Manufacturing & Service Operations Management

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

Shared Mobility for Last-Mile Delivery: Design, Operational Prescriptions, and Environmental Impact

Wei Qi, Lefei Li, Sheng Liu, Zuo-Jun Max Shen

MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Articles in Advance, pp. 1–15
http://pubsonline.informs.org/journal/msom/
ISSN 1523-4614 (print), ISSN 1526-5498 (online)

Shared Mobility for Last-Mile Delivery: Design, Operational Prescriptions, and Environmental Impact

Wei Qi,[a] Lefei Li,[b] Sheng Liu,[c] Zuo-Jun Max Shen[c,d]

[a] Desautels Faculty of Management, McGill University, Montreal, Quebec, Canada H3A 1G5; [b] Department of Industrial Engineering, Tsinghua University, Beijing 100084, China; [c] Department of Industrial Engineering and Operations Research, University of California, Berkeley, Berkeley, California 94720; [d] Department of Civil and Environmental Engineering, University of California, Berkeley, Berkeley, California 94720

Contact: wei.qi@mcgill.ca, http://orcid.org/0000-0003-3948-835X (WQ); lilefei@tsinghua.edu.cn (LL); lius10@berkeley.edu, http://orcid.org/0000-0003-2365-6013 (SL); maxshen@berkeley.edu, http://orcid.org/0000-0003-4538-8312 (Z-JMS)

Received: January 11, 2016
Revised: August 3, 2016; May 16, 2017; September 27, 2017
Accepted: September 28, 2017
Published Online in Articles in Advance: June 7, 2018

https://doi.org/10.1287/msom.2017.0683

Abstract. Two socioeconomic transformations, namely, the booms in the sharing economy and retail e-commerce, lead to the prospect where shared mobility of passenger cars prevails throughout urban areas for home delivery services. Logistics service providers as well as local governments are in need of evaluating the potentially substantial impacts of this mode shift, given their economic objectives and environmental concerns. This paper addresses this need by providing new logistics planning models and managerial insights. These models characterize open-loop car routes, car drivers' wage-response behavior, interplay with the ride-share market, and optimal sizes of service zones within which passenger vehicles pick up goods and fulfill the last-mile delivery. Based on theoretical analysis and empirical estimates in a realistic setting, the findings suggest that crowdsourcing shared mobility is not as scalable as the conventional truck-only system in terms of the operating cost. However, a transition to this paradigm has the potential for creating economic benefits by reducing the truck fleet size and exploiting additional operational flexibilities (e.g., avoiding high-demand areas and peak hours, adjusting vehicle loading capacities, etc.). These insights are insignificantly affected by the dynamic adjustment of wages and prices of the ride-share market. If entering into this paradigm, greenhouse gas emissions may increase because of prolonged car trip distance; on the other hand, even exclusively minimizing operating costs incurs only slightly more emissions than exclusively minimizing emissions.

Funding: This study was partially supported by the National Science Foundation [Grant CMMI 1265671].
Supplemental Material: The online appendix is available at https://doi.org/10.1287/msom.2017.0683.

Keywords: logistics planning • shared mobility • open vehicle routing problem • sustainability

1. Introduction

California, along with many other regions in the world, is experiencing two socioeconomic transformations: (1) *The sharing economy*, in which people share access to goods and services, is prospering. The market size of the sharing economy has the potential for increasing global revenues from roughly $15 billion today to around $335 billion by 2025 (PWC 2015). One of the key sectors being impacted is transportation, with leading companies (e.g., Uber, Lyft, and Car2Go) disrupting the traditional business of people-riding transit. (2) Meanwhile, *retail e-commerce* sales is expanding rapidly worldwide, with an annual growth rate of around 20%, and is projected to reach $2.4 trillion in 2017 according to eMarketer (2014). To advance their logistics competitiveness, some major players piloted new delivery services (e.g., Amazon Prime Now and Google Express, etc.) over the past several years.

A question that naturally arises from these phenomena but has not been well addressed is the following: Is it possible for retail e-commerce to rely on crowdsourced shared mobility for delivery services on a large scale? Botsman (2014) anticipates that the power of such a mode is that

> it does not require the asset-heavy infrastructure of warehouses, vehicle fleets, fuel costs, and employed drivers that traditional logistics companies have to pay for and manage . . . it's an asset-light model, akin to the likes of Uber and Airbnb, with low overheads meaning it can scale relatively fast depending on demand.

Can this marriage of the sharing economy and e-commerce really have all these claimed merits? Until recently, this practice is still in its infancy. On the one hand, most of the existing leading companies and promising start-ups in the courier industry primarily rely on professional people and dedicated fleets (e.g., 95% of Zipments' couriers are professionals). On the other hand, some companies have begun to experiment with crowdsourced delivery (e.g., Amazon Flex, UberEATS, and start-ups including Postmates and Deliv), yet on a small scale and/or focusing on express deliveries.

The goal of this paper is to evaluate whether and how to foster the large-scale adoption of shared mobility for last-mile home delivery services. The evaluation is mainly from the perspective of a logistics service provider, who minimizes the costs of both operating a fleet of short-haul trucks for bulk transportation and outsourcing shared mobility for the last-mile delivery. Also considered is the perspective of a local government, which aims to abate greenhouse gas (GHG) emissions. This responsibility, for example, is particularly relevant to the California state government, given its strategies to develop multimodel integrated freight systems and to create car sharing programs (CFMP 2014) as well as for its commitment to a 15% reduction (from 2010 levels) of GHG emissions to achieve 1990 levels by 2020 as dictated in Assembly Bill 32 (2006).

The first part of the paper develops planning models for this prospective mode of sharing logistics. The models are based on a one-transshipment logistics setting: a fleet of short-haul trucks are dispatched from a depot and unload goods at terminals of service zones. Passenger cars nearby with available mobility are attracted to each terminal, each picking up a ration of goods and delivering them within the service zone to their destinations. Contrasting the conventional mode where dedicated vehicles travel closed-loop and back-to-back routes, a salient feature of shared mobility is its one-way and one-shot nature: a car starts an outbound trip by approaching its first demand destination and the service ends once it drops off the last package. To characterize these properties, a continuous approximation (CA) model is developed, along with a lower-bound analysis to establish the suboptimality of the approximation. Then a wage response model and a model of synergy and competition with the ride-share market are derived to determine the wages paid to car drivers and the level of the supply of shared mobility. Combining these models results in service zone deployment in the form of the optimal service zone sizes as well as the associated operating cost. The second part of the paper presents managerial implications of this sharing logistics paradigm, concerning its economic scalability, asset lightness, operational prescriptions, robustness to dynamic adjustment of wages and prices of the ride-share market, and environmental impacts. A case study calibrates the optimal design with empirical parameter estimates in a setting of 15 zip-code areas in the east San Francisco Bay Area, considering various operating scenarios.

The contributions of this paper are as follows:

(1) To the best of the authors' knowledge, this work is the first attempt to design and analyze the logistics system that features large-scale integration of shared mobility for home delivery services, using analytical models and empirical parameter estimates. In particular, the asymmetry of passenger cars' one-way routes invalidates existing CA models for vehicle routing problems (VRP); the CA model in this work and the lower-bound construction fill this gap in the literature. The wage response model and the model of interplay with the ride-share market can be useful for studies of dynamically compensating drivers for delivery services.

(2) The paper finds that crowdsourcing shared mobility is not as economically scalable as the conventional truck-only system in terms of the operating cost, unless the pool size of shared mobility keeps pace with the increase in demand density. The difference in scalability is driven by their different payment structure. Unlike the conventional truck-only system where drivers are hired and paid a fixed rate, the payment to crowdsourced drivers is route specific and accounts for the competition with the ride-share service market.

(3) A major finding is that the value of shared mobility is not immediate operating cost savings, but rather its ability to reduce the truck fleet size and the additional operational flexibilities it creates that can significantly enhance the cost-efficiency of the whole package distribution system. Specifically, the paper identifies four operational prescriptions: (i) to pilot this business model in low-demand-density areas since it is not as economically scalable as the conventional truck-only system; (ii) to replace regular-size delivery vans with heavy-duty trucks to enhance inbound-trucking efficiency; (iii) to deliver during off-peak hours to pay lower wages to car drivers; and (iv) to load more packages in cars to exploit the economies of scale of last-mile trips and also dampen the competition with the ride-share market. The case study shows that collectively implementing the last three prescriptions can significantly reduce the operating cost. Moreover, these insights are insignificantly affected by dynamically adjusting wages and prices in the ride-share market.

(4) The paper identifies a twofold environmental impact: (i) This paradigm may incur more GHG emissions, because the emission efficiency gains from cars' open-loop routes and smaller per-km emission rate are offset by the demerit that the total car trip distance is much longer than that by trucks because of cars' smaller capacity. (ii) Even exclusively minimizing operating costs does not significantly increase emissions relative to the minimum level of emissions, because the cost objective function and the emission objective function share a similar transshipment structure that is robust to the service zone sizing decisions.

The remainder of this paper is organized as follows: Section 2 reviews the related literature; Section 3 introduces the basic model settings; Section 4 develops routing and wage response models of shared mobility and the model of synergy and competition effects of the ride-share market; Section 5 presents the shared-mobility logistics planning model; Section 6 discusses managerial implications; Section 7 concludes the paper. Additional technical proofs, empirical parameter estimates, and numerical results are provided in the appendices of this paper (provided in the online appendix).

