(k) := \inf\{t \in \mathbb{R}_+ : E_j(\Gamma_j(t)) = k\}$ is the arrival time of the kth type $j \in \mathcal{J}$ customer, $k \in \mathbb{N}_+$, then

$$D^\pi_{ij}(t) := \sum_{k=1}^{E_j(\Gamma_j(t))} \sum_{m\in\mathbb{N}} \mathbb{I}\left(\nu_j(k) \in [\tau_m, \tau_{m+1}),\ a^k_j(m) \geq t_{ij}(m),\ \pi_j(k) = i\right), \tag{4}$$

for all $i \in \mathcal{N}$, $j \in \mathcal{J}$, $t \in \mathbb{R}_+$, and under any policy π. Our objective is to find a policy π that maximizes the total cumulative number of matchings made in a finite time horizon over a specified class of admissible matching policies Π, that is, to solve

$$\max_{\pi\in\Pi} \sum_{i\in\mathcal{N}, j\in\mathcal{J}} D^\pi_{ij}(T,\omega), \tag{5}$$

for all $\omega \in \Omega$. In studying this problem, we obtain some results on the more general objective

$$\max_{\pi \in \Pi} \sum_{i \in \mathcal{N}, j \in \mathcal{J}} w_{ij} D_{ij}^{\pi}(T, \omega) \tag{6}$$

that also allows the system manager to provide weights $\{w_{ij}, i \in \mathcal{N}, j \in \mathcal{J}\}$ on the possible matchings. The objectives (5) and (6) are very strong objectives because solving either requires specifying a policy that maximizes the number of matchings on every sample path.

Remark 1. In our model formulation, the driver types are formed only based on the arrival locations of the drivers. However, we can extend our results to an arbitrary (but finite) number of driver types $I \in \mathbb{N}_+$ by updating $a_j^k(m)$ to $a_{ij}^k(m)$ for all $k \in \mathbb{N}_+$, $j \in \mathcal{J}$, $m \in \mathbb{N}$, and $i \in \mathcal{I} := \{1, 2, \ldots, I\}$. Then, $a_{ij}^k(m)$ denotes the patience time of the kth type j customer for a type i driver given that she arrived in the system on the time interval $[\tau_m, \tau_{m+1})$, for all $k \in \mathbb{N}_+$, $j \in \mathcal{J}$, $m \in \mathbb{N}$, and $i \in \mathcal{I}$. All of our results hold under this extension.

Remark 2. In reality, customer and driver arrival rates and driver departure rates in an area can depend on the decisions of the system controller; that is, the customer and driver behaviors are endogenous. For simplicity, we assume exogenous customer and driver behaviors. There is some support for such an assumption in the work of Zhong et al. (2019), who use data from the ride-sharing company Didi to show that a Markovian queueing model with exogenous customer and driver behaviors provides good estimates for system performance measures during rush hours in China.

2.1. Admissible Policies

Roughly speaking, an admissible matching policy cannot anticipate the future. This is formalized mathematically by defining the filtration $\mathbb{F} := \{\mathcal{F}(t), t \in \mathbb{R}_+\}$ such that

$$\mathcal{F}(t) := \sigma\Bigg\{A_i(\Lambda_i(s)), E_j\big(\Gamma_j(s)\big), R_i\bigg(\int_0^{s-} \theta_i(u) Q_i(u) \mathrm{d}u\bigg), D_{ij}(s-), Q_i(s-), a_j^k(m), \quad \forall s \in [0, t],\ i \in \mathcal{N},\ j \in \mathcal{J}, \\ \forall m \in \mathbb{N},\ k \in \big\{1, 2, \ldots, E_j\big(\Gamma_j(t-)\big)\big\}\Bigg\}. \tag{7}$$

The information in (7) includes past observations on the amount of time customers have been willing to wait for pickup [the $a_j^k(m)$'s], which is generally not available in practice. In that case, any proposed policy should not rely on those observations (and our proposed policies in the next section do not). The reason we include such information in the filtration is to make our asymptotic optimality proof stronger, because the upper bound result we prove later (in Theorem 1) is with respect to a larger policy class. However, the information in (7) cannot include how long a customer arriving in the system at time t will wait for driver pickup, because that would allow the ability to cherry pick certain customers to offer faraway drivers without consequence. The technical implication is that the filtration $\mathbb{F}$ is not right continuous.

The information available to an admissible policy at the arrival epoch of the kth type j customer is

$$\mathcal{F}_j(k) := \sigma\Bigg\{A_i\big(\Lambda_i(s \wedge \nu_j(k))\big), E_{j'}\big(\Gamma_{j'}(s \wedge \nu_j(k))\big), R_i\bigg(\int_0^{(s \wedge \nu_j(k))-} \theta_i(u) Q_i(u) \mathrm{d}u\bigg), D_{ij'}\big((s \wedge \nu_j(k))-\big), Q_i\big((s \wedge \nu_j(k))-\big), \\ \forall s \in \mathbb{R}_+,\ i \in \mathcal{N},\ j' \in \mathcal{J}, \quad a_{j'}^r(m), r \in \big\{1, \ldots, E_{j'}\big(\Gamma_{j'}(\nu_j(k)-)\big)\big\}, \forall j' \in \mathcal{J} \backslash \{j\},\ a_j^r(m), r \in \{1, \ldots, k-1\},\ \forall m \in \mathbb{N}\Bigg\}, \tag{8}$$

for all $k \in \mathbb{N}_+$ and $j \in \mathcal{J}$. Because each of the stochastic processes that generate the σ-field in (7) is either right or left continuous and $\nu_j(k)$ is a stopping time with respect to the filtration $\mathbb{F}$ for all $j \in \mathcal{J}$ and $k \in \mathbb{N}_+$, the σ-field in (8) is well defined. Because $\nu_j(k) \le \nu_j(k+1)$ for all $k \in \mathbb{N}_+$, $\mathbb{F}_j := \{\mathcal{F}_j(k), k \in \mathbb{N}_+\}$ is a filtration for all $j \in \mathcal{J}$.

Definition 1 (Admissible Policies). For all $j \in \mathcal{J}$, let Π_j denote the set of discrete-time stochastic processes with domain $\mathbb{N}_+ \times \Omega$ and range $\mathcal{N} \cup \{0\}$, such that for all $\pi_j \in \Pi_j$, π_j is $\mathbb{F}_j$-adapted (i.e., $\pi_j(k) \in \mathcal{F}_j(k)$ for all $k \in \mathbb{N}_+$), and if $Q_i(\nu_j(k)-) = 0$ for some $i \in \mathcal{N}$, then $\pi_j(k) \ne i$. Let Π be the set of J-dimensional discrete-time stochastic processes such that for all $\pi \in \Pi$, we have $\pi = (\pi_1, \pi_2, \ldots, \pi_J)$, where $\pi_j \in \Pi_j$ for all $j \in \mathcal{J}$. Then, Π is the set of admissible policies.

Lemma 1. *The CD policy is admissible; that is,* $\pi_{CD} \in \Pi$.

The proof of Lemma 1 is presented in Appendix E, Section E.1.

Remark 3. Although the system controller cannot anticipate the arrival times of the customers and drivers, how long customers will wait for driver pickup, or how long drivers will remain in their current area, he can accurately forecast the arrival rates of the customer and driver types and the driver acceptance probabilities associated with the customer types. In other words, he knows the functions Λ_i, Γ_j, and $\bar{F}_{ij}$ for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$.

Remark 4. The filtration $\mathbb{F}$ can be augmented to include additional stochastic processes $\Upsilon_m : \mathbb{R}_+ \times \Omega \to \mathbb{R}$ with right- or left-continuous sample paths for all $m \in \mathbb{N}$, provided that the sequence $\{\Upsilon_m, m \in \mathbb{N}\}$ does not contain any future information related to the processes that generate $\mathcal{F}(t)$ for all $t \in \mathbb{R}_+$. For example, if the system controller randomly chooses a driver to offer an arriving customer, then $\{\Upsilon_m, m \in \mathbb{N}\}$ includes a sequence of i.i.d. random variables that reflect the outcome of an N-sided die roll.

Remark 5. Because the arrival process of each customer type is a nonhomogeneous Poisson process, the probability that more than one customer arrives in the system simultaneously at some time epoch is zero. Therefore, the range of all $\pi_j \in \Pi_j$ is chosen as $\mathcal{N} \cup \{0\}$ instead of the m-fold Cartesian product of $\mathcal{N} \cup \{0\}$ for some $m \geq 2$ for all $j \in \mathcal{J}$.

Remark 6. Let $\mathcal{F}(\nu_j(k))$ denote the sigma algebra defined by the stopping time $\nu_j(k)$ as in definition 1.2.12 of Karatzas and Shreve (1988). Another way to define Π_j is such that for all $\pi_j \in \Pi_j$, $\pi_j(k) \in \mathcal{F}(\nu_j(k))$ for all $k \in \mathbb{N}_+$. We do not choose this option because proving $a_j^k(m) \perp \mathcal{F}(\nu_j(k))$ for all $j \in \mathcal{J}$, $m \in \mathbb{N}$, and $k \in \mathbb{N}_+$ is difficult,[1] but $a_j^k(m) \perp \mathcal{F}_j(k)$ is by construction (see (8)). This result is exactly what prevents the system controller knowing how long an arriving customer is willing to wait for driver pickup, and so is crucial to our model and analysis.

3. An Asymptotically Optimal CLP-Based Matching Policy

It is very difficult to solve the optimization problem (6) exactly. Even if we can accomplish this very challenging task, the optimal matching policy will most likely be sample path dependent and will be very complicated. Hence, we consider a large market where the arrival rates of the customers and drivers grow without bound and solve (6) under fluid scaling in that limiting regime. Section 3.1 specifies the large market limiting regime and the fluid scaling. Section 3.2 establishes that the solution to a CLP serves as an asymptotic upper bound on the objective (6) under fluid scaling. Section 3.3 provides a simple policy that can asymptotically mimic the performance of any feasible matching process for the CLP and, therefore, can be used to specify an asymptotically optimal policy when the CLP is solvable.

3.1. A Large Matching Market

We consider a sequence of matching systems indexed by n, $n \in \mathbb{N}_+$. Each matching system has the same primitives with the one introduced in Section 2 except that the arrival rates of the drivers and customers, and departure rates of unmatched drivers from their current areas depend on n. In the nth matching system, for all $i \in \mathcal{N}$, $j \in \mathcal{J}$, and $t \in \mathbb{R}_+$, let

$$\Lambda_i^n(t) := \int_0^t \lambda_i^n(s)\mathrm{d}s, \qquad \Gamma_j^n(t) := \int_0^t \mu_j^n(s)\mathrm{d}s, \tag{9}$$

where $\lambda_i^n : \mathbb{R}_+ \to \mathbb{R}_+$ and $\mu_j^n : \mathbb{R}_+ \to \mathbb{R}_+$ are Lebesgue-measurable rate functions for all $i \in \mathcal{N}$, $j \in \mathcal{J}$, and $n \in \mathbb{N}_+$. Moreover, $\theta_i^n \in \mathbb{D}$ and $\theta_i^n \geq \mathbf{0}$ for all $i \in \mathcal{N}$ and $n \in \mathbb{N}_+$. A policy $\pi = \{\pi^n, n \in \mathbb{N}_+\}$ is a sequence that specifies a policy for each n, and π is admissible if π^n is admissible for all $n \in \mathbb{N}_+$. For a policy such as CD that does not change with n, in a slight abuse of notation, we specify π (i.e., $\pi = \pi_{CD}$) and assume $\pi^n = \pi$ for all $n \in \mathbb{N}_+$. Our notational convention is to denote a process (or random variable) X in the nth system under the admissible policy π by $X^{\pi,n}$.

Increasing the arrival rates without a bound and keeping the departure rates of unmatched drivers from their areas bounded results in a crowded matching system where we can use law of large numbers type of results to obtain tractable approximations for the processes of interest.

Assumption 1 (Large Market). *We assume that* $\Lambda_i^n/n \to \Lambda_i$, $\mu_j^n/n \to \mu_j$, *and* $\theta_i^n \to \theta_i$ *u.o.c. as* $n \to \infty$ *for all* $i \in \mathcal{N}$ *and* $j \in \mathcal{J}$ *such that* $\Lambda_i = \int_0^t \lambda_i(t)\mathrm{d}t$ *for all* $t \in \mathbb{R}_+$, *the functions* $\lambda_i : \mathbb{R}_+ \to \mathbb{R}_+$ *and* $\mu_j : \mathbb{R}_+ \to \mathbb{R}_+$ *are Lebesgue measurable, and the function* θ_i *is defined such that* $\theta_i \in \mathbb{D}$ *and* $\theta_i \geq \mathbf{0}$ *for all* $i \in \mathcal{N}$ *and* $j \in \mathcal{J}$. *We also assume that* $\sup_{t\in\mathbb{R}_+} \lambda_i(t) < \infty$,

Published in Stochastic Systems on February 25, 2020 as DOI: 10.1287/stsy.2019.0037.
This article has not been copyedited or formatted. The final version may differ from this version.

$\sup_{t\in\mathbb{R}_+}\theta_i(t)<\infty$, and $\sup_{t\in\mathbb{R}_+}\mu_j(t)<\infty$ for all $i\in\mathcal{N}$ and $j\in\mathcal{J}$. The unit rate Poisson processes A_i, R_i, and E_j are mutually independent for all $i\in\mathcal{N}$ and $j\in\mathcal{J}$.

The reader will have noticed that Assumption 1 reuses the notation Λ_i, λ_i, μ_j, and θ_i for all $i\in\mathcal{N}$ and $j\in\mathcal{J}$. In Section 2, the notation refers to parameters associated with a particular prelimit matching system having index n. In this section and the next, the parameters without the superscript n refer to the limiting quantities in Assumption 1.

Assumption 1 does not necessarily imply that the arrival rates of all customer and driver types grow to infinity for all $t\in\mathbb{R}_+$ as $n\to\infty$. In particular, arrival rates can be zero in some areas during some time periods (e.g., $\lambda_i^n(t)=\mu_j^n(t)=\lambda_i(t)=\mu_j(t)=0$ for some $i\in\mathcal{N}$, $j\in\mathcal{J}$, and $t\in\mathbb{R}_+$ and for all $n\in\mathbb{N}_+$), as may be true in parts of the city during some time intervals.

We focus on the first-order imbalances between the driver supply and customer demand by considering fluid scaling. For all $i\in\mathcal{N}$, $j\in\mathcal{J}$, $t\in\mathbb{R}_+$, $n\in\mathbb{N}_+$, and admissible policy π, define

$$\bar{A}_i^n(t):=A_i(nt)/n,\quad \bar{R}_i^n(t):=R_i(nt)/n,\quad \bar{E}_j^n(t):=E_j(nt)/n, \tag{10a}$$

$$\bar{\Lambda}_i^n:=\Lambda_i^n/n,\quad \bar{\Gamma}_j^n:=\Gamma_j^n/n,\quad \bar{Q}_i^{\pi,n}:=Q_i^{\pi,n}/n,\quad \bar{D}_{ij}^{\pi,n}:=D_{ij}^{\pi,n}/n. \tag{10b}$$

By (1), (10a), and (10b), for all $i\in\mathcal{N}$, $t\in\mathbb{R}_+$, $n\in\mathbb{N}_+$, and admissible policy π,

$$\bar{Q}_i^{\pi,n}(t)=\bar{Q}_i^n(0)+\bar{A}_i^n\big(\bar{\Lambda}_i^n(t)\big)-\bar{R}_i^n\left(\int_0^t\theta_i^n(s)\bar{Q}_i^{\pi,n}(s)\mathrm{d}s\right)-\sum_{j\in\mathcal{J}}\bar{D}_{ij}^{\pi,n}(t). \tag{11}$$

We make the following assumption about the initial number of drivers.

Assumption 2 (Initial Conditions). *For all $i\in\mathcal{N}$, $\bar{Q}_i^n(0)\xrightarrow{a.s.}\bar{Q}_i(0)$ as $n\to\infty$, where $\bar{Q}_i(0)\in\mathbb{R}_+$.*

Assumptions 1 and 2 are in force throughout this paper.

3.2. An Asymptotic CLP Upper Bound

In this section, we derive an asymptotic upper bound on the fluid scaled objective (6) by solving a CLP. The decision variables are $\{\boldsymbol{q},\boldsymbol{x}\}:=\{q_i,x_{ij},i\in\mathcal{N},j\in\mathcal{J}\}$ such that $q_i,x_{ij}:[0,T]\to\mathbb{R}_+$, $q_i(t)$ denotes the number of type i drivers at time t, and $x_{ij}(t)$ denotes the fraction of type j customers who are offered type i drivers at time t, so that $\mu_j(t)\bar{F}_{ij}(t)x_{ij}(t)$ approximates the instantaneous matching rate between type i drivers and type j customers for all $i\in\mathcal{N}$ and $j\in\mathcal{J}$. The relevant CLP is as follows:

$$\max_{\boldsymbol{q},\boldsymbol{x}}\ \sum_{i\in\mathcal{N},j\in\mathcal{J}}w_{ij}\int_0^T\mu_j(s)\bar{F}_{ij}(s)x_{ij}(s)\mathrm{d}s \tag{12a}$$

$$\text{s.t. } q_i(t)=\bar{Q}_i(0)+\Lambda_i(t)-\int_0^t\theta_i(s)q_i(s)\mathrm{d}s-\sum_{j\in\mathcal{J}}\int_0^t\mu_j(s)\bar{F}_{ij}(s)x_{ij}(s)\mathrm{d}s,\quad \forall i\in\mathcal{N},t\in[0,T], \tag{12b}$$

$$\sum_{i\in\mathcal{N}}x_{ij}(t)\le 1,\quad \forall j\in\mathcal{J},t\in[0,T], \tag{12c}$$

$$q_i(t)\ge 0,\ x_{ij}(t)\ge 0,\quad \forall i\in\mathcal{N},j\in\mathcal{J},t\in[0,T], \tag{12d}$$

$$q_i \text{ and } x_{ij} \text{ are Lebesgue measurable for all } i\in\mathcal{N} \text{ and } j\in\mathcal{J}. \tag{12e}$$

Constraint (12b) is the queue length equation (see (11)). Constraint (12c) implies that a customer cannot be offered more than one driver. Any q_i feasible for CLP (12) is Lipschitz continuous by (12b) and Assumption 1.

Lemma 2. (1) *Suppose that both $\{\boldsymbol{q}^{(1)},\boldsymbol{x}\}$ and $\{\boldsymbol{q}^{(2)},\boldsymbol{x}\}$ are feasible process pairs for the CLP (12). Then $\boldsymbol{q}^{(1)}=\boldsymbol{q}^{(2)}$, that is, a feasible matching process $\boldsymbol{x}$ is associated with a unique feasible queue length process.* (2) *There exists an optimal solution of the CLP (12).*

The proof of Lemma 2 is presented in Appendix E, Section E.3. We denote an optimal solution of the CLP (12) by the process $\{\tilde{\boldsymbol{q}},\tilde{\boldsymbol{x}}\}:=\{\tilde{q}_i(t),\tilde{x}_{ij}(t),i\in\mathcal{N},j\in\mathcal{J},t\in[0,T]\}$. The following theorem establishes that the optimal objective function value of the CLP (12) is an asymptotic upper bound on the fluid scaled objective (6) under any admissible policy for almost all sample paths.

Theorem 1. *Under any admissible policy π,*

$$\limsup_{n\to\infty}\sum_{i\in\mathcal{N},j\in\mathcal{J}}w_{ij}\bar{D}_{ij}^{\pi,n}(T)\le\sum_{i\in\mathcal{N},j\in\mathcal{J}}w_{ij}\int_0^T\mu_j(s)\bar{F}_{ij}(s)\tilde{x}_{ij}(s)\mathrm{d}s,\quad a.s.$$

The proof of Theorem 1 is presented in Appendix A. The key challenge is to show that there exists a feasible matching process $\{x_{ij}(t), i \in \mathcal{N}, j \in \mathcal{J}, t \in [0,T]\}$ (not necessarily optimal) such that the derivative of the limit of the fluid scaled cumulative matching process $\{\bar{D}^{\pi,n}_{ij}(t), i \in \mathcal{N}, j \in \mathcal{J}, t \in [0,T]\}$ corresponds to the process $\{\mu_j(t)\bar{F}_{ij}(t)x_{ij}(t), i \in \mathcal{N}, j \in \mathcal{J}, t \in [0,T]\}$. If the aforementioned feasible matching process is $\tilde{\boldsymbol{x}}$, then the upper bound in Theorem 1 is attained, meaning the policy is asymptotically optimal.

3.3. A CLP-Based Randomized Policy

We would like to find an admissible matching policy that can reproduce any feasible queue length process and matching process $\{\boldsymbol{q}, \boldsymbol{x}\}$ for the CLP (12) (including an optimal one, $\{\tilde{\boldsymbol{q}}, \tilde{\boldsymbol{x}}\}$). This is important because finding a feasible queue length and matching process pair that improves on any myopic matching policy (such as CD) may be possible even when finding an optimal CLP solution is not. The question is, How do we translate between a feasible process pair $\{\boldsymbol{q}, \boldsymbol{x}\}$ and the decision of which driver to offer an arriving customer?

Definition 2 (Randomized Policy). If a type j customer arrives in the system at time t, the system controller makes a random selection from the set $\mathcal{N} \cup \{0\}$ such that the probability that the outcome is i is $x_{ij}(t)$ for all $i \in \mathcal{N}$ and the probability that the outcome is 0 is $1 - \sum_{i\in\mathcal{N}} x_{ij}(t)$. If the outcome is i for some $i \in \mathcal{N}$ and there is a type i driver in the system, then the system controller offers a type i driver to the customer. If the outcome is i for some $i \in \mathcal{N}$ but there is no type i driver in the system or if the outcome is 0, then no driver is offered to the customer (and so the customer is lost).

Because the randomized policy does not use the queue length process $\boldsymbol{q}$ and by Lemma 2, part 1, we denote it by π_R and write $\pi_R(\boldsymbol{x})$ when we want to emphasize its dependence on the particular feasible matching process $\boldsymbol{x}$. Under a technical condition on the associated feasible matching process for the CLP (12), π_R is admissible.

Lemma 3. *Suppose that the feasible matching process for the CLP* (12), $\boldsymbol{x}$, *is such that* x_{ij} *is a Borel-measurable simple function for all* $i \in \mathcal{N}$ *and* $j \in \mathcal{J}$. *Then,* $\pi_R(\boldsymbol{x})$ *is admissible.*

The proof of Lemma 3 is presented in Appendix E, Section E.2.

The randomized policy is a simple policy, because no state information is needed for its implementation. Still, its performance replicates any feasible process pair for the CLP (12).

Theorem 2. *Let* $\{\boldsymbol{q}, \boldsymbol{x}\}$ *be a feasible process pair for the CLP* (12) *such that* $\boldsymbol{x}$ *satisfies the condition in Lemma* 3. *For all* $i \in \mathcal{N}$, *as* $n \to \infty$,

$$\sup_{0\le t\le T}\left|\sum_{j\in\mathcal{J}}\left(\bar{D}^{\pi_R(\boldsymbol{x}),n}_{ij}(t) - \int_0^t \mu_j(s)\bar{F}_{ij}(s)x_{ij}(s)\mathrm{d}s\right)\right| \xrightarrow{a.s.} 0 \quad \textit{and} \quad \left\|Q^{\pi_R(\boldsymbol{x}),n}_i - q_i\right\|_T \xrightarrow{a.s.} 0.$$

The proof of Theorem 2 is presented in Appendix D.

Theorem 2 shows that when an optimal matching process $\tilde{\boldsymbol{x}}$ of the CLP (12) is a Borel-measurable simple function, then $\pi_R(\tilde{\boldsymbol{x}})$ attains the upper bound in Theorem 1. However, $\tilde{\boldsymbol{x}}$ may not satisfy the aforementioned condition. Then, we do not know the associated randomized policy is admissible, and so we cannot use Theorem 2 to ensure the upper bound in Theorem 1 is achieved. However, we can approximate an optimal matching process of CLP (12) with a sequence of Borel-measurable simple functions to ensure the upper bound is nearly achieved. Specifically, by theorem 2.10 and proposition 2.12 of Folland (1999), there exists a sequence $\{\tilde{\boldsymbol{x}}^r, r \in \mathbb{N}_+\}$ such that $\tilde{\boldsymbol{x}}^r := \{\tilde{x}^r_{ij}(t), i \in \mathcal{N}, j \in \mathcal{J}, t \in [0,T]\}$; for all $i \in \mathcal{N}$, $j \in \mathcal{J}$, and $r \in \mathbb{N}_+$, $\tilde{x}^r_{ij}$ is a Borel-measurable simple function; $0 \le \tilde{x}^r_{ij}(t) \le \tilde{x}^{r+1}_{ij}(t) \le \tilde{x}_{ij}(t)$ for all $t \in [0,T]$ except on a set of zero measure; and $\tilde{x}^r_{ij}(t) \to \tilde{x}_{ij}(t)$ as $r \to \infty$ for all $t \in [0,T]$ except on a set of zero measure. By Lemma 2, part 1, and Lemma C.1, part 3, for each $\tilde{\boldsymbol{x}}^r$, there exists a unique $\tilde{\boldsymbol{q}}^r$ such that $\{\tilde{\boldsymbol{q}}^r, \tilde{\boldsymbol{x}}^r\}$ is a feasible process pair for the CLP (12). Then, by Theorem 2 and the bounded convergence theorem used on the sequence $\{\mu_j(t)\bar{F}_{ij}(t)\tilde{x}^r_{ij}(t), t \in [0,T], r \in \mathbb{N}_+\}$, we have the following corollary.

Corollary 1. *For any* $\epsilon > 0$, *there exists an* $r_0(\epsilon) \in \mathbb{N}_+$ *such that if* $r \ge r_0(\epsilon)$,

$$\lim_{n\to\infty} \sum_{i\in\mathcal{N}, j\in\mathcal{J}} \bar{D}^{\pi_R(\tilde{\boldsymbol{x}}^r),n}_{ij}(T) \ge \sum_{i\in\mathcal{N}, j\in\mathcal{J}} \int_0^T \mu_j(s)\bar{F}_{ij}(s)\tilde{x}_{ij}(s)\mathrm{d}s - \epsilon, \quad a.s.$$

Solving the CLP (12) or accurately approximating an optimal solution of it is a very challenging task (see Perold 1981). If $\mu_j(t)\bar{F}_{ij}(t)$ was a constant function of t and $\theta_i = \boldsymbol{0}$ for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$, then the CLP (12) would become a separated continuous linear program (SCLP; see Anderson et al. 1983). Although an SCLP is solvable

Published in Stochastic Systems on February 25, 2020 as DOI: 10.1287/stsy.2019.0037.
This article has not been copyedited or formatted. The final version may differ from this version.

(see Anderson et al. 1983, Weiss 2008), it is an NP-hard problem (see Bertsimas et al. 2015). Hence, we will derive conditions under which the CLP (12) can be simplified into an LP in the following section.

4. An Asymptotically Optimal LP-Based Randomized Policy

LP-based matching policies arise when the system controller does not need to "save" drivers for future customers, that is, when the system controller can be myopic. Section 4.1 shows a myopic LP-based randomized policy is asymptotically optimal when the drivers are always busy, or fully utilized. A myopic LP-based randomized policy is also asymptotically optimal when parameters are time homogeneous, regardless of whether drivers are fully utilized, and that LP can be modified to include pricing. We show in Section 4.2 that when pricing affects driver supply and customer demand, jointly optimizing over pricing and matching decisions results in fully utilized drivers, which provides a partial justification of the aforementioned "fully utilized driver" condition.

4.1. An LP-Based Randomized Matching Policy

The following LP is relevant at each fixed $t \in [0, T]$:

$$\max_{\boldsymbol{x}(t)} \sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij}\mu_j(t)\bar{F}_{ij}(t)x_{ij}(t) \tag{13a}$$

$$\text{s.t.} \sum_{j\in\mathcal{J}} \mu_j(t)\bar{F}_{ij}(t)x_{ij}(t) \le \lambda_i(t), \quad \forall i \in \mathcal{N}, \tag{13b}$$

$$\sum_{i\in\mathcal{N}} x_{ij}(t) \le 1, \quad \forall j \in \mathcal{J}, \tag{13c}$$

$$x_{ij}(t) \ge 0, \quad \forall i \in \mathcal{N},\ j \in \mathcal{J}, \tag{13d}$$

where the decision variables are $\{x_{ij}(t), i \in \mathcal{N}, j \in \mathcal{J}\}$. The main difference between the LP (13) and the CLP (12) is that (13a) is the derivative of (12a) and (13b) is a capacity constraint instead of a queue length equation (meaning we do not need $\boldsymbol{q}(t)$ as a decision variable). By Assumption 1 and because $x_{ij}(t) = 0$ for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$ is feasible for the LP (13) for all $t \in [0, T]$, there exists an optimal solution of the LP (13) for all $t \in [0, T]$. We denote an optimal solution of the LP (13) at time t by $\{x^*_{ij}(t), i \in \mathcal{N}, j \in \mathcal{J}\}$ for all $t \in [0, T]$ and $\boldsymbol{x}^* := \{x^*_{ij}(t), i \in \mathcal{N}, j \in \mathcal{J}, t \in [0, T]\}$.

Assumption 3 (Measurability). *There exists an $\boldsymbol{x}^*$ such that $\{x^*_{ij}(t), t \in [0, T]\}$ is Lebesgue measurable for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$.*

If each of λ_i and μ_j is a step function for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$, then there exists a function which satisfies Assumption 3, and that function can be chosen as a step function. Assumption 3 is valid in Section 4.1.

The CLP (12) can be solved using the LP (13) when the initial driver mass is zero and drivers are fully utilized.

Lemma 4. *Suppose that*

$$\bar{Q}_i(0) = 0, \quad \forall i \in \mathcal{N}, \tag{14}$$

$$\sum_{i\in\mathcal{N}, j\in\mathcal{J}} \int_0^t \mu_j(s)\bar{F}_{ij}(s)x_{ij}(s)\mathrm{d}s = \sum_{i\in\mathcal{N}} \Lambda_i(t) \quad \text{for all } t \in [0, T], \tag{15}$$

for some matching process $\boldsymbol{x}$ that is feasible for the CLP (12).[2] *If $w_{ij} = 1$ for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$, then there exists a process $\boldsymbol{q}^*$ such that the pair $\{\boldsymbol{q}^*, \boldsymbol{x}^*\}$ is an optimal solution of the CLP (12).*

The proof of Lemma 4 is presented in Appendix E, Section E.4.

The following asymptotic LP-based upper bound on the fluid scaled objective (5) follows from Lemma 4 combined with Theorem 1.