2. Literature Review

Large-scale adoption of shared mobility for retail e-commerce is an important prospect, but has not received wide attention in literature. The vast majority of the studies concerning the impact of the sharing economy on the transportation sector have focused on the transportation of passengers instead of packages. One group of these studies investigates vehicle sharing, which refers to the case where people access mobility services by temporary usage without ownership of the vehicles. For example, Bellos et al. (2017) study the economic and environmental implications of the car sharing business and find that its overall environmental impact can be negative because of aggregate vehicle usage. Shu et al. (2013) and Kabra et al. (2018) study the deployment and operations problems arising in bike-sharing systems, employing network flow models, and econometric methods, respectively. He et al. (2017) notice the recent development of one-way electric vehicle sharing (e.g., Car2Go) and formulate a service region design problem. Another stream of literature is concerned with ride sharing, which refers to a different transportation mode where a passenger accesses a vehicle for a trip with its driver already en route. See Furuhata et al. (2013) for a comprehensive review of ride-sharing management and Agatz et al. (2012) for a review of an important operations problem—the dynamic matching of shared rides. Recently, Cachon et al. (2017) explore surge pricing in ride sharing and find that all stakeholders can be better-off with surge pricing.

By contrast, studies on goods logistics with shared mobility are few and mostly qualitative. Carbone et al. (2015) identify different types of logistics coexisting in the sharing economy based on an exploratory analysis of 32 cases. Rougès and Montreuil (2014) examine 18 businesses. They argue that the two major obstacles to the development of crowdsourced delivery are trust building and a "chicken-and-egg" dilemma: a critical mass of couriers is needed to insure quality delivery and to attract customers, but a critical mass of customers is needed to attract couriers. However, the analysis in this paper shows that a major obstacle may be that the operating costs are not competitive without further exploitation of the operational flexibilities. Recently, Li et al. (2014 and 2016) propose a conceptual taxi-based people-freight sharing system. They formulate a routing problem as a mixed-integer linear program and propose a neighborhood search heuristic solution algorithm. Kafle et al. (2017) design a crowdsource-enabled parcel delivery system where cyclists and pedestrians are undertaking delivery and pick-up jobs. They assume a bidding procedure on the crowdsourced side and model the carrier's decision as a mixed-integer nonlinear program, which is solved with a Tabu search-based algorithm. By comparison, this paper develops continuous approximation models for a home delivery logistics system, with emphasis on its overall economic and environmental implications.

Logistics planning with environmental sustainability considerations has been studied in contexts other than the sharing economy. Cachon (2014) models the dependence of operating costs and GHG emissions on the density of retail stores and finds that improving consumer fuel efficiency can be an effective way of mitigating emissions. Carlsson et al. (2016) find that household-level economies of scale in transportation (i.e., a person performs many errands in a single trip) may actually increase the overall carbon footprint relative to the case of home delivery services. The service region deployment problem considered in this paper is based on CA models in a one-transshipment setting. CA methods for one-transshipment logistics systems are investigated in Xie and Ouyang (2015) with detailed discussion on the approximation error, and in Schaefer and Konur (2015) with environmental considerations. Carlsson and Jia (2013, 2014) find the asymptotically optimal configurations for one-transshipment and multitransshipment logistics systems considering both fixed costs and transportation costs. Again, the key distinction of this paper is the inclusion of shared mobility.

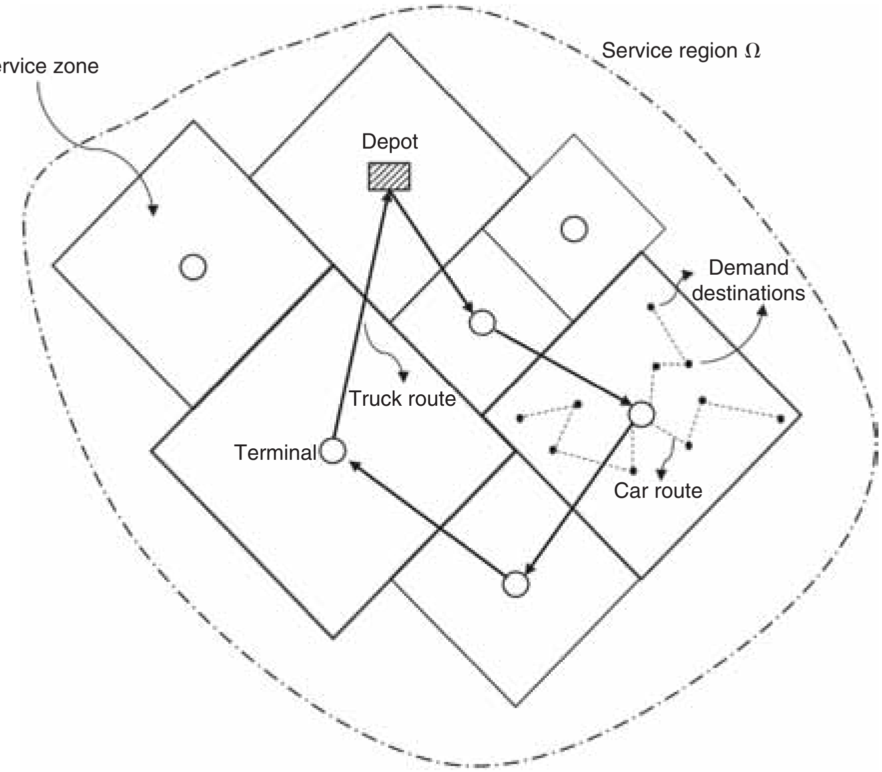
3. Model Settings

3.1. Topology

Consider a one-transshipment system that a logistics service provider operates. A *depot* is located in a *service region* Ω. At the *inbound* stage, a fleet of short-haul trucks or vans (hereafter referred to as trucks) are dispatched from the depot, each carrying a truck load of packages, and finally return to the depot. A truck stops at multiple *terminals*, where it unloads packages to be delivered in *service zones*. Each terminal is located at the center of a service zone. Service zone sizes and terminal locations can change from dispatch to dispatch in response to realized package delivery demands. Figure1 illustrates the route of a truck that visits four service zones. The zones are tiled to cover the whole service region and are of squared diamond shape.

At the *outbound* stage, multiple passenger cars nearby with available mobility (hereafter referred to as cars) are attracted to each terminal in real time to pick up all the packages. Then they deliver them to their *destinations* within a service zone. Manhattan metric of

Figure 1. An Illustration of a One-Transshipment Logistics System for Home Delivery Services with Shared Mobility of Passenger Cars (Only One of Multiple Truck Routes Is Displayed)



distance (l_1 norm) is used throughout the paper. Such a transshipment setting captures the fundamental tradeoff: to carry a truck load of goods to travel a certain distance, a truck is more cost-efficient than a cohort of cars of equal capacity; however, to carry one kilogram of goods to travel a certain distance, a car is more cost-efficient than a truck.

3.2. Demands and Mobility Availability

The demand destinations constitute a nonhomogenous point process. The mean density of the destinations is $n(x)$ at the point coordinates $x = (x_1, x_2)$. The random package weight for a destination at x is independently distributed with mean $g(x)$. Also assume that available cars are distributed with density $m(x)$. Summaries of parameter notation and estimates are available in Appendices A and E in the online appendix, respectively. Symbols throughout this paper are subscripted with a vehicle type ("t" for trucks and "c" for cars) wherever necessary.

3.3. Objective

The logistics service provider wants to minimize its operating cost. The operating cost consists of the wages paid to car drivers and the cost of dispatching its own truck fleet (including the depreciation cost as the amortized truck fleet investment cost). Therefore, the logistics service provider needs to evaluate the vehicle routing decisions and the payment scheme to attract shared mobility from the ride-share service market. The logistics service provider also needs to determine the cost-minimizing service zone sizes. Integrating these considerations forms the basis of understanding whether and how to foster large-scale adoption of shared mobility for last-mile delivery. The following sections elaborate on these aspects.

4. Shared Mobility: Routing and Wage Response

4.1. Terminal-Bound Car Trips

Assume that the profiles of demands and shared mobility availability (i.e., $n(x)$, $g(x)$ and $m(x)$) are slow varying continuous functions of x. That is, although $n(x)$, $g(x)$, and $m(x)$ may vary significantly over the entire service region, their values at locations within a service zone can be approximated by their values at the terminal of that zone. Then each service zone of area $A(x)$ with its terminal at x on average requires $m_z(x) = g(x)A(x)n(x)/v_c$ nearest cars to deliver packages within the zone, since $g(x)A(x)n(x)$ is the average weight of packages of the zone and the cars are assumed to be loaded to capacity v_c. These cars are approximately uniformly scattered in a local squared diamond of area $m_z(x)/m(x)$ centering the terminal.

Their mean terminal-bound distance, d_{TB}, is the mean distance between a point in this diamond and the center, computed by integration as follows (location point index x is suppressed wherever the discussion holds for all locations in the remainder of the paper):

$$d_{TB} = \int_0^{(1/2)\sqrt{m_z/m}} \sqrt{2}r \frac{8r}{m_z/m}\,dr = \frac{\sqrt{2}}{3}\sqrt{\frac{m_z}{m}}. \qquad (1)$$

4.2. Outbound Car Trips

Assume that the time window for picking up packages is short, such that car drivers do not take back-to-back delivery trips. In fact, truck drivers staying for too long at terminals can be costly, because the wages paid to them is a considerable portion of the operating cost (see Appendix E.2 in the online appendix). Also implicitly assumed is that enough cars are available to pick up packages with each making one trip, that is, $m > m_z/A = gn/v_c$. This assumption holds for practical parameter values (see Appendix E.3 in the online appendix). As a result, cars' outbound trips are open loop: they do not return to the terminal once finishing delivering the last package. The optimal open-loop routes are the solution to the open vehicle routing problem (OVRP) (Li et al. 2007). Aggregating all the open-loop trips in a zone, an approximation of the total outbound trip distance is given as follows:

Result 1. The total outbound trip distance of shared mobility in a service zone is given by

$$d_{OVRP} = \left(\frac{\sqrt{2}}{3} - \alpha\left(\frac{v_c}{g}\right)^{\beta}\left(1 - \frac{1}{\gamma\sqrt{An}}\right)\right)\frac{gA^{1.5}n}{v_c} + \lambda A\sqrt{n}, \qquad (2)$$

where $\alpha = 0.055$, $\beta = 0.470$, $\gamma = 0.374$, and $\lambda = 1.005$.

In (2), the first term is associated with the *line-haul* portion of the car routes. That is, a car travels the shortest path from the terminal to the first destination of those on its route. The second term is associated with the *detour* portion of the route to visit all the remaining destinations. Developing (2) is based on solving 9,600 numerical instances. The details are available in Appendix B in the online appendix.

4.3. Structural Properties of OVRP

The OVRP is a key component of the shared mobility model. The nonreturn property distinguishes the OVRP from the classical vehicle routing problem (VRP). In the previous subsection, the CA form of the OVRP solution approximates an upper bound of the optimal distance. In what follows, two lower bounds are proposed, assuming vehicles are loaded to capacity. Although this assumption loses some generality, it reflects the setting where drivers are induced to be fully engaged in delivery services. With slight abuse of notation, let X^n denote the set of n demand points in a plane of unit area, where the terminal is located at the center. Let $TSP(X^n)$, $VRP(X^n)$, and $OVRP(X^n)$ denote the lengths of the optimal traveling salesman problem (TSP), VRP, and OVRP tours traversing points in X^n, respectively. Subscript k is used to denote the maximum number of packages loaded by each vehicle. And $\lceil\cdot\rceil$ is the ceiling function.

Proposition 1. *If vehicles are loaded to capacity in the OVRP,*

$$TSP(X^n) - \left\lceil\frac{n}{k}\right\rceil r_{\max}^n \le VRP_k(X^n) - \left\lceil\frac{n}{k}\right\rceil r_{\max}^n \le OVRP_k(X^n),$$

where $r_{\max}^n$ is the maximum distance from points in X^n to the center.

Proposition 2. *If vehicles are loaded to capacity in the OVRP,*

$$\begin{aligned} &TSP(X^n) - \tfrac{1}{2}TSP(X^{\lceil n/k\rceil}) - r_{\max}^n \\ &\quad \le VRP_{2k}(X^n) - \tfrac{1}{2}TSP(X^{\lceil n/k\rceil}) - r_{\max}^n \le OVRP_k(X^n), \end{aligned}$$

where $X^{\lceil n/k\rceil}$ is the set of end demand points in the OVRP solution.

Proofs of these two propositions construct feasible VRP solutions from the optimal OVRP solution (see Appendix C in the online appendix). Numerical tests with wide range of parameter values find that the median relative optimality gap is 23.45%, with a 75th percentile of 24.2%. Following Proposition 2, an asymptotic result of the OVRP is developed below when demand points in X^n are independent and identically distributed with finite expected distance $E(r)$ to the center.

Proposition 3. *If vehicles are fully loaded in the OVRP and k is constant,*

$$\lim_{n\to\infty}\frac{1}{n}OVRP_k(X^n) = \frac{E(r)}{k} \quad (a.s.).$$

This limit is a half of its counterpart for the VRP derived by Haimovich and Kan (1985).

4.4. Wage Response

The logistics service provider crowdsources shared mobility by paying trip-specific wages to car drivers, who bear vehicle operating costs (of fuel, maintenance, depreciation, etc.) by themselves. A car driver is willing to participate in a delivery service only if the payment, after discounted by considerations of faster vehicle depreciation and more physical effort, is at least the amount that he or she can otherwise expect to earn by providing shared rides to passengers. Therefore, efficient payment should be based on the expected revenue out of ride share services, as derived as follows.

Packages uploaded at each terminal are to be delivered along open-loop car routes, which are offered to nearby idling car drivers. Consider an offer of payment $w(d_o)$ that entails d_o km of outbound trip distance. If a car driver who is r km away from the terminal decides to accept this offer, the estimated total time of fulfilling the delivery service is $\tau = r/s + d_o/s_o$, where s denotes the normal cruise speed of a car and s_o denotes the speed during outbound delivery services ($s_o < s$). Alternatively, if the driver is engaged in ride-sharing services during the time window $(0, \tau)$, then the driver is either carrying passengers or waiting for a new passenger. When carrying passengers, the driver receives payments from a ride-share company. Typically, the ride fare consists of three parts: base fare, per-mile fare, and per-minute fare. A passenger ride of duration τ_0 thus generates income of $w_b + w_m\tau_0$, where w_b is the base fare and w_m is in terms of dollars per minute but incorporates both the per-minute fare and the per-mile fare, since the travel time is assumed to be proportional to the travel distance. Also assume that the waiting time and passenger ride time are exponentially distributed with means $1/\mu$ and $1/\nu$, respectively. The driver's expected income of sharing rides $w_p(\tau)$ over $(0, \tau)$ is given by the following (the proof is available in Appendix D in the online appendix):

$$w_p(\tau) = \frac{\mu(w_b + w_m/\nu)}{\mu+\nu} \cdot \left(\nu\tau + \frac{\mu}{\mu+\nu}(1-\exp(-(\mu+\nu)\tau))\right). \qquad (3)$$

A driver's choice between delivering packages and waiting for ride requests boils down to comparing $w_p(\tau)$ and the offered wage $w(d_o)$. Since $\tau = r/s + d_o/s_o$, a driver will accept the offer if $w(d_o)/\eta \geq w_p(r/s + d_o/s_o)$, where $\eta > 1$ is the discount factor that captures the additional disutility shouldered by the driver because of faster vehicle depreciation and more physical effort when engaged in delivery services. Therefore, for each pair of (r, d_o), the minimum payment is $\eta w_p(r/s + d_o/s_o)$.

Consider the practical setting where the logistics service provider broadcasts offers simultaneously to nearby drivers with a uniform pricing scheme. In other words, independent from drivers' individual distances to the terminal, each offer with outbound trip length d_o is with a wage $w(d_o) = \eta w_p(r_c/s + d_o/s_o)$, where r_c is the terminal-bound trip distance such that drivers accept the offer if being within r_c km from the terminal, or reject the offer otherwise. To determine r_c, notice that the total number of drivers attracted to the terminal is $m\int_0^{r_c/\sqrt{2}} 8r\,dr = 2m(r_c)^2$. To ensure the supply of $m_z = gAn/v_c$ drivers requires $2m(r_c)^2 \geq m_z$. It follows immediately that the payment-minimizing $r_c^* = \sqrt{m_z/(2m)} = \sqrt{gAn/(2mv_c)}$, when the constraint is binding. Consequently, the following formula specifies the payment scheme to induce sufficient drivers for delivery services:

Result 2. For a delivery service with outbound trip length d_o, the wage paid to a car driver is

$$\begin{aligned} w^*(d_o) &= \eta w_p\left(\frac{r_c^*}{s} + \frac{d_o}{s_o}\right) = \eta\frac{\mu(w_b\nu + w_m)}{\mu+\nu}\sqrt{\frac{gAn}{2mv_c}}\frac{1}{s} \\ &\quad + \eta\frac{\mu(w_b\nu + w_m)}{\mu+\nu}\frac{d_o}{s_o} + \eta\frac{\mu^2(w_b\nu + w_m)}{\nu(\mu+\nu)^2} \\ &\quad \cdot \left(1 - \exp\left(-(\mu+\nu)\left(\sqrt{\frac{gAn}{2mv_c}}\frac{1}{s} + \frac{d_o}{s_o}\right)\right)\right). \qquad (4) \end{aligned}$$

In the wage response model, the first term and the third term excluding the exponential part represent the base fare for the delivery offer; the second term is proportional to the outbound trip length d_o and the exponential part of the third term is a nonlinear fare as a function of d_o.

4.5. Synergy and Competition with the Ride-Share Market

The operations of crowdsourcing shared mobility is by no means in isolation. In fact, the ride-share market and the logistics service provider compete for car drivers, while car drivers compete for providing both ride-share and delivery services.

To characterize such interplay among drivers and the two service markets, two key parameters that are thus far exogenous should instead be endogenous to the model. The first key parameter is m, the spacial density of available cars. Parameter m is bounded from above by the potential pool size of shared mobility, that is, the density of cars subscribed for shared use, denoted by $\bar{m}$. Meanwhile, m is the real-time supply of shared mobility on the road, and is thus induced by a driver's average earning rate (which is the same across the two service markets), measured by $\mu(\nu w_b + w_m)/(\mu+\nu)$ (since $(\nu w_b + w_m)$ is the mean earning rate in a ride-share service and $\mu/(\mu+\nu)$ is the proportion of the time that a driver is in service). Therefore, m is endogenously determined by

$$m = \bar{m}F\left(\frac{\mu(\nu w_b + w_m)}{\mu+\nu}\right), \qquad (5)$$

where $F(y)$ is a nondecreasing continuous function and satisfies $\lim_{y\to 0^+} F(y) = 0$ and $\lim_{y\to+\infty} F(y) = 1$.