Corollary 2. *If the conditions (14) and (15) hold, under any admissible policy π, we have*

$$\limsup_{n\to\infty} \sum_{i\in\mathcal{N}, j\in\mathcal{J}} \bar{D}^{\pi,n}_{ij}(T) \le \sum_{i\in\mathcal{N}, j\in\mathcal{J}} \int_0^T \mu_j(s)\bar{F}_{ij}(s)x^*_{ij}(s)\mathrm{d}s, \quad a.s.$$

In order to obtain the upper bound in Corollary 2, the LP (13) must be solved infinitely many times (for all $t \in [0, T]$). However, in practice, the LP (13) can be solved at finitely many time epochs, and the remaining $x^*_{ij}(t)$ values can be approximated by, for example, linear interpolation or assuming that x^*_{ij} is a step function.

Moreover, the time-varying parameters of the LP (13) can be estimated in real-time, because the matching decisions do not depend on future estimated arrival rates of the customers and drivers.

If the initial driver mass is zero and drivers are fully utilized, there is no need to be forward looking. Consequently, the myopic LP-based randomized policy $\pi_R(\boldsymbol{x}^*)$ is asymptotically optimal.

Theorem 3. *Suppose $\boldsymbol{x}^*$ satisfies the condition in Lemma 3, so $\pi_R(\boldsymbol{x}^*)$ is admissible. If conditions (14) and (15) hold and $w_{ij} = 1$ for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$, then $\pi_R(\boldsymbol{x}^*)$ achieves the asymptotic upper bound given in Corollary 2, that is,*

$$\lim_{n\to\infty} \sum_{i\in\mathcal{N}, j\in\mathcal{J}} \bar{D}_{ij}^{\pi_R(\boldsymbol{x}^*),n}(T) = \sum_{i\in\mathcal{N}, j\in\mathcal{J}} \int_0^T \mu_j(s)\bar{F}_{ij}(s)x_{ij}^*(s)\mathrm{d}s, \quad a.s.$$

The proof of Theorem 3 is presented in Appendix D.

4.2. Jointly Optimizing Pricing and Matching When Parameters Are Time Homogeneous

One natural intuition is that the required balance in (15) arises naturally when the system controller can use pricing to influence customer and driver behavior. This is because a “smart” system controller will not raise prices beyond what is needed to match driver supply with customer demand. This leads to a joint pricing and matching problem. However, any time-varying problem formulation is very difficult to solve. Therefore, we focus on a static formulation to show how (15) can be viewed as a consequence on good pricing decisions.

We begin with the observation that we do not need the condition (15) to show that the randomized policy based on the solution to the LP (13) is asymptotically optimal when parameters are time homogeneous.

Theorem 4. *If λ_i, μ_j, and $\bar{F}_{ij}$ are constant functions of t for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$, we can choose $\boldsymbol{x}^*$ as a constant function of t. Suppose that condition (14) holds. Then, there exists a process $\boldsymbol{q}^*$ such that the pair $\{\boldsymbol{q}^*, \boldsymbol{x}^*\}$ is an optimal solution of the CLP (12) and $\pi_R(\boldsymbol{x}^*)$ is admissible. Moreover, if $w_{ij} = 1$ for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$, then $\pi_R(\boldsymbol{x}^*)$ achieves the asymptotic upper bound given in Theorem 1; that is,*

$$\lim_{n\to\infty} \sum_{i\in\mathcal{N}, j\in\mathcal{J}} \bar{D}_{ij}^{\pi_R(\boldsymbol{x}^*),n}(T) = \sum_{i\in\mathcal{N}, j\in\mathcal{J}} \int_0^T \mu_j(s)\bar{F}_{ij}(s)\tilde{x}_{ij}(s)\mathrm{d}s, \quad a.s.$$

The proof of Theorem 4 is presented in Appendix D.

The question we address is, When the LP (13) formulation does not vary with time, and is expanded to include pricing, does an optimal solution satisfy (15)? To do this, suppose that, for all $i \in \mathcal{N}$, $j \in \mathcal{J}$, and $n \in \mathbb{N}_+$, λ_i^n and μ_j^n are constant functions that depend on the prices determined by the system controller at time 0, $\bar{F}_{ij}$ is a time-homogeneous function independent of the prices, and θ_i^n can depend on both the prices and the time. Specifically, first, the system controller sets prices $\boldsymbol{p} = \{p_{ij}, i \in \mathcal{N}, j \in \mathcal{J}\}$ at time 0. The prices can depend on both the customer's type $j \in \mathcal{J}$ (and so can be based on the customer's origin and destination area) and the area $i \in \mathcal{N}$ where the assigned driver is currently located. Second, the time-homogeneous arrival rates of the customers and drivers (λ_i^n and μ_j^n) and time-varying departure rates of unmatched drivers from their current areas (θ_i^n) are realized, and matchings are performed continuously over the time horizon $[0, T]$. We extend Assumption 1 in the following way:

Assumption 4 (Technicalities). *Let $\mathcal{P} := \{\boldsymbol{p} \in \mathbb{R}_+^{NJ} : p_{ij} \in [0, \bar{p}], \forall i \in \mathcal{N}, j \in \mathcal{J}\}$ be the set of possible price vectors, where $\bar{p} \in \mathbb{R}_+$ is the maximum chargeable price. For all $i \in \mathcal{N}$, $j \in \mathcal{J}$, $n \in \mathbb{N}_+$, and $\boldsymbol{p} \in \mathcal{P}$, we have $\lambda_i, \mu_j, \lambda_i^n, \mu_j^n : \mathcal{P} \to \mathbb{R}_+$, and the functions $\theta_i, \theta_i^n : \mathbb{R}_+ \times \mathcal{P} \to \mathbb{R}_+$ are defined such that $\theta_i(\cdot, \boldsymbol{p}) \in \mathbb{D}$, $\theta_i^n(\cdot, \boldsymbol{p}) \in \mathbb{D}$, $\sup_{t\in\mathbb{R}_+} \theta_i(t, \boldsymbol{p}) < \infty$, and*

$$\left|\lambda_i^n(\boldsymbol{p})/n - \lambda_i(\boldsymbol{p})\right| \vee \left|\mu_j^n(\boldsymbol{p})/n - \mu_j(\boldsymbol{p})\right| \vee \sup_{t\in[0,T_1]} \left|\theta_i^n(t, \boldsymbol{p}) - \theta_i(t, \boldsymbol{p})\right| \to 0,$$

as $n \to \infty$ for all $T_1 \in \mathbb{R}_+$.

The system controller sets prices at time 0 by solving the following optimization problem:

$$\max_{x,p} \sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij}\mu_j(\boldsymbol{p})\bar{F}_{ij}x_{ij} \tag{16a}$$

$$\text{s.t.} \sum_{j\in\mathcal{J}} \mu_j(\boldsymbol{p})\bar{F}_{ij}x_{ij} \le \lambda_i(\boldsymbol{p}), \quad \forall i \in \mathcal{N}, \tag{16b}$$

$$\sum_{i\in\mathcal{N}} x_{ij} \le 1, \quad \forall j \in \mathcal{J}, \tag{16c}$$

$$x_{ij} \ge 0, \quad p_{ij} \in [0, \bar{p}], \quad \forall i \in \mathcal{N},\ j \in \mathcal{J}, \tag{16d}$$

where the decision variables are x_{ij} and p_{ij} for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$. The optimization problem (16) is the LP (13), which is time homogeneous but modified to include pricing decisions. Because $x_{ij} = 0$ and $p_{ij} \in [0, \bar{p}]$ for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$ is feasible for (16), the feasible region is nonempty.

We cannot know whether there exists an optimal solution without additional assumptions on the functions λ_i and μ_j, $i \in \mathcal{N}$, $j \in \mathcal{J}$. For example, if λ_i and μ_j are continuous functions of $\boldsymbol{p}$ for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$, then the feasible region is compact and so an optimal solution of (16) exists. If the price(s) in an area are equal to 0, then one can expect the total customer arrival rate to be very large, but the driver arrival rate to be 0 and increasing in price. Thus, an optimal solution should have nonzero prices.

We associate the set of admissible policies given in Definition 1 with a price vector $\boldsymbol{p} \in \mathcal{P}$, which determines the rates of the processes generating the filtration $\mathbb{F} = \{\mathcal{F}(t), t \in \mathbb{R}_+\}$ representing the available information as time moves forward (see (7)). Specifically, for a given $\boldsymbol{p} \in \mathcal{P}$, a policy $\pi(\boldsymbol{p}) = \{\pi^n(\boldsymbol{p}), n \in \mathbb{N}_+\}$ is a sequence that specifies a policy for each n, and $\pi(\boldsymbol{p})$ is admissible if $\pi^n(\boldsymbol{p})$ is admissible for all $n \in \mathbb{N}_+$. Let $\{\boldsymbol{x}, \boldsymbol{p}\}$ be feasible matching fractions and prices for (16), and let $\pi_{R,\boldsymbol{p}}(\boldsymbol{x})$ denote the randomized policy associated with the matching fractions $\boldsymbol{x}$ (see Definition 2) under the price vector $\boldsymbol{p}$; that is, the system controller sets the price vector $\boldsymbol{p}$ at time 0 and then makes the matching decisions using the randomized policy with matching fractions $\boldsymbol{x}$. Then, we have the following corollary to Theorems 1 and 4.

Corollary 3. *Suppose that condition (14) holds and $\{\boldsymbol{x}^*, \boldsymbol{p}^*\}$ is an optimal solution of (16). Then, under any admissible policy $\pi(\boldsymbol{p})$, $\boldsymbol{p} \in \mathcal{P}$,*

$$\limsup_{n\to\infty} \sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij} \bar{D}_{ij}^{\pi(\boldsymbol{p}),n}(T) \le \sum_{i\in\mathcal{N}, j\in\mathcal{J}} T w_{ij} \mu_j(\boldsymbol{p}^*) \bar{F}_{ij} x_{ij}^*, \quad a.s.$$

Moreover, if $w_{ij} = 1$ for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$, then $\pi_{R,\boldsymbol{p}^*}(\boldsymbol{x}^*)$ is asymptotically optimal; that is,

$$\lim_{n\to\infty} \sum_{i\in\mathcal{N}, j\in\mathcal{J}} \bar{D}_{ij}^{\pi_{R,\boldsymbol{p}^*}(\boldsymbol{x}^*),n}(T) = \sum_{i\in\mathcal{N}, j\in\mathcal{J}} T \mu_j(\boldsymbol{p}^*) \bar{F}_{ij} x_{ij}^*, \quad a.s.$$

If the constraint (16b) is binding for all driver types under given feasible matching fractions and prices, then condition (15) holds. Hence, we would like to know whether the constraint (16b) binds under "good" pricing decisions, for example, under an optimal solution of (16) (if it exists). The intuition that a good pricing policy is one under which the constraint (16b) binds is natural when lowering prices results in fewer driver arrivals but more customer arrivals. Hence, there is no need to have idle drivers in the system at any time. That intuition is consistent with a result proved by Bimpikis et al. (2019, proposition 2) and results proved by Özkan (2018, propositions 1 and 3) showing that drivers never idle. We provide sufficient conditions in Example 1 below, under which the constraint (16b) binds under an optimal solution of (16).

Example 1. There exists a customer-specific baseline price for riding, denoted by $c_j > 0$, and an area specific surge multipliers $s_i \in [0, \bar{s}]$ for $i \in \mathcal{N}$, where $\bar{s} \in \mathbb{R}_+$ is the maximum possible surge multiplier. Suppose that c_j is constant but the system controller determines the surge multipliers at time 0. Let $\alpha : \mathcal{J} \to \mathcal{N}$ be a function that specifies the arrival location of each customer type. Then, the price that a type j customer needs to pay for a ride provided by a type i driver is $p_{ij} = c_j s_{\alpha(j)}$ for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$. Suppose that there exists $C_2 \ge C_1 > 0$ such that

$$-C_2 \le \sum_{j\in\mathcal{J}:\alpha(j)=i} \frac{\partial \mu_j}{\partial s_i} \le -C_1, \quad 0 \le \frac{\partial \lambda_i}{\partial s_i} \le C_2, \quad \text{for all } i \in \mathcal{N}, \tag{17a}$$

$$\sum_{j\in\mathcal{J}:\alpha(j)=i} \frac{\partial \mu_j}{\partial s_k} = 0, \quad \text{for all } i, k \in \mathcal{N} \text{ such that } i \ne k, \tag{17b}$$

$$\left| \sum_{j\in\mathcal{J}:\alpha(j)=i} \frac{\partial \mu_j}{\partial s_i} \right| - \sum_{k\in\mathcal{N}\setminus\{i\}} \left| \frac{\partial \lambda_k}{\partial s_i} \right| \ge C_1 \quad \text{for all } i \in \mathcal{N}, \tag{17c}$$

$$\text{for all } i \in \mathcal{N}, \text{ when } s_i = 0, \lambda_i = 0. \tag{17d}$$

Condition (17a) states that as the surge multiplier in an area decreases, the total customer arrival rate in that area increases, and the driver arrival rate in that area decreases. Condition (17b) implies that the total customer arrival rate in an area is not affected by surge multipliers in other areas. Condition (17c) roughly requires that the change in the surge multiplier in an area affects that area more than it affects the other areas. Condition (17d) enforces that drivers do not work for free.

Lemma 5. *Under the conditions* (17a)–(17d), *there exists an optimal solution of* (16) *in which the constraint* (16b) *is binding for all driver types.*

The proof of Lemma 5 is presented in Appendix E, Section E.5.

5. Performance Evaluation

We begin by observing that the performance of the randomized policy can likely be improved by incorporating state information. This is because the randomized policy can match a customer with an area in which there are no drivers, leading to that customer being lost. To correct this, in Section 5.1, we introduce state-dependent LP- and CLP-based policies that require knowledge of driver locations. Then, in Section 5.2, we compare the performance of those policies and the LP- and CLP-based randomized policy against the benchmark CD policy. We do this first when parameters are time homogeneous and second when they vary with time. In the first case, the LP- and CLP-based policies coincide (see Theorem 4), whereas in the second, when the importance of considering future customer and driver arrival rates becomes important, they do not.

5.1. Additional Proposed Matching Policies

We drop the superscript n in this section to be clear that any matching policy we propose must be well-defined and admissible for the model given in Section 2. We associate the proposed matching policies with $\boldsymbol{x}$, which is a feasible matching process for the CLP (12) for an appropriately defined $\boldsymbol{q}$ (see Lemma 2, part 1). We begin by modifying the randomized policy to incorporate information regarding which areas have no drivers.

5.1.1. Randomized Weighted Queue Policy. If a type j customer arrives in the system at time t, the system controller offers a type i driver with probability

$$\frac{x_{ij}(t)Q_i(t-)}{\sum_{i\in\mathcal{N}} x_{ij}(t)Q_i(t-)}, \tag{18}$$

if the denominator in (18) is strictly positive. Otherwise, the system controller does not offer any driver to the customer.

Under the randomized and the randomized weighted queue (RWQ) policies, if $x_{ij}(t) = 0$ for some $i \in \mathcal{N}$ and $j \in \mathcal{J}$, then the system controller will never offer a type i driver to a type j customer, even if the only drivers in the system are of type i. In comparison, the benchmark CD policy in (3) is *match conserving* in the sense that a customer arriving in a system with at least one driver available is always offered a driver, regardless of the driver types present. This motivates us to introduce two additional policies that are match conserving, one deterministic and one not.

5.1.2. Deterministic Policy. If a type $j \in \mathcal{J}$ customer arrives in the system at time t, the system controller offers a driver type from the set

$$\underset{\{i\in\mathcal{N}:\ Q_i(t-)>0\}}{\operatorname{argmin}} \left\{ D_{ij}(t-) - \int_0^t \mu_j(s)\bar{F}_{ij}(s)x_{ij}(s)\mathrm{d}s \right\}. \tag{19}$$

If the set in (19) is not a singleton, then the system controller can use any tie breaking rule that does not use any future information. If there is no driver in the system, then no driver is offered.

5.1.3. Hybrid Policy. Suppose that a type j customer arrives in the system and there is a driver in the system. Then the system controller makes a random selection in the same way explained under the randomized policy (see Definition 2). If the outcome is $i \in \mathcal{N}$ and there is a type i driver in the system, the system controller offers a type i driver to the customer. If the outcome is $i \in \mathcal{N}$ but there is no type i driver in the system or if the outcome is 0, the customer is offered the driver type specified in (3) by the CD policy. If there is no driver in the system, then the customer is lost.

Remark 7. If $x_{ij} = \mathbf{0}$ for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$ such that $i \neq \alpha(j)$, then the Hybrid policy is the CD policy defined in (3).

Lemma 6. *The RWQ, deterministic, and hybrid policies are admissible under the condition on* $\boldsymbol{x}$ *in Lemma* 3.

The proof of Lemma 6 is presented in Appendix E, Section E.2.

The RWQ, deterministic, and hybrid policies can be implemented regardless of whether they are admissible. However, in order to prove asymptotic optimality results parallel to those for the randomized policy (specifically, Theorems 2–4), admissibility is required. We conjecture that the statements of Theorems 2–4 hold under the RWQ, deterministic, and hybrid policies. However, the proofs are more difficult because of the state dependence of the aforementioned policies.

5.2. Simulation Experiments

In this section, we present simulation experiments where we test the performances of the CD policy and the LP- and CLP-based matching policies. We associate the LP- and CLP-based policies based on an optimal solution of the LP (13) and a solution of the CLP (12) with at most 1.02% optimality gap [that we find by solving CLP (12) when $\theta_i = \mathbf{0}$ for all $i \in \mathcal{N}$], respectively. We present an experiment with time-homogeneous parameters in Section 5.2.2 and an experiment with time-varying parameters in Section 5.2.3.

The term "offered driver" refers to the driver offered to an arriving customer, which is determined by the matching policy. The customer may or may not accept being matched with the driver offered at the arrival time $t \in [0, T]$, depending on the acceptance probability $\bar{F}_{ij}(t)$. Any variation of the word "match" means that the relevant customer has both been offered a driver and has accepted that driver, meaning that driver picks up that customer.

We assume that the larger the $\bar{F}_{ij}(t)$ at time $t \in [0, T]$, the smaller the pickup time of a type i driver for a type j customer. Then, under the CD policy, the offered driver set for a type j customer arriving at time t in (3) is exactly $\mathrm{argmax}_{\{i \in \mathcal{N}:\, Q_i(t-)>0\}} \bar{F}_{ij}(t)$ for all $j \in \mathcal{J}$ and $t \in \mathbb{R}_+$ (and the match occurs if also the resulting pickup time is lower than the time that customer is willing to wait for pickup).

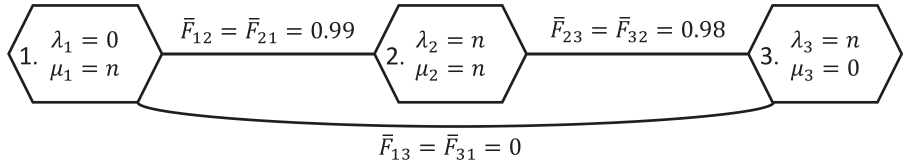
We choose $N = J = 3$ in both simulation experiments (so the customer type represents the customer's arrival location). This is the smallest network size possible that allows for us to illustrate the more general insight behind when and why LP-based policies outperform the CD policy. The reason a two-area network will not work is that as long as $\bar{F}_{ii}(t) \geq \bar{F}_{ij}(t)$ for all $i, j \in \{1, 2\}$ and $t \in [0, T]$, the LP-based policies give more priority to within-area matchings than they give to between-area matchings, which coincides with the CD policy. In contrast, the reason CLP-based policies outperform the CD policy (as well as any other myopic policy such as an LP-based one) has to do with their potential to be forward-looking (which can be seen in a two-area network).

5.2.1. Implementation Details. We used Omnet++ discrete-event simulation freeware. In each experiment, we started with an empty system, that is, $Q_i(0) = 0$ for all $i \in \mathcal{N}$. At each simulation instance associated with each matching policy, we did Rep number of replications. The maximum of the margin of errors associated with the 95% confidence interval for the percentage of all customers matched ($= t_{0.025, Rep-1} \times$ sample standard deviation $/\sqrt{Rep}$) among all policies in all instances is less than 0.42.

5.2.2. Time-Homogeneous Parameters. We present a simulation experiment that shows that our LP-based policies can potentially match more customers and drivers in a finite time horizon than the CD policy does. Figure 3 provides all input parameters. Because all parameters are time homogeneous, the LP- and CLP-based policies coincide (see Theorem 4). We omit (t) from the notation. Moreover, $n \in \{1, 10, 100\}$ measures the market size determined by the number of arrivals per time unit, as in Assumption 1.

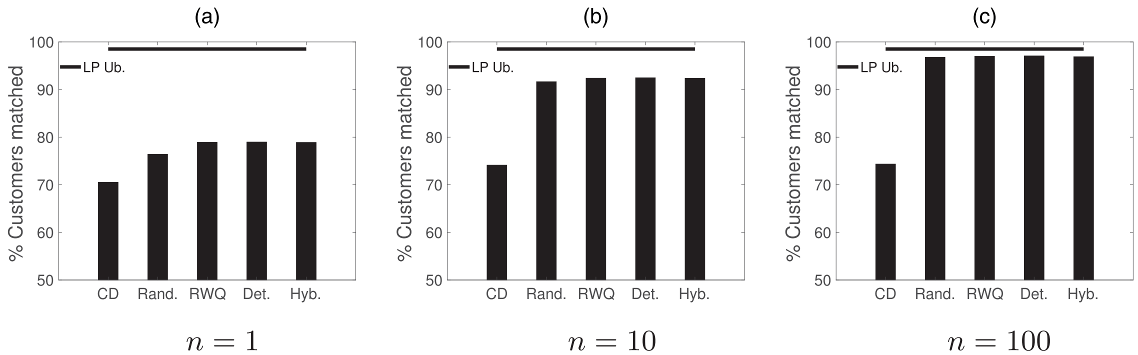
The intuition for why we do not expect the CD policy to perform well in this example is as follows. Under the CD policy, when a type 1 or 2 customer arrives in the system, the offered driver is of type 2 (if there is one in the system) because $\lambda_1 = 0$, $\bar{F}_{21} > \bar{F}_{31}$, and $\bar{F}_{22} > \bar{F}_{32}$. Because $\mu_1 + \mu_2 = 2n > \lambda_2 = n$, some of the type 1 or 2 customers cannot be offered type 2 drivers and must be offered type 3 drivers. The type 2 customers who are not offered type 2 drivers are offered type 3 drivers, and so 98% of those customers are matched. However, the type 1 customers offered type 3 drivers are lost because $\bar{F}_{31} = 0$. In comparison, the optimal solution of the LP (13) has $x^*_{21} = 1$, $x^*_{22} = 0.01$, and $x^*_{32} = 0.99$. In words, the LP-based policies match type 1 customers with type 2

Figure 3. The Parameters of the First Simulation Experiment Are $Q_i(0) = 0$, $1/\theta_i = 10$ Time Units, $\bar{F}_{ii} = 1$, $w_{ij} = 1$ for All $i, j \in \{1, 2, 3\}$, and $n \in \{1, 10, 100\}$



Published in Stochastic Systems on February 25, 2020 as DOI: 10.1287/stsy.2019.0037.
This article has not been copyedited or formatted. The final version may differ from this version.

Figure 4. The Percentage of All Customers Matched in the Simulation Experiment Corresponding to Figure 3



Note. Rand., Randomized; Det., deterministic; Hyb., hybrid.

drivers and type 2 customers with type 3 drivers, to prevent losing a nontrivial percentage of the type 1 customers.

Figure 4 shows the percentage of all customers matched with drivers in the simulation experiment (i.e., the objective (5), because $w_{ij} = 1$ for all $i, j \in \{1, 2, 3\}$) under the CD policy and the LP-based policies. “LP Ub.” shows 100 times the optimal objective function value of the LP (13) divided by the total customer arrival rate to the system, which is an approximate upper bound on the percentage of all customers matched under any admissible policy by Theorems 1 and 4. The key observations are that

i. the LP-based policies outperform the CD policy in all traffic intensities, and

ii. the LP-based policies are very close to the approximate upper bound based on Theorem 1 as the arrival rates become large, which verifies Theorem 4 numerically.

5.2.3. Time-Varying Parameters. Motivated by the example in Figure 2 in Section 1, our next simulation experiment is designed to show that CLP-based policies can significantly outperform LP-based policies and the CD policy by taking into account the future customer and driver arrival rates. We assume that the customer arrival rate in one area will spike; in particular, we assume $\mu_1(t) = 0$ for all $t \in [0, T/2]$ and $\mu_1(t) = 2n$ for all $t \in [T/2, T]$, where $n \in \{0.1, 1, 10, 100\}$ is used to determine the customer and driver arrival rates to the system. Figure 5 provides the other input parameters, which are time homogeneous, and so (t) is omitted. The average driver patience time is $1/\theta \in \{1.667, 16.67, 166.7, 1{,}667\}$ minutes, and so noting that $T = 30$ minutes, $1/\theta = 1.667$ minutes represents impatient drivers, and $1/\theta = 1{,}667$ minutes represents very patient drivers.

Because the total customer and driver arrival rates in the first half of the simulation ($t \in [0, T/2]$) are n and $2n$, respectively, at least half of the arriving drivers will be idle in the first half of the simulation. This implies that the condition (15) requiring drivers to be fully utilized does not hold in this experiment.

In the first half of the simulation, that is, in the time interval $[0, T/2]$, the CD policy offers type 1 drivers to type 2 customers. In $[T/2, T]$, the CD policy offers type 1 drivers to both type 1 and type 2 customers. However, in $[T/2, T]$, the arrival rate of the type 1 drivers is n, but the arrival rate of the customers who will accept a match with type 1 drivers if offered is $2n + 0.99n = 2.99n$. Therefore, only 1/2.99 of the type 1 customers are offered and matched with type 1 drivers and the rest will be lost.

The optimal solution of the LP (13) is such that $x^*_{12}(t) = 1$ for all $t \in [0, T/2]$ and $x^*_{11}(t) = 0.5$ and $x^*_{32}(t) = 1$ for all $t \in [T/2, T]$, and all other decision variables at all other times are 0. Hence, in $[0, T/2]$, type 2 customers are

Figure 5. The Parameters of the Second Simulation Experiment Are $Q_i(0) = 0$, $\bar{F}_{ii} = 1$, $w_{ij} = 1$ for All $i, j \in \{1, 2, 3\}$, $n \in \{0.1, 1, 10, 100\}$, $T = 1{,}800$ Seconds = 30 Minutes, $\theta_i(t) = \theta$ Is Constant in i and t and $1/\theta \in \{1.667, 16.67, 166.7, 1{,}667\}$ Minutes, $\mu_1(t) = 0$ Customers/Second for All $t \in [0, T/2]$, and $\mu_1(t) = 2n$ Customers/Second for All $t \in [T/2, T]$

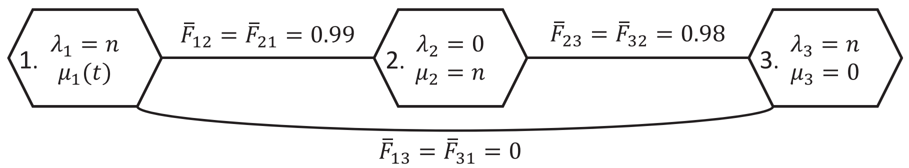


Table 1. The Percentage of All Customers Matched in the Simulation Experiment Corresponding to Figure 5, for $T = 30$ Minutes

			LP-based policies				CLP-based policies			
$1/\theta$ (min)	n	CD	Rand.	RWQ	Determ.	Hybrid	Rand.	RWQ	Determ.	Hybrid
1.667	0.1	63.8	57.9	62.0	67.1	67.5	62.5	62.8	68.1	67.9
	1	66.1	69.1	71.0	72.6	72.9	72.3	72.1	73.9	74.1
	10	66.1	72.9	73.6	73.9	74.1	75.2	75.2	75.5	75.5
	100	66.1	74.0	74.2	74.2	74.3	75.9	75.9	75.9	75.9
16.67	0.1	67.0	69.4	72.7	73.5	75.4	82.0	82.2	83.4	83.7
	1	66.4	72.7	73.9	74.1	74.8	85.8	85.7	86.1	86.3
	10	66.3	73.9	74.3	74.2	74.5	86.8	86.8	86.8	86.9
	100	66.2	74.2	74.4	74.2	74.4	86.9	86.9	86.9	86.9
166.7	0.1	68.0	70.3	74.6	73.7	77.0	92.1	92.5	92.8	93.5
	1	66.7	73.0	74.5	74.2	75.2	96.0	96.0	96.0	96.1
	10	66.3	74.0	74.5	74.2	74.7	96.8	96.8	96.8	96.8
	100	66.2	74.2	74.5	74.2	74.5	96.9	96.9	96.9	96.9
1,667	0.1	67.8	70.4	74.9	73.9	77.0	93.6	93.0	93.7	94.3
	1	66.8	73.2	74.4	74.1	75.4	97.4	97.3	97.4	97.6
	10	66.4	74.0	74.5	74.2	74.7	98.5	98.5	98.5	98.6
	100	66.2	74.2	74.5	74.3	74.5	98.8	98.8	98.7	98.8

Note. Rand., Randomized; Determ., deterministic.

offered type 1 drivers, and in $[T/2, T]$, type 2 customers are offered type 3 drivers so that type 1 customers can be offered type 1 drivers. Therefore, unlike the CD policy, the LP-based policy considers the fact that type 1 customers accept only type 1 drivers. However, in $[T/2, T]$, the arrival rate of type 1 customers is $2n$, but the arrival rate of the type 1 drivers is n, which implies that half of the type 1 customers will be lost.

When $\theta_i = \mathbf{0}$ for all $i \in \mathcal{N}$, an optimal matching process for the CLP (12) has $x_{32}(t) = 1$ for all $t \in [0, T]$, $x_{11}(t) = 0$ for all $t \in [0, T/2]$, and $x_{11}(t) = 1$ for all $t \in [T/2, T]$, and all other decision variables are $\mathbf{0}$. (To see that that process is optimal, notice that as many drivers as possible are matched.) We consider the CLP-based matching policy that always offers type 3 drivers to type 2 customers and offers type 1 drivers to type 1 customers as much as possible. This solution is forward-looking because the type 1 drivers arriving before time $T/2$ are "saved" to be offered to the type 1 customers arriving after time $T/2$. Furthermore, for any given $\theta \in \mathbb{R}_+$, the optimality gap under this policy is at most 1.02%. To see this, under this policy, type 1 customers are matched as much as possible, and the number of matched type 2 customers is $0.98nT$ and bounded above by $0.99nT$. Therefore, the loss compared with an optimal CLP solution is at most $(0.99 - 0.98)/0.98 = 1.02\%$.