The second key parameter is μ, the ride-share service request rate that each driver faces. Parameter μ depends on three quantities: (1) $\bar{\mu}$, the total ride-share demand rate per-unit area; (2) m, the density of available cars competing for providing services; and (3)

ng/v_c, the density of those available cars that delivery services cannibalize. Therefore, this relationship is modeled as

$$\mu = \frac{\bar{\mu}}{m - ng/v_c}, \tag{6}$$

where $m > ng/v_c$ always holds under the assumption that the potential pool size of shared mobility is sufficient for delivery services, that is, $\bar{m} > ng/v_c$.

Given the exogenous parameters $\{\bar{m}, \nu, w_b, w_m, \bar{\mu}, n, g, v_c\}$, as well as the functional form of $F(\cdot)$, simultaneously solving (5) and (6) yields the values of m and μ at equilibrium, denoted by $\tilde{m}$ and $\tilde{\mu}$, respectively. The continuity and the monotonicity of $F(\cdot)$ ensure the existence and the uniqueness of $\tilde{m}$ and $\tilde{\mu}$.

As a result, the logistics service provider should use these equilibrium values to account for the effects that the ride-share service market has on the delivery service market. These effects have elements of both competition and synergy, as the following proposition formally states:

Proposition 4. *If* $F(\cdot)$ *is differentiable, then* (i) $\mathrm{d}\tilde{\mu}/\mathrm{d}\bar{\mu} > 0$ (*competition effect*) *and* (ii) $\mathrm{d}\tilde{m}/\mathrm{d}\bar{\mu} > 0$ (*synergy effect*).

To explain, the *competition* effect is that, as the total ride-share demand ($\bar{\mu}$) increases, the equilibrium rate of ride-share service requests $\tilde{\mu}$ will also increase, resulting in higher expected earning rate of drivers. It will be more expensive for the logistics service provider to compete with ride-share companies to attract drivers for delivery services. On the other hand, the *synergy* effect offsets the competition effect: a higher expected earning rate will induce more drivers on the road. The logistics service provider favors this increase of mobility supply, because the expected inbound car trip (which the wage payment covers) will be shorter, and because the competition among drivers for services partially offsets the increase of $\tilde{\mu}$ and their earning rate. The proof of Proposition 4 as well as numerical tests of these effects on the total operating cost is available in Appendix D in the online appendix.

5. The Shared-Mobility Logistics Planning Problem

The following sections describe the cost models of the shared-mobility component (Section 5.1) and the trucking component (Section 5.2) of this transshipment logistics system. Striking a balance between these cost components yields a closed-form expression of the optimal zone sizes and the associated operating cost (Section 5.3), which can be compared with the optimal cost of the conventional truck-only scenario (Section 5.4).

5.1. Cost Density of Shared Mobility

Integrating the routing and the wage response models in Section 4 leads to an approximation of the cost of crowdsoucing shared mobility for last-mile delivery. Specifically, define the cost density of shared mobility, denoted by $W_c(x, A(x))$, to be the ratio of the cost of hiring shared mobility in the service zone that has a terminal at x and area $A(x)$ to $A(x)$. Mathematically,

$$W_c(x, A(x)) = \frac{\int_{A(x)} w^*(d_o, x) n_o(d_o, x)\, dd_o}{A(x)}, \tag{7}$$

where $n_o(d_o, x)$ is the product of $m_z(x)$, the number of cars hired in the zone centering around x, and the probability density function of cars' outbound trip lengths d_o in the zone. Directly estimating $n_o(d_o, x)$ for all pairs of d_o and x is computationally challenging, entailing solving large amounts of OVRP problems with randomized realizations of demand destinations. Instead, consider linearizing the wage formula of $w^*(d_o, x)$ in (4) with respect to d_o by dropping the exponential part (i.e., $\exp(-(\mu+\nu)(\sqrt{gAn/(2mv_c)}(1/s) + d_o/s_o)))$ in the third term. This linearization increases $w^*(d_o, x)$, yet by a negligible degree (numerical justification is available in Appendix E.4 in the online appendix). With the linearized wage response model, only the integral values of $\int_{A(x)} n_o(d_o, x)\, dd_o$ and $\int_{A(x)} d_o n_o(d_o, x)\, dd_o$ are necessary, which the following two identities give:

$$\int_{A(x)} n_o(d_o, x)\, dd_o = m_z(x) = \frac{g(x)A(x)n(x)}{v_c}, \tag{8a}$$

$$\int_{A(x)} d_o n_o(d_o, x)\, dd_o = d_{OVRP}, \tag{8b}$$

where the closed-form expression of d_{OVRP} is given by Result 1. The cost density can thus be computed in closed-form as follows:

Result 3. Based on identities (8a) and (8b) and the linearization of (4), the cost density $W_c(x, A(x))$ in (7) can be expressed as

$$W_c(x, A(x)) = \Phi_c(x)\sqrt{A(x)} + \Psi_c(x), \tag{9}$$

where

$$\Phi_c(x) = \eta\left(\frac{g(x)\tilde{\mu}}{v_c}\sqrt{\frac{g(x)n(x)}{2\tilde{m}v_c}}\frac{w_b\nu + w_m}{s(\tilde{\mu}+\nu)} + \frac{\tilde{\mu}(w_b\nu + w_m)}{s_o(\tilde{\mu}+\nu)} \cdot \left(\frac{\sqrt{2}}{3} - \alpha\left(\frac{v_c}{g(x)}\right)^{\beta}\right)\frac{g(x)}{v_c}\right)n(x), \tag{10a}$$

$$\Psi_c(x) = \frac{\eta g(x)\tilde{\mu}^2(w_b\nu + w_m)n(x)}{v_c\nu(\tilde{\mu}+\nu)^2} + \frac{\eta\tilde{\mu}(w_b\nu + w_m)}{s_o(\tilde{\mu}+\nu)} \cdot \left(\alpha\left(\frac{v_c}{g(x)}\right)^{\beta}\frac{\sqrt{n(x)}g(x)}{\gamma v_c} + \lambda\sqrt{n(x)}\right). \tag{10b}$$

5.2. Cost Density of Inbound Trucking

At the inbound stage, each truck carries a full truck load (i.e., v_t kilograms) of packages, unloads them at multiple terminals, and then returns to the depot. The detour portion is thus a trip on diamond-shape lattices. The detour trip length in a zone of area $A(x)$ is $\sqrt{2A(x)}$. Hence, the local detour trip density at x is $\sqrt{2A(x)}/A(x) = \sqrt{2/A(x)}$, and the total detour trip length over the service region is $d_{t,d} = \int_\Omega \sqrt{2/A(x)}\, dx$.

The line-haul portion consists of the trips both before and after the detour trip. Following a similar discussion in Daganzo (1984) for this routing scenario, to minimize the total line-haul truck trip length $d_{t,l}$ in the region, arrange zones assigned to each truck into an area such that this area is elongated toward the depot. A truck first visits the terminal that is the nearest to the depot, and then moves forward to visit the rest of the first half of the terminals. On its return trip, the truck visits the second half of the terminals. Consequently, the average departing or returning line-haul trip (i.e., one half of the line-haul round-trip) plus a one-way trip halfway into this elongated area (i.e. a quarter of the detour round-trip) approximates an average distance from the depot to a point in this area. This relationship still holds if aggregating all those elongated areas, that is, $\frac{1}{2} \times$ (total line-haul trip length $d_{t,l}$) $+ \frac{1}{4} \times$ (total detour trip length) $\approx$ (number of dispatched trucks) $\times$ (package-weight-weighted mean distance from the depot to a point in the service region), or, mathematically, $\frac{1}{2} d_{t,l} + \frac{1}{4}\int_\Omega \sqrt{2/A(x)}\, dx = ((\int_\Omega n(x)g(x)\,dx)/v_t) \times ((\int_\Omega r_d(x)n(x)g(x)\,dx)/(\int_\Omega n(x)g(x)\,dx))$, where $r_d(x)$ is the distance from the depot to point x. It immediately follows that $d_{t,l} = \int_\Omega (2g(x)r_d(x)n(x))/v_t\, dx - \frac{1}{2}\int_\Omega \sqrt{2/A(x)}\, dx$.

Let the average per-km operating cost of trucks be c_t. Define the cost density of inbound trucking, denoted by $W_t(x, A(x))$, to be the ratio of inbound trucking cost in the service zone with its terminal at x and area $A(x)$ to $A(x)$. Then $\int_\Omega W_t(x, A(x))\, dx = c_t(d_{t,l} + d_{t,d})$. Derived from this equation, $W_t(x, A(x))$ is formally given by the following result:

Result 4. The cost density of inbound trucking can be expressed as

$$W_t(x, A(x)) = \Phi_t(x)/\sqrt{A(x)} + \Psi_t(x), \tag{11}$$

where

$$\Phi_t(x) = \frac{c_t}{\sqrt{2}}, \qquad \Psi_t(x) = \frac{2g(x)r_d(x)n(x)c_t}{v_t}. \tag{12}$$

5.3. Service Zone Design

The planning of delivery service zones is represented by the decisions of zone sizes $A(x)$ over the service region, striking a balance between the truck mode and the shared mobility mode. The goal is to minimize the total operating cost, which can be expressed in terms of cost densities of shared mobility and inbound trucking:

$$\begin{aligned} C &= \int_\Omega (W_c(x, A(x)) + W_t(x, A(x)))\, dx \\ &= \int_\Omega \left(\Phi_c(x)\sqrt{A(x)} + \Psi_c(x) + \Phi_t(x)/\sqrt{A(x)} + \Psi_t(x)\right) dx. \end{aligned} \tag{13}$$

Minimizing the total operating cost C is equivalent to minimizing the integrand on the right-hand side of (13) at each point x. Applying the first-order necessary condition yields the optimal service zone design stated in the following result:

Result 5. The optimal size of the service zone with its terminal at point x and the optimal total operating cost of one dispatch are given, respectively, by

$$A^*(x) = \frac{\Phi_t(x)}{\Phi_c(x)}, \tag{14a}$$

$$C^* = \int_\Omega \left(2\sqrt{\Phi_c(x)\Phi_t(x)} + \Psi_c(x) + \Psi_t(x)\right) dx. \tag{14b}$$

The continuous service zone design in Result 5 needs to be discretized for real implementation (see Ouyang and Daganzo 2006 for a discretization procedure). Also notice that this design does not consider the cost of package handling in the transshipment process, with three reasons: (1) First, omitting the term associated with the fixed charge leads to a closed-form expression of $A^*(x)$. (2) This cost, which is mainly because of fuel consumption of vehicles while idling and the time consumption of drivers while loading/unloading packages, is small relative to the en route cost. (3) The estimate of per-km truck cost c_t in Appendix E.1 in the online appendix already takes into account different states (i.e., traveling and idling).

5.4. Benchmark: A Truck-Only System

The benchmark for evaluating the logistics system with shared mobility is a conventional logistics system where only trucks fulfill home delivery services from the depot to demand destinations. Denote the average per-km operating cost of trucks during the last-mile deliveries by c_t'. $c_t' > c_t$, since trucks move more slowly than during bulk transport trips, incurring higher truck driver wage payment. Following the approach in Daganzo (2005), Result 6 states the VRP approximation model for the benchmark:

Result 6. The total operating cost of the truck-only system is given by

$$C_b = \int_\Omega \left(0.82\sqrt{n(x)}c_t' + \Psi_t(x)\right) dx; \tag{15}$$

C^* in Result 5 and the above C_b share the same element $\int \Psi_t(x)\, dx$, which captures the bulk trucking cost.

Their expressions only differ in terms that characterize the last-mile cost performance. Therefore, benchmarking the shared-mobility system against the truck-only system enables an exclusive evaluation of the last-mile cost competitiveness of shared-mobility paradigm. Section 6 presents managerial insights of this paradigm through this evaluation in various cases.

6. Managerial Implications

Based on the planning models in Section 5, a logistics service provider can evaluate whether and how to carry out large-scale crowdsourcing of shared mobility. For this purpose, this section examines several aspects: the scalability and the asset lightness of the business model, operational prescriptions, dynamic wages, and environmental impacts.

6.1. Scalability

A key question that drives this research is whether shared mobility is scalable (in terms of the operating cost) to meet the increasing last-mile delivery demand for e-commerce.

One may conjecture that shared mobility for last-mile deliveries can be increasingly favorable when the demand density increases, reasoning that cars are both more fuel efficient and more time efficient in delivering small loads with many stops. However, this conjecture is not true. First notice that, given the market conditions (e.g., $\bar{m}$, $F(\cdot)$, w_b and w_m), the wage payment by the logistics service provider to car drivers is not directly related to the fuel economy of cars, but instead matches car drivers' earning rate in the ride-share service market. More deeply, this question can be addressed by directly comparing the optimal cost density of shared mobility, $\rho_s = 2\sqrt{\Phi_c \Phi_t} + \Psi_c + \Psi_t$, and the corresponding cost density of the benchmark truck-only system, $\rho_t = 0.82\sqrt{n}c_t' + \Psi_t$. The following proposition presents the relative behaviors of ρ_s and ρ_t as the demand density n increases:

Proposition 5. *There exists a threshold $\bar{n}$ such that $\rho_s \le \rho_t$ for $n < \bar{n}$ and $\rho_s > \rho_t$ otherwise, where $\bar{n}$ can be negative or nonnegative.*

The result shows that shared mobility is less cost-efficient than the truck-only mode as delivery demand soars. Shared mobility may lead to cost savings for small n but cannot keep the edge over the truck-only system when n exceeds a certain threshold. This result can be explained by the different economies of scale of the two systems. Whereas ρ_t only involves a square-root term $0.82c_t'\sqrt{n}$, ρ_s includes a linear term $(\eta g \tilde{\mu}^2(w_b \nu + w_m)/(v_c \nu(\tilde{\mu}+\nu)^2))n$. Hence, when n becomes large, the linear component of ρ_s will dominate ρ_t.

The inefficient linear cost component of ρ_S results from the wage response behavior of shared mobility. In the truck-only system, truck drivers are hired and paid with a fixed hourly wage. Their wage payment is proportional to the trip distance. By contrast, car drivers providing shared mobility receive route-specific payments (Equation (4)), which incorporates stochastic waiting time and ride time. The derived wage payment mainly consists of a base fare and a per-mile fare. Under this payment structure, the payment/distance ratio is higher for short delivery routes than for long delivery routes. The higher payment/distance ratio compensates drivers for the lower utilization level when taking short routes (as measured by the expected fraction of time when a driver is in service). If the payment/distance ratio was constant, short-distance delivery routes would be less profitable than long-distance trips from the ride-sharing market. Instead of taking the short route, drivers would prefer to wait for a potentially longer trip from which the driver can make more money (Janssen and Parakhonyak 2011). The (base + mile) payment structure is prevalent in the ride-sharing market and taxi industry, and is generally used in the literature (e.g., see Yang et al. 2005, 2010 for details). For the shared mobility system, as n grows, the optimal route assigned to each driver becomes shorter, and the resulting payment/distance ratio is higher. Therefore, the total wage payment for the shared-mobility system increases faster than the truck-only system, wherein the per-km payment is constant for professional drivers. Proposition 5 suggests that the demand-dense areas should be avoided when planning a shared-mobility system for last-mile delivery. This insight is further validated in the case study in Section 6.4. To be noted, this insight may not hold if the ride-share market does not serve as car drivers' outside option, or if the logistics service provider schedules shared mobility far in advance for an extended service duration (which is the case of Amazon Flex). In the former case, the (base + mile) payment structure is not necessarily the least-cost payment structure. In the latter case, the delivery routes are likely to be back-to-back VRP routes. Using shared mobility is likely to be more costly, but more scalable in the mean time.

The aforementioned discussion takes the pool size of shared-mobility $\bar{m}$ as fixed, which is suitable for short-run operations. In the long-run, if the demand for both ride-share and delivery services increases, more drivers will be likely to register to provide shared mobility services, that is, $\bar{m}$ will grow. Numerical analysis shows that growing $\bar{m}$ induces more real-time supply of shared mobility, thereby compensating for the cost inefficiency of the shared-mobility logistics system by reducing the wage payment to car drivers. In other words, shared mobility may be favorable in demand-dense areas if the pool size grows sufficiently large. However, on-demand ride-share service platforms need to carefully control the pool size of shared mobility to ensure that the individual car drivers' expected earning rate is above a certain level (Cachon et al. 2017). See more discussion on scaling up the delivery demand and $\bar{m}$ after the proof of Proposition 5 in Appendix D in the online appendix.

6.2. Asset Lightness

While the preceding analysis focuses on the operating cost, an important feature and a potential advantage of the shared mobility system over the truck-only system is asset lightness. That is, with shared mobility, the logistics service provider operates a smaller fleet of trucks and thus needs less capital investment. Specifically, in the designed one-transshipment logistics system, trucks only visit terminals at the inbound stage. By contrast, trucks are dispatched to visit all demand destinations in the truck-only system. The truck fleet size can be calculated as total truck working hours/# of working hours per driver, where the total truck working hours is the total truck travel distance divided by the truck speed. Then the fleet size ratio for the two systems is $\int_\Omega((\sqrt{\Phi_c(x)\Phi_t(x)} + \Psi_t(x))/(c_t s_t))\,dx / \int_\Omega(0.82\sqrt{n(x)}/s'_t + \Psi_t(x)/(c_t s_t))\,dx$, where s_t and s'_t are the speeds of trucks during inbound trips and last-mile deliveries, respectively. In the case study, the asset lightness of the shared mobility system is presented in terms of this fleet size ratio (the infrastructure of the two systems may differ in other aspects, for example, the information management system and communication system, but the quantification of those differences is out of the scope of this paper). The shared mobility model can benefit from asset lightness as it allows quick expansion, which can help the company capture the whole market. In addition, asset lightness also implies less depreciation from the capital assets. In the case study presented in the next section, the operating cost includes the depreciation cost.

6.3. Case Study: East San Francisco Bay Area

6.3.1. Service Region. Consider a service region that consists of 15 zip-code areas in the east San Francisco Bay Area in California, as Figures 2(a)–(d) depict. Choose the Fedex Ship Center in Emeryville (area 8) as the depot. Appendix E.1 in the online appendix lists their zip-codes along with their population densities (from ESRI 2015) and mean distances to the depot. The area of the entire region is 80.8 km^2. The region is representative of a nonhomogenous setting, spanning densely populated areas, such as downtown and south Berkeley (in the center, area 6) and Albany (northwest, area 11), and sparsely populated areas, such as east Berkeley (northeast, hilly, areas 5 and 13) and west Oakland (southwest, a port, area 14). For simplicity of exposition, the setting within each zip-code area is assumed to be homogenous.

6.3.2. Parameters. Table 1 summarizes the main vehicle-related parameter values. Detailed calibration with empirical evidence is available in Appendix E.2 in the online appendix. In addition, assume the demand destination density $n(x)$ to be 1% of the population density of the zip-code area that covers point x (sensitivity analysis of $n(x)$ is in Sections 6.1 and 6.4). Other exogenous information pertaining to shared-mobility supply and wages, $\{w_b, w_m, \nu, \bar{m}, \bar{\mu}, F(\cdot)\}$, is available in Appendix E.3 in the online appendix.