Table 1 shows the percentage of all customers that are matched with drivers in the simulation experiment (i.e., the objective (5), because $w_{ij} = 1$ for all $i, j \in \{1, 2, 3\}$) under the CD policy, the LP-based policies, and the CLP-based policies.

Table 1 confirms that the CLP-based policies outperform the LP-based policies, and the LP-based policies outperform the CD policy, although condition (15) does not hold. Furthermore, the performance of the CLP-based policies improves as θ becomes small, whereas the performance of the CD and LP-based policies is independent of θ. This is because the CLP-based policies are forward-looking but the CD and LP-based policies are not.

6. Concluding Remarks

The decisions on which driver to offer to each arriving customer in a ride-sharing system impact the overall number of customers matched. This is because those decisions determine whether future available drivers will be close to the locations of arriving customers. We have formulated an optimization problem whose solution serves as an asymptotic upper bound on the cumulative number of matchings as the market becomes large. That optimization problem accounts for (i) the differing arrival rates of customers and drivers in different areas of the city, (ii) how long customers are willing to wait for driver pickup, (iii) how long drivers are willing to wait for a customer, and (iv) the time-varying nature of all the aforementioned parameters.

The aforementioned optimization problem is in general a CLP, which can be difficult to solve. We establish that a simple randomized matching policy can asymptotically mimic the performance of any feasible matching process for the CLP, so that there is potential to develop "good" CLP-based matching policies, even when an optimal CLP solution is unknown. When an optimal CLP solution is known, then a CLP-based randomized policy asymptotically achieves the aforementioned optimization problem upper bound.

Under the assumption that drivers are fully utilized or when the CLP parameters are time homogeneous, the CLP solution can be specified through LP solutions. Then, an LP-based randomized policy asymptotically achieves the aforementioned optimization problem upper bound. In the time-homogeneous setting, when customer and driver arrival rates depend on price, we establish an asymptotic upper bound on the cumulative number of matchings by solving an optimization problem that jointly optimizes over prices and matchings, and provide a joint pricing and matching policy that achieves that upper bound. Excellent questions for future research include better understanding the joint pricing and matching problem in time-varying settings and how to incorporate customer and driver behaviors.

We do not track drivers after they drop off customers, or after they choose to relocate from one area to another while waiting to be matched. Tracking drivers is an interesting future research topic. This is clearly doable when the parameters are time homogeneous (see Banerjee et al. 2016, 2018; Braverman et al. 2016), but is challenging when the parameters are time varying. First, let us consider time-homogeneous parameters. Then, the CLP (12) simplifies into the LP (13) without requiring the condition (15) (see Theorem 4). Let $d: \mathcal{J} \to \mathcal{N}$ be a function denoting the destination area of the customer types. In order to track the drivers after being matched, we can modify the constraint (13b) as follows:

$$\sum_{j\in\mathcal{J}} \mu_j \bar{F}_{ij} x_{ij} \le \lambda_i + \sum_{k\in\mathcal{N}} \sum_{j\in\mathcal{J}:d(j)=i} \mu_j \bar{F}_{kj} x_{kj}, \quad \forall i \in \mathcal{N}, \tag{20}$$

where λ_i denotes the external idle driver arrival rate in area $i \in \mathcal{N}$ (i.e., λ_i denotes the rate of drivers becoming online at area i), and the second term on the right-hand side of (20) denotes the rate of drivers who drop customers and become available in area $i \in \mathcal{N}$. Notice that (20) is a linear constraint; thus, the resulting optimization problem is still an LP. We conjecture that the randomized policy associated with an optimal solution of the modified LP is asymptotically optimal under the objective of maximizing the steady-state total matching rate.

Next, let us consider time-varying parameters. Let f_{ij} denote the deterministic travel time of a driver from area i to j for all $i, j \in \mathcal{N}$ and recall that $\alpha(j)$ denotes the arrival area of type j customers. Then, the constraint (12b) is modified as follows:

$$\begin{aligned} q_i(t) = \bar{Q}_i(0) + \Lambda_i(t) - \int_0^t \theta_i(s) q_i(s) \mathrm{d}s + \sum_{k\in\mathcal{N}} \sum_{j\in\mathcal{J}:d(j)=i} \int_0^{\left(t - f_{k\alpha(j)} - f_{\alpha(j)d(j)}\right)^+} \mu_j(s) \bar{F}_{kj}(s) x_{kj}(s) \mathrm{d}s \\ - \sum_{j\in\mathcal{J}} \int_0^t \mu_j(s) \bar{F}_{ij}(s) x_{ij}(s) \mathrm{d}s, \qquad \forall i \in \mathcal{N}, t \in \mathbb{R}_+, \end{aligned} \tag{21}$$

where $\Lambda_i(t)$ denotes the expected cumulative number of external idle driver arrivals in area $i \in \mathcal{N}$ up to time $t \in \mathbb{R}_+$ and the fourth term in the right-hand side of (22) denotes the cumulative number of drivers who drop customers and become available in area $i \in \mathcal{N}$ up to time $t \in \mathbb{R}_+$. Notice that solving the CLP under the constraint (21) is very challenging independent of the system being in transient or steady state.

Acknowledgments
The authors thank Siddhartha Banerjee and Kimon Drakopoulos for their valuable comments that helped them to write Section 4.2. They also thank the associate editor and two anonymous referees, whose comments significantly improved this paper.

Appendix A. Proof of Theorem 1
Appendix A provides the proof of Theorem 1, which provides an asymptotic upper bound on the fluid-scaled objective (6). That proof relies on the fluid equations, and the details of their derivation are given in Appendix B. Appendix C provides a new regulator mapping result, which is used in the proofs for asymptotic optimality results in Appendix D. The proofs of Lemmas 1–6 are in Appendix E. Finally, we present a relative compactness result in space $\mathbb{D}$, which is used in the derivation of fluid equations, in Appendix F.

The proof of Theorem 1 relies on understanding the behavior of fluid limits. In particular, the proof uses the fact that all fluid limits satisfy a set of equations that can be connected to the constraints of the CLP (12). The implication is that any fluid limit gives a feasible matching process for the CLP (12), which implies the optimal objective function value of the CLP (12) is an asymptotic upper bound for the objective (6) under fluid scale.

Published in Stochastic Systems on February 25, 2020 as DOI: 10.1287/stsy.2019.0037.
This article has not been copyedited or formatted. The final version may differ from this version.

For all $i \in \mathcal{N}$, $j \in \mathcal{J}$, $t \in \mathbb{R}_+$, $n \in \mathbb{N}_+$, and admissible policy $\pi = \{\pi^n, n \in \mathbb{N}_+\}$, let us define

$$G_{ij}^{\pi,n}(t) := \sum_{k=1}^{E_j(\Gamma_j^n(t))} \mathbb{I}\left(\pi_j^n(k) = i\right), \tag{A.1}$$

which is the cumulative number of times the system controller offers a type i driver to a type j customer up to time t, and let $\bar{G}_{ij}^{\pi,n} := G_{ij}^{\pi,n}/n$ be the associated fluid scaled process. Clearly, $D_{ij}^{\pi,n} \le G_{ij}^{\pi,n}$ for all $i \in \mathcal{N}$, $j \in \mathcal{J}$, $n \in \mathbb{N}_+$, and admissible policy π by (4) and (A.1). For all $n \in \mathbb{N}_+$ and admissible policy π, let

$$\mathbb{X}^{\pi,n} := \left(A_i \circ \Lambda_i^n,\ E_j \circ \Gamma_j^n,\ Q_i^{\pi,n},\ D_{ij}^{\pi,n},\ G_{ij}^{\pi,n}, R_i^n\left(\int_0^\cdot \theta_i^n(s) Q_i^{\pi,n}(s)\mathrm{d}s\right),\quad \forall i \in \mathcal{N},\ j \in \mathcal{J}\right), \tag{A.2}$$

where "$\circ$" denotes the composition map. Then, $\mathbb{X}^{\pi,n}(\cdot,\omega) \in \mathbb{D}^{2NJ+3N+J}$ for all $\omega \in \Omega$. Let $\bar{\mathbb{X}}^{\pi,n} := \mathbb{X}^{\pi,n}/n$ be the fluid scaled version of $\mathbb{X}^{\pi,n}$. We define fluid limit(s) of $\{\mathbb{X}^{\pi,n}, n \in \mathbb{N}_+\}$ similarly to the definition in Dai and Tezcan (2011, p. 299) in the following way.

Definition A.1 (Fluid Limit). Let us fix an arbitrary admissible policy $\pi = \{\pi^n, n \in \mathbb{N}_+\}$. Then, $\bar{\mathbb{X}}^\pi$ is a fluid limit of $\{\mathbb{X}^{\pi,n}, n \in \mathbb{N}_+\}$ if there exists an $\omega \in \Omega$ and a subsequence $\{n_k, k \in \mathbb{N}_+\}$ such that $\bar{\mathbb{X}}^{\pi,n_k}(\cdot,\omega) \to \bar{\mathbb{X}}^\pi$ u.o.c. as $k \to \infty$.

Proposition A.1 (Fluid Equations). *For any admissible policy $\pi = \{\pi^n, n \in \mathbb{N}_+\}$, there exists a full set $\mathcal{A}^\pi \subset \Omega$ [i.e., $\mathbf{P}(\mathcal{A}^\pi) = 1$] such that for any $\omega \in \mathcal{A}^\pi$, $\{\bar{\mathbb{X}}^{\pi,n}(\cdot,\omega), n \in \mathbb{N}_+\}$ is relatively compact (i.e., every subsequence has a convergent subsequence) in the Skorokhod space $\mathbb{D}^{2NJ+3N+J}$ endowed with the u.o.c. topology. Thus, fluid limits exist in almost all sample paths and any fluid limit*

$$\bar{\mathbb{X}}^\pi = \left(\Lambda_i,\ \Gamma_j,\ \bar{Q}_i^\pi,\ \bar{D}_{ij}^\pi,\ \bar{G}_{ij}^\pi,\ \int_0^\cdot \theta_i(s)\bar{Q}_i^\pi(s)\mathrm{d}s,\quad \forall i \in \mathcal{N},\ j \in \mathcal{J}\right)$$

and satisfies the following equations for all $t \in \mathbb{R}_+$:

$$\bar{Q}_i^\pi(t) = \bar{Q}_i(0) + \Lambda_i(t) - \int_0^t \theta_i(s)\bar{Q}_i^\pi(s)\mathrm{d}s - \sum_{j\in\mathcal{J}} \bar{D}_{ij}^\pi(t) \ge 0,\ \forall i \in \mathcal{N}, \tag{A.3a}$$

$$\bar{D}_{ij}^\pi(0) = \bar{G}_{ij}^\pi(0) = 0,\quad \forall i \in \mathcal{N},\ j \in \mathcal{J}, \tag{A.3b}$$

$$\bar{D}_{ij}^\pi \text{ and } \bar{G}_{ij}^\pi \text{ are nondecreasing for all } i \in \mathcal{N} \text{ and } j \in \mathcal{J}, \tag{A.3c}$$

$$\bar{D}_{ij}^\pi, \bar{G}_{ij}^\pi, \text{ and } \bar{Q}_i^\pi \text{ are Lipschitz continuous for all } i \in \mathcal{N} \text{ and } j \in \mathcal{J}, \tag{A.3d}$$

$$\sum_{i\in\mathcal{N}}\left(\bar{G}_{ij}^\pi(t_2) - \bar{G}_{ij}^\pi(t_1)\right) \le \Gamma_j(t_2) - \Gamma_j(t_1),\quad \forall j \in \mathcal{J} \text{ and } t_1, t_2 \in \mathbb{R}_+ \text{ such that } t_2 \ge t_1, \tag{A.3e}$$

$$\left[\bar{D}_{ij}^\pi(t) - \bar{D}_{ij}^\pi(\tau_m) - \bar{F}_{ij}(\tau_m)\left(\bar{G}_{ij}^\pi(t) - \bar{G}_{ij}^\pi(\tau_m)\right)\right] \times \mathbb{I}(t \in [\tau_m, \tau_{m+1})) = 0 \quad \text{for all } m \in \mathbb{N},\ i \in \mathcal{N}, \text{ and } j \in \mathcal{J}. \tag{A.3f}$$

The proof of Proposition A.1 is presented in Appendix B. Let us fix an arbitrary admissible policy π and an arbitrary $\omega \in \mathcal{A}^\pi$. Then there exists a subsequence $\{n_k, k \in \mathbb{N}_+\}$ such that

$$\limsup_{n\to\infty} \sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij}\bar{D}_{ij}^{\pi,n}(T,\omega) = \lim_{k\to\infty} \sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij}\bar{D}_{ij}^{\pi,n_k}(T,\omega).$$

Because $\bar{D}_{ij}^{\pi,n}(\cdot,\omega)$ is relatively compact (see Proposition A.1) for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$, there exists a subsequence of $\{n_k, k \in \mathbb{N}_+\}$ denoted by $\{n_l, l \in \mathbb{N}_+\}$ such that $\bar{D}_{ij}^{\pi,n_l}(\cdot,\omega)$ converges to a limit $\bar{D}_{ij}^\pi(\cdot)$ u.o.c. as $l \to \infty$ for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$. Hence,

$$\limsup_{n\to\infty} \sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij}\bar{D}_{ij}^{\pi,n}(T,\omega) = \lim_{k\to\infty} \sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij}\bar{D}_{ij}^{\pi,n_k}(T,\omega) = \lim_{l\to\infty} \sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij}\bar{D}_{ij}^{\pi,n_l}(T,\omega) = \sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij}\bar{D}_{ij}^\pi(T).$$

Then, Theorem 1 follows by the following result.

Proposition A.2. *Let $\pi = \{\pi^n, n \in \mathbb{N}_+\}$ denote an arbitrary admissible policy. Under any fluid limit of $\{\mathbb{X}^{\pi,n}, n \in \mathbb{N}_+\}$, we have*

$$\sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij}\bar{D}_{ij}^\pi(T) \le \sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij} \int_0^T \mu_j(s)\bar{F}_{ij}(s)\tilde{x}_{ij}(s)\mathrm{d}s.$$

Next, we will prove Proposition A.2. Let us consider the following CLP, which has the decision variables $\{\boldsymbol{q}, \boldsymbol{z}\} := \{q_i, z_{ij}, i \in \mathcal{N}, j \in \mathcal{J}\}$ such that $z_{ij} : [0, T] \to \mathbb{R}_+$ replaces $\mu_j(t)\bar{F}_{ij}(t)x_{ij}(t)$ in the CLP (12) for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$:

$$\max_{\boldsymbol{q},\boldsymbol{z}} \sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij} \int_0^T z_{ij}(s)\mathrm{d}s \tag{A.4a}$$

$$\text{s.t. } q_i(t) = \bar{Q}_i(0) + \Lambda_i(t) - \int_0^t \theta_i(s)q_i(s)\mathrm{d}s - \sum_{j\in\mathcal{J}} \int_0^t z_{ij}(s)\mathrm{d}s, \forall i \in \mathcal{N}, t \in [0, T], \tag{A.4b}$$

$$\sum_{i\in\mathcal{N}:\bar{F}_{ij}(t)>0} \frac{z_{ij}(t)}{\bar{F}_{ij}(t)} \le \mu_j(t), \quad \forall j \in \mathcal{J}, t \in [0, T], \tag{A.4c}$$

$$z_{ij}(t)\mathbb{I}\big(\bar{F}_{ij}(t) = 0\big) = 0, \quad \forall i \in \mathcal{N}, j \in \mathcal{J}, t \in [0, T], \tag{A.4d}$$

$$z_{ij}(t)\mathbb{I}\big(\mu_j(t) = 0\big) = 0, \quad \forall i \in \mathcal{N}, j \in \mathcal{J}, t \in [0, T], \tag{A.4e}$$

$$q_i(t) \ge 0, z_{ij}(t) \ge 0, \quad \forall i \in \mathcal{N}, j \in \mathcal{J}, t \in [0, T], \tag{A.4f}$$

$$q_i \text{ and } z_{ij} \text{ are Lebesgue measurable for all } i \in \mathcal{N} \text{ and } j \in \mathcal{J}. \tag{A.4g}$$

By an argument similar to Lemma 2, there exists an optimal solution of the CLP (A.4), which is denoted by $\{q_i^*(t), z_{ij}^*(t), i \in \mathcal{N}, j \in \mathcal{J}, t \in [0, T]\}$. We consider the CLP (A.4) instead of the CLP (12) for three main reasons. First, the feasible region of the CLP (A.4) is smaller than the one of the CLP (12) in the sense that there exist a surjective but not bijective mapping from the feasible region of the CLP (12) to the one of the CLP (A.4). Second, it is easier to connect a fluid limit of $\{\bar{D}_{ij}^{\pi,n}, i \in \mathcal{N}, j \in \mathcal{J}, n \in \mathbb{N}_+\}$ with the decision variables of the CLP (A.4) than the decision variables of the CLP (12). Third, we have the following result.

Lemma A.1. *The optimal objective function values of the CLP* (12) *and the CLP* (A.4) *are equal.*

Proof. Recall that $\{\tilde{q}_i(t), \tilde{x}_{ij}(t), i \in \mathcal{N}, j \in \mathcal{J}, t \in [0, T]\}$ is an optimal solution of the CLP (12). Consider the process pair $\{\tilde{q}_i(t), \mu_j(t)\bar{F}_{ij}(t)\tilde{x}_{ij}(t), i \in \mathcal{N}, j \in \mathcal{J}, t \in [0, T]\}$. It is easy to see that this process is feasible for the CLP (A.4). Thus,

$$\sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij} \int_0^T \mu_j(s)\bar{F}_{ij}(s)\tilde{x}_{ij}(s)\mathrm{d}s \le \sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij} \int_0^T z_{ij}^*(s)\mathrm{d}s. \tag{A.5}$$

Second, for all $i \in \mathcal{N}$, $j \in \mathcal{J}$, and $t \in [0, T]$, let

$$y_{ij}(t) := \begin{cases} \dfrac{z_{ij}^*(t)}{\mu_j(t)\bar{F}_{ij}(t)} & \text{if } \bar{F}_{ij}(t) > 0 \textit{ and } \mu_j(t) > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{A.6}$$

It is easy to see that the process $\{q_i^*(t), y_{ij}(t), i \in \mathcal{N}, j \in \mathcal{J}, t \in [0, T]\}$ is feasible for the CLP (12); thus,

$$\begin{aligned} \sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij} \int_0^T \mu_j(s)\bar{F}_{ij}(s)y_{ij}(s)\mathrm{d}s &= \sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij} \int_0^T z_{ij}^*(s)\mathbb{I}\big(\bar{F}_{ij}(s) > 0, \mu_j(s) > 0\big)\mathrm{d}s \\ &= \sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij} \int_0^T z_{ij}^*(s)\mathrm{d}s \le \sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij} \int_0^T \mu_j(s)\bar{F}_{ij}(s)\tilde{x}_{ij}(s)\mathrm{d}s, \end{aligned} \tag{A.7}$$

where the equality in (A.7) is by (A.4d) and (A.4e). Hence, the optimal objective function values of the CLP (12) and the CLP (A.4) are equal to each other by (A.5) and (A.7). □

Let us consider an arbitrary fluid limit of an arbitrary admissible policy π. By (A.3d) and theorem 3.35 of Folland (1999), both $\bar{D}_{ij}^{\pi}$ and $\bar{G}_{ij}^{\pi}$ are differentiable almost everywhere on $[0, T]$ for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$. By (A.3c), let d_{ij}^{π} and g_{ij}^{π} be the nonnegative derivatives of $\bar{D}_{ij}^{\pi}$ and $\bar{G}_{ij}^{\pi}$, respectively, on the points where they are differentiable and without loss of generality be equal to 0 on the points where they are not differentiable for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$ on the interval $[0, T]$. By (A.3b), (A.3d), and the fundamental theorem of calculus for Lebesgue integrals (see Folland 1999, theorem 3.35), for all $i \in \mathcal{N}$, $j \in \mathcal{J}$, and $t \in [0, T]$,

$$\bar{D}_{ij}^{\pi}(t) = \int_0^t d_{ij}^{\pi}(s)\mathrm{d}s, \qquad \bar{G}_{ij}^{\pi}(t) = \int_0^t g_{ij}^{\pi}(s)\mathrm{d}s. \tag{A.8}$$

Then, $\{\bar{Q}_i^{\pi}(t), d_{ij}^{\pi}(t), i \in \mathcal{N}, j \in \mathcal{J}, t \in [0, T]\}$ satisfies (A.4b) by (A.3a) and (A.8), and (A.4f) and (A.4g) by (A.3a), (A.3c), (A.3d), theorem 3.35 of Folland (1999), and construction.

Next, fix an arbitrary customer type $j \in \mathcal{J}$. By taking the derivatives of the both sides in (A.3e), we have

$$\sum_{i\in\mathcal{N}} g^{\pi}_{ij}(t) \le \mu_j(t), \tag{A.9}$$

for all $t \in [0,T]$. Let us fix an arbitrary $t \in [0,T]$, and let $t \in [\tau_m, \tau_{m+1})$ for some $m \in \mathbb{N}$. By (A.3f) and (A.8), for all $i \in \mathcal{N}$,

$$d^{\pi}_{ij}(t) = \begin{cases} \bar{F}_{ij}(\tau_m) g^{\pi}_{ij}(t) & \text{if } \bar{F}_{ij}(t) > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{A.10}$$

where (A.10) holds for all $t \in [\tau_m, \tau_{m+1})$ except on a set of zero measure, and we modify d^{π}_{ij} such that $d^{\pi}_{ij}(t) = 0$ in that set of zero measure. Then, by (A.9) and (A.10) and the nonnegativity of g^{π}_{ij} for all $i \in \mathcal{N}$,

$$\sum_{i\in\mathcal{N}:\bar{F}_{ij}(t)>0} \frac{d^{\pi}_{ij}(t)}{\bar{F}_{ij}(t)} \le \mu_j(t),$$

so that $\{\bar{Q}^{\pi}_i(t), d^{\pi}_{ij}(t), i \in \mathcal{N}, j \in \mathcal{J}, t \in [0,T]\}$ satisfies (A.4c). Moreover, it satisfies (A.4d) by (A.10). Last, by (A.9), we have

$$g^{\pi}_{ij}(t)\mathbb{I}\big(\mu_j(t) = 0\big) = 0, \qquad \forall i \in \mathcal{N}. \tag{A.11}$$

Thus, if $\bar{F}_{ij}(t) > 0$ for some $i \in \mathcal{N}$, then $d^{\pi}_{ij}(t)\mathbb{I}\big(\mu_j(t) = 0\big) = 0$ by (A.10) and (A.11). Otherwise, $d^{\pi}_{ij}(t) = 0$ by (A.10). Hence, $\{\bar{Q}^{\pi}_i(t), d^{\pi}_{ij}(t), i \in \mathcal{N}, j \in \mathcal{J}, t \in [0,T]\}$ satisfies (A.4e) as well, so it is feasible for the CLP (A.4). Then,

$$\sum_{i\in\mathcal{N},j\in\mathcal{J}} w_{ij}\int_0^T z^*_{ij}(s)\mathrm{d}s \ge \sum_{i\in\mathcal{N},j\in\mathcal{J}} w_{ij}\int_0^T d^{\pi}_{ij}(s)\mathrm{d}s = \sum_{i\in\mathcal{N},j\in\mathcal{J}} w_{ij} D^{\pi}_{ij}(T), \tag{A.12}$$

where the equality in (A.12) is by (A.8). Therefore, (A.12) together with Lemma A.1 proves Proposition A.2.

Appendix B. Proof of Proposition A.1 (Properties of Fluid Limits)

Let us fix an arbitrary $T_1 \in \mathbb{R}_+$. For all $k \in \mathbb{N}_+$, let $\mathbb{D}^k[0,T_1]$ denote the space of functions with domain $[0,T_1]$ and range $\mathbb{R}^k$, which are right continuous with left limits. We will first prove Proposition A.1 except the property (A.3f) with respect to processes defined in $\mathbb{D}^{2NJ+3N+J}[0,T_1]$, then we will extend these results to the processes in $\mathbb{D}^{2NJ+3N+J}$. Last, we will prove (A.3f). By the functional strong law of large numbers, random time-change theorem (see Chen and Yao 2001, theorems 5.10 and 5.3, respectively), and Assumption 1, we have

$$\bar{A}^n_i \circ \bar{\Lambda}^n_i \xrightarrow{a.s.} \Lambda_i \text{ u.o.c.}, \quad \bar{R}^n_i \xrightarrow{a.s.} e \text{ u.o.c.}, \quad \bar{E}^n_j \circ \bar{\Gamma}^n_j \xrightarrow{a.s.} \Gamma_j \text{ u.o.c.} \tag{B.1}$$

as $n \to \infty$ for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$. Let

$$\begin{aligned} \mathcal{A}_1 := \{\omega \in \Omega : \quad & \bar{Q}^n_i(0,\omega) \to \bar{Q}_i(0), \quad \bar{A}^n_i \circ \bar{\Lambda}^n_i(\cdot,\omega) \to \Lambda_i \text{ u.o.c.}, \\ & \bar{R}^n_i(\cdot,\omega) \to e \text{ u.o.c.}, \quad \bar{E}^n_j \circ \bar{\Gamma}^n_j(\cdot,\omega) \to \Gamma_j \text{ u.o.c.}, \\ & \text{as } n \to \infty \text{ for all } i \in \mathcal{N} \text{ and } j \in \mathcal{J}\}. \end{aligned} \tag{B.2}$$

Then, $\mathcal{A}_1$ is a full set [i.e., $\mathbf{P}(\mathcal{A}_1) = 1$] by (B.1) and Assumption 2. Let us fix an arbitrary $\omega \in \mathcal{A}_1$ and an arbitrary admissible policy $\pi = \{\pi^n, n \in \mathbb{N}_+\}$. We omit π and ω from the notation up to Lemma B.1 below for notational convenience. Clearly, for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$, $\{\bar{A}^n_i \circ \bar{\Lambda}^n_i, n \in \mathbb{N}_+\}$ and $\{\bar{E}^n_j \circ \bar{\Gamma}^n_j, n \in \mathbb{N}_+\}$ are relatively compact in the u.o.c. topology by (B.2). Moreover, Λ_i and Γ_j are Lipschitz continuous by Assumption 1, and for all $i \in \mathcal{N}$, $j \in \mathcal{J}$, and $t_1, t_2 \in \mathbb{R}_+$ such that $t_2 \ge t_1$, we have

$$\sum_{i\in\mathcal{N}} \Big(\bar{G}^n_{ij}(t_2) - \bar{G}^n_{ij}(t_1)\Big) \le \bar{E}^n_j \circ \bar{\Gamma}^n_j(t_2) - \bar{E}^n_j \circ \bar{\Gamma}^n_j(t_1), \tag{B.3}$$

$$\bar{D}^n_{ij}(t_2) - \bar{D}^n_{ij}(t_1) \le \bar{G}^n_{ij}(t_2) - \bar{G}^n_{ij}(t_1), \tag{B.4}$$

where (B.3) is by the fact that the system controller can offer a driver to a customer only at the customer arrival epochs, and (B.4) is by (4) and (A.1). Then, by Lemma F.1 (see Appendix F), $\{\bar{G}^n_{ij}, n \in \mathbb{N}_+\}$ and $\{\bar{D}^n_{ij}, n \in \mathbb{N}_+\}$ are relatively compact in $\mathbb{D}[0,T_1]$ endowed with the u.o.c. topology for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$ such that all of their subsequential limits are Lipschitz continuous. Moreover, because $\bar{G}^n_{ij}$ and $\bar{D}^n_{ij}$ are nondecreasing and $\bar{G}^n_{ij}(0) = \bar{D}^n_{ij}(0) = 0$ for all $i \in \mathcal{N}$, $j \in \mathcal{J}$, and $n \in \mathbb{N}_+$, any fluid limits of these processes satisfy (A.3b), (A.3c), and (A.3d). Furthermore, (A.3e) follows by (B.2) and (B.3).

By Assumption 1, let $\bar{\lambda}_i := \sup_{t\in\mathbb{R}_+} \lambda_i(t) < \infty$ and $\bar{\mu}_j := \sup_{t\in\mathbb{R}_+} \mu_j(t) < \infty$ for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$. Moreover, there exists a constant $\bar{\theta} \in \mathbb{R}_+$ and $n_0 \in \mathbb{N}_+$ such that if $n \ge n_0$, $\theta^n_i(t) \le \bar{\theta}$ for all $t \in \mathbb{R}_+$ and $i \in \mathcal{N}$ by Assumption 1. Let us fix an arbitrary

$i \in \mathcal{N}$ and consider the process $\int_0^{\cdot} \theta_i^n(s)\bar{Q}_i^n(s)ds$. By (B.2), $\bar{A}_i^n \circ \bar{\Lambda}_i^n(T_1) \to \Lambda_i(T_1) \le \bar{\lambda}_i T_1 < \infty$ and $\bar{Q}_i^n(0) \to \bar{Q}_i(0)$ as $n \to \infty$. Hence, there exists an $n_1 \in \mathbb{N}_+$ and a constant $C < \infty$ such that if $n \ge n_1$, $\bar{Q}_i^n(0) + \bar{A}_i^n \circ \bar{\Lambda}_i^n(T_1) \le C$. Then

$$\begin{aligned}\int_0^{t_2} \theta_i^n(s)\bar{Q}_i^n(s)ds - \int_0^{t_1} \theta_i^n(s)\bar{Q}_i^n(s)ds &= \int_{t_1}^{t_2} \theta_i^n(s)\bar{Q}_i^n(s)ds\\ &\le \bar{\theta}\int_{t_1}^{t_2} \bar{Q}_i^n(s)ds \le \bar{\theta}\int_{t_1}^{t_2} \big(\bar{Q}_i^n(0) + \bar{A}_i^n \circ \bar{\Lambda}_i^n(s)\big)ds\\ &\le \bar{\theta}\big(\bar{Q}_i^n(0) + \bar{A}_i^n \circ \bar{\Lambda}_i^n(T_1)\big)(t_2 - t_1) \le \bar{\theta}C(t_2 - t_1)\end{aligned} \tag{B.5}$$

for all $n \ge n_0 \vee n_1$ and $t_1, t_2 \in [0, T_1]$ such that $t_2 \ge t_1$, and the second inequality in (B.5) is by (11). Then, $\{\int_0^{\cdot} \theta_i^n(s)\bar{Q}_i^n(s)ds, n \in \mathbb{N}_+\}$ is relatively compact in the u.o.c. topology with Lipschitz continuous subsequential limits by Lemma F.1 (see Appendix F). By the fact that $\bar{R}_i^n \to e$ u.o.c. (see (B.2)) and lemma 11 of Ata and Kumar (2005), $\{\bar{R}_i^n(\int_0^{\cdot} \theta_i^n(s)\bar{Q}_i^n(s)ds), n \in \mathbb{N}_+\}$ is relatively compact in the u.o.c. topology, and all of its limits are Lipschitz continuous. Hence, $\{\bar{Q}_i^n, n \in \mathbb{N}_+\}$ is also relatively compact in the u.o.c. topology with Lipschitz continuous and nonnegative limits by (11) and (B.2). Hence, fluid limits of $\{\bar{Q}_i^n, n \in \mathbb{N}_+\}$ satisfy (A.3d).