6.3.3. Scenarios. In addition to the baseline scenario, the following three alternative scenarios address three key factors that influence operational performance:

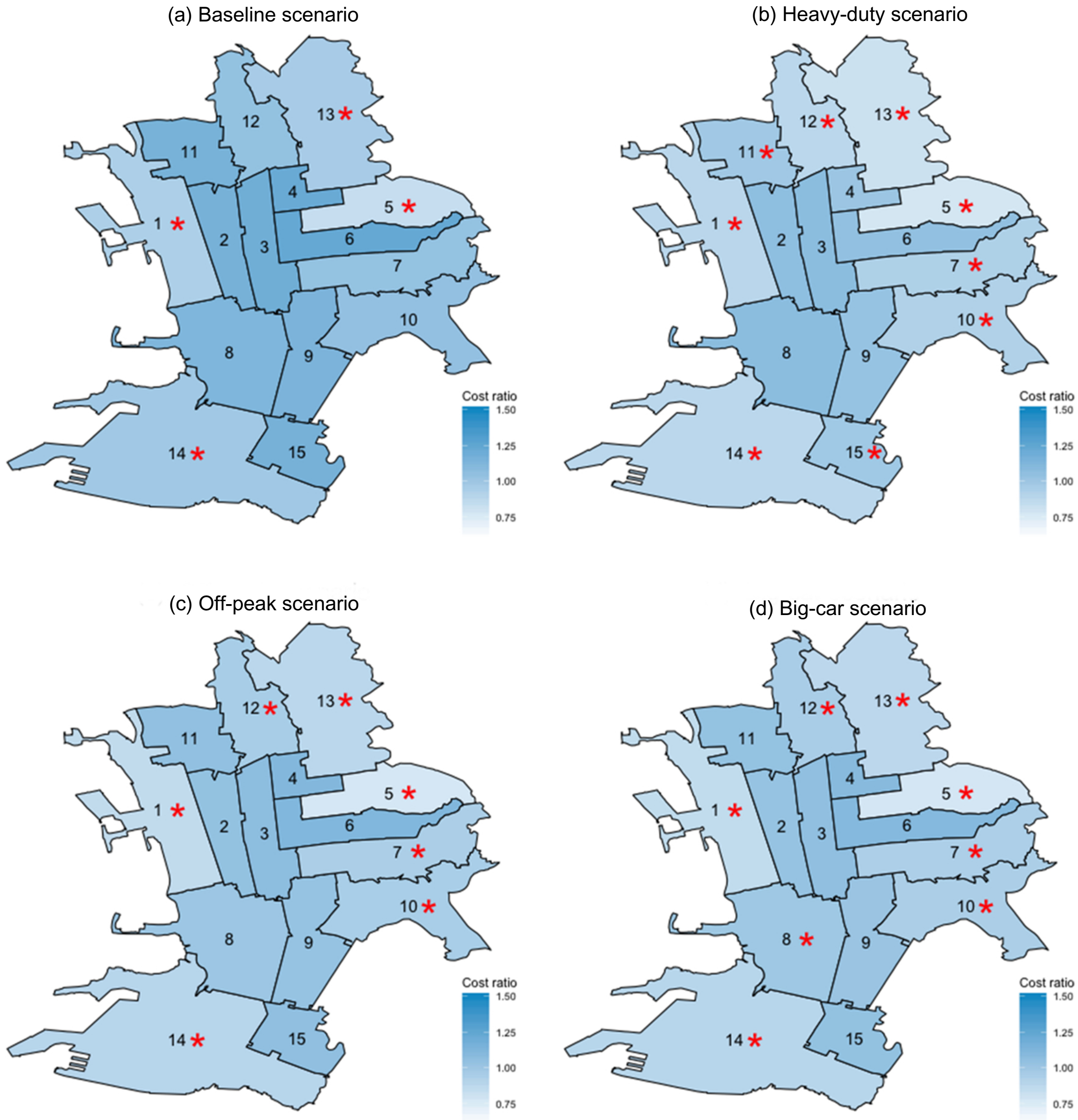
- *Heavy-duty* scenario, where heavy-duty trucks (Freightliner M2-106) are in place of the baseline truck model (Freightliner P70) for the inbound-stage trucking in the shared-mobility system (the benchmark truck-only system still uses the baseline truck model). The load capacity of this heavy-duty model is 6,000 kg, about three times as large as that of the baseline truck model.
- *Off-peak* scenario, where the deliveries are during off-peak hours when demands for passenger rides are lower. The rate of total ride-share demand per unit area, $\bar{\mu}$, declines by 25%.
- *Big-car* scenario, where cars are assumed to have higher rated loading capacity and are thus filled with more packages. v_c/g increases from 15 to 18.

Applying Result 5 to the four scenarios yields logistic system designs with shared mobility; applying Result 6 yields designs for the benchmark truck-only system. Figures 2(a)–2(d) show whether or not shared-mobility results in lower operating cost than the benchmark in each zip-code area in those scenarios. Detailed results for the baseline scenario, including zip-code-level zone densities and costs, are available in Appendix E.1 in the online appendix. On the whole-region level, Table 2 summarizes the results of comparing operating costs, truck fleet sizes, and GHG emissions. Note that model assumptions, parameter estimation errors, and the fast-evolving dynamics of the ride-share industry may bring inaccuracies to the results. Still, the case study provides a realistic setting to reasonably infer the viability of using shared mobility for last-mile delivery with the factors considered in this paper. The nonhomogeneous setting also helps understand how the viability changes under various conditions. The following sections use these results to discuss more managerial implications.

6.4. Operational Prescriptions

Whether using shared mobility is favorable depends on performance metrics and operational conditions. The second column of Table 2 shows that crowdsourcing shared mobility significantly reduces the truck fleet size. This asset-light nature is likely to represent

Figure 2. (Color online) Zip-Code Level Comparisons of Operating Costs in the Four Scenarios: (a) Baseline, (b) Heavy-Duty, (c) Off-Peak, and (d) Big-Car



Notes. Patches represent zip-code areas. Star marks indicate the areas where the shared-mobility system incurs smaller cost than the truck-only system.

an advantage over the truck-only system. However, the logistic system with shared mobility incurs 8.2% more operating cost than the truck-only system in the baseline scenario, as the first entry of Table 2 shows. This result indicates that the wages paid to car drivers for their mobility can exceed the savings by reducing the fleet size and keeping trucks from last-mile trips. On the other hand, adopting shared mobility creates additional operational flexibilities. In what follows, four operational prescriptions are identified for the logistics service provider to exploit those flexibilities so that crowdsourcing shared mobility can be more cost-efficient.

Table 1. Main Vehicle-Related Parameters

	Operating cost c ($\$\cdot\text{km}^{-1}$)	Vehicle capacity v (kg)	Emission c^e ($\text{kg CO}_2\cdot\text{km}^{-1}$)
Trucks	1.257 (3.182)†	2,000	0.597
Heavy-duty trucks	1.355	6,000	1.003
Cars	—	150	0.369

†For last-mile delivery in the benchmark truck-only system.

1. *Avoid areas with too high demand densities.* Figure 2(a) shows that shared-mobility outperforms conventional trucking only in areas 1, 5, 13, and 14, which

Table 2. Region-Level Comparisons of Operating Costs, Fleet Sizes, and GHG Emissions

	Cost ratio (shared mobility vs. truck only)	Fleet size ratio (shared mobility vs. truck only)	Emission ratio (shared mobility vs. truck only)	Emission suboptimality (cost minimizing vs. emission minimizing)
Baseline	1.082	0.276	1.067	1.002
Heavy duty	0.956	0.135	0.987	1.019
Off peak	0.980	0.273	1.067	1.001
Big car	0.972	0.266	1.040	1.002
Combined	0.760	0.124	0.949	1.011

have the sparsest demand for delivery services in the region. This prescription validates and is explained by the scaling behavior of the wage-response structure of shared mobility, as Section 6.1 presents.

2. *Use heavy-duty trucks for inbound trucking.* If heavy-duty trucks replace regular-size delivery vans for inbound-stage trucking, Figure 2(b) shows that five more areas favor shared mobility, and Table 2 shows that the total operating cost of the shared-mobility system decreases by 12.6% from the baseline and becomes less than the total operating cost of the truck-only system. This is because, with shared mobility, using heavy-duty trucks enhances the inbound trucking efficiency without compromising the efficiency of last-mile delivery. In fact, trucks are significantly less cost-efficient for last-mile delivery than during the bulk-trucking trips. The underlying reason is that trucks move much more slowly in the last-mile trips, incurring higher per-km driver wages; trucks also have higher per-km vehicle-related cost than cars.

3. *Deliver during off-peak hours.* In the off-peak scenario with 25% less ride-share demand $\bar{\mu}$, Figure 2(c) shows that the shared-mobility system costs less than the truck-only benchmark in 7 out of the 15 zip-code areas, and Table 2 shows that the total operating cost of the shared-mobility system becomes 98.0% of that of the truck-only system. As Section 4.5 discusses, reducing ride-share demand attenuates both the synergy and the competition effects of the ride-share market on the delivery service market. That is, both the mobility supply $\tilde{m}$ and the ride-share service request rate $\tilde{\mu}$ decline. The overall effect is a smaller expected earning rate of car drivers and less payment to them. Appendix D in the online appendix reports additional numerical experiments on the cost effect of perturbing ride-share demand $\bar{\mu}$. In general, this prescription applies to normal operating hours; in other periods (e.g., midnight), $\tilde{\mu}$ may surge occasionally with lack of shared mobility supply and truck drivers are costly to dispatch.