Let us fix an arbitrary subsequence $\{n_l, l \in \mathbb{N}_+\}$ such that $\{\int_0^{\cdot} \theta_i^{n_l}(s)\bar{Q}_i^{n_l}(s)ds, l \in \mathbb{N}_+\}$ converges to a Lipschitz continuous limit under the uniform norm. Then, there exists a subsequence of $\{n_l, l \in \mathbb{N}_+\}$, denoted by $\{n_k, k \in \mathbb{N}_+\}$, such that $\|\bar{Q}_i^{n_k} - \bar{Q}_i\|_{T_1} \to 0$ as $k \to \infty$ for all $i \in \mathcal{N}$, where $\bar{Q}_i \in \mathbb{D}[0, T_1]$ is nonnegative and Lipschitz continuous for all $i \in \mathcal{N}$. Then, if $n_k \ge n_0$,

$$\begin{aligned}&\sup_{t\in[0,T_1]}\left|\int_0^t \theta_i^{n_k}(s)\bar{Q}_i^{n_k}(s)ds - \int_0^t \theta_i(s)\bar{Q}_i(s)ds\right|\\ &\quad\le \sup_{t\in[0,T_1]}\left|\int_0^t \theta_i^{n_k}(s)\bar{Q}_i^{n_k}(s)ds - \int_0^t \theta_i(s)\bar{Q}_i^{n_k}(s)ds\right| + \sup_{t\in[0,T_1]}\left|\int_0^t \theta_i(s)\bar{Q}_i^{n_k}(s)ds - \int_0^t \theta_i(s)\bar{Q}_i(s)ds\right|\\ &\quad\le T_1\|\bar{Q}_i^{n_k}\|_{T_1}\|\theta_i^{n_k} - \theta_i\|_{T_1} + T_1\bar{\theta}\|\bar{Q}_i^{n_k} - \bar{Q}_i\|_{T_1} \to 0, \quad \text{as } k \to \infty,\end{aligned}$$

by Assumption 1. Therefore, each convergent subsequence of the sequence $\{\int_0^{\cdot} \theta_i^n(s)\bar{Q}_i^n(s)ds, n \in \mathbb{N}_+\}$ converges to $\int_0^{\cdot} \theta_i(s)\bar{Q}_i(s)ds$ u.o.c. where $\bar{Q}_i$ is a fluid limit of $\{\bar{Q}_i^n, n \in \mathbb{N}_+\}$ for all $i \in \mathcal{N}$. Then, by lemma 11 of Ata and Kumar (2005), each convergent subsequence of $\{\bar{R}_i^n(\int_0^{\cdot} \theta_i^n(s)\bar{Q}_i^n(s)ds), n \in \mathbb{N}_+\}$ converges to $\int_0^{\cdot} \theta_i(s)\bar{Q}_i(s)ds$ u.o.c., where $\bar{Q}_i$ is a fluid limit of $\{\bar{Q}_i^n, n \in \mathbb{N}_+\}$ for all $i \in \mathcal{N}$. This proves (A.3a).

By theorems 16.2 and 16.4 of Billingsley (1999) and the fact that $T_1 \in \mathbb{R}_+$ is arbitrarily chosen, we can extend the results above to the processes in $\mathbb{D}^{2NJ+3N+J}$, which proves Proposition A.1 except (A.3f). Last, in order to prove (A.3f), it is enough to prove the following result. From this point forward, we keep π and ω in the notation.

Lemma B.1. *For all $i \in \mathcal{N}$, $j \in \mathcal{J}$, $m \in \mathbb{N}$, and admissible policy $\pi = \{\pi^n, n \in \mathbb{N}_+\}$, as $n \to \infty$,*

$$\sup_{t\in[\tau_m,\tau_{m+1})}\left|\bar{D}_{ij}^{\pi,n}(t) - \bar{D}_{ij}^{\pi,n}(\tau_m) - \bar{F}_{ij}(\tau_m)\Big(\bar{G}_{ij}^{\pi,n}(t) - \bar{G}_{ij}^{\pi,n}(\tau_m)\Big)\right| \xrightarrow{a.s.} 0. \tag{B.6}$$

Proof. Let us fix arbitrary $i \in \mathcal{N}$, $j \in \mathcal{J}$, $m \in \mathbb{N}$, and an admissible policy $\pi = \{\pi^n, n \in \mathbb{N}_+\}$. Then the converging number in (B.6) is equal to

$$\sup_{t\in[\tau_m,\tau_{m+1})}\left|\frac{1}{n}\sum_{k=E_j\circ\Gamma_j^n(\tau_m)+1}^{E_j\circ\Gamma_j^n(t)}\Big(\mathbb{I}\Big(a_j^k(m) \ge t_{ij}(m)\Big) - \bar{F}_{ij}(\tau_m)\Big)\mathbb{I}\Big(\pi_j^n(k) = i\Big)\right|, \tag{B.7}$$

by (4), (10b), and (A.1). For notational completeness, for any $n, m \in \mathbb{N}$ and sequence of real numbers $\{x_k, k \in \mathbb{N}\}$, if $n > m$, then $\sum_{k=n}^m x_k := 0$. Let

$$Y_k := \mathbb{I}\Big(a_j^k(m) \ge t_{ij}(m)\Big) - \bar{F}_{ij}(\tau_m),\ \ Z_k^n := \mathbb{I}\Big(\pi_j^n(k) = i\Big),\ \ X_k^n := Y_k \times Z_k^n.$$

Then, (B.7) is equal to

$$\sup_{t\in[\tau_m,\tau_{m+1})}\left|\frac{1}{n}\sum_{k=E_j\circ\Gamma_j^n(\tau_m)+1}^{E_j\circ\Gamma_j^n(t)} X_k^n\right|. \tag{B.8}$$

By the definition of the admissible policies (see Definition 1), $\pi_j^n(k)$ is $\mathscr{F}_j^n(k)$-measurable for all $k, n \in \mathbb{N}_+$, so is Z_k^n. Moreover, $a_j^k(m)$ is independent of $\mathscr{F}_j^n(k)$ by construction (see (8)), so does Y_k. Then, $\mathbf{E}\big[X_k^n | \mathscr{F}_j^n(k)\big] = 0$ for all $k, n \in \mathbb{N}_+$ because $\mathbf{E}[Y_k] = 0$. Therefore, we expect to have a Martingale strong law of large numbers result for triangular arrays:

$$\frac{1}{n} \sum_{k=E_j \circ \Gamma_j^n(\tau_m)+1}^{E_j \circ \Gamma_j^n(\tau_m)+n} X_k^n \xrightarrow{a.s.} 0, \qquad \text{as } n \to \infty. \tag{B.9}$$

We present the formal proof of (B.9) done by the technique introduced by de Jong (1996) in Section B.1.

Let us first choose an arbitrary $\epsilon > 0$ and then choose arbitrary $\epsilon_1 > 0$ and $\epsilon_2 > 0$ such that $\epsilon_1(\epsilon_2 + \bar{\mu}_j \tau_{m+1}) < \epsilon$. Let the set of $\omega \in \Omega$ that satisfy (B.9) be denoted by $\mathscr{A}_2^\pi$. Then $\mathbf{P}(\mathscr{A}_2^\pi) = 1$. Let us choose an arbitrary $\omega \in \mathscr{A}_1 \cap \mathscr{A}_2^\pi$, where $\mathscr{A}_1$ is defined as in (B.2). Then, there exists an $n_2(\epsilon_2, \omega) \in \mathbb{N}$ such that $|\bar{E}_j^n \circ \bar{\Gamma}_j^n(\tau_{m+1}, \omega) - \Gamma_j(\tau_{m+1})| < \epsilon_2$ for all $n \geq n_2(\epsilon_2, \omega)$ by (B.2). By the definition of Γ_j, (10a), (10b), and Assumption 1,

$$E_j^n \circ \Gamma_j^n(\tau_{m+1}, \omega)/n < \epsilon_2 + \bar{\mu}_j \tau_{m+1}, \qquad \forall n \geq n_2(\epsilon_2, \omega). \tag{B.10}$$

By (B.9), there exists an $n_1(\epsilon_1, \omega) \in \mathbb{N}$ such that

$$\frac{1}{n} \left| \sum_{k=E_j \circ \Gamma_j^n(\tau_m, \omega)+1}^{E_j \circ \Gamma_j^n(\tau_m, \omega)+n} X_k^n(\omega) \right| < \epsilon_1, \qquad \forall n \geq n_1(\epsilon_1, \omega). \tag{B.11}$$

Let us fix an arbitrary $t \in [\tau_m, \tau_{m+1})$. On the one hand, if $E_j \circ \Gamma_j^n(t, \omega) - E_j \circ \Gamma_j^n(\tau_m, \omega) \geq n_1(\epsilon_1, \omega)$, then

$$\frac{1}{n} \left| \sum_{k=E_j \circ \Gamma_j^n(\tau_m, \omega)+1}^{E_j \circ \Gamma_j^n(t, \omega)} X_k^n(\omega) \right| < \epsilon_1 \frac{E_j \circ \Gamma_j^n(t, \omega) - E_j \circ \Gamma_j^n(\tau_m, \omega)}{n} \leq \epsilon_1 \frac{E_j \circ \Gamma_j^n(\tau_{m+1}, \omega)}{n}, \tag{B.12}$$

by (B.11). On the other hand, if $E_j \circ \Gamma_j^n(t, \omega) - E_j \circ \Gamma_j^n(\tau_m, \omega) < n_1(\epsilon_1, \omega)$, then

$$\frac{1}{n} \left| \sum_{k=E_j \circ \Gamma_j^n(\tau_m, \omega)+1}^{E_j \circ \Gamma_j^n(t, \omega)} X_k^n(\omega) \right| \leq \frac{1}{n} \Big(E_j \circ \Gamma_j^n(t, \omega) - E_j \circ \Gamma_j^n(\tau_m, \omega) \Big) < \frac{n_1(\epsilon_1, \omega)}{n}. \tag{B.13}$$

Hence, for all $t \in [\tau_m, \tau_{m+1})$, if $n \geq n_2(\epsilon_2, \omega)$,

$$\frac{1}{n} \left| \sum_{k=E_j \circ \Gamma_j^n(\tau_m, \omega)+1}^{E_j \circ \Gamma_j^n(t, \omega)} X_k^n(\omega) \right| < \big(\epsilon_1 (\epsilon_2 + \bar{\mu}_j \tau_{m+1}) \big) \vee \frac{n_1(\epsilon_1, \omega)}{n} < \epsilon \vee \frac{n_1(\epsilon_1, \omega)}{n}, \tag{B.14}$$

by (B.10), (B.12), (B.13), and the fact that $\epsilon_1(\epsilon_2 + \bar{\mu}_j \tau_{m+1}) < \epsilon$. Let

$$n_0(\epsilon, \omega) := \max\{n_2(\epsilon_2, \omega), n_1(\epsilon_1, \omega)/\epsilon\}. \tag{B.15}$$

Then, for all $\epsilon > 0$, there exists an $n_0(\epsilon, \omega)$ such that if $n \geq n_0(\epsilon, \omega)$,

$$\sup_{t \in [\tau_m, \tau_{m+1})} \left| \frac{1}{n} \sum_{k=E_j \circ \Gamma_j^n(\tau_m, \omega)+1}^{E_j \circ \Gamma_j^n(t, \omega)} X_k^n(\omega) \right| < \epsilon, \tag{B.16}$$

by (B.14) and (B.15). Because (B.16) is true for all $\omega \in \mathscr{A}_1 \cap \mathscr{A}_2^\pi$, the proof is complete. □

We complete the proof of Proposition A.1 by defining the full set $\mathscr{A}^\pi := \mathscr{A}_1 \cap \mathscr{A}_2^\pi$.

B.1. Proof of (B.9)

Because the indices in the sum in (B.9) are random, we first prove that $Y_{E_j \circ \Gamma_j^n(\tau_m)+k} \perp Z^n_{E_j \circ \Gamma_j^n(\tau_m)+k}$ and $\mathbf{E}[X^n_{E_j \circ \Gamma_j^n(\tau_m)+k}] = \mathbf{E}[Y_{E_j \circ \Gamma_j^n(\tau_m)+k}] = 0$ for all $n \in \mathbb{N}_+$ and $k \in \mathbb{N}_+$, so that we can use a Martingale strong law of large numbers result for triangular arrays in order to prove (B.9).

For notational convenience, we omit the superscript π from the notation. Remember that we fix arbitrary $i \in \mathcal{N}$, $j \in \mathcal{J}$, and $m \in \mathbb{N}$ in Lemma B.1. Let $\tilde{K}_k^n := K^n_{E_j \circ \Gamma_j^n(\tau_m)+k}$ for all $K \in \{X, Z\}$ and $k \in \mathbb{N}_+$, and $\tilde{Y}_k := Y_{E_j \circ \Gamma_j^n(\tau_m)+k}$ for all $k \in \mathbb{N}_+$. Then, proving (B.9) is equivalent to proving

$$\frac{1}{n} \sum_{k=1}^{n} \tilde{X}_k^n \xrightarrow{a.s.} 0, \qquad \text{as } n \to \infty. \tag{B.17}$$

Published in Stochastic Systems on February 25, 2020 as DOI: 10.1287/stsy.2019.0037.
This article has not been copyedited or formatted. The final version may differ from this version.

Let $\tilde{\mathcal{F}}_j^n(k) := \mathcal{F}_j^n(E_j \circ \Gamma_j^n(\tau_m) + k)$ for all $k \in \mathbb{N}_+$ (see (8)) and $\tilde{\mathcal{F}}_j^n(k) := \{\emptyset, \Omega\}$ for all $k \in \mathbb{Z} \backslash \mathbb{N}_+$, where $\mathbb{Z}$ denotes the set of integers. Because $\nu_j^n(E_j \circ \Gamma_j^n(\tau_m) + k)$ is a stopping time with respect to the filtration $\mathbb{F}^n$ for all $k \in \mathbb{N}_+$ and is increasing in k, $\tilde{\mathcal{F}}_j^n(k)$ is well defined, and $\{\tilde{\mathcal{F}}_j^n(k), k \in \mathbb{Z}\}$ is a filtration.

Let $\mathbb{S}_k^n : \mathbb{R}_+ \times \Omega \to \mathbb{R}^{\mathbb{N}}$ be defined such that

$$\mathbb{S}_k^n(t,\omega) := \Bigg(A_i \circ \Lambda_i^n\Big(t \wedge \nu_j^n(k,\omega), \omega\Big), E_{j'} \circ \Gamma_{j'}^n\Big(t \wedge \nu_j^n(k,\omega), \omega\Big), D_{ij'}^n\Big(\Big(t \wedge \nu_j^n(k,\omega)\Big)-, \omega\Big), Q_i^n\Big(\Big(t \wedge \nu_j^n(k,\omega)\Big)-, \omega\Big),$$
$$R_i\left(\int_0^{(t\wedge \nu_j^n(k,\omega))-} \theta_i^n(u) Q_i^n(u,\omega) du,\ \omega\right),\ \forall i \in \mathcal{N}, j' \in \mathcal{J},\ a_{j'}^{r \wedge E_{j'} \circ \Gamma_{j'}^n\left(\nu_j^n(k,\omega)-,\omega\right)}(m,\omega),\ a_j^{r\wedge(k-1)}(m,\omega),\ \forall r, m \in \mathbb{N}\Bigg), \tag{B.18}$$

for all $k, n \in \mathbb{N}_+$, $t \in \mathbb{R}_+$, and $\omega \in \Omega$. By (8),

$$\mathcal{F}_j^n(k) = \sigma\{\mathbb{S}_k^n(t), t \in \mathbb{R}_+\}. \tag{B.19}$$

Let $(\mathbb{R}^{\mathbb{N}})^{\mathbb{N}} := \mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}} \times \ldots$. Then we have the following result.

Lemma B.2. *Fix arbitrary $k, n \in \mathbb{N}_+$ and $j \in \mathcal{J}$. Let g be an arbitrary $\mathcal{F}_j^n(k)$-measurable function. Then, there exist a $\mathcal{B}((\mathbb{R}^{\mathbb{N}})^{\mathbb{N}})$-measurable function f and a sequence of nonnegative real numbers $\{t_l, l \in \mathbb{N}\}$ such that $g(\omega) = f(\mathbb{S}_k^n(t_l,\omega), l \in \mathbb{N})$ for all $\omega \in \Omega$.*

Proof. The proof follows by exercise 1.5.6 of Stroock and Varadhan (2006) and the fact that $\mathbb{R}$, $\mathbb{R}^{\mathbb{N}}$, and $(\mathbb{R}^{\mathbb{N}})^{\mathbb{N}}$ are all Polish spaces (see Aliprantis and Border 2006, corollary 3.39). □

Lemma B.3. *We have $\pi_j^n(E_j \circ \Gamma_j^n(\tau_m) + k) \in \tilde{\mathcal{F}}_j^n(k)$ and $a_j^{E_j \circ \Gamma_j^n(\tau_m)+k}(m) \in \tilde{\mathcal{F}}_j^n(k+l)$ for all $j \in \mathcal{J}$, $k, l, n \in \mathbb{N}_+$, and $m \in \mathbb{N}$.*

Proof. We have $a_j^{E_j \circ \Gamma_j^n(\tau_m)+k}(m) \in \tilde{\mathcal{F}}_j^n(k+l)$ by definition of $\tilde{\mathcal{F}}_j^n(k+l)$ [see (8)]. Let us fix arbitrary $j \in \mathcal{J}$, $k, n \in \mathbb{N}_+$, and $m \in \mathbb{N}$. For notational convenience, let $\tilde{E} := E_j \circ \Gamma_j^n(\tau_m)$. Then

$$\begin{aligned}\pi_j^n\Big(E_j \circ \Gamma_j^n(\tau_m) + k\Big) &= \sum_{r=0}^{\infty} \pi_j^n(r+k)\mathbb{I}(\tilde{E} = r)\\ &= \sum_{r=0}^{\infty} f_r(\mathbb{S}_{r+k}^n(t_l^r), l \in \mathbb{N})\mathbb{I}(\tilde{E} = r) &\text{(B.20)}\\ &= \sum_{r=0}^{\infty} f_r(\mathbb{S}_{\tilde{E}+k}^n(t_l^r), l \in \mathbb{N})\mathbb{I}(\tilde{E} = r), &\text{(B.21)}\end{aligned}$$

where the equality in (B.20) is by Lemma B.2. First, $\mathbb{I}(\tilde{E} = r) \in \tilde{\mathcal{F}}_j^n(k)$ for all $r \in \mathbb{N}$ by definition of $\tilde{\mathcal{F}}_j^n(k)$. Second, let $y : \Omega \to (\mathbb{R}^{\mathbb{N}})^{\mathbb{N}}$ be such that $y(\omega) := (\mathbb{S}_{\tilde{E}(\omega)+k}^n(t_l^r, \omega), l \in \mathbb{N})$. Because $\mathcal{B}((\mathbb{R}^{\mathbb{N}})^{\mathbb{N}}) = \mathcal{B}(\mathbb{R}^{\mathbb{N}}) \otimes \mathcal{B}(\mathbb{R}^{\mathbb{N}}) \otimes \ldots$ (see Kallenberg 1997, lemma 1.2), $\mathcal{B}((\mathbb{R}^{\mathbb{N}})^{\mathbb{N}})$ is generated by the sets of type $B := \{\prod_{l\in\mathbb{N}} B_l : B_l \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})\}$ (see Folland 1999, proposition 1.3). Then, because $\tilde{\mathcal{F}}_j^n(k) = \sigma\{\mathbb{S}_{\tilde{E}+k}^n(t), t \in \mathbb{R}_+\}$ (see (B.19)), we have

$$\{\omega \in \Omega : y(\omega) \in B\} = \Big\{\omega \in \Omega : \mathbb{S}_{\tilde{E}(\omega)+k}^n(t_l^r, \omega) \in B_l, l \in \mathbb{N}\Big\} = \bigcap_{l\in\mathbb{N}} \Big\{\omega \in \Omega : \mathbb{S}_{\tilde{E}(\omega)+k}^n(t_l^r, \omega) \in B_l\Big\} \in \tilde{\mathcal{F}}_j^n(k).$$

Because $f_r \in \mathcal{B}((\mathbb{R}^{\mathbb{N}})^{\mathbb{N}})$ by definition, the mapping in (B.21) is $\tilde{\mathcal{F}}_j^n(k)$-measurable. □

Then, we have the following result.

Lemma B.4. *We have $a_j^{E_j \circ \Gamma_j^n(\tau_m)+k}(m) \perp \tilde{\mathcal{F}}_j^n(k)$ for all $j \in \mathcal{J}$, $k, n \in \mathbb{N}_+$, and $m \in \mathbb{N}$.*

Proof. For notational convenience, again let $\tilde{E} := E_j \circ \Gamma_j^n(\tau_m)$. Let us fix arbitrary $j \in \mathcal{J}$, $k, n \in \mathbb{N}_+$, $m \in \mathbb{N}$, and $c \in \mathbb{R}$ and an arbitrary set $B \in \tilde{\mathcal{F}}_j^n(k)$. Then

$$\begin{aligned}&\mathbf{P}\Big(a_j^{\tilde{E}+k}(m) < c,\ B\Big)\\ &\quad= \sum_{r=0}^{\infty} \mathbf{P}\Big(a_j^{r+k}(m) < c,\ B,\ \tilde{E} = r\Big)\\ &\quad= \sum_{r=0}^{\infty} \mathbf{E}\Big[\mathbb{I}\Big(a_j^{r+k}(m) < c\Big)\mathbb{I}(B)\mathbb{I}(\tilde{E} = r)\Big]\\ &\quad= \sum_{r=0}^{\infty} \mathbf{E}\Big[\ \mathbb{I}\Big(a_j^{r+k}(m) < c\Big)\ f(\mathbb{S}_{\tilde{E}+k}^n(t_l), l \in \mathbb{N})\mathbb{I}(\tilde{E} = r)\Big] &\text{(B.22)}\end{aligned}$$

$$
\begin{aligned}
&= \sum_{r=0}^{\infty} \mathbf{E}\Big[\,\mathbb{I}\Big(a_j^{r+k}(m) < c\Big) f\big(\mathbb{S}_{r+k}^n(t_l), l \in \mathbb{N}\big)\mathbb{I}\big(\tilde{E} = r\big)\Big] \\
&= \sum_{r=0}^{\infty} \mathbf{P}\Big(a_j^{r+k}(m) < c\Big) \mathbf{E}\big[f\big(\mathbb{S}_{r+k}^n(t_l), l \in \mathbb{N}\big)\mathbb{I}\big(\tilde{E} = r\big)\big] && \text{(B.23)} \\
&= \mathbf{P}\Big(a_j^1(m) < c\Big) \sum_{r=0}^{\infty} \mathbf{E}\big[f\big(\mathbb{S}_{r+k}^n(t_l), l \in \mathbb{N}\big)\mathbb{I}\big(\tilde{E} = r\big)\big] && \text{(B.24)} \\
&= \mathbf{P}\Big(a_j^1(m) < c\Big) \sum_{r=0}^{\infty} \mathbf{P}\big(B,\ \tilde{E} = r\big) = \mathbf{P}\Big(a_j^1(m) < c\Big)\mathbf{P}(B) \\
&= \sum_{r=0}^{\infty} \Big(\mathbf{P}\Big(a_j^1(m) < c\Big)\mathbf{P}\big(\tilde{E} = r\big)\Big)\mathbf{P}(B) \\
&= \sum_{r=0}^{\infty} \Big(\mathbf{P}\Big(a_j^{r+k}(m) < c\Big)\mathbf{P}\big(\tilde{E} = r\big)\Big)\mathbf{P}(B) \\
&= \sum_{r=0}^{\infty} \mathbf{P}\Big(a_j^{r+k}(m) < c,\ \tilde{E} = r\Big)\mathbf{P}(B) && \text{(B.25)} \\
&= \sum_{r=0}^{\infty} \mathbf{P}\Big(a_j^{\tilde{E}+k}(m) < c, \tilde{E} = r\Big)\mathbf{P}(B) \\
&= \mathbf{P}\Big(a_j^{\tilde{E}+k}(m) < c\Big)\mathbf{P}(B), && \text{(B.26)}
\end{aligned}
$$

where (B.22) is by Lemma B.2; (B.23) is by (B.19) and the fact that $a_j^{r+k}(m) \perp \mathcal{F}_j^n(r+k)$ (see (8)), $\mathbb{I}(\tilde{E} = r) \in \mathcal{F}_j^n(r+k)$ for all $r \in \mathbb{N}$, and $f(\mathbb{S}_{r+k}^n(t_l), l \in \mathbb{N}) \in \mathcal{F}_j^n(r+k)$ (see proof of Lemma B.3 for a similar argument explained in detail); (B.24) is by the i.i.d. property of the sequence $\{a_j^r(m), r \in \mathbb{N}\}$; and the equality in (B.25) is by the fact that $a_j^{r+k}(m) \perp \mathcal{F}_j^n(r+k)$ and $\mathbb{I}(\tilde{E} = r) \in \mathcal{F}_j^n(r+k)$ for all $r \in \mathbb{N}$. By definition of independence (see Durrett 2010, p. 41), the equality in (B.26) proves the desired result. □

Remark B.1. If the filtration defined in (7) is generated by the sequence of processes $\{\Upsilon_m, m \in \mathbb{N}\}$ defined in Remark 4, we can still prove Lemmas B.3 and B.4 by updating the definition in (B.18) such that the infinite dimensional process $\mathbb{S}_k^n$ includes the sequence $\{\Upsilon_m, m \in \mathbb{N}\}$.

Now we are ready to prove (B.9) (or, equivalently, (B.17)) by the technique introduced by de Jong (1996). For any $x \in \mathbb{R}$, let $\lfloor x \rfloor$ denote the greatest integer that is smaller than or equal to x. Then,

$$
\begin{aligned}
\frac{1}{n}\sum_{k=1}^{n} \tilde{X}_k^n &= \frac{1}{n}\sum_{k=1}^{n}\Big(\tilde{X}_k^n - \mathbf{E}\Big[\tilde{X}_k^n | \tilde{\mathcal{F}}_j^n\big(k + \lfloor n^{0.25}\rfloor - 1\big)\Big]\Big) \\
&\quad + \frac{1}{n}\sum_{k=1}^{n} \mathbf{E}\Big[\tilde{X}_k^n | \tilde{\mathcal{F}}_j^n\big(k - \lfloor n^{0.25}\rfloor\big)\Big] \\
&\quad + \frac{1}{n}\sum_{k=1}^{n}\Big(\mathbf{E}\Big[\tilde{X}_k^n | \tilde{\mathcal{F}}_j^n\big(k + \lfloor n^{0.25}\rfloor - 1\big)\Big] - \mathbf{E}\Big[\tilde{X}_k^n | \tilde{\mathcal{F}}_j^n\big(k - \lfloor n^{0.25}\rfloor\big)\Big]\Big) \\
&=: A_1^n + A_2^n + A_3^n. && \text{(B.27)}
\end{aligned}
$$

We will consider each of A_1^n, A_2^n, and A_3^n separately. Let us choose an arbitrary $\epsilon > 0$. First, let us consider

$$
\sum_{n=1}^{\infty} \mathbf{P}\big(|A_1^n| > \epsilon\big) \ = \ \sum_{n=1}^{\infty} \mathbf{P}\left(\left|\sum_{k=1}^{n}\Big(\tilde{X}_k^n - \mathbf{E}\Big[\tilde{X}_k^n | \tilde{\mathcal{F}}_j^n\big(k + \lfloor n^{0.25}\rfloor - 1\big)\Big]\Big)\right| > n\epsilon\right). \quad \text{(B.28)}
$$

Notice that $\tilde{Z}_k^n$ is $\tilde{\mathcal{F}}_j^n(k)$-measurable by Lemma B.3 and $\tilde{Y}_k \perp \tilde{\mathcal{F}}_j^n(k)$ by Lemma B.4. Moreover,

$$
\begin{aligned}
\mathbf{E}\big[\tilde{Y}_k\big] &= \mathbf{E}\Big[\mathbb{I}\Big(a_j^{E_j \circ \Gamma_j^n(\tau_m)+k}(m) \ge t_{ij}(m)\Big) - \bar{F}_{ij}(\tau_m)\Big] \\
&= \mathbf{E}\left[\sum_{r=0}^{\infty} \mathbb{I}\Big(a_j^{r+k}(m) \ge t_{ij}(m)\Big)\mathbb{I}\Big(E_j \circ \Gamma_j^n(\tau_m) = r\Big)\right] - \bar{F}_{ij}(\tau_m) \\
&= \sum_{r=0}^{\infty} \mathbf{E}\Big[\mathbb{I}\Big(a_j^{r+k}(m) \ge t_{ij}(m)\Big)\mathbb{I}\Big(E_j \circ \Gamma_j^n(\tau_m) = r\Big)\Big] - \bar{F}_{ij}(\tau_m) \\
&= \sum_{r=0}^{\infty} \mathbf{E}\Big[\mathbb{I}\Big(a_j^{r+k}(m) \ge t_{ij}(m)\Big)\Big]\mathbf{E}\Big[\mathbb{I}\Big(E_j \circ \Gamma_j^n(\tau_m) = r\Big)\Big] - \bar{F}_{ij}(\tau_m) && \text{(B.29)} \\
&= \sum_{r=0}^{\infty} \bar{F}_{ij}(\tau_m)\mathbf{E}\Big[\mathbb{I}\Big(E_j \circ \Gamma_j^n(\tau_m) = r\Big)\Big] - \bar{F}_{ij}(\tau_m) && \text{(B.30)}
\end{aligned}
$$

$$= \bar{F}_{ij}(\tau_m)\left(\sum_{r=0}^{\infty}\mathbf{E}\left[\mathbb{I}\left(E_j \circ \Gamma_j^n(\tau_m) = r\right)\right] - 1\right) = 0, \tag{B.31}$$

where (B.29) is by the fact that $a_j^{r+k}(m) \perp \mathcal{F}_j^n(r+k)$ (see (8)) and $\mathbb{I}(E_j \circ \Gamma_j^n(\tau_m) = r) \in \mathcal{F}_j^n(r+k)$ for all $k \in \mathbb{N}_+$ and $r \in \mathbb{N}$, and (B.30) is by definition of $\bar{F}_{ij}(\tau_m)$ (see (2)). Hence, on the one hand,

$$\mathbf{E}\left[\tilde{X}_k^n|\tilde{\mathcal{F}}_j^n(k)\right] = \mathbf{E}\left[\tilde{Y}_k\tilde{Z}_k^n|\tilde{\mathcal{F}}_j^n(k)\right] = \tilde{Z}_k^n\mathbf{E}\left[\tilde{Y}_k\right] = 0,\ \forall k, n \in \mathbb{N}_+, \tag{B.32a}$$

$$\mathbf{E}\left[\tilde{X}_k^n|\tilde{\mathcal{F}}_j^n(k-l)\right] = \mathbf{E}\left[\mathbf{E}\left[\tilde{X}_k^n|\tilde{\mathcal{F}}_j^n(k)\right]|\tilde{\mathcal{F}}_j^n(k-l)\right] = 0, \text{for all } k,\ l,\ n \in \mathbb{N}_+, \tag{B.32b}$$

by (B.31). On the other hand, $\tilde{X}_k^n$ is $\tilde{\mathcal{F}}_j^n(k+l)$-measurable for all $k \in \mathbb{N}_+$ and $l \in \mathbb{N}_+$ by Lemma B.3. Hence, for all $k, n \in \mathbb{N}_+$,

$$\mathbf{E}\left[\tilde{X}_k^n|\tilde{\mathcal{F}}_j^n\left(k + \lfloor n^{0.25}\rfloor - 1\right)\right] = \begin{cases} 0 & \text{if } n < 16, \\ \tilde{X}_k^n & n \geq 16. \end{cases} \tag{B.33}$$

Because $\left|\tilde{X}_k^n\right| \leq 1$ for all $k \in \mathbb{N}_+$ and $n \in \mathbb{N}_+$, the sum on the right-hand side in (B.28) is less than or equal to $15 < \infty$ by (B.33). Second, by (B.32), we have

$$\sum_{n=1}^{\infty}\mathbf{P}\left(|A_2^n| > \epsilon\right) = \sum_{n=1}^{\infty}\mathbf{P}\left(\left|\sum_{k=1}^{n}\mathbf{E}\left[\tilde{X}_k^n\middle|\tilde{\mathcal{F}}_j^n\left(k - \lfloor n^{0.25}\rfloor\right)\right]\right| > n\epsilon\right) = 0. \tag{B.34}$$

Third,

$$\begin{aligned}
\sum_{n=1}^{\infty}\mathbf{P}\left(|A_3^n| > \epsilon\right) &= \sum_{n=1}^{\infty}\mathbf{P}\left(\left|\sum_{k=1}^{n}\left(\mathbf{E}\left[\tilde{X}_k^n\middle|\tilde{\mathcal{F}}_j^n\left(k + \lfloor n^{0.25}\rfloor - 1\right)\right] - \mathbf{E}\left[\tilde{X}_k^n\middle|\tilde{\mathcal{F}}_j^n\left(k - \lfloor n^{0.25}\rfloor\right)\right]\right)\right| > n\epsilon\right) \\
&= \sum_{n=1}^{\infty}\mathbf{P}\left(\left|\sum_{k=1}^{n}\sum_{l=-\lfloor n^{0.25}\rfloor+1}^{\lfloor n^{0.25}\rfloor-1}\left(\mathbf{E}\left[\tilde{X}_k^n\middle|\tilde{\mathcal{F}}_j^n(k+l)\right] - \mathbf{E}\left[\tilde{X}_k^n\middle|\tilde{\mathcal{F}}_j^n(k+l-1)\right]\right)\right| > n\epsilon\right) \\
&\leq \sum_{n=1}^{\infty}\mathbf{P}\left(\sum_{l=-\lfloor n^{0.25}\rfloor+1}^{\lfloor n^{0.25}\rfloor-1}\left|\sum_{k=1}^{n}\left(\mathbf{E}\left[\tilde{X}_k^n\middle|\tilde{\mathcal{F}}_j^n(k+l)\right] - \mathbf{E}\left[\tilde{X}_k^n\middle|\tilde{\mathcal{F}}_j^n(k+l-1)\right]\right)\right| > n\epsilon\right) \\
&\leq \sum_{n=1}^{\infty}\sum_{l=-\lfloor n^{0.25}\rfloor+1}^{\lfloor n^{0.25}\rfloor-1}\mathbf{P}\left(\left|\sum_{k=1}^{n}\left(\mathbf{E}\left[\tilde{X}_k^n\middle|\tilde{\mathcal{F}}_j^n(k+l)\right] - \mathbf{E}\left[\tilde{X}_k^n\middle|\tilde{\mathcal{F}}_j^n(k+l-1)\right]\right)\right| > \frac{n\epsilon}{2\lfloor n^{0.25}\rfloor}\right).
\end{aligned} \tag{B.35}$$

For all $r \in \mathbb{N}_+$, let

$$M_l^n(r) := \sum_{k=1}^{r}\left(\mathbf{E}\left[\tilde{X}_k^n\middle|\tilde{\mathcal{F}}_j^n(k+l)\right] - \mathbf{E}\left[\tilde{X}_k^n\middle|\tilde{\mathcal{F}}_j^n(k+l-1)\right]\right).$$

First, $\mathbf{E}\left[|M_l^n(r)|\right] \leq 2r < \infty$ and $M_l^n(r) \in \tilde{\mathcal{F}}_j^n(l+r)$ [even when $l + r \leq 0$ in which $M_l^n(r) = 0$ because $\tilde{\mathcal{F}}_j^n(k) = \{\emptyset, \Omega\}$ for $k \in \mathbb{Z}\backslash\mathbb{N}_+$] for all $r \in \mathbb{N}_+$. Second,

$$\begin{aligned}
&\mathbf{E}\left[M_l^n(r+1)|\tilde{\mathcal{F}}_j^n(l+r)\right] \\
&= \mathbf{E}\left[\mathbf{E}\left[\tilde{X}_{r+1}^n|\tilde{\mathcal{F}}_j^n(l+r+1)\right] - \mathbf{E}\left[\tilde{X}_{r+1}^n|\tilde{\mathcal{F}}_j^n(l+r)\right] + M_l^n(r)|\tilde{\mathcal{F}}_j^n(l+r)\right] \\
&= 0 + M_l^n(r).
\end{aligned}$$

Therefore, for all fixed $n \in \mathbb{N}_+$ and $l \in \mathbb{Z}$, $\{M_l^n(r), \tilde{\mathcal{F}}_j^n(l+r), r \in \mathbb{N}_+\}$ is a martingale sequence such that $\mathbf{E}[M_l^n(r)] = 0$. Moreover, $\left|M_l^n(r) - M_l^n(r-1)\right| \leq 2$ for all $r \in \mathbb{N}_+$; that is, the martingale differences are bounded by 2. Then, the sum in (B.35) is equal to

$$\begin{aligned}
&\sum_{n=1}^{\infty}\sum_{l=-\lfloor n^{0.25}\rfloor+1}^{\lfloor n^{0.25}\rfloor-1}\mathbf{P}\left(\left|M_l^n(n)\right| > \frac{n\epsilon}{2\lfloor n^{0.25}\rfloor}\right) \\
&\quad\leq \sum_{n=1}^{\infty}\sum_{l=-\lfloor n^{0.25}\rfloor+1}^{\lfloor n^{0.25}\rfloor-1}2\exp\left(-2\left(\frac{n\epsilon}{2\lfloor n^{0.25}\rfloor}\right)^2\left(\sum_{k=1}^{n}4^2\right)^{-1}\right) \\
&\quad= \sum_{n=1}^{\infty}2\left(2\lfloor n^{0.25}\rfloor - 1\right)\exp\left(-\frac{n\epsilon^2}{32\lfloor n\rfloor^{0.5}}\right) < \infty,
\end{aligned} \tag{B.36}$$

where the inequality is by Azuma's inequality (see Ross 1996, theorem 6.3.3).

Finally, (B.17) follows by the fact that $\epsilon > 0$ is arbitrary and

$$\mathbf{P}\left(\limsup\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\tilde{X}_k^n\right| > \epsilon\right\}\right) = \mathbf{P}\left(\limsup\left\{\left|A_1^n + A_2^n + A_3^n\right| > \epsilon\right\}\right) = 0,$$

where the second equality is by the fact that the sum in (B.28) is finite, (B.34), (B.36), and the Borel–Cantelli lemma (see Durrett 2010, theorem 2.3.1).

Appendix C. A Regulator Mapping Result

We need regulator mapping results in order to prove Theorems 2, 3, and 4 in Appendix D. Because the generalized one-sided regulator mappings defined in the literature (see Reed and Ward 2004, 2008; Ward and Kumar 2008) are not applicable to our case, we introduce a new one-sided and nonlinear regulator mapping in this section. Let $\phi, \psi : \mathbb{D} \to \mathbb{D}$ be such that for all $x \in \mathbb{D}$ and $t \in \mathbb{R}_+$,

$$\psi(x)(t) := \sup_{0\le s\le t}(-x(s))^+, \qquad \phi(x)(t) := x(t) + \psi(x)(t). \tag{C.1}$$

Then, ϕ is the conventional one-sided and one-dimensional regulator mapping (see Whitt 2002, chapter 13.5). We define the following regulator mapping.

Definition C.1. (A Time-Dependent, One-Sided, and Nonlinear Regulator Mapping). Let $x, y \in \mathbb{D}$ be such that $x(0) \ge 0$ and $\sup_{t\in\mathbb{R}_+} |y(t)| < \infty$. The time-dependent, one-sided, and nonlinear regulator mapping $(\phi^{\mathcal{M}}, \psi^{\mathcal{M}}) : \mathbb{D}^2 \to \mathbb{D}^2$ is defined by $(\phi^{\mathcal{M}}, \psi^{\mathcal{M}})(x, y) = (z, \ell)$, where

C1. $z(t) = x(t) - \int_0^t y(s)z(s)\mathrm{d}s + \ell(t) \ge 0$ for all $t \in \mathbb{R}_+$;
C2. $\ell(0) = 0$, ℓ is nondecreasing, and $\int_0^\infty z(t)\mathrm{d}\ell(t) = 0$.

If $y = \mathbf{0}$ in Definition C.1, then the time-dependent, one-sided, and nonlinear regulator mapping becomes the conventional one-sided and one-dimensional regulator mapping defined in (C.1) (see Chen and Yao 2001, theorem 6.1); if $y \in \mathbb{D}$ is a constant function, then it becomes the linearly generalized one-sided mapping described in Ward and Kumar (2008).

In order to write the time-dependent, one-sided, and nonlinear regulator mapping in terms of the conventional one-sided and one-dimensional regulator mapping defined in (C.1), we make the following definition.

Definition C.2 (Integral Equation). Let $x, y \in \mathbb{D}$ be such that $\sup_{t\in\mathbb{R}_+} |y(t)| < \infty$. Let $\mathcal{M} : \mathbb{D}^2 \to \mathbb{D}$ be a mapping such that $u := \mathcal{M}(x, y)$ solves the integral equation

$$u(t) = x(t) - \int_0^t y(s)\phi(u)(s)\mathrm{d}s, \quad \forall t \in \mathbb{R}_+. \tag{C.2}$$

A mapping $g : \mathbb{D} \to \mathbb{D}$ is Lipschitz continuous with respect to the uniform norm if for all $T_1 \in \mathbb{R}_+$, there exists a constant $\kappa \in \mathbb{R}_+$ that may depend on T_1 such that $\|g(x) - g(y)\|_{T_1} \le \kappa\|x - y\|_{T_1}$ for all $x, y \in \mathbb{D}$. Then, the first main result of this section is the following.

Proposition C.1. *For any given $x, y \in \mathbb{D}$ such that $x(0) \ge 0$ and $\sup_{t\in\mathbb{R}_+} |y(t)| < \infty$, there exists a unique pair of functions $(\phi^{\mathcal{M}}, \psi^{\mathcal{M}})(x, y)$ that satisfies Conditions* C1 *and* C2 *in Definition* C.1. *Moreover,*

1. *$\phi^{\mathcal{M}}(x, y) = \phi(\mathcal{M}(x, y))$ and $\psi^{\mathcal{M}}(x, y) = \psi(\mathcal{M}(x, y))$;*
2. *both $\phi^{\mathcal{M}}(\cdot, y) : \mathbb{D} \to \mathbb{D}$ and $\psi^{\mathcal{M}}(\cdot, y) : \mathbb{D} \to \mathbb{D}$ are Lipschitz continuous with respect to the uniform norm.*

The following lemma will be useful in the proof of Proposition C.1.

Lemma C.1. *Suppose that $x, y \in \mathbb{D}$ such that $\sup_{t\in\mathbb{R}_+} |y(t)| < \infty$. Then we have the following:*

1. *There exists a unique $u \in \mathbb{D}$ that solves the integral equation* (C.2). *Moreover, the mapping $\mathcal{M}(\cdot, y) : \mathbb{D} \to \mathbb{D}$ is Lipschitz continuous with respect to the uniform norm.*
2. *If x is nondecreasing and absolutely continuous and $x(0) \ge 0$, then $\mathcal{M}(x, y) \ge \mathbf{0}$.*
3. *Let x_1 and x_2 be two absolutely continuous functions such that $x_1 \ge x_2$ and $x_1' \ge x_2'$, where $'$ denotes the derivative. If $\mathcal{M}(x_2, y) \ge \mathbf{0}$, then $\mathcal{M}(x_1, y) \ge \mathcal{M}(x_2, y)$.*

Proof.

Part 1. Fix arbitrary $x, y \in \mathbb{D}$ such that $\bar{y} := \sup_{t\in\mathbb{R}_+} |y(t)| < \infty$. Let $\eta_y : \mathbb{D} \to \mathbb{D}$ be such that for all $u \in \mathbb{D}$ and $t \in \mathbb{R}_+$, $\eta_y(u)(t) := y(t)\phi(u)(t)$. Then, for all $u_1, u_2 \in \mathbb{D}$ and $t \in \mathbb{R}_+$,

$$\left\|\eta_y(u_1) - \eta_y(u_2)\right\|_t \le \bar{y}\left\|\phi(u_1) - \phi(u_2)\right\|_t \le 2\bar{y}\|u_1 - u_2\|_t,$$

where the last inequality is by the fact that the mapping ϕ is Lipschitz continuous with respect to the uniform norm with Lipschitz constant 2 (see Whitt 2002, lemma 13.5.1). Thus, η_y is Lipschitz continuous with respect to the uniform norm. Because (C.2) is equivalent to

$$u(t) = x(t) - \int_0^t \eta_y(u)(s)\mathrm{d}s, \quad \forall t \in \mathbb{R}_+,$$

there exists a unique $u \in \mathbb{D}$ that solves (C.2), and the mapping $\mathcal{M}(\cdot, y) : \mathbb{D} \to \mathbb{D}$ is Lipschitz continuous with respect to the uniform norm by lemma 1 of Reed and Ward (2004).

Part 2. Because x is nondecreasing, $x' \geq \mathbf{0}$. Consider

$$u(t) = x(t) - \int_0^t y(s)u(s)\mathrm{d}s, \qquad \forall t \in \mathbb{R}_+. \tag{C.3}$$

By algebra, one can see that

$$u(t) := \frac{x(0) + \int_0^t x'(s)\exp\left\{\int_0^s y(r)\mathrm{d}r\right\}\mathrm{d}s}{\exp\left\{\int_0^t y(s)\mathrm{d}s\right\}}, \quad \forall t \in \mathbb{R}_+,$$

is a solution of the equality (C.3). By (C.1) and the fact that $u \geq \mathbf{0}$, we have $\phi(u) = u$. Moreover, by the fact that $\phi(u) = u$ and (C.3), u is the unique solution of the integral equation

$$u(t) = x(t) - \int_0^t y(s)\phi(u)(s)\mathrm{d}s, \qquad \forall t \in \mathbb{R}_+.$$

Hence, $\mathcal{M}(x, y) = u \geq \mathbf{0}$.

Part 3. Let $u_i := \mathcal{M}(x_i, y)$ for all $i \in \{1, 2\}$, $\bar{x} := x_1 - x_2$, and $\bar{x}' := x_1' - x_2'$. Consider the equation

$$v(t) = \bar{x}(t) - \int_0^t y(s)v(s)\mathrm{d}s, \qquad \forall t \in \mathbb{R}_+. \tag{C.4}$$

By algebra, one can see that

$$v(t) := \frac{\bar{x}(0) + \int_0^t \bar{x}'(s)\exp\left\{\int_0^s y(r)\mathrm{d}r\right\}\mathrm{d}s}{\exp\left\{\int_0^t y(s)\mathrm{d}s\right\}}, \quad \forall t \in \mathbb{R}_+,$$

is a solution of the equality (C.4). Because $\bar{x} \geq \mathbf{0}$ and $\bar{x}' \geq \mathbf{0}$, we have $v \geq \mathbf{0}$. If $u_2 \geq \mathbf{0}$, we have $\phi(u_2) = u_2$, and thus

$$v(t) + u_2(t) = x_1(t) - \int_0^t y(s)(v(s) + u_2(s))\mathrm{d}s, \qquad \forall t \in \mathbb{R}_+. \tag{C.5}$$

Because $v + u_2$ solves (C.5) and $v + u_2 \geq \mathbf{0}$, we have $\phi(v + u_2) = v + u_2$. Therefore, $u_1 = \mathcal{M}(x_1, y) = v + u_2$ by part 1 and (C.5), and so $u_1 - u_2 = v \geq \mathbf{0}$. □

Proof of Proposition C.1. Let us fix an arbitrary pair $x, y \in \mathbb{D}$ such that $x(0) \geq 0$ and $\sup_{t \in \mathbb{R}_+} |y(t)| < \infty$. Let $u := \mathcal{M}(x, y)$ (notice that such a $u \in \mathbb{D}$ uniquely exists by Lemma C.1, part 1). Moreover, let $(z, \ell) := (\phi, \psi)(u)$. Then, by (C.1), (C.2), and the fact that $x(0) = u(0) \geq 0$,

$$\begin{aligned} z(t) = u(t) + \ell(t) &= x(t) - \int_0^t y(s)\phi(u)(s)\mathrm{d}s + \ell(t) \\ &= x(t) - \int_0^t y(s)z(s)\mathrm{d}s + \ell(t) \geq 0, \quad \forall t \in \mathbb{R}_+, \end{aligned}$$

and thus Condition C1 in Definition C.1 is satisfied by (z, ℓ). Because $u(0) = x(0) \geq 0$, then $\ell(0) = \psi(u)(0) = 0$, and $\ell(\cdot)$ is nondecreasing by the definition of the mapping ψ (see (C.1)). Last,

$$\int_0^\infty z(t)\mathrm{d}\ell(t) = \int_0^\infty \phi(u)(t)\mathrm{d}\psi(u)(t) = 0$$

by definition of the conventional one-sided, one-dimensional regulator mapping (see Chen and Yao 2001, theorem 6.1). Therefore, the pair $(z, \ell) = (\phi, \psi)(u)$ satisfies Conditions C1 and C2 in Definition C.1.

Next, we will prove uniqueness. Let (z_1, ℓ_1) be another pair that satisfies Conditions C1 and C2 and $g \in \mathbb{D}$ be such that

$$g(t) := x(t) - \int_0^t y(s)z_1(s)\mathrm{d}s, \qquad \forall t \in \mathbb{R}_+.$$

Then, $z_1(t) = g(t) + \ell_1(t)$ for all $t \in \mathbb{R}_+$. By Condition C2 and the uniqueness of the Skorokhod mapping (see Chen and Yao 2001, theorem 6.1), $(z_1, \ell_1) = (\phi, \psi)(g)$, so

$$g(t) = x(t) - \int_0^t y(s)\phi(g)(s)\mathrm{d}s, \qquad \forall t \in \mathbb{R}_+. \tag{C.6}$$

By Lemma C.1, part 1, there exists a unique solution of (C.6) that is $\mathcal{M}(x,y) = u = g$. Therefore, $(z_1, \ell_1) = (\phi, \psi)(g) = (\phi, \psi)(u) = (z, \ell)$, which proves uniqueness. Moreover, $(\phi^{\mathcal{M}}, \psi^{\mathcal{M}})(x,y) = (z,\ell) = (\phi,\psi)(u) = (\phi,\psi)(\mathcal{M}(x,y))$.

Next, let us consider arbitrary $x_1, x_2 \in \mathbb{D}$ such that $x_1(0) \geq 0$ and $x_2(0) \geq 0$. Let $u_1 := \mathcal{M}(x_1, y)$ and $u_2 := \mathcal{M}(x_2, y)$. By Lemma C.1, part 1, the mapping $\mathcal{M}(\cdot, y) : \mathbb{D} \to \mathbb{D}$ is Lipschitz continuous with respect to the uniform norm. Let $\kappa_y(t)$ be the corresponding Lipschitz constant for $t \in \mathbb{R}_+$. Then, for all $t \in \mathbb{R}_+$,

$$\left\|\phi^{\mathcal{M}}(x_1, y) - \phi^{\mathcal{M}}(x_2, y)\right\|_t = \left\|\phi(u_1) - \phi(u_2)\right\|_t \leq 2\|u_1 - u_2\|_t \leq 2\kappa_y(t)\|x_1 - x_2\|_t,$$

where the first inequality is by the fact that the mapping ϕ is Lipschitz continuous with respect to the uniform norm with Lipschitz constant 2 (see Whitt 2002, lemma 13.5.1). Hence, $\phi^{\mathcal{M}}(\cdot, y)$ is Lipschitz continuous with respect to the uniform norm. Last, for all $t \in \mathbb{R}_+$,

$$\left\|\psi^{\mathcal{M}}(x_1, y) - \psi^{\mathcal{M}}(x_2, y)\right\|_t = \left\|\psi(u_1) - \psi(u_2)\right\|_t \leq \|u_1 - u_2\|_t \leq \kappa_y(t)\|x_1 - x_2\|_t,$$

where the first inequality is by the fact that the mapping ψ is Lipschitz continuous with respect to the uniform norm with Lipschitz constant 1 (see Whitt 2002, lemma 13.4.1). Hence, $\psi^{\mathcal{M}}(\cdot, y)$ is also Lipschitz continuous with respect to the uniform norm. □

Next, we present the following preliminary result.

Lemma C.2. *Let $T_1 \in \mathbb{R}_+$ be an arbitrary constant, $x, y \in \mathbb{D}$ be such that $y \geq \mathbf{0}$, and $\sup_{t\in\mathbb{R}_+} y(t) \leq K$ for some $K \in \mathbb{R}_+$. Then, there exists a constant $C = C(K, T_1) \in \mathbb{R}_+$ such that $\|\mathcal{M}(x,y)\|_{T_1} \leq C\|x\|_{T_1}$.*

Proof. Let $u := \mathcal{M}(x,y)$ be the unique solution of the integral equation (C.2). Notice that $\phi(u) \geq \mathbf{0}$ by (C.1). Then, by (C.2) and the fact that $y \geq \mathbf{0}$ and $\phi(u) \geq \mathbf{0}$,

$$u(t) \leq \|x\|_{T_1}, \qquad \forall t \in [0, T_1]. \tag{C.7}$$

Let $\lceil a \rceil$ denote the smallest integer that is greater than or equal to a for all $a \in \mathbb{R}$. Let us partition the interval $[0, T_1]$ into subintervals with length $1/(4K)$ except the last interval, which is $[(\lceil 4KT_1 \rceil - 1)/(4K), T_1]$. Then there are $\lceil 4KT_1 \rceil$ subintervals. Let

$$\begin{aligned} f_n &:= \sup\left\{|u(t)| : t \in \left[\frac{n-1}{4K}, \frac{n}{4K}\right]\right\}, \quad \forall n \in \{1, \ldots, \lceil 4KT_1 \rceil - 1\}, \\ f_{\lceil 4KT_1 \rceil} &:= \sup\left\{|u(t)| : t \in \left[\frac{\lceil 4KT_1 \rceil - 1}{4K}, T_1\right]\right\}. \end{aligned} \tag{C.8}$$

Then,

$$\|\mathcal{M}(x,y)\|_{T_1} = \|u\|_{T_1} = \max_{n\in\{1,\ldots,\lceil 4KT_1 \rceil\}} f_n.$$

By (C.2),

$$-u(t) = -x(t) + \int_0^t y(s)\phi(u)(s)\mathrm{d}s, \quad \forall t \in \mathbb{R}_+. \tag{C.9}$$

By (C.9) and the fact that the mapping ϕ is Lipschitz continuous with respect to the uniform norm with Lipschitz constant 2 (see Whitt 2002, lemma 13.5.1),

$$-u(t) \leq \|x\|_{T_1} + \frac{1}{4K}2K\|u\|_{1/(4K)}, \quad \forall t \in \left[0, \frac{1}{4K}\right] \quad \Rightarrow \quad f_1 \leq 2\|x\|_{T_1}, \tag{C.10}$$

where the second inequality above is by (C.7) and (C.8). By (C.9), we have, for all $t \in [1/4K, 1/2K]$,

$$\begin{aligned} -u(t) &= -x(t) + x\left(\frac{1}{4K}\right) - u\left(\frac{1}{4K}\right) + \int_{1/(4K)}^t y(s)\phi(u)(s)\mathrm{d}s, \\ &\Rightarrow -u(t) \leq 2\|x\|_{T_1} + f_1 + \frac{1}{4K}2K\|u\|_{1/(2K)}, \quad \forall t \in \left[\frac{1}{4K}, \frac{1}{2K}\right] \\ &\Rightarrow -u(t) \leq 2\|x\|_{T_1} + f_1 + \frac{1}{2}(f_1 + f_2), \quad \forall t \in \left[\frac{1}{4K}, \frac{1}{2K}\right] \\ &\Rightarrow f_2 \leq 4\|x\|_{T_1} + 3f_1, \end{aligned}$$

where the last inequality follows by (C.7) and (C.8). By induction, we can show that

$$f_n \leq 4\|x\|_{T_1} + 2f_{n-1} + \sum_{k=1}^{n-1} f_k, \quad \forall n \in \{2, \ldots, \lceil 4KT_1 \rceil\}. \tag{C.11}$$

By (C.10) and (C.11), one can show that

$$\|u\|_{T_1} = \max_{n\in\{1,\dots,\lceil 4KT_1\rceil\}} f_n \le (\lceil 4KT_1\rceil + 4)!\ \|x\|_{T_1}.$$

We complete the proof by defining $C(K,T_1) := (\lceil 4KT_1\rceil + 4)!$. □

The second main result of this section is the following.

Lemma C.3. *Let $\{x_n, n\in\mathbb{N}\}$ and $\{y_n, n\in\mathbb{N}\}$ be sequences in $\mathbb{D}$ such that $x_n(0)\ge 0$ and $y_n \ge \mathbf{0}$ for all $n\in\mathbb{N}$, there exists a constant $K\in\mathbb{R}_+$ such that $y_n(t)\le K$ for all $n\in\mathbb{N}$ and $t\in\mathbb{R}_+$, and there exist $x\in\mathbb{D}$ and $y\in\mathbb{D}$ such that $x_n\to x$ u.o.c. and $y_n\to y$ u.o.c. as $n\to\infty$. Then, $(\phi^{\mathcal{M}},\psi^{\mathcal{M}})(x_n,y_n)\to(\phi^{\mathcal{M}},\psi^{\mathcal{M}})(x,y)$ u.o.c. as $n\to\infty$.*

Proof. For each $n\in\mathbb{N}$, let $u_n\in\mathbb{D}$ be the unique solution of the equation

$$u_n(t) = x_n(t) - \int_0^t y_n(s)\phi(u_n)(s)\mathrm{d}s, \quad \forall t\in\mathbb{R}_+, \tag{C.12}$$

which exists by Lemma C.1, part 1. For each $n\in\mathbb{N}$, let us define

$$f_n(t) := \int_0^t y_n(s)\phi(u_n)(s)\mathrm{d}s, \quad \forall t\in\mathbb{R}_+. \tag{C.13}$$

Notice that f_n is continuous for all $n\in\mathbb{N}$ and $u_n = x_n - f_n$ for all $n\in\mathbb{N}$ by (C.12) and (C.13). Let T_1 be an arbitrary constant in $\mathbb{R}_+$. We will show that the sequence $\{f_n, n\in\mathbb{N}\}$ restricted to the compact domain $[0,T_1]$ is relatively compact by the Arzelà–Ascoli Theorem (see Billingsley 1999, theorem 7.2). First, $\sup_{n\in\mathbb{N}}|f_n(0)| = 0 < \infty$. Second,

$$\begin{aligned}
&\lim_{\delta\to 0}\sup_{n\in\mathbb{N}}\sup_{|t_1-t_2|\le\delta}\left|f_n(t_2)-f_n(t_1)\right| \\
&\quad= \lim_{\delta\to 0}\sup_{n\in\mathbb{N}}\sup_{|t_1-t_2|\le\delta}\left|\int_{t_1}^{t_2} y_n(s)\phi(u_n)(s)\mathrm{d}s\right| \\
&\quad\le \lim_{\delta\to 0}\delta\left(2\sup_{n\in\mathbb{N}}\|y_n\|_{T_1}\sup_{n\in\mathbb{N}}\|u_n\|_{T_1}\right) \le \lim_{\delta\to 0}\delta\left(2K\sup_{n\in\mathbb{N}}\|u_n\|_{T_1}\right), &\text{(C.14)}\\
&\quad\le \lim_{\delta\to 0}\delta\left(2KC(K,T_1)\sup_{n\in\mathbb{N}}\|x_n\|_{T_1}\right) = 0, &\text{(C.15)}
\end{aligned}$$

where the first inequality in (C.14) is by the fact that the mapping ϕ is Lipschitz continuous with respect to the uniform norm with Lipschitz constant 2 (see Whitt 2002, lemma 13.5.1). The inequality in (C.15) is by Lemma C.2. Notice that $x_n\to x$ u.o.c. where x is a bounded function in $[0,T_1]$; thus there exists a sufficiently large $n_0\in\mathbb{N}$ such that $\sup_{n\ge n_0}\|x_n\|_{T_1}<\infty$, and without loss of generality, we assume $\sup_{n\in\mathbb{N}}\|x_n\|_{T_1}<\infty$. Hence, we obtain the convergence result in (C.15). Therefore, $\{f_n, n\in\mathbb{N}\}$ restricted to the compact domain $[0,T_1]$ is relatively compact.