4. *Fill more packages in cars.* If possible, increasing the rated loading capacity of cars to fill more packages can also reduce the operating cost, as Figure 2(d) and Table 2 show. This is because filling more packages prolongs cars' individual outbound trips, which has two scale-economy implications: (1) First, a longer trip dilutes the fixed wage component as identified in Section 6.1. (2) Second, the length of an outbound trip itself increases less than linearly in the number of destinations to visit. In addition, filling more packages reduces the number of cars for delivery services, thus reducing wage payment by dampening the competition with the ride-share market.

Collectively, Table 2 shows that simultaneously implementing the last three prescriptions can reduce the total operating cost of the shared-mobility system to be only 76.0% of that of the conventional truck-only system. This is a considerable saving potential, but some other factors should also be evaluated before real implementation. Among them, the following sections discuss dynamic wages and an environmental perspective.

6.5. Surge Pricing

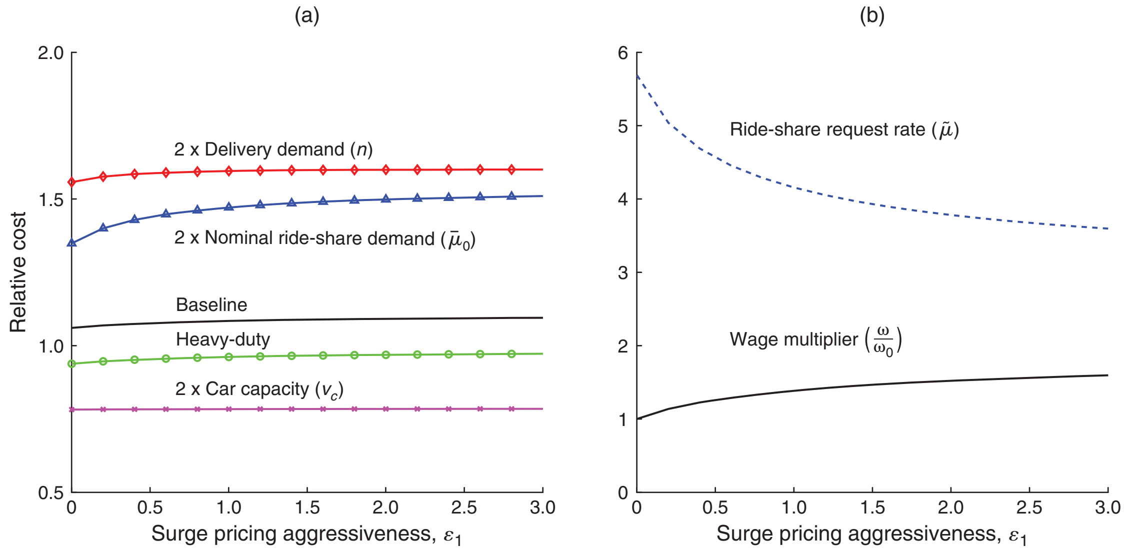
Uber and Lyft have been dynamically adjusting prices and wages based on the real-time state of supply and demand. How does such practice, hereafter referred to as *surge pricing* as Uber calls it, affect the cost performance of the delivery services with shared mobility? Designing surge pricing contracts has been a key research effort of those prominent ride-share companies. The underlying surge pricing algorithms being implemented are evolving over time and may not be accessible by external entities. Therefore, the logistics service provider is primarily interested in whether this new delivery business model withstands surge pricing, rather than inferring the optimal surge pricing contract (which is the focus in Cachon et al. 2017, for example). For this purpose, consider the following equations to represent a surge pricing setting, where the wage parameters w_b and w_m as well as the rate of the total ride-share demand $\bar{\mu}$ become endogenous:

$$\frac{w}{w_0} = 1 + \epsilon_1\left(\frac{\tilde{\mu}}{\mu_0} - 1\right), \tag{16a}$$

$$\frac{\bar{\mu}}{\bar{\mu}_0} = 1 - \epsilon_2\left(\frac{w}{w_0} - 1\right), \tag{16b}$$

where w stands for both w_b and w_m. $\{w_0, \mu_0\}$ are given baseline values of w and $\tilde{\mu}$, respectively. (16a) characterizes how the wage parameters adjust according to the request rate $\tilde{\mu}$, which measures the tension between

Figure 3. (Color online) (a) Surge-Pricing-Adjusted Operating Cost (Divided by the Operating Cost of the Truck-Only System) for the Baseline and for Scenarios with Heavy-Duty Trucks, with Double Delivery Demand Density, with Double Nominal Total Ride-Share Demand and with Double Car Capacity, Respectively; (b) Endogenous Ride-Share Request Rate and Wage Multiplier in the Scenario with Double Nominal Total Ride-Share Demand



the supply and the demand of ride-share services. $\epsilon_1 \geq 0$ represents the aggressiveness of surge pricing, that is, how forcefully the ride-share company leverages surge pricing to push the supply demand tension toward the reference level μ_0. (16b) characterizes how the wage adjustment affects the total ride-share demand $\bar{\mu}$ to deviate from the given nominal level $\bar{\mu}_0$, assuming that the ride prices change along with the wages. $\epsilon_2 \geq 0$ is the demand elasticity. The surge pricing effects that (16a) and (16b) characterize are consistent with the empirical evidence of Uber's surge pricing effects (Hall et al. 2015).

Simultaneously solving Equations (16a) and (16b) along with (5) and (6) generates surge-pricing-adjusted parameters $\{w, \tilde{m}, \tilde{\mu}, \bar{\mu}\}$. Again, it can be verified that the continuity and the monotonicity of $F(\cdot)$ in (5) ensure the existence and the uniqueness of the solution. Plugging the values of $\{w, \tilde{m}, \tilde{\mu}\}$ into (14b) generates the adjusted total operating cost of using shared mobility for last-mile delivery. Figure 3(a) shows this operating cost for the service region considered in Section 6.3. The curves represent five operating scenarios and are with respect to a wide range of surge pricing aggressiveness ϵ_1. In these numerical experiments, the demand elasticity ϵ_2 is 0.57 according to the empirical study in Cohen et al. (2016), the reference level of the ride-share request rate $\mu_0 = 3\ \text{hr}^{-1}$, and the baseline values of wage w_0 and the total ride-share demand $\bar{\mu}_0$ are the same as in the preceding setting.

Figure 3(a) shows that the operating costs are insignificantly affected by surge pricing in the considered numerical settings. In all those scenarios, perturbing surge pricing aggressiveness ϵ_1 in a wide range (up to 300%) does not significantly affect the operating costs, compared with the case in absence of surge pricing ($\epsilon_1 = 0$). This is an interesting result, considering that the operating cost is sensitive to the drivers' expected earning rate, which is driven by the wage parameter w and the ride-share request rate $\tilde{\mu}$. In fact, although the drivers' expected earning rate is sensitive to w and $\tilde{\mu}$, the values of w and $\tilde{\mu}$ move in opposite directions in a surge pricing mechanism, counteracting each other's impact. In other words, surge pricing forms a negative-feedback loop between w and $\tilde{\mu}$. As Figure 3(b) illustrates for the scenario with doubled nominal total ride-share demand $\bar{\mu}_0$, reinforcing the surge pricing aggressiveness increases the wage parameter w in response to the high supply demand tension of ride-share services (Equation (16a)). This change in turn dampens the ride-share demand $\bar{\mu}$ (Equation (16b)), induces more supply of self-scheduling shared mobility $\tilde{m}$ (Equation (5)), and eases the supply demand tension $\tilde{\mu}$ (Equation (6)), which weakens the effect of increased w on the car drivers' expected earning rate. Hence, under realistic parameter settings, the preceding managerial implications and prescriptions remain largely valid in the presence of surge pricing. The drivers of the cost-effectiveness remain to be (1) the inherent tension between the supply and demand for shared mobility and (2) the vehicle capacity configurations.

6.6. Environmental Impact

Suppose that the service region adopts shared mobility for home delivery services. Does it reduce GHG emissions? It turns out that (unless low-emission cars are assumed) neither the baseline nor the three alternative scenarios see significant emission reduction relative to the emission level of the truck-only system, as the third column of Table 2 shows. In fact, the emission efficiency of using cars is twofold: (1) Their open-loop routes are in general shorter than conventional closed-loop routes for vehicles of the same capacity, and their per-km emission rate is lower than that of trucks. (2) However, these efficiency gains are offset by a demerit: cars are of much smaller loading capacity than trucks and consequently the share-mobility system incurs over 35% longer total trip distance than the truck-only system in the baseline scenario. In addition, using heavy-duty trucks and filling more packages in cars reduce emissions, but not as much as they reduce operating costs. This is because the wage effects of those measures do not affect emissions.

Next, to what extent does the design that minimizes operating costs compromise the goal of abating GHG emissions? It turns out that this logistics system has remarkably small tension between the economic and environmental preferences: minimizing operating costs while ignoring GHG emissions results in a service zone deployment that increases emissions from the minimum level by a negligible percentage (e.g., 0.2% in the baseline scenario), as the fourth column of Table 2 shows; this compatibility of objectives is robust to a wide range of parameter settings.