Because both $\{x_n, n\in\mathbb{N}\}$ and $\{f_n, n\in\mathbb{N}\}$ are relatively compact in $\mathbb{D}[0,T_1]$ endowed with the u.o.c. topology, so is $\{u_n, n\in\mathbb{N}\}$ by (C.12) and (C.13). Let us consider an arbitrary subsequence of $\{u_n, n\in\mathbb{N}\}$, denoted by $\{u_{n_k}, k\in\mathbb{N}\}$, such that $u_{n_k}\to u$ u.o.c. as $k\to\infty$, where $u\in\mathbb{D}[0,T_1]$. Then,

$$\begin{aligned}
&\sup_{t\in[0,T_1]}\left|u(t)-x(t)+\int_0^t y(s)\phi(u)(s)\mathrm{d}s\right| &\text{(C.16)}\\
&\quad= \sup_{t\in[0,T_1]}\left|u(t)-u_{n_k}(t)+u_{n_k}(t)-x(t)+\int_0^t y(s)\phi(u)(s)\mathrm{d}s\right| \\
&\quad\le \left\|u-u_{n_k}\right\|_{T_1}+\left\|x_{n_k}-x\right\|_{T_1}+\sup_{t\in[0,T_1]}\left|\int_0^t \left(y(s)\phi(u)(s)-y_{n_k}(s)\phi(u_{n_k})(s)\right)\mathrm{d}s\right| \\
&\quad\le \left\|u-u_{n_k}\right\|_{T_1}+\left\|x_{n_k}-x\right\|_{T_1}+\sup_{t\in[0,T_1]}\left|\int_0^t \left(y(s)\phi(u)(s)-y_{n_k}(s)\phi(u)(s)\right)\mathrm{d}s\right|+\sup_{t\in[0,T_1]}\left|\int_0^t \left(y_{n_k}(s)\phi(u)(s)-y_{n_k}(s)\phi(u_{n_k})(s)\right)\mathrm{d}s\right| \\
&\quad\le \left\|u-u_{n_k}\right\|_{T_1}+\left\|x_{n_k}-x\right\|_{T_1}+2T_1\|u\|_{T_1}\left\|y_{n_k}-y\right\|_{T_1}+2T_1K\left\|u_{n_k}-u\right\|_{T_1}. &\text{(C.17)}
\end{aligned}$$

As $k\to\infty$, all of the terms in (C.17) converge to 0, so (C.16) is equal to 0. Thus, $u=\mathcal{M}(x,y)$, and u is the unique solution of (C.2) by Lemma C.1, part 1. Therefore, each subsequence of $\{u_n, n\in\mathbb{N}\}$ has a convergent subsequence that converges to the same limit, which implies that $u_n \to u$ u.o.c. as $n \to \infty$, where $u = \mathcal{M}(x, y)$. Notice that the process $(\phi^{\mathcal{M}}, \psi^{\mathcal{M}})(x_n, y_n) \in \mathbb{D}^2$ is well defined for all $n \in \mathbb{N}$ by Proposition C.1. Because the mappings ϕ and ψ are Lipschitz continuous with respect to the uniform norm,

$$(\phi^{\mathcal{M}}, \psi^{\mathcal{M}})(x_n, y_n) = (\phi, \psi)(\mathcal{M}(x_n, y_n)) = (\phi, \psi)(u_n) \to (\phi, \psi)(u) = (\phi, \psi)(\mathcal{M}(x, y)) = (\phi^{\mathcal{M}}, \psi^{\mathcal{M}})(x, y), \quad \text{u.o.c. as } n \to \infty,$$

which completes the proof. □

Appendix D. Proofs of Theorems 2, 3, and 4

We prove Theorems 2, 3, and 4 by representing the queue length process of each driver type with the time-dependent, one-sided, and nonlinear regulator mapping defined in Definition C.1. We first present some preliminary results that will be used in the proofs of all of the three theorems in Section D.1. Then, we prove Theorems 2, 3, and 4 in Sections D.2, D.3, and D.4, respectively.

D.1. Preliminary Results

In this section, we present some preliminary results that will be used in the proofs of Theorems 2, 3, and 4. We consider $\pi_R(\boldsymbol{x})$ where $\{\boldsymbol{q}, \boldsymbol{x}\} = \{q_i(t), x_{ij}(t), i \in \mathcal{N}, j \in \mathcal{J}, t \in [0, T]\}$ is a feasible process pair for the CLP (12) such that x_{ij} is a Borel-measurable simple function (see Folland 1999, p. 46) for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$ in this section. We start with the formal definition of the randomized policy given in Definition 1.

Let $\{\hat{x}_{ij}(m), i \in \mathcal{N}, j \in \mathcal{J}, m \in \mathcal{U}\}$ be a sequence of real numbers in $\mathbb{R}_+$ such that $\mathcal{U}$ is a finite subset of $\mathbb{N}$. Let $\{B_m, m \in \mathcal{U}\}$ be a disjoint partition of the interval $[0, T]$ such that B_m is a Borel-measurable set for all $m \in \mathcal{U}$. Then,

$$x_{ij}(t) := \sum_{m \in \mathcal{U}} \hat{x}_{ij}(m) \mathbb{I}(t \in B_m), \qquad \forall t \in [0, T], i \in \mathcal{N}, j \in \mathcal{J}. \tag{D.1}$$

Observe that $\bar{F}_{ij}$ has finite range on the interval $[0, T]$ for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$ by (2) and the fact that $\tau_m \to \infty$ as $m \to \infty$. Thus, without loss of generality, we choose the partition $\{B_m, m \in \mathcal{U}\}$ such that for each $m \in \mathcal{U}$, there exists $l \in \mathbb{N}$ such that $B_m \subseteq [\tau_l, \tau_{l+1})$. By (12c), $\sum_{i \in \mathcal{N}} \hat{x}_{ij}(m) \le 1$ for all $j \in \mathcal{J}$ and $m \in \mathcal{U}$. Let us define the sequence of independent random variables $\{p_j^k(m), k \in \mathbb{N}_+, j \in \mathcal{J}, m \in \mathcal{U}\}$, which is independent of all other stochastic primitives and $\mathcal{F}(0)$-measurable such that $\mathbf{P}(p_j^k(m) = i) = \hat{x}_{ij}(m)$ and $\mathbf{P}(p_j^k(m) = 0) = 1 - \sum_{i \in \mathcal{N}} \hat{x}_{ij}(m)$ for all $i \in \mathcal{N}$, $j \in \mathcal{J}$, $k \in \mathbb{N}_+$, and $m \in \mathcal{U}$. Notice that the sequence $\{p_j^k(m), k \in \mathbb{N}_+, j \in \mathcal{J}, m \in \mathcal{U}\}$ corresponds to $\{\Upsilon_m, m \in \mathbb{N}\}$ defined in Remark 4 associated with the randomized policy. With a slight abuse of notation, let $\pi_R(\boldsymbol{x}) = (\pi_1^n, \pi_2^n, \ldots, \pi_J^n)$ in the nth system. Then,

$$\pi_j^n(k) = \sum_{i \in \mathcal{N}} \sum_{m \in \mathcal{U}} i\, \mathbb{I}\left(\nu_j^n(k) \in B_m,\ p_j^k(m) = i,\ Q_i^{\pi_R(\boldsymbol{x}),n}\left(\nu_j^n(k)-\right) > 0\right), \tag{D.2}$$

for all $j \in \mathcal{J}$ and $k, n \in \mathbb{N}_+$. Then, clearly, $\pi_j^n(k)$ is $\mathcal{F}_j^n(k)$-measurable for all $j \in \mathcal{J}$ and $k, n \in \mathbb{N}_+$, and thus $\pi_R(\boldsymbol{x})$ is admissible by Definition 1.

Let $\beta : \mathbb{N} \to \mathbb{N}$ be a function such that $\beta(m) := \{l \in \mathbb{N} : B_m \subseteq [\tau_l, \tau_{l+1})\}$. Under $\pi_R(\boldsymbol{x})$, for all $i \in \mathcal{N}$, $j \in \mathcal{J}$, $t \in [0, T]$, and $n \in \mathbb{N}_+$, we have

$$D_{ij}^{\pi_R(\boldsymbol{x}),n}(t) = \sum_{k=1}^{E_j \circ \Gamma_j^n(t)} \sum_{m \in \mathcal{U}} \mathbb{I}\left(\nu_j^n(k) \in B_m,\ a_j^k(\beta(m)) \ge t_{ij}(\beta(m)), p_j^k(m) = i,\ Q_i^{\pi_R(\boldsymbol{x}),n}\left(\nu_j^n(k)-\right) > 0\right).$$

Let us define the following stochastic process such that for all $i \in \mathcal{N}$, $j \in \mathcal{J}$, $n \in \mathbb{N}_+$, and $t \in [0, T]$,

$$\begin{aligned} H_{ij}^{\pi_R(\boldsymbol{x}),n}(t) &:= \sum_{k=1}^{E_j \circ \Gamma_j^n(t)} \sum_{m \in \mathcal{U}} \bar{F}_{ij}\left(\tau_{\beta(m)}\right) \hat{x}_{ij}(m) \mathbb{I}\left(\nu_j^n(k) \in B_m,\ Q_i^{\pi_R(\boldsymbol{x}),n}\left(\nu_j^n(k)-\right) > 0\right), \\ &= \int_0^t \sum_{m \in \mathcal{U}} \bar{F}_{ij}\left(\tau_{\beta(m)}\right) \hat{x}_{ij}(m) \mathbb{I}\left(s \in B_m,\ Q_i^{\pi_R(\boldsymbol{x}),n}(s-) > 0\right) \mathrm{d}E_j \circ \Gamma_j^n(s), \\ &= \int_0^t \bar{F}_{ij}(s) x_{ij}(s) \mathbb{I}\left(Q_i^{\pi_R(\boldsymbol{x}),n}(s-) > 0\right) \mathrm{d}E_j \circ \Gamma_j^n(s), \end{aligned}$$

where the second equality is the Lebesgue–Stieltjes integral (see Folland 1999, p. 107), and the third equality is by (2) and (D.1). Let $\bar{H}_{ij}^{\pi_R(\boldsymbol{x}),n} := (1/n) H_{ij}^{\pi_R(\boldsymbol{x}),n}$. Then we have the following result whose, proof is presented in Section D.1.1.

Lemma D.1. *For all $i \in \mathcal{N}$ and $j \in \mathcal{J}$, we have*

$$\left\| \bar{D}_{ij}^{\pi_R(\boldsymbol{x}),n} - \bar{H}_{ij}^{\pi_R(\boldsymbol{x}),n} \right\|_T \xrightarrow{a.s.} 0, \qquad \text{as } n \to \infty.$$

We omit the superscript $\pi_R(x)$ from the notation for convenience in presentation in the rest of Section D.1. Let us define the following stochastic process such that for all $i \in \mathcal{N}$, $j \in \mathcal{J}$, $n \in \mathbb{N}_+$, and $t \in [0, T]$,

$$\begin{aligned} I_{ij}^n(t) &:= \sum_{k=1}^{E_j \circ \Gamma_j^n(t)} \sum_{m \in \mathcal{U}} \bar{F}_{ij}(\tau_{\beta(m)})\hat{x}_{ij}(m)\mathbb{I}\left(\nu_j^n(k) \in B_m,\ Q_i^n\left(\nu_j^n(k)-\right) = 0\right), \\ &= \int_0^t \bar{F}_{ij}(s)x_{ij}(s)\mathbb{I}(Q_i^n(s-) = 0)\mathrm{d}E_j \circ \Gamma_j^n(s). \end{aligned} \tag{D.3}$$

Let $\bar{I}_{ij}^n := (1/n)I_{ij}^n$ for all $i \in \mathcal{N}$, $j \in \mathcal{J}$, and $n \in \mathbb{N}_+$. Then, by (11) and some algebra,

$$\bar{Q}_i^n(t) = \bar{X}_i^n(t) - \int_0^t \theta_i^n(s)\bar{Q}_i^n(s)\mathrm{d}s + \sum_{j\in\mathcal{J}} \bar{I}_{ij}^n(t), \tag{D.4}$$

for all $i \in \mathcal{N}$, $n \in \mathbb{N}_+$, and $t \in [0, T]$, where

$$\bar{X}_i^n(t) := \bar{Q}_i^n(0) + \bar{A}_i^n \circ \bar{\Lambda}_i^n(t) - \sum_{j\in\mathcal{J}}\left(\bar{D}_{ij}^n(t) - \bar{H}_{ij}^n(t)\right) - \left(\bar{R}_i^n\left(\int_0^t \theta_i^n(s)\bar{Q}_i^n(s)\mathrm{d}s\right) - \int_0^t \theta_i^n(s)\bar{Q}_i^n(s)\mathrm{d}s\right) - \frac{1}{n}\sum_{j\in\mathcal{J}}\int_0^t \bar{F}_{ij}(s)x_{ij}(s)\mathrm{d}E_j \circ \Gamma_j^n(s)$$

for all $i \in \mathcal{N}$, $n \in \mathbb{N}_+$, and $t \in [0, T]$. Without loss of generality, we define

$$\bar{Q}_i^n(t) := \bar{Q}_i^n(T),\ \bar{X}_i^n(t) := \bar{X}_i^n(T),\ \sum_{j\in\mathcal{J}} \bar{I}_{ij}^n(t) := \sum_{j\in\mathcal{J}} \bar{I}_{ij}^n(T),\ \theta_i^n(t) := 0 \tag{D.5}$$

for all $i \in \mathcal{N}$, $n \in \mathbb{N}_+$, and $t \geq T$ for mathematical completeness, so (D.4) is well defined for all $t \in \mathbb{R}_+$.

We approximate $\bar{X}_i^n$ by a deterministic process in the following result.

Lemma D.2. *For all $i \in \mathcal{N}$ and $j \in \mathcal{J}$, as $n \to \infty$, we have*

$$\sup_{0\le t\le T}\left|\frac{1}{n}\int_0^t \bar{F}_{ij}(s)x_{ij}(s)\mathrm{d}E_j \circ \Gamma_j^n(s) - \int_0^t \bar{F}_{ij}(s)x_{ij}(s)\mu_j(s)\mathrm{d}s\right| \xrightarrow{a.s.} 0,$$
$$\sup_{0\le t<\infty}\left|\bar{X}_i^n(t) - \bar{X}_i(t)\right| \xrightarrow{a.s.} 0,$$

where

$$\bar{X}_i(t) := \bar{Q}_i(0) + \Lambda_i(t) - \sum_{j\in\mathcal{J}}\int_0^t \bar{F}_{ij}(s)x_{ij}(s)\mu_j(s)\mathrm{d}s,$$

for all $t \in [0, T]$ and $\bar{X}_i(t) := \bar{X}_i(T)$ for all $t \geq T$. Moreover, $\bar{X}_i$ is nonnegative and Lipschitz continuous for all $i \in \mathcal{N}$.

The proof of Lemma D.2 is presented in Section D.1.2. Notice that for any $t \in [0, T]$, $I_{ij}^n(t)$ can increase only if there is a type j customer arrival at time t and $Q_i^n(t-) = 0$ by (D.3). Because the probability that there is also a type i driver arrival at time t is 0, then $Q_i^n(t) = 0$ with probability 1. Hence, for all $i \in \mathcal{N}$ and $n \in \mathbb{N}_+$,

$$\sum_{j\in\mathcal{J}} \bar{I}_{ij}^n(0) = 0, \quad \sum_{j\in\mathcal{J}} \bar{I}_{ij}^n(\cdot) \text{ is nondecreasing,} \tag{D.6a}$$

$$\int_0^\infty \bar{Q}_i^n(t)\mathrm{d}\left(\sum_{j\in\mathcal{J}} \bar{I}_{ij}^n(t)\right) = 0 \quad \text{a.s.,} \tag{D.6b}$$

by (D.3) and (D.5). By Assumption 1, the fact that $\bar{X}_i^n(0) \geq 0$, (D.4), (D.5), (D.6), Definition C.1, and Proposition C.1, we have

$$\left(\bar{Q}_i^n, \sum_{j\in\mathcal{J}} \bar{I}_{ij}^n\right) = \left(\phi^{\mathcal{M}}, \psi^{\mathcal{M}}\right)\left(\bar{X}_i^n, \theta_i^n\right) \quad \text{a.s.}$$

for all $i \in \mathcal{N}$ and $n \in \mathbb{N}_+$.

There exists a constant $\bar{\theta} \in \mathbb{R}_+$ and $n_0 \in \mathbb{N}_+$ such that $\theta_i^n(t) \leq \bar{\theta}$ for all $t \in \mathbb{R}_+$, $i \in \mathcal{N}$, and $n \geq n_0$ by Assumption 1 and (D.5). Then, by Assumption 1, Lemma C.3, and Lemma D.2,

$$\left\|\bar{Q}_i^n - \phi^{\mathcal{M}}(\bar{X}_i, \theta_i)\right\|_T \xrightarrow{a.s.} 0, \quad \left\|\sum_{j\in\mathcal{J}} \bar{I}_{ij}^n - \psi^{\mathcal{M}}(\bar{X}_i, \theta_i)\right\|_T \xrightarrow{a.s.} 0, \tag{D.7}$$

as $n \to \infty$ for all $i \in \mathcal{N}$. Last, we have the following result.

Published in Stochastic Systems on February 25, 2020 as DOI: 10.1287/stsy.2019.0037.
This article has not been copyedited or formatted. The final version may differ from this version.

Lemma D.3. *Fix an arbitrary* $i \in \mathcal{N}$. *If* $\psi^{\mathcal{M}}(\bar{X}_i, \theta_i) = \mathbf{0}$, *then*

$$\sup_{0\le t\le T}\left|\sum_{j\in\mathcal{J}}\left(\bar{D}^n_{ij}(t) - \int_0^t \mu_j(s)\bar{F}_{ij}(s)x_{ij}(s)\mathrm{d}s\right)\right| \xrightarrow{a.s.} 0, \quad \textit{as } n\to\infty.$$

Proof. By (D.7) and the fact that $\psi^{\mathcal{M}}(\bar{X}_i, \theta_i) = \mathbf{0}$,

$$\left\|\sum_{j\in\mathcal{J}} \bar{I}^n_{ij}\right\|_T \xrightarrow{a.s.} 0, \quad \textit{as } n\to\infty. \tag{D.8}$$

Then,

$$\begin{aligned}
&\sup_{0\le t\le T}\left|\sum_{j\in\mathcal{J}}\left(\bar{D}^n_{ij}(t) - \int_0^t \mu_j(s)\bar{F}_{ij}(s)x_{ij}(s)\mathrm{d}s\right)\right| \\
&= \sup_{0\le t\le T}\left|\sum_{j\in\mathcal{J}}\left(\bar{D}^n_{ij}(t) - \bar{H}^n_{ij}(t) - \bar{I}^n_{ij}(t) + \frac{1}{n}\int_0^t \bar{F}_{ij}(s)x_{ij}(s)\mathrm{d}E_j\circ\Gamma^n_j(s) - \int_0^t \mu_j(s)\bar{F}_{ij}(s)x_{ij}(s)\mathrm{d}s\right)\right| \\
&\le \sum_{j\in\mathcal{J}}\left\|\bar{D}^n_{ij} - \bar{H}^n_{ij}\right\|_T + \left\|\sum_{j\in\mathcal{J}}\bar{I}^n_{ij}\right\|_T + \sum_{j\in\mathcal{J}}\sup_{0\le t\le T}\left|\frac{1}{n}\int_0^t \bar{F}_{ij}(s)x_{ij}(s)\mathrm{d}E_j\circ\Gamma^n_j(s) - \int_0^t \bar{F}_{ij}(s)x_{ij}(s)\mu_j(s)\mathrm{d}s\right| \\
&\xrightarrow{a.s.} 0, \quad \text{as } n\to\infty,
\end{aligned}$$

where the inequality is by triangular inequality and the convergence result is by Lemma D.1, (D.8), and Lemma D.2. □

D.1.1. Proof of Lemma D.1. The proof of Lemma D.1 is very similar to the one of Lemma B.1. For notational convenience, we omit the superscript $\pi_R(x)$ from the notation in this section. Let us fix arbitrary $i\in\mathcal{N}$ and $j\in\mathcal{J}$. Then, by definition, we have

$$\left\|\bar{D}^n_{ij} - \bar{H}^n_{ij}\right\|_T = \sup_{0\le t\le T}\left|\frac{1}{n}\sum_{k=1}^{E_j\circ\Gamma^n_j(t)}\sum_{m\in\mathcal{U}}\left(\mathbb{I}\left(a^k_j(\beta(m)) \ge t_{ij}(\beta(m)),\ p^k_j(m) = i\right) - \bar{F}_{ij}(\tau_{\beta(m)})\hat{x}_{ij}(m)\mathbb{I}\left(\nu^n_j(k)\in B_m,\ Q^n_i\left(\nu^n_j(k)-\right) > 0\right)\right|. \tag{D.9}$$

For all $k, n\in\mathbb{N}_+$ and $m\in\mathcal{U}$, let

$$\begin{aligned}
Y_k(m) &:= \mathbb{I}\left(a^k_j(\beta(m)) \ge t_{ij}(\beta(m)),\ p^k_j(m) = i\right) - \bar{F}_{ij}(\tau_{\beta(m)})\hat{x}_{ij}(m), \\
Z^n_k(m) &:= \mathbb{I}\left(\nu^n_j(k)\in B_m,\ Q^n_i\left(\nu^n_j(k)-\right) > 0\right), \\
W^n_k(m) &:= Y_k(m)\times Z^n_k(m), \\
\tilde{W}^n_k &:= \sum_{m\in\mathcal{U}} W^n_k(m).
\end{aligned}$$

Then, the right-hand side of (D.9) is equal to

$$\sup_{0\le t\le T}\left|\frac{1}{n}\sum_{k=1}^{E_j\circ\Gamma^n_j(t)}\tilde{W}^n_k\right|. \tag{D.10}$$

Notice that, if we can prove that

$$\frac{1}{n}\left|\sum_{k=1}^{n}\tilde{W}^n_k\right| \xrightarrow{a.s.} 0, \qquad \text{as } n\to\infty, \tag{D.11}$$

then we can prove that the term in (D.10) converges to 0 as $n\to\infty$ by the same way we prove Lemma B.1. Hence, in the remaining part of this section, we will prove (D.11), whose proof is very similar to the one of (B.9).

For all $k\in\mathbb{Z}$ and $n\in\mathbb{N}_+$, if $k\in\mathbb{Z}\backslash\mathbb{N}_+$, we let $\mathcal{G}^n_k := \{\emptyset, \Omega\}$, and if $k\in\mathbb{N}_+$, we let

$$\begin{aligned}
\mathcal{G}^n_k := \sigma\Bigg\{&A_i\circ\Lambda^n_i\left(s\wedge\nu^n_j(k)\right), E_{j'}\circ\Gamma^n_{j'}\left(s\wedge\nu^n_j(k)\right), R_i\left(\int_0^{(s\wedge\nu^n_j(k))-}\theta^n_i(u)Q^n_i(u)\mathrm{d}u\right), D^n_{ij'}\left(\left(s\wedge\nu^n_j(k)\right)-\right), \\
&Q^n_i\left(\left(s\wedge\nu^n_j(k)\right)-\right), \quad \forall i\in\mathcal{N}, j'\in\mathcal{J}, s\in\mathbb{R}_+, a^r_{j'}(\beta(m)), p^r_{j'}(m),\ r\in\left\{1,\ldots,E_{j'}\circ\Gamma^n_{j'}\left(\nu^n_j(k)-\right)\right\}, \forall j'\in\mathcal{J}\backslash\{j\}, m\in\mathcal{U} \\
&a^r_j(\beta(m)), p^r_j(m),\ r\in\{1,\ldots,k-1\}, \quad \forall m\in\mathcal{U}\Bigg\}.
\end{aligned}$$

Then, $\{\mathcal{G}_k^n, k \in \mathbb{Z}\}$ is a filtration for all $n \in \mathbb{N}_+$. Moreover, $Y_k(m) \perp \mathcal{G}_k^n$ and $Z_k^n(m) \in \mathcal{G}_k^n$ for all $k, n \in \mathbb{N}_+$ and $m \in \mathcal{U}$, and $W_k^n(m) \in \mathcal{G}_{k+l}^n$ for all $k, n, l \in \mathbb{N}_+$ by construction. First, for all $k, n \in \mathbb{N}_+$,

$$\begin{aligned}\mathbf{E}\big[\tilde{W}_k^n|\mathcal{G}_k^n\big] &= \mathbf{E}\Bigg[\sum_{m\in\mathcal{U}} W_k^n(m)|\mathcal{G}_k^n\Bigg] = \sum_{m\in\mathcal{U}} \mathbf{E}\big[W_k^n(m)|\mathcal{G}_k^n\big] \\ &= \sum_{m\in\mathcal{U}} \mathbf{E}\big[Y_k(m)Z_k^n(m)|\mathcal{G}_k^n\big] = \sum_{m\in\mathcal{U}}\big(Z_k^n(m)\mathbf{E}\big[Y_k(m)|\mathcal{G}_k^n\big]\big) \\ &= \sum_{m\in\mathcal{U}}\big(Z_k^n(m)\mathbf{E}[Y_k(m)]\big) = 0. \end{aligned} \tag{D.12}$$

Second, for all $k, n, l \in \mathbb{N}_+$,

$$\mathbf{E}\big[\tilde{W}_k^n|\mathcal{G}_{k-l}^n\big] = \mathbf{E}\big[\mathbf{E}\big[\tilde{W}_k^n|\mathcal{G}_k^n\big]|\mathcal{G}_{k-l}^n\big] = 0, \tag{D.13}$$

by (D.12). Third, for all $k, n \in \mathbb{N}_+$,

$$\mathbf{E}\Big[\tilde{W}_k^n|\mathcal{G}_{k+\lfloor n^{0.25}\rfloor-1}^n\Big] = \begin{cases} 0 & \text{if } n < 16, \\ \tilde{W}_k^n & n \geq 16. \end{cases} \tag{D.14}$$

Therefore, by (D.12), (D.13), and (D.14), we can prove (D.11) by the same method that we use in Appendix B, Section B.1 (starting from (B.27)) in order to prove (B.9).

D.1.2. Proof of Lemma D.2. By triangular inequality, for all $i \in \mathcal{N}$,

$$\begin{aligned}&\|\bar{X}_i^n - \bar{X}_i\|_T \\ &\quad\leq |\bar{Q}_i^n(0) - \bar{Q}_i(0)| + \|\bar{A}_i^n \circ \bar{\Lambda}_i^n - \Lambda_i\|_T + \sum_{j\in\mathcal{J}} \Big\|\bar{D}_{ij}^n - \bar{H}_{ij}^n\Big\|_T &\text{(D.15a)}\\ &\quad+ \sup_{0\leq t\leq T}\left|\bar{R}_i^n\left(\int_0^t \theta_i^n(s)\bar{Q}_i^n(s)\mathrm{d}s\right) - \int_0^t \theta_i^n(s)\bar{Q}_i^n(s)\mathrm{d}s\right| &\text{(D.15b)}\\ &\quad+ \sum_{j\in\mathcal{J}}\sup_{0\leq t\leq T}\left|\int_0^t \bar{F}_{ij}(s)x_{ij}(s)\Big(\mathrm{d}\bar{E}_j^n \circ \bar{\Gamma}_j^n(s) - \mathrm{d}\Gamma_j(s)\Big)\right|. &\text{(D.15c)}\end{aligned}$$

We will consider each term on the right-hand side of (D.15) separately. The right-hand side of (D.15a) converges to 0 a.s. by Assumption 2, (B.1), and Lemma D.1, respectively.

Next, let us consider the term in (D.15b). Let us fix an arbitrary $\omega \in \mathcal{A}_1$ (see (B.2)). From the proof of Proposition A.1, we know that each subsequence of the term in (D.15b) has a subsequence that converges to 0 on the sample path ω, which implies that the term in (D.15b) itself converges to 0 as $n \to \infty$ on the sample path ω. Because $\mathbf{P}(\mathcal{A}_1) = 1$, the term in (D.15b) converges to 0 a.s as $n \to \infty$.

Next let us consider the term in (D.15c) and fix an arbitrary $i \in \mathcal{N}$ and $j \in \mathcal{J}$. By (D.1) and the construction of the sequence of Borel-measurable sets $\{B_m, m \in \mathcal{U}\}$,

$$\bar{F}_{ij}(t)x_{ij}(t) = \sum_{m\in\mathcal{U}} \bar{F}_{ij}\big(\tau_{\beta(m)}\big)\hat{x}_{ij}(m)\mathbb{I}(t \in B_m), \qquad \forall t \in [0, T].$$

Let ℓ denote the Lebesgue measure on $\mathbb{R}$. Then, for all $\epsilon > 0$, there exists a sequence of sets $\{B_m^{(1)}, m \in \mathcal{U}\}$ such that $B_m^{(1)}$ is finite union of open intervals and $\ell(B_m\backslash B_m^{(1)}) + \ell(B_m^{(1)}\backslash B_m) < \epsilon$ for all $m \in \mathcal{U}$ by proposition 1.20 of Folland (1999). Hence, for all $\epsilon > 0$, there exists a simple function ξ such that $\xi(t) := \sum_{m\in\mathcal{U}^{(2)}} z_m\mathbb{I}(t \in B_m^{(2)})$ for all $t \in [0, T]$, $\mathcal{U}^{(2)}$ is a finite subset of $\mathbb{N}$, $\{z_m, m \in \mathcal{U}^{(2)}\}$ is a nonnegative sequence of real numbers bounded above by 1, $B_m^{(2)}$ is finite union of open intervals for all $m \in \mathcal{U}^{(2)}$, and $\ell(t \in [0, T] : \xi(t) \neq \bar{F}_{ij}(t)x_{ij}(t)) < \epsilon$. Moreover, for all $\epsilon > 0$, there exists a continuous function $g : \mathbb{R} \to [0, 1]$ such that $g(t) = 0$ for all $t \notin [0, T]$, g has bounded variation, and $\ell(t \in [0, T] : g(t) \neq \xi(t)) < \epsilon$ (see the proof of theorem 2.26 of Folland 1999 for a specific construction, and the reason why g has bounded variation is that ξ has finite range and $B_m^{(2)}$ is finite union of open intervals for all $m \in \mathcal{U}^{(2)}$). Let $\mathcal{U}^{(3)} := \{t \in [0, T] : g(t) \neq \bar{F}_{ij}(t)x_{ij}(t)\}$. Then, $\ell(\mathcal{U}^{(3)}) < 2\epsilon$. Let $\bar{L}_j^n := \bar{E}_j^n \circ \bar{\Gamma}_j^n - \Gamma_j$. Then,

$$\sup_{0\leq t\leq T}\left|\int_0^t \bar{F}_{ij}(s)x_{ij}(s)\Big(\mathrm{d}\bar{E}_j^n \circ \bar{\Gamma}_j^n(s) - \mathrm{d}\Gamma_j(s)\Big)\right| \leq \sup_{0\leq t\leq T}\left|\int_0^t \big(\bar{F}_{ij}(s)x_{ij}(s) - g(s)\big)\mathrm{d}\bar{L}_j^n(s)\right| + \sup_{0\leq t\leq T}\left|\int_0^t g(s)\mathrm{d}\bar{L}_j^n(s)\right|. \tag{D.16}$$

First, let us consider the first term in (D.16), which is less than or equal to

$$\begin{aligned}&\sup_{0\leq t\leq T}\left|\int_0^t 2\mathbb{I}\Big(s \in \mathcal{U}^{(3)}\Big)\mathrm{d}\bar{E}_j^n \circ \bar{\Gamma}_j^n(s)\right| + \sup_{0\leq t\leq T}\left|\int_0^t 2\mathbb{I}\Big(s \in \mathcal{U}^{(3)}\Big)\mathrm{d}\Gamma_j(s)\right| \\ &= 2\int_0^T \mathbb{I}\Big(s \in \mathcal{U}^{(3)}\Big)\mathrm{d}\bar{E}_j^n \circ \bar{\Gamma}_j^n(s) + 2\int_0^T \mathbb{I}\Big(s \in \mathcal{U}^{(3)}\Big)\mu_j(s)\mathrm{d}s \\ &\leq \frac{2}{n}\int_0^T \mathbb{I}\Big(s \in \mathcal{U}^{(3)}\Big)\mathrm{d}E_j^n \circ \Gamma_j^n(s) + 4\bar{\mu}_j\epsilon, \end{aligned} \tag{D.17}$$

where (D.17) is by (10) and the fact that $\ell(\mathcal{U}^{(3)}) < 2\epsilon$ and $\bar{\mu}_j = \sup_{t\in\mathbb{R}_+} \mu_j(t)$. Let us consider the first term in (D.17). By Assumption 1, there exists a sufficiently large $n_0 \in \mathbb{N}_+$ such that if $n \geq n_0$, then $\bar{\mu}_j^n := \sup_{t\in[0,T]} \mu_j^n(t) < \infty$, and $\sup_{n\geq n_0}(\bar{\mu}_j^n/n) < C$, where $C \in \mathbb{R}_+$ is a constant. Let us fix an arbitrary $n \geq n_0$ and consider a time-homogeneous Poisson process, denoted by M_j^n, with rate $\bar{\mu}_j^n$ and associated arrival times denoted by the sequence $\{\bar{\nu}_j^n(k), k \in \mathbb{N}_+\}$. Because we can think of the nonhomogeneous Poisson process $E_j^n \circ \Gamma_j^n$ as being a random sample from the homogeneous Poisson process M_j^n (see Ross 1996, p. 80), the first term in (D.17) is less than or equal to

$$\frac{2}{n}\int_0^T \mathbb{I}\left(s \in \mathcal{U}^{(3)}\right) \mathrm{d}M_j^n(s) = \frac{2}{n}\sum_{k=1}^{M_j^n(T)} \mathbb{I}\left(\bar{\nu}_j^n(k) \in \mathcal{U}^{(3)}\right). \tag{D.18}$$

Notice that

$$\begin{aligned}
&\mathbf{P}\left(\frac{2}{n}\sum_{k=1}^{M_j^n(T)} \mathbb{I}\left(\bar{\nu}_j^n(k) \in \mathcal{U}^{(3)}\right) > \sqrt{\epsilon}\right)\\
&\quad\leq \frac{2}{n\sqrt{\epsilon}}\mathbf{E}\left[\sum_{k=1}^{M_j^n(T)} \mathbb{I}\left(\bar{\nu}_j^n(k) \in \mathcal{U}^{(3)}\right)\right]\\
&\quad= \frac{2}{n\sqrt{\epsilon}}\sum_{K\in\mathbb{N}} \mathbf{E}\left[\sum_{k=1}^{K} \mathbb{I}\left(\bar{\nu}_j^n(k) \in \mathcal{U}^{(3)}\right)\middle| M_j^n(T) = K\right]\mathbf{P}\left(M_j^n(T) = K\right)\\
&\quad= \frac{2}{n\sqrt{\epsilon}}\sum_{K\in\mathbb{N}} \mathbf{E}\left[\mathrm{Binomial}\left(K, \frac{\ell(\mathcal{U}^{(3)})}{T}\right)\right]\mathbf{P}\left(M_j^n(T) = K\right)\\
&\quad= \frac{2}{n\sqrt{\epsilon}}\sum_{K\in\mathbb{N}} K\frac{\ell(\mathcal{U}^{(3)})}{T}\mathbf{P}\left(M_j^n(T) = K\right)\\
&\quad\leq \frac{4\sqrt{\epsilon}\bar{\mu}_j^n T}{nT} \leq 4C\sqrt{\epsilon},
\end{aligned} \tag{D.19}$$

where the first inequality is by Markov's inequality, and the second equality is by the fact that, given that $M_j^n(T) = K$, the K arrival times $\bar{\nu}_j^n(1), \ldots, \bar{\nu}_j^n(K)$ have the same distribution as order statistics corresponding to K independent random variables uniformly distributed on the interval $[0, T]$ (see Ross 1996, theorem 2.3.1), and the probability that a random variable that is uniformly distributed on $[0, T]$ is in the Borel-measurable set $\mathcal{U}^{(3)}$ is $\ell(\mathcal{U}^{(3)})/T$.

Therefore, (D.18) and (D.19) imply that (D.17) is bounded above by $\sqrt{\epsilon} + 4\bar{\mu}_j\epsilon$ with probability $1 - 4C\sqrt{\epsilon}$. Because $\epsilon > 0$ is arbitrary, the first term in (D.16) is equal to 0 with probability 1 for all $n \geq n_0$.

Next, let us consider the second term in (D.16) and fix an arbitrary $\omega \in \mathcal{A}_1$ [see (B.2)]. Then, $\bar{L}_j^n(\cdot, \omega) \to \mathbf{0}$ u.o.c. as $n \to \infty$. By theorem 3.36 of Folland (1999) (integration by parts), the second term in (D.16) is equal to

$$\sup_{0\leq t\leq T}\left|\bar{L}_j^n(t,\omega)g(t) - \int_0^t \bar{L}_j^n(s,\omega)\mathrm{d}g(s)\right| \leq \sup_{t\in[0,T]}\left|\bar{L}_j^n(t,\omega)\right| + \sup_{0\leq t\leq T}\left|\int_0^t \bar{L}_j^n(s,\omega)\mathrm{d}g(s)\right|. \tag{D.20}$$

Notice that the Lebesgue–Stieltjes measure induced by g can be a signed measure. By theorem 3.3 of Folland (1999) (the Hahn decomposition theorem), there exists a positive set P and negative set N for the Lebesgue–Stieltjes measure induced by g such that $P \cup N = [0, T]$ and $P \cap N = \emptyset$. Let $\mathcal{V}(g) < \infty$ denote the total variation of g. Then, the sum of the terms in (D.20) is less than or equal to

$$\begin{aligned}
&\sup_{t\in[0,T]}\left|\bar{L}_j^n(t,\omega)\right| + \sup_{0\leq t\leq T}\left|\int_{P\cap[0,t]} \bar{L}_j^n(s,\omega)\mathrm{d}g(s)\right| + \sup_{0\leq t\leq T}\left|\int_{N\cap[0,t]} \bar{L}_j^n(s,\omega)\mathrm{d}g(s)\right|\\
&\quad\leq \sup_{t\in[0,T]}\left|\bar{L}_j^n(t,\omega)\right| + 2\mathcal{V}(g)\sup_{t\in[0,T]}\left|\bar{L}_j^n(t,\omega)\right| = \sup_{t\in[0,T]}\left|\bar{L}_j^n(t,\omega)\right|\left(1 + 2\mathcal{V}(g)\right)\\
&\quad\to 0 \quad \text{as } n \to \infty.
\end{aligned}$$

Therefore, the sum in (D.16) converges to 0 as $n \to \infty$ a.s., and so do the term in (D.15c) and $\left\|\bar{X}_i^n - \bar{X}_i\right\|_T$. Because $\bar{X}_i^n(t) = \bar{X}_i^n(T)$ for all $t \geq T$ (see (D.5)),

$$\sup_{0\leq t<\infty}\left|\bar{X}_i^n(t) - \bar{X}_i(t)\right| \xrightarrow{a.s.} 0, \quad \text{as } n \to \infty \text{ for all } i \in \mathcal{N}.$$

Last, $\bar{X}_i$ is nonnegative by (12b) and (12d) and is Lipschitz continuous by Assumption 1.

D.2. Proof of Theorem 2

By (12d), (C.1), and the fact that q_i is Lipschitz continuous, we have $\phi(q_i) = q_i$ for all $i \in \mathcal{N}$. Hence, by (12b),

$$q_i(t) = \bar{X}_i(t) - \int_0^t \theta_i(s)\phi(q_i)(s)ds, \quad \forall i \in \mathcal{N}, t \in [0,T].$$

Therefore, $q_i = \mathcal{M}(\bar{X}_i, \theta_i)$ by Lemma C.1, part 1. Then, we have $\psi^{\mathcal{M}}(\bar{X}_i, \theta_i) = \psi(\mathcal{M}(\bar{X}_i, \theta_i)) = \psi(q_i) = \mathbf{0}$ for all $i \in \mathcal{N}$ by Proposition C.1, (12d), and (C.1). Then, the first part of the proof follows by Lemma D.3. Furthermore, by Proposition C.1, (12d), and (C.1), for all $i \in \mathcal{N}$, we have

$$\|\bar{Q}_i^n - q_i\|_T = \|\bar{Q}_i^n - \phi(q_i)\|_T = \|\bar{Q}_i^n - \phi(\mathcal{M}(\bar{X}_i, \theta_i))\|_T = \|\bar{Q}_i^n - \phi^{\mathcal{M}}(\bar{X}_i, \theta_i)\|_T \xrightarrow{a.s.} 0,$$

where the convergence is by (D.7).

D.3. Proof of Theorem 3

The process $\boldsymbol{x}^*$ is a feasible matching process for the CLP (12) by Lemma 4, and it is a Borel-measurable simple function for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$ by assumption. Then, the results of Section D.1 apply to this process. Moreover, we can extend the results of Lemma D.2 such that $\bar{X}_i$ is also nondecreasing for all $i \in \mathcal{N}$ by (13b). Then, by Lemma C.1, part 2, $\mathcal{M}(\bar{X}_i, \theta_i) \geq \mathbf{0}$ for all $i \in \mathcal{N}$. By (C.1) and the fact that $\psi^{\mathcal{M}}(\bar{X}_i, \theta_i) = \psi(\mathcal{M}(\bar{X}_i, \theta_i))$ (see Proposition C.1), we have $\psi^{\mathcal{M}}(\bar{X}_i, \theta_i) = \mathbf{0}$ for all $i \in \mathcal{N}$. Then, by Lemma D.3, for all $i \in \mathcal{N}$,

$$\sup_{0\leq t\leq T}\left|\sum_{j\in\mathcal{J}}\left(\bar{D}_{ij}^{\pi_R(\boldsymbol{x}^*),n}(t) - \int_0^t \mu_j(s)\bar{F}_{ij}(s)x_{ij}^*(s)ds\right)\right| \xrightarrow{a.s.} 0, \quad \text{as } n \to \infty,$$

which gives us the desired result.

D.4. Proof of Theorem 4

Because the parameters of LP (13) are time homogeneous, $\boldsymbol{x}^*$ can be chosen as a constant function of time, and so we let $x_{ij}^*(t) = x_{ij}^*$ for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$. Then, for all $i \in \mathcal{N}$,

$$\bar{X}_i(t) = \lambda_i t - \sum_{j\in\mathcal{J}} \mu_j \bar{F}_{ij} x_{ij}^* t \quad \text{and let} \quad q_i^* := \mathcal{M}(\bar{X}_i, \theta_i),$$

Then, $\bar{X}_i$ is nonnegative and nondecreasing by (13b), and it is Lipschitz continuous. By Lemma C.1, part 2, $q_i^* \geq \mathbf{0}$, and so $\phi(q_i^*) = q_i^*$ for all $i \in \mathcal{N}$ by (C.1). Thus, $\{\boldsymbol{q}^*, \boldsymbol{x}^*\}$ satisfies (12b), satisfies (12e) because $\boldsymbol{x}^*$ is a constant, and satisfies (12c) and (12d) by satisfying (13c) and (13d), respectively. Therefore, $\{\boldsymbol{q}^*, \boldsymbol{x}^*\}$ is a feasible process pair for the CLP (12). Thus,

$$\sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij}\mu_j\bar{F}_{ij}x_{ij}^*T \leq \sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij}\mu_j\bar{F}_{ij}\int_0^T \tilde{x}_{ij}(s)ds. \tag{D.21}$$

Let $\bar{x}_{ij} := (1/T)\int_0^T \tilde{x}_{ij}(s)ds$ for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$. Then it is easy to see that $\{\bar{x}_{ij}, i \in \mathcal{N}, j \in \mathcal{J}\}$ is feasible for the LP (13) when the parameters are time homogeneous, and thus

$$\sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij}\mu_j\bar{F}_{ij}\bar{x}_{ij} = \sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij}\mu_j\bar{F}_{ij}\left(\frac{1}{T}\int_0^T \tilde{x}_{ij}(s)ds\right) \leq \sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij}\mu_j\bar{F}_{ij}x_{ij}^*. \tag{D.22}$$

By (D.21) and (D.22), we see that $\{\boldsymbol{q}^*, \boldsymbol{x}^*\}$ is an optimal solution of the CLP (12). Because $\boldsymbol{x}^*$ is constant, $\pi_R(\boldsymbol{x}^*)$ is admissible by Lemma 3.

The rest of the proof of Theorem 4 is very similar to the one of Theorem 3. Because the process $\boldsymbol{x}^*$ is a feasible matching process for the CLP (12) and x_{ij}^* is a constant function of time and thus a Borel-measurable simple function for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$, the results of Section D.1 apply to this process. Moreover, $\psi^{\mathcal{M}}(\bar{X}_i, \theta_i) = \psi(\mathcal{M}(\bar{X}_i, \theta_i)) = \psi(q_i^*) = \mathbf{0}$ for all $i \in \mathcal{N}$ by (C.1). Then, by Lemma D.3, (D.21), and (D.22), for all $i \in \mathcal{N}$,

$$\sup_{0\leq t\leq T}\left|\sum_{j\in\mathcal{J}}\left(\bar{D}_{ij}^{\pi_R(\boldsymbol{x}^*),n}(t) - \int_0^t \mu_j(s)\bar{F}_{ij}(s)\tilde{x}_{ij}(s)ds\right)\right| \xrightarrow{a.s.} 0, \quad \text{as } n \to \infty,$$

which gives us the desired result.

Appendix E. Lemma Proofs

We first prove Lemma 1, second prove Lemmas 3 and 6 (because they are related to Lemma 1), third prove Lemma 2, then prove Lemma 4, and finally prove Lemma 5.

E.1. Proof of Lemma 1

Let $\pi_{CD} = (\pi_1, \pi_2, \ldots, \pi_J)$. For notational convenience, we omit the superscript π_{CD} from the notation in this proof. Let us define the sequence of independent random variables $\{q_j^k(l), k \in \mathbb{N}_+, j \in \mathcal{J}, l \in \mathbb{N}^N\}$, which is independent of all other stochastic primitives and $\mathscr{F}(0)$-measurable, such that $q_j^k(0,0,\ldots,0)=0$, and if $\sum_{i=1}^N l_i>0$, then $\mathbf{P}(q_j^k(l_1,l_2,\ldots,l_N)=i)=l_i/\sum_{i=1}^N l_i$ for all $k\in\mathbb{N}_+$, $j\in\mathcal{J}$, and $l\in\mathbb{N}^N$ such that $l=(l_1,l_2,\ldots,l_N)$. For notational convenience, let us define the set

$$S_j(k):=\underset{\{i\in\mathcal{N}:\,Q_i(\nu_j(k)-)>0\}}{\operatorname{argmin}}\sum_{m\in\mathbb{N}}t_{ij}(m)\mathbb{I}(\nu_j(k)\in[\tau_m,\tau_{m+1})) \tag{E.1}$$

for all $j\in\mathcal{J}$ and $k\in\mathbb{N}_+$. Let $|S_j(k)|$ denote the cardinality of the set $S_j(k)$. Then, under the CD policy, for all $j\in\mathcal{J}$ and $k\in\mathbb{N}_+$, let

$$\pi_j(k):=\mathbb{I}(|S_j(k)|=1)\sum_{i\in\mathcal{N}}\big(i\,\mathbb{I}(i\in S_j(k))\big)+\mathbb{I}(|S_j(k)|>1)\times q_j^k\big(Q_1(\nu_j(k)-)\mathbb{I}(1\in S_j(k)),\ldots,Q_N(\nu_j(k)-)\mathbb{I}(N\in S_j(k))\big).$$

Notice that $Q_i(\nu_j(k)-)=\lim_{t\to\infty}Q_i((t\wedge\nu_j(k))-)\in\mathscr{F}_j(k)$ by proposition 2.7 of Folland (1999) and definition of $\mathscr{F}_j(k)$ [see (8)] for all $i\in\mathcal{N}$, $j\in\mathcal{J}$, and $k\in\mathbb{N}_+$. By (8), (E.1), and the fact that $\nu_j(k)\in\mathscr{F}_j(k)$, the random variables $\mathbb{I}(|S_j(k)|=1)$, $\mathbb{I}(|S_j(k)|>1)$, $\mathbb{I}(i\in S_j(k))$, and $\mathbb{I}(S_j(k)=\mathcal{N}_1)$ are $\mathscr{F}_j(k)$-measurable for all $i\in\mathcal{N}$, $j\in\mathcal{J}$, $k\in\mathbb{N}_+$, and $\mathcal{N}_1\subset\mathcal{N}$. Finally,

$$\begin{aligned}&q_j^k\big(Q_1(\nu_j(k)-)\mathbb{I}(1\in S_j(k)),\ldots,Q_N(\nu_j(k)-)\mathbb{I}(N\in S_j(k))\big)\\&\quad=\sum_{\mathcal{N}_1\subset\mathcal{N}}\mathbb{I}(S_j(k)=\mathcal{N}_1)\times\left(\sum_{\{l\in\mathbb{N}^N:l_i=0,\forall i\notin\mathcal{N}_1\}}q_j^k(l_1,l_2,\ldots,l_N)\mathbb{I}\big(Q_i(\nu_j(k)-)=l_i,\forall i\in\mathcal{N}_1\big)\right)\in\mathscr{F}_j(k).\end{aligned}$$

Therefore, $\pi_j\in\mathbb{F}_j$ for all $j\in\mathcal{J}$ under the CD policy, so it is an admissible policy. Notice that the sequence $\{q_j^k(l),k\in\mathbb{N}_+,j\in\mathcal{J},l\in\mathbb{N}^N\}$ corresponds to $\{\Upsilon_m,m\in\mathbb{N}\}$ defined in Remark 4 associated with the CD policy.

E.2. Proofs of Lemmas 3 and 6

First, the randomized policy given in Definition 2 is admissible by the fact that (D.2) is $\mathscr{F}_j^n(k)$-measurable for all $j\in\mathcal{J}$ and $k,n\in\mathbb{N}_+$. Second, the admissibility proof of the hybrid policy follows by the fact that it is a hybrid of the randomized policy given in Definition 2 and the CD policy defined in (3), and both of the latter two policies are admissible. Third, the deterministic policy is admissible because both $Q_i^n(\nu_j^n(k)-)$ and $D_{ij}^n(\nu_j^n(k)-)$ are $\mathscr{F}_j^n(k)$-measurable for all $i\in\mathcal{N}$, $j\in\mathcal{J}$, and $k,n\in\mathbb{N}_+$ (see (8)), and the tie breaking rule does not use any future information.

Last, we will prove that the RWQ policy is admissible. We let $\pi=(\pi_1,\pi_2,\ldots,\pi_J)$ denote the RWQ policy, and we omit the superscript π from the notation for notational convenience in this proof. Recall from Appendix D, Section D.1, that $\{\hat{x}_{ij}(m),i\in\mathcal{N},j\in\mathcal{J},m\in\mathcal{U}\}$ is a sequence of real numbers in $\mathbb{R}_+$, $\mathcal{U}$ is a finite subset of $\mathbb{N}$, and $\{B_m,m\in\mathcal{U}\}$ is a disjoint partition of the interval $[0,T]$ such that B_m is a Borel-measurable set for all $m\in\mathcal{U}$. Let us define the sequence of independent random variables $\{\tilde{q}_{j,m}^k(l),k\in\mathbb{N}_+,j\in\mathcal{J},m\in\mathcal{U},l\in\mathbb{N}^N\}$, which is independent of all other stochastic primitives and $\mathscr{F}(0)$-measurable, such that

$$\tilde{q}_{j,m}^k(l_1,l_2,\ldots,l_N)=\begin{cases}i & \text{with probability } l_i\hat{x}_{ij}(m)\Big/\left(\sum_{i=1}^N l_i\hat{x}_{ij}(m)\right)\\ & \text{if } \sum_{i=1}^N l_i\hat{x}_{ij}(m)>0\\ 0 & \text{if } \sum_{i=1}^N l_i\hat{x}_{ij}(m)=0\end{cases}$$

for all $k\in\mathbb{N}_+$, $j\in\mathcal{J}$, $m\in\mathcal{U}$, and $l\in\mathbb{N}^N$ such that $l=(l_1,l_2,\ldots,l_N)$.

Under the RWQ policy, for all $j\in\mathcal{J}$ and $k\in\mathbb{N}_+$, let

$$\pi_j(k):=\sum_{m\in\mathcal{U}}\mathbb{I}(\nu_j(k)\in B_m)\left(\sum_{l\in\mathbb{N}^N}\tilde{q}_{j,m}^k(l)\mathbb{I}\big(Q_i(\nu_j(k)-)=l_i,\ \forall i\in\mathcal{N}\big)\right)$$

by (8), (18), and (D.1) for all $j\in\mathcal{J}$ and $k\in\mathbb{N}_+$. By (8) and the fact that B_m is Borel measurable for all $m\in\mathcal{U}$ (by definition), $\pi_j(k)\in\mathscr{F}_j(k)$ for all $j\in\mathcal{J}$ and $k\in\mathbb{N}_+$. Therefore, $\pi_j\in\mathbb{F}_j$ for all $j\in\mathcal{J}$ under the RWQ policy, so it is an admissible policy. Notice that, the sequence $\{\tilde{q}_{j,m}^k(l),k\in\mathbb{N}_+,j\in\mathcal{J},m\in\mathcal{U},l\in\mathbb{N}^N\}$ corresponds to $\{\Upsilon_m,m\in\mathbb{N}\}$ defined in Remark 4 associated with the RWQ policy.

E.3. Proof of Lemma 2

E.3.1. Proof of Part 1. For all $i\in\mathcal{N}$ and $t\in[0,T]$, let

$$x_i(t):=\bar{Q}_i(0)+\Lambda_i(t)-\sum_{j\in\mathcal{J}}\int_0^t\mu_j(s)\bar{F}_{ij}(s)x_{ij}(s)\mathrm{d}s.$$

By (12d) and (C.1), $q_i^{(k)}=\phi(q_i^{(k)})$ for all $i\in\mathcal{N}$ and $k\in\{1,2\}$. By Assumption 1 and (12b), both $q_i^{(1)}$ and $q_i^{(2)}$ are solutions of the integral equation in Definition C.2 associated with the pair $\{x_i,\theta_i\}$ for all $i\in\mathcal{N}$. By Lemma C.1, part 1, $q_i^{(1)}=q_i^{(2)}=\mathcal{M}(x_i,\theta_i)$ for all $i\in\mathcal{N}$, which completes the proof.

E.3.2. Proof of Part 2. We use the proof technique introduced by Levinson (1966). We first present a preliminary result from Levinson (1966) and then prove Lemma 2. Let $L^2([0,T])$ denote the space of Lebesgue-measurable functions with domain $[0,T]$ and range $\mathbb{R}$ such that if $f \in L^2([0,T])$, then $\int_0^T f(t)^2 dt < \infty$. We let $\xrightarrow{w}$ denote weak convergence in $L^2([0,T])$ as defined in proposition 6 in Royden and Fitzpatrick (2010, section 8.2).

Lemma E.1 (Levinson 1966, lemma 2.1). *Let $\{f_r, r \in \mathbb{N}\}$ be a uniformly bounded sequence of functions in $L^2([0,T])$ such that $f_r \xrightarrow{w} f$ for some $f \in L^2([0,T])$. Let $f_u, f_l : [0,T] \to \mathbb{R}$ be defined as $f_u(t) := \limsup_{r\to\infty} f_r(t)$ and $f_l(t) := \liminf_{r\to\infty} f_r(t)$ for all $t \in [0,T]$. Then, $f(t) \le f_u(t)$ and $f(t) \ge f_l(t)$ for all $t \in [0,T]$ except on a set of zero measure.*

By (12b), (12d), and Assumption 1, $\|q_i\|_T \le \bar{Q}_i(0) + \bar{\lambda}T$ for all $i \in \mathcal{N}$. Second, x_{ij} is nonnegative and bounded for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$ by (12c) and (12d). Third, both μ_j and Λ_i are bounded processes on $[0,T]$ by Assumption 1, and w_{ij} and $\bar{F}_{ij}$ are bounded by definition for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$.

Let $x_{ij} = \mathbf{0}$ for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$, and consider the equation

$$q_i(t) = \bar{Q}_i(0) + \Lambda_i(t) - \int_0^t \theta_i(s)\phi(q_i)(s)ds, \qquad \forall i \in \mathcal{N}, t \in [0,T]. \tag{E.2}$$

By Lemma C.1, parts 1 and 2, there exists a unique solution of (E.2) such that $q_i = \mathcal{M}(\bar{Q}_i(0) + \Lambda_i, \theta_i) \ge \mathbf{0}$. Then, $q_i = \phi(q_i)$ for all $i \in \mathcal{N}$ and so $q_i = \mathcal{M}(\bar{Q}_i(0) + \Lambda_i, \theta_i)$ and $x_{ij} = \mathbf{0}$ for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$ is a feasible pair for the CLP (12). Hence, the feasible region of the CLP (12), denoted by $\mathcal{H} := \{(\boldsymbol{q}, \boldsymbol{x}) : (\boldsymbol{q}, \boldsymbol{x})$ satisfies (12b)-(12e)$\}$, is nonempty. Let

$$M := \sup_{(\boldsymbol{q},\boldsymbol{x})\in\mathcal{H}} \sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij} \int_0^T \mu_j(s)\bar{F}_{ij}(s)x_{ij}(s)ds < \infty.$$

Then, there exists a sequence of feasible matching processes for the CLP (12), denoted by $\{(\boldsymbol{q}^r, \boldsymbol{x}^r), r \in \mathbb{N}\}$, such that $(\boldsymbol{q}^r, \boldsymbol{x}^r) = \{q_i^r(t), x_{ij}^r(t), i \in \mathcal{N}, j \in \mathcal{J}, t \in [0,T]\}$ for all $r \in \mathbb{N}$ and

$$\lim_{r\to\infty} \sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij} \int_0^T \mu_j(s)\bar{F}_{ij}(s)x_{ij}^r(s)ds = M.$$

By theorem 14 in Royden and Fitzpatrick (2010, section 8.3), there exists a subsequence of $\{(\boldsymbol{q}^r, \boldsymbol{x}^r), r \in \mathbb{N}\}$ that is again denoted by $\{(\boldsymbol{q}^r, \boldsymbol{x}^r), r \in \mathbb{N}\}$ for notational convenience such that $q_i^r \xrightarrow{w} \check{q}_i \in L^2([0,T])$ and $x_{ij}^r \xrightarrow{w} \check{x}_{ij} \in L^2([0,T])$ for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$. Let $(\check{\boldsymbol{q}}, \check{\boldsymbol{x}}) := \{\check{q}_i, \check{x}_{ij}(t), i \in \mathcal{N}, j \in \mathcal{J}, t \in [0,T]\}$. By definition of weak convergence in $L^2([0,T])$, we have

$$\sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij} \int_0^T \mu_j(s)\bar{F}_{ij}(s)\check{x}_{ij}(s)ds = \lim_{r\to\infty} \sum_{i\in\mathcal{N}, j\in\mathcal{J}} w_{ij} \int_0^T \mu_j(s)\bar{F}_{ij}(s)x_{ij}^r(s)ds = M. \tag{E.3}$$

Hence, it is enough to prove that a modification of $(\check{\boldsymbol{q}}, \check{\boldsymbol{x}})$ is in $\mathcal{H}$.

First, let us consider (12b). Let us fix an arbitrary $i \in \mathcal{N}$ and $t \in [0,T]$. By definition of weak convergence, we have

$$\lim_{r\to\infty}\left(\int_0^t \theta_i(s)q_i^r(s)ds + \sum_{j\in\mathcal{J}} \int_0^t \mu_j(s)\bar{F}_{ij}(s)x_{ij}^r(s)ds\right) = \int_0^t \theta_i(s)\check{q}_i(s)ds + \sum_{j\in\mathcal{J}} \int_0^t \mu_j(s)\bar{F}_{ij}(s)\check{x}_{ij}(s)ds. \tag{E.4}$$

By (12b) and (E.4), $\boldsymbol{q}^r$ converges pointwise to some nonnegative function $\hat{\boldsymbol{q}}$ as $r \to \infty$, and $\hat{\boldsymbol{q}} = \check{\boldsymbol{q}}$ almost everywhere on $[0,T]$ by Lemma E.1.

Next, let us consider (12c) and (12d). Notice that $x_{ij}^r \xrightarrow{w} \check{x}_{ij}$ for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$ implies $\sum_{i\in\mathcal{N}} x_{ij}^r \xrightarrow{w} \sum_{i\in\mathcal{N}} \check{x}_{ij}$ for all $j \in \mathcal{J}$. Then, by Lemma E.1, for all $i \in \mathcal{N}$, $j \in \mathcal{J}$, and $t \in [0,T]$ except on a set of zero measure,

$$\sum_{i\in\mathcal{N}} \check{x}_{ij}(t) \le \limsup_{r\to\infty} \sum_{i\in\mathcal{N}} x_{ij}^r(t) \le 1, \qquad \check{x}_{ij}(t) \ge \liminf_{r\to\infty} x_{ij}^r(t) \ge 0. \tag{E.5}$$

Hence, $\check{\boldsymbol{x}}$ satisfies (12c) and (12d) for all $t \in [0,T]$ except on a set of zero measure, and we denote this set of zero measure by H and its complement by H^c, that is, $H^c := [0,T]\backslash H$. Let $\hat{\boldsymbol{x}} = \{\hat{x}_{ij}(t), i \in \mathcal{N}, j \in \mathcal{J}, t \in [0,T]\}$ be such that $\hat{x}_{ij}(t) := \check{x}_{ij}(t)$ for all $t \in H^c$, $i \in \mathcal{N}$, and $j \in \mathcal{J}$, and $\hat{x}_{ij}(t) := 0$ for all $t \in H$, $i \in \mathcal{N}$, and $j \in \mathcal{J}$.

Notice that $(\hat{\boldsymbol{q}}, \hat{\boldsymbol{x}})$ is a feasible pair for the CLP (12). To see this, $(\hat{\boldsymbol{q}}, \hat{\boldsymbol{x}})$ satisfies (12b), (12c), and (12d) by construction, and satisfies (12e) by proposition 2.11 of Folland (1999). Furthermore, $(\hat{\boldsymbol{q}}, \hat{\boldsymbol{x}})$ is an optimal solution of the CLP (12) by (E.3) and the fact that it is equal to $(\check{\boldsymbol{q}}, \check{\boldsymbol{x}})$ almost everywhere on $[0,T]$.

E.4. Proof of Lemma 4

Suppose that $w_{ij} = 1$ for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$. By (12b), (12d), and (14), we have

$$\sum_{j\in\mathcal{J}} \int_0^t \mu_j(s)\bar{F}_{ij}(s)x_{ij}(s)ds \le \Lambda_i(t), \quad \forall i \in \mathcal{N}, t \in [0,T]. \tag{E.6}$$

Then, by summing (E.6) in $i \in \mathcal{N}$, we can see that an upper bound on the objective function value of the CLP (12) is $\sum_{i\in\mathcal{N}} \Lambda_i(T)$. Any feasible matching process that satisfies condition (15) has the objective function value $\sum_{i\in\mathcal{N}} \Lambda_i(T)$ and thus is an optimal CLP solution. Moreover, by (E.6), condition (15) implies that (E.6) binds for all $i \in \mathcal{N}$ and $t \in [0,T]$ under at least one optimal CLP (12) solution.

Let $\{\tilde{\boldsymbol{q}}, \tilde{\boldsymbol{x}}\}$ be an optimal solution of the CLP (12) under which (E.6) is binding for all driver types at all times and so $\tilde{q}_i = \mathbf{0}$ for all $i \in \mathcal{N}$ by (12b) and (12d). Notice that $\tilde{\boldsymbol{x}}$ satisfies (13c) and (13d) because these constraints are exactly the same as (12c) and (12d), respectively. Because (E.6) is binding for all driver types at all times under $\tilde{\boldsymbol{x}}$, then by taking the derivatives of both the right- and left-hand sides of (E.6) with respect to t, we can see that $\{\tilde{x}_{ij}(t), i \in \mathcal{N}, j \in \mathcal{J}\}$ (or a modification of it) satisfies (13b) for all $t \in [0,T]$, so it is feasible for the LP (13) for all $t \in [0,T]$. Then,

$$\sum_{i\in\mathcal{N}, j\in\mathcal{J}} \mu_j(t)\bar{F}_{ij}(t)\tilde{x}_{ij}(t) \le \sum_{i\in\mathcal{N}, j\in\mathcal{J}} \mu_j(t)\bar{F}_{ij}(t)x^*_{ij}(t), \quad \forall t \in [0,T],$$

which implies

$$\sum_{i\in\mathcal{N}, j\in\mathcal{J}} \int_0^T \mu_j(s)\bar{F}_{ij}(s)\tilde{x}_{ij}(s)\mathrm{d}s \le \sum_{i\in\mathcal{N}, j\in\mathcal{J}} \int_0^T \mu_j(s)\bar{F}_{ij}(s)x^*_{ij}(s)\mathrm{d}s. \tag{E.7}$$

Second, for all $i \in \mathcal{N}$ and $t \in [0,T]$, let

$$\bar{X}^*_i(t) := \Lambda_i(t) - \sum_{j\in\mathcal{J}} \int_0^t \mu_j(s)\bar{F}_{ij}(s)x^*_{ij}(s)\mathrm{d}s, \quad q^*_i := \mathcal{M}(\bar{X}^*_i, \theta_i).$$

Then, $\bar{X}^*_i$ is nonnegative and nondecreasing by (13b) and Lipschitz continuous by Assumption 1. By Lemma C.1, part 2, $q^*_i \ge \mathbf{0}$ and so $\phi(q^*_i) = q^*_i$ for all $i \in \mathcal{N}$ by (C.1). Thus, $\{\boldsymbol{q}^*, \boldsymbol{x}^*\}$ satisfies (12b), satisfies (12e) by Assumption 3, and satisfies (12c) and (12d) by satisfying (13c) and (13d), respectively. Therefore, $\{\boldsymbol{q}^*, \boldsymbol{x}^*\}$ is a feasible process pair for the CLP (12). Then,

$$\sum_{i\in\mathcal{N}, j\in\mathcal{J}} \int_0^T \mu_j(s)\bar{F}_{ij}(s)x^*_{ij}(s)\mathrm{d}s \le \sum_{i\in\mathcal{N}, j\in\mathcal{J}} \int_0^T \mu_j(s)\bar{F}_{ij}(s)\tilde{x}_{ij}(s)\mathrm{d}s. \tag{E.8}$$

Hence, $\{\boldsymbol{q}^*, \boldsymbol{x}^*\}$ is an optimal solution of the CLP (12) by (E.7) and (E.8).

E.5. Proof of Lemma 5

Because λ_i and μ_j are continuous functions of the surge multipliers and s_i has a compact domain (so does p_{ij}) for all $i \in \mathcal{N}$ and $j \in \mathcal{J}$, an optimal solution of (16) exists. Let $\{p_{ij}, x_{ij}, i \in \mathcal{N}, j \in \mathcal{J}\}$ be feasible prices and matching fractions for the optimization problem (16) such that the constraint (16b) is not binding for some driver type(s). Let $\{\mathrm{s}_i, i \in \mathcal{N}\}$ be the surge multipliers corresponding to the $\{p_{ij}, i \in \mathcal{N}, j \in \mathcal{J}\}$. Let $z_i := \sum_{j\in\mathcal{J}} \mu_j \bar{F}_{ij} x_{ij}$, $z := \sum_{i\in\mathcal{N}} z_i$, $G_i := \lambda_i - z_i$, and $G := \sum_{i\in\mathcal{N}} G_i$ for all $i \in \mathcal{N}$. Then, z_i denotes the matching rate of type i drivers, z denotes the total matching rate, $G_i \ge 0$ for all $i \in \mathcal{N}$, G denotes the gap between the total driver arrival rate and the total matching rate, and $G > 0$ because the constraint (16b) is not binding for some driver type(s). We will show that the system controller can increase z and decrease G to 0 by changing the surge multipliers. This implies that there exists an optimal solution of (16) in which the constraint (16b) is binding for all driver types.

Consider an area $i \in \mathcal{N}$ such that $G_i > 0$. (Because $G > 0$, such an area exists.) If we decrease s_i, then $\sum_{j\in\mathcal{J}:\alpha(j)=i} \mu_j$ will increase and λ_i may decrease by condition (17a). Let us decrease s_i and match the additional customers arriving at area i with the excess type i drivers so that z_i increases. We should decrease s_i until all type i drivers are matched with customers because decreasing s_i more results in excess customers, which lowers z_i; that is, we should decrease s_i until G_i becomes equal to 0. Because we match the additional customers arriving at area i with the excess type i drivers, for $j \in \mathcal{J}$ such that $\alpha(j) = i$, we should increase x_{ij} and decrease x_{kj} for all $k \in \mathcal{N}\backslash\{i\}$ such that the constraints (16c) and (16d) hold.

Next, let us consider the change in z and G. If there exists $k \in \mathcal{N}\backslash\{i\}$ such that $\partial\lambda_k/\partial\mathrm{s}_i > 0$, then λ_k decreases in area k, which can decrease z_k. However, the total decrease in matching rates in all other areas is less than the increase in z_i by condition (17c). Hence, z strictly increases. When λ_k decreases, G_k does not increase, and it is possible that some of the customers matched with type k drivers may not find an available type k driver anymore and so leave the system without being matched. In such a case, we should decrease x_{kj} for some $j \in \mathcal{J}$ such that constraints (16c) and (16d) hold.

If there exists $k \in \mathcal{N}\backslash\{i\}$ such that $\partial\lambda_k/\partial\mathrm{s}_i \le 0$, then λ_k is nondecreasing. If λ_k is strictly increasing and there are unmatched customers in the system, we prefer to not to match those customers with the additional type k drivers arriving in the system for mathematical simplicity; thus, z_k stays constant, but G_k increases. However, the increase in $\sum_{k\in\mathcal{N}\backslash\{i\}} G_k$ is less than the decrease in G_i by condition (17c). Hence, G strictly decreases.

In summary, we can increase z and decrease G by decreasing s_i until G_i becomes equal to 0. By condition (17d), such an s_i exists. We propose the following algorithm, which decreases G to 0 and increases z:

Let $G_i = \max_{k\in\mathcal{N}} G_k > 0$. Decrease s_i until G_i becomes equal to 0. Then, update the surge multipliers, customer and driver arrival rates, and feasible matching fractions. If G decreases to 0, stop. Otherwise, repeat the procedure.

Under this algorithm, at each step, z strictly increases and G strictly decreases. Let Δs_i and ΔG denote the total change in s_i and G, respectively, in a step where $G_i = \max_{k\in\mathcal{N}} G_k$. Then, $G_i/(2C_2) \le |\Delta s_i| \le G_i/C_1$ by condition (17a). Moreover, $\partial G/\partial s_i < 0$ and

$$\left|\frac{\partial G}{\partial s_i}\right| \ge \left|\frac{\partial \lambda_i}{\partial s_i}\right| + \left|\sum_{j\in\mathcal{J}:\alpha(j)=i} \frac{\partial \mu_j}{\partial s_i}\right| - \sum_{k\in\mathcal{N}\setminus\{i\}} \left|\frac{\partial \lambda_k}{\partial s_i}\right| \ge C_1$$

by (17c). Because $G_i \ge G/N$,

$$|\Delta G| \ge C_1|\Delta s_i| \ge \frac{C_1 G_i}{2C_2} \ge \frac{C_1 G}{2NC_2}.$$

Therefore, G decreases $C_1/(2NC_2) \times 100\% > 0\%$ at each step, which implies that G can be made arbitrarily close to 0 in finite steps and converges to 0 as the number of steps increases to infinity.

Appendix F. Relative Compactness in Space $\mathbb{D}$

In this section, we present a relative compactness result in space $\mathbb{D}$ that we use in the proof of Proposition A.1 (see Appendix B). Although this result is known in folklore, we could not find a specific theorem to refer to; thus, we provide one.

Let $T_1 \in \mathbb{R}_+$ be an arbitrary constant. We consider $\mathbb{D}[0, T_1]$ endowed with the usual Skorokhod J_1 topology (see Billingsley 1999, chapter 3). For some $x, y \in \mathbb{D}[0, T_1]$, let $d(x, y)$ denote the Skorokhod J_1 distance between these processes (see Billingsley 1999, equation (12.13)).

Lemma F.1. *Let $\{Y^n, n \in \mathbb{N}_+\}$ be a relatively compact sequence in $\mathbb{D}[0, T_1]$ endowed with the u.o.c. topology such that all of its subsequential limits are uniformly continuous. Let $\{X^n, n \in \mathbb{N}_+\}$ be a sequence in $\mathbb{D}[0, T_1]$ such that*

$$\sup_{n\in\mathbb{N}_+} |X^n(0)| < \infty, \tag{F.1a}$$

$$|X^n(t_2) - X^n(t_1)| \le K|Y^n(t_2) - Y^n(t_1)|, \text{ for all } t_1, t_2 \in [0, T_1] \text{ and } n \in \mathbb{N}_+, \tag{F.1b}$$

where $K \in \mathbb{R}_+$ is a constant. Then, $\{X^n, n \in \mathbb{N}_+\}$ is relatively compact in $\mathbb{D}[0, T_1]$ endowed with the u.o.c. topology, and all of its subsequential limits are uniformly continuous.

Moreover, if all of the subsequential limits of $\{Y^n, n \in \mathbb{N}_+\}$ are absolutely (Lipschitz) continuous, then all of the subsequential limits of $\{X^n, n \in \mathbb{N}_+\}$ are also absolutely (Lipschitz) continuous.

Proof. Because $\{Y^n, n \in \mathbb{N}_+\}$ is relatively compact with respect to the u.o.c. topology, it is also relatively compact with respect to the Skorokhod J_1 topology. Then, by theorem 12.3 of Billingsley (1999),

$$\sup_{n\in\mathbb{N}_+} \|Y^n\|_{T_1} < \infty, \qquad \lim_{\delta\to 0} \sup_{n\in\mathbb{N}_+} w'(Y^n, \delta) = 0, \tag{F.2}$$

where w' is defined in equation (12.6) of Billingsley (1999). Then,

$$\begin{aligned}
\sup_{n\in\mathbb{N}_+} \|X^n\|_{T_1} &= \sup_{n\in\mathbb{N}_+} \sup_{0\le t\le T_1} |X^n(t)| \\
&\le \sup_{n\in\mathbb{N}_+} |X^n(0)| + \sup_{n\in\mathbb{N}_+} \sup_{0\le t\le T_1} |X^n(t) - X^n(0)| \\
&\le \sup_{n\in\mathbb{N}_+} |X^n(0)| + K \sup_{n\in\mathbb{N}_+} \sup_{0\le t\le T_1} |Y^n(t) - Y^n(0)| \qquad \text{(F.3)} \\
&\le \sup_{n\in\mathbb{N}_+} |X^n(0)| + 2K \sup_{n\in\mathbb{N}_+} \|Y^n\|_{T_1} < \infty, \qquad \text{(F.4)}
\end{aligned}$$

where the inequality in (F.3) is by (F.1b) and the strict inequality in (F.4) is by (F.1a) and (F.2). Next,

$$0 \le \lim_{\delta\to 0} \sup_{n\in\mathbb{N}_+} w'(X^n, \delta) \le K \lim_{\delta\to 0} \sup_{n\in\mathbb{N}_+} w'(Y^n, \delta) = 0, \tag{F.5}$$

where the second inequality is by the definition of w' (see Billingsley 1999, equation (12.6)) and (F.1b), and the equality is by (F.2). Therefore, $\{X^n, n \in \mathbb{N}_+\}$ is a relatively compact sequence in $\mathbb{D}[0, T_1]$ endowed with the Skorokhod J_1 topology by (F.4), (F.5), and theorem 12.3 of Billingsley (1999).

Let $\{X^{n_l}, l \in \mathbb{N}_+\}$ be an arbitrary convergent subsequence of $\{X^n, n \in \mathbb{N}_+\}$ such that $d(X^{n_l}, X) \to 0$ for some $X \in \mathbb{D}[0, T_1]$ as $l \to \infty$. Then, there exists a subsequence of $\{n_l, l \in \mathbb{N}_+\}$, denoted by $\{n_k, k \in \mathbb{N}_+\}$, such that $d(Y^{n_k}, Y) \to 0$ as $k \to \infty$, where $Y \in \mathbb{D}[0, T_1]$ and Y is uniformly continuous. Let us fix an arbitrary $\epsilon > 0$. There exits a $\delta := \delta(\epsilon, K) > 0$ such that if $|t_2 - t_1| < \delta$, $|Y(t_2) - Y(t_1)| < \epsilon/(6K)$. Because convergence in the Skorokhod J_1 metric implies u.o.c. convergence when the limit is continuous (see Billingsley 1999, p. 124), we also have $\|Y^{n_k} - Y\|_{T_1} \to 0$ as $k \to \infty$. Let Δ denote the set of continuous, strictly increasing, and bijective mappings from the domain $[0, T_1]$ onto itself. Then, there exists a $k_0 \in \mathbb{N}_+$ and a sequence $\{\alpha^k, k \in \mathbb{N}_+\}$ in the set Δ such that if $k \geq k_0$,

$$\|\alpha^k - e\|_{T_1} \vee \|X^{n_k} \circ \alpha^k - X\|_{T_1} < \frac{\epsilon}{8} \wedge \delta, \qquad \|Y^{n_k} - Y\|_{T_1} < \frac{\epsilon}{8K}, \tag{F.6}$$

where the first inequality is by the definition of the Skorokhod J_1 metric (see Billingsley 1999, equation (12.13)) and the fact that $d(X^{n_k}, X) \to 0$ as $k \to \infty$. Let us fix an arbitrary $k \geq k_0$. Then, for all $t_1, t_2 \in [0, T_1]$ such that $|t_2 - t_1| \leq \delta$,

$$\begin{aligned}
|X(t_2) - X(t_1)| &\leq \left|X(t_2) - X^{n_k}(\alpha^k(t_2))\right| + \left|X^{n_k}(\alpha^k(t_2)) - X^{n_k}(\alpha^k(t_1))\right| + \left|X^{n_k}(\alpha^k(t_1)) - X(t_1)\right| \\
&\leq \sum_{l=1}^{2}\left|X(t_l) - X^{n_k}(\alpha^k(t_l))\right| + K\left|Y^{n_k}(\alpha^k(t_2)) - Y^{n_k}(\alpha^k(t_1))\right| \\
&\leq \sum_{l=1}^{2}\left(\left|X(t_l) - X^{n_k}(\alpha^k(t_l))\right| + K\left|Y^{n_k}(\alpha^k(t_l)) - Y(\alpha^k(t_l))\right| + K\left|Y(\alpha^k(t_l)) - Y(t_l)\right|\right) + K|Y(t_2) - Y(t_1)| \\
&< 2\|X^{n_k} \circ \alpha^k - X\|_{T_1} + 2K\|Y^{n_k} - Y\|_{T_1} + \frac{\epsilon}{3} + \frac{\epsilon}{6} \\
&< \frac{\epsilon}{4} + \frac{\epsilon}{4} + \frac{\epsilon}{2} = \epsilon,
\end{aligned}$$

where the second inequality is by (F.1b), the fourth inequality is by the definition of δ and (F.6), and the last inequality is by (F.6). Therefore, X is uniformly continuous, and thus $X^{n_l} \to X$ u.o.c. as $l \to \infty$ (see Billingsley 1999, p. 124). This implies that $\{X^n, n \in \mathbb{N}_+\}$ is also relatively compact in the u.o.c. topology and all of its subsequential limits are uniformly continuous.

Next, suppose that Y is absolutely continuous. We will prove that X is also absolutely continuous. Let us fix an arbitrary $\epsilon_1 > 0$. Let $\{(a_l, b_l)\}_{l=1}^m$ be an arbitrary finite set of disjoint intervals such that $(a_l, b_l) \subset [0, T_1]$ for all $l \in \{1, 2, \ldots, m\}$. Because Y is absolutely continuous, for any $\epsilon_1 > 0$, there exists a $\delta_1 = \delta_1(\epsilon_1, K) > 0$ such that if $\sum_{l=1}^m (b_l - a_l) < \delta_1$, then $\sum_{l=1}^m |Y(b_l) - Y(a_l)| < \epsilon_1 / K$. Then

$$\begin{aligned}
&\sum_{l=1}^{m}|X(b_l) - X(a_l)| \\
&\leq \sum_{l=1}^{m}\left(|X(b_l) - X^{n_k}(b_l)| + |X^{n_k}(b_l) - X^{n_k}(a_l)| + \left|X^{n_k}(a_l) - X(a_l)\right|\right) \\
&\leq \sum_{l=1}^{m}\left(|X(b_l) - X^{n_k}(b_l)| + |X^{n_k}(a_l) - X(a_l)| + K|Y^{n_k}(b_l) - Y(b_l)| + K|Y^{n_k}(a_l) - Y(a_l)| + K|Y(b_l) - Y(a_l)|\right) \\
&\leq \sum_{l=1}^{m}\left(2\|X - X^{n_k}\|_{T_1} + 2K\|Y^{n_k} - Y\|_{T_1} + K|Y(b_l) - Y(a_l)|\right) \\
&= 2m\left(\|X - X^{n_k}\|_{T_1} + K\|Y^{n_k} - Y\|_{T_1}\right) + K\sum_{l=1}^{m}|Y(b_l) - Y(a_l)|,
\end{aligned} \tag{F.7}$$

where the second inequality is by (F.1b). By letting $k \to \infty$, the first term in (F.7) converges to 0; thus, the sum of the terms in (F.7) becomes less than ϵ_1. Hence, X is absolutely continuous.

Finally, suppose that Y is Lipschitz continuous with Lipschitz constant $\kappa \in \mathbb{R}_+$. We will prove that X is also Lipschitz continuous. For all $t_1, t_2 \in \mathbb{R}_+$,

$$\begin{aligned}
|X(t_2) - X(t_1)| &\leq \limsup_{k\to\infty}\left(|X(t_2) - X^{n_k}(t_2)| + |X^{n_k}(t_2) - X^{n_k}(t_1)| + |X^{n_k}(t_1) - X(t_1)|\right) \\
&\leq \limsup_{k\to\infty}\left(|X(t_2) - X^{n_k}(t_2)| + |X^{n_k}(t_1) - X(t_1)| + K|Y^{n_k}(t_2) - Y(t_2)| + K|Y^{n_k}(t_1) - Y(t_1)|\right) + K|Y(t_2) - Y(t_1)| \\
&\leq \limsup_{k\to\infty}\left(2\|X - X^{n_k}\|_{T_1} + 2K\|Y^{n_k} - Y\|_{T_1}\right) + K|Y(t_2) - Y(t_1)| \\
&\leq K\kappa|t_2 - t_1|,
\end{aligned} \tag{F.8}$$

where the second inequality is by (F.1b). Because the first term in (F.8) converges to 0 and Y is Lipschitz continuous, we obtain the last inequality, so X is Lipschitz continuous. □

Endnotes

[1] Intuitively, one can expect that $\mathcal{F}_j(k) = \mathcal{F}(\nu_j(k))$ for all $j \in \mathcal{J}$ and $k \in \mathbb{N}_+$. Although such a result is proved under specific measure spaces (see Karatzas and Shreve 1988, lemma 5.4.18; Stroock and Varadhan 2006, lemma 1.3.3) or under some assumptions (see Shiryaev 2008, theorem 1.6), none of them is applicable to our case.

[2] Condition (15) is parallel to assumption 1 in Harrison (2000), which introduces an optimization problem to define fully utilized resources in the queueing literature when parameters do not vary with time.

References

Aliprantis CD, Border KC (2006) *Infinite Dimensional Analysis: A Hitchhiker's Guide*, 3rd ed. (Springer, Berlin).
Anderson EJ, Nash P, Perold AF (1983) Some properties of a class of continuous linear programs. *SIAM J. Control Optim*. 21(5): 758–765.
Arnosti N, Johari R, Kanoria Y (2016) Managing congestion in matching markets. Working paper, Columbia University, New York. http://ssrn.com/abstract=2427960.
Ata B, Kumar S (2005) Heavy traffic analysis of open processing networks with complete resource pooling: Asymptotic optimality of discrete review policies. *Ann. Appl. Probab*. 15(1A):331–391.
Azevedo EM, Weyl EG (2016) Matching markets in the digital age. *Science* 352(6289):1056–1057.
Banerjee S, Freund D, Lykouris T (2016) Pricing and optimization in shared vehicle systems: An approximation framework. Preprint arXiv: 1608.06819v3, submitted August 24, 2016, https://arxiv.org/abs/1608.06819v3.
Banerjee S, Kanoria Y, Qian P (2018) The value of state dependent control in ridesharing systems. Preprint arXiv:1803.04959v1, submitted March 13, https://arxiv.org/abs/1803.04959.
Bertsimas D, Nasrabadi E, Paschalidis IC (2015) Robust fluid processing networks. *IEEE Trans. Automatic Control* 60(3):715–728.
Besbes O, Castro F, Lobel I (2018) Surge pricing and its spatial supply response. Working paper, Columbia University, New York. https://ssrn.com/abstract=3124571.
Billingsley P (1999) *Convergence of Probability Measures*, 2nd ed. (Wiley, Hoboken, NJ).
Bimpikis K, Candoğan O, Saban D (2019) Spatial pricing in ride-sharing networks. *Oper. Res*. 67(3):744–769.
Braverman A, Dai JG, Liu X, Ying L (2016) Empty-car routing in ridesharing systems. Preprint arXiv:1609.07219v3, submitted September 23, https://arxiv.org/abs/1609.07219v3.
Cachon GP, Daniels KM, Lobel R (2017) The role of surge pricing on a service platform with self-scheduling capacity. *Manufacturing Service Oper. Management* 19(3):368–384.
Castillo JC, Knoepfle DT, Weyl EG (2016) Surge pricing solves the wild goose chase. Working paper, Stanford University, Stanford, CA. https://ssrn.com/abstract=2890666.
Chaudhari HA, Byers JW, Terzi E (2018) Putting data in the driver's seat: Optimizing earnings for on-demand ride-hailing. *Proc. 11th ACM Internat. Conf. Web Search Data Mining* (Association for Computing Machinery, New York), 90–98.
Chen K, Sheldon M (2015) Dynamic pricing in a labor market: Surge pricing and flexible work on the Uber platform. Working paper, University of California, Los Angeles, Los Angeles.
Chen L, Mislove A, Wilson C (2015) Peeking beneath the hood of Uber. *Proc. 2015 Internet Measurement Conf*. (Association for Computing Machinery, New York), 495–508.
Chen H, Yao DD (2001) *Fundamentals of Queueing Networks: Performance, Asymptotics, and Optimization* (Springer, Berlin).
Dai JG, Tezcan T (2011) State space collapse in many-server diffusion limits of parallel server systems. *Math. Oper. Res*. 36(2):271–320.
de Jong RM (1996) A strong law of large numbers for triangular mixingale arrays. *Statist. Probab. Lett*. 27(1):1–9.
Durrett R (2010) *Probability: Theory and Examples*, 4th ed. (Cambridge University Press, Cambridge, UK).
Economist, The (2016) A fare shake. *The Economist* (May 14), http://www.economist.com/news/finance-and-economics/21698656-jacking-up-prices-may-not-be-only-way-balance-supply-and-demand-taxis.
Folland GB (1999) *Real Analysis Modern Techniques and Their Applications*, 2nd ed. (Wiley, Hoboken, NJ).
Gopalakrishnan R, Mukherjee K, Tulabandhula T (2016) The costs and benefits of ridesharing: Sequential individual rationality and sequential fairness. Preprint, arXiv:1607.07306v2, submitted July 25, https://arxiv.org/abs/1607.07306v2.
Guda H, Subramanian U (2019) Your Uber is arriving: Managing on-demand workers through surge pricing, forecast communication, and worker incentives. *Management Sci*. 65(5):1995–2014.
Gurvich I, Ward A (2014) On the dynamic control of matching queues. *Stochastic Systems* 4(2):479–523.
Hall J, Kendrick C, Nosko C (2016) The effects of Uber's surge pricing: A case study. Working paper, Uber, San Francisco.
Harrison JM (2000) Brownian models of open processing networks: Canonical representation of workload. *Ann. Appl. Probab*. 10(1): 75–103.
Hu M, Zhou Y (2015) Dynamic type matching. Working paper, University of Toronto, Toronto. http://ssrn.com/abstract=2592622.
Kallenberg O (1997) *Foundations of Modern Probability* (Springer, Berlin).
Karatzas I, Shreve SE (1988) *Brownian Motion and Stochastic Calculus*, 1st ed. (Springer-Verlag, Berlin).
Leshno JD (2016) Dynamic matching in overloaded waiting lists. Working paper, Columbia University, New York.
Levinson N (1966) A class of continuous linear programming problems. *J. Math. Anal. Appl*. 16(1):7–83.
Michallon C (2016) "Shame on you": Uber comes under fire as users complain of price surge in the aftermath of Manhattan explosion—with rides costing up to three times their usual prices. *Daily Mail* (September 18), http://www.dailymail.co.uk/news/article-3795474/Uber-comes-fire-users-complain-price-surge-aftermath-Manhattan-explosion.html.
Nikzad A (2017) Thickness and competition in ride-sharing markets. Working paper, University of Southern California, Los Angeles.
Özkan E (2018) Joint pricing and matching in ridesharing systems. Working paper, Koç University, Istanbul. https://ssrn.com/abstract=3217642.
Perold AF (1981) Extreme points and basic feasible solutions in continuous time linear programming. *SIAM J. Control Optim*. 19(1):52–63.
Plambeck EL, Ward AR (2006) Optimal control of a high-volume assemble-to-order system. *Math. Oper. Res*. 31(3):453–477.
Reed JE, Ward AR (2004) A diffusion approximation for a generalized Jackson network with reneging. *Proc. 42nd Annual Allerton Conf. Comm. Control Comput*. (University of Illinois, Monticello), 983–995.
Reed JE, Ward AR (2008) Approximating the GI/GI/1+GI queue with a nonlinear drift diffusion: Hazard rate scaling in heavy traffic. *Math. Oper. Res*. 33(3):606–644.
Reiman MI, Wang Q (2015) Asymptotically optimal inventory control for assemble-to-order systems with identical lead times. *Oper. Res*. 63(3):716–732.

Riquelme C, Banerjee S, Johari R (2015) Pricing in ride-share platforms: A queueing-theoretic approach. Working paper, Google, Mountain View, CA. http://ssrn.com/abstract=2568258.

Ross SM (1996) *Stochastic Processes*, 2nd ed. (Wiley, Hoboken, NJ).

Royden HL, Fitzpatrick PM (2010) *Real Analysis*, 4th ed. (Pearson, London).

Shiryaev AN (2008) *Optimal Stopping Rules* [reprint of the 1978 ed.] (Springer, Berlin).

Stroock DW, Varadhan SRS (2006) *Multidimensional Diffusion Processes* [reprint of the 1997 ed.] (Springer, Berlin).

Ünver MU (2010) Dynamic kidney exchange. *Rev. Econom. Stud.* 77(1):372–414.

Ward AR, Kumar S (2008) Asymptotically optimal admission control of a queue with impatient customers. *Math. Oper. Res.* 33(1): 167–202.

Weiss G (2008) A simplex based algorithm to solve separated continuous linear programs. *Math. Programming* 115(1):151–198.

White D (2016) Uber users are complaining about pricey New Year's Eve rides. *Time* (January 1), http://time.com/4165410/uber-new-years-eve-price-surge-rides/.

Whitt W (2002) *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues* (Springer, Berlin).

Zhong Y, Wan Z, Shen ZJM (2019) Balancing supply and demand: Queuing vs. surge pricing mechanisms. Working paper, University of Chicago, Chicago.