Minimizing operating costs and minimizing emissions are compatible and robust because both objective functions take on the same form, expressed as $\Phi_c\sqrt{A} + \Psi_c + \Phi_t/\sqrt{A} + \Psi_t$ and $\Phi_c^e\sqrt{A} + \Psi_c^e + \Phi_t^e/\sqrt{A} + \Psi_t^e$, respectively. In both functions, the terms that depend on zone size A constitute a remarkably flat function of A. Such structure is common to transshipment systems in other contexts (see detailed discussion in Daganzo 2005 and Cachon 2014). The emissions inefficiency in the operating cost minimizing design relative to the minimum emissions level is bounded as

$$\begin{aligned}
&\frac{\Phi_c^e\sqrt{A^*} + \Phi_t^e/\sqrt{A^*} - 2\sqrt{\Phi_c^e\Phi_t^e}}{2\sqrt{\Phi_c^e\Phi_t^e} + \Psi_c^e + \Psi_t^e}\\
&\quad < \frac{\Phi_c^e\sqrt{A^*} + \Phi_t^e/\sqrt{A^*} - 2\sqrt{\Phi_c^e\Phi_t^e}}{2\sqrt{\Phi_c^e\Phi_t^e}}\\
&\quad = \frac{1}{2}\left(\frac{\sqrt{\Phi_c^e/\Phi_c}}{\sqrt{\Phi_t^e/\Phi_t}} + \frac{\sqrt{\Phi_t^e/\Phi_t}}{\sqrt{\Phi_c^e/\Phi_c}}\right) - 1,
\end{aligned}$$

in which $A^* = \Phi_t/\Phi_c$. This upper bound is small and robust. Significantly enlarging the discrepancy between Φ_c^e/Φ_c and Φ_t^e/Φ_t can only slightly increase this upper bound.

Moreover, the objectives' compatibility is much further enhanced by the fact that A-independent terms, $\{\Psi_c, \Psi_t\}$ and $\{\Psi_c^e, \Psi_t^e\}$, are dominant. In the baseline scenario, for example, these two components account for 88.3% of the total operating costs and 80.7% of the total emissions. Consequently, emissions inefficiency is even much smaller than the above upper bound, as the inequality results from dropping $\Psi_c^e + \Psi_t^e$ from the denominator. In fact, Ψ_c and Ψ_c^e are associated with the detour trips of cars; Ψ_t and Ψ_t^e are associated with the mean distance from the depot to a point in the service region times the number of trucks. These two parts account for a large share of routes and are irrelevant to adjusting service zone deployment for different economic/environmental preferences.

7. Conclusion

This paper studies whether and how to integrate shared mobility for home delivery services, given the rapid development of the sharing economy and retail e-commerce in recent years. The logistics setting is a one-transshipment system: a fleet of short-haul trucks are dispatched from a depot and unload packages at terminals of service zones. Passenger cars with available mobility are attracted to the terminals to pick up and deliver packages to demand destinations. The paper presents planning and operations models for this sharing logistics system. The proposed CA model based on numerically solving OVRP instances characterizes the one-way and one-shot nature of shared-mobility routes. The wage response model characterizes the cost of crowdsourcing shared mobility. The model of synergy and competition with the ride-share market determines the supply of shared mobility. The optimal service zone design is in the form of service zone sizes and the associated operating cost.

The models, analysis, and the case study lead to several findings: (1) In terms of the operating cost, crowdsourcing shared mobility is not as scalable as the conventional truck-only system when the delivery demand increases. (2) However, this logistics business model has the potential to generate economic benefits, because it maintains much lighter truck fleet asset and creates additional operational flexibilities. Four operational prescriptions are (i) to start with low-demand-density areas, (ii) to adopt heavy-duty trucks, (iii) to deliver during off-peak hours, and (iv) to increase car loads of packages. These managerial implications also withstand dynamic adjustment of wages and prices in the ride-share market. (3) Entering this sharing paradigm may not help abate GHG emissions, because of additional trips incurred; on the other hand, even exclusively minimizing operating costs does not significantly increase emissions relative to the minimum level of emissions. (4) Finally, the CA model that characterizes open-loop vehicle routes, the construction of the lower bounds for the optimal OVRP solution, the shared mobility wage response model, and the model of interaction with the ride-share market can all be useful for logistics services providers and governments to evaluate the routing behavior and cost-effectiveness of this logistics paradigm under various conditions.

This paper is the first attempt to design and analyze the prospective sharing logistics system based on analytical models and empirical parameter estimates. To further investigate the impacts that this combination of the sharing economy and retail e-commerce can potentially bring about, two directions can be explored. First, home delivery services can utilize shared mobility in different ways with additional complexity to analyze. For example, car drivers can provide ride-share services to passengers along the trip of delivering packages. Second, whereas this paper focuses on operating costs, truck fleet sizes, and GHG emissions as performance metrics, it is worth considering other implications of sharing logistics, such as social trust building, service agility, and unemployment rate, which may or may not justify additional government regulation and incentives.

Acknowledgments

The authors thank Hau Lee (the department editor), the associate editor, and two anonymous referees for their valuable comments. The authors also thank Dr. Wilko Schulz-Mahlendorf and Dr. Chen Peng for their helpful discussion on the relevance of this research idea.

References

Agatz N, Erera A, Savelsbergh M, Wang X (2012) Optimization for dynamic ride-sharing: A review. *Eur. J. Oper. Res.* 223(2):295–303.

Assembly Bill 32 (2006) The California global warming solutions act of 2006. Accessed on November 10, 2015, https://www.arb.ca.gov/cc/ab32/ab32.htm.

Bellos I, Ferguson M, Toktay LB (2017) The car sharing economy: Interaction of business model choice and product line design. *Manufacturing Service Oper. Management* 19(2):185–201.

Botsman R (2014) Crowdshipping: Using the crowd to transform delivery. *AFR Boss Magazine* (September 12), http://www.afr.com/it-pro/crowdshipping-using-the-crowd-to-transform-delivery-20140911-jyk63.

Cachon GP (2014) Retail store density and the cost of greenhouse gas emissions. *Management Sci.* 60(8):1907–1925.

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.

Carbone V, Rouquet A, Roussat C (2015) "Carried away by the crowd": What types of logistics characterise collaborative consumption? *1st Internat. Workshop Sharing Econom.*, 4–5.

Carlsson JG, Jia F (2013) Euclidean hub-and-spoke networks. *Oper. Res.* 61(6):1360–1382.

Carlsson JG, Jia F (2014) Continuous facility location with backbone network costs. *Transportation Sci.* 49(3):433–451.

Carlsson JG, Behroozi M, Devulapalli R, Meng X (2016) Household-level economies of scale in transportation. *Oper. Res.* 64(6): 1372–1387.

CFMP (2014) California freight mobility plan. Accessed September 28, 2015, http://www.dot.ca.gov/hq/tpp/offices/ogm/california_freight_mobility_plan.html.

Cohen P, Hahn R, Hall J, Levitt S, Metcalfe R (2016) Using big data to estimate consumer surplus: The case of Uber. Technical report, National Bureau of Economic Research, Cambridge, MA.

Daganzo CF (1984) The distance traveled to visit n points with a maximum of c stops per vehicle: An analytic model and an application. *Transportation Sci.* 18(4):331–350.

Daganzo CF (2005) *Logistics Systems Analysis*, 4th ed. (Springer, New York).

eMarketer (2014) Global B2C ecommerce sales to hit $1.5 trillion this year driven by growth in emerging markets. Accessed September 28, 2015, http://www.dot.ca.gov/hq/tpp/offices/ogm/california_freight_mobility_plan.html.

ESRI (2015) ESRI data and maps of U.S. zip-code areas (ArcGIS layer). Accessed April 15, 2016, http://www.arcgis.com/home/item.html?id=5203d33eed8b4611859a2d9dca8b604e.

Furuhata M, Dessouky M, Ordóñez F, Brunet ME, Wang X, Koenig S (2013) Ridesharing: The state-of-the-art and future directions. *Transportation Res. Part B: Methodological* 57(November):28–46.

Haimovich M, Kan AR (1985) Bounds and heuristics for capacitated routing problems. *Math. Oper. Res.* 10(4):527–542.

Hall J, Kendrick C, Nosko C (2015) The effects of Uber's surge pricing: A case study. Working paper, Booth School of Business, University of Chicago, Chicago.

He L, Mak HY, Rong Y, Shen ZJM (2017) Service region design for urban electric vehicle sharing systems. *Manufacturing Service Oper. Management* 19(2):309–327.

Janssen M, Parakhonyak A (2011) Service refusal in regulated markets for credence goods. Higher School of Economics Research Paper No. 08/EC/2011, Higher School of Economics, Moscow.

Kabra A, Belavina E, Girotra K (2018) Bike-share systems: Accessibility and availability. Chicago Booth Research Paper No. 15-04, University of Chicago, Chicago.

Kafle N, Zou B, Lin J (2017) Design and modeling of a crowdsource-enabled system for urban parcel relay and delivery. *Transportation Res. Part B: Methodological* 99(May):62–82.

Li B, Krushinsky D, Reijers HA, Van Woensel T (2014) The share-a-ride problem: People and parcels sharing taxis. *Eur. J. Oper. Res.* 238(1):31–40.

Li B, Krushinsky D, Van Woensel T, Reijers HA (2016) An adaptive large neighborhood search heuristic for the share-a-ride problem. *Comput. Oper. Res.* 66(February):170–180.

Li F, Golden B, Wasil E (2007) The open vehicle routing problem: Algorithms, large-scale test problems, and computational results. *Comput. Oper. Res.* 34(10):2918–2930.

Ouyang Y, Daganzo CF (2006) Discretization and validation of the continuum approximation scheme for terminal system design. *Transportation Sci.* 40(1):89–98.

PWC (2015) The sharing economy—Consumer intelligence series. PricewaterhouseCoopers. Accessed September 25, 2015, https://www.pwc.fr/fr/assets/files/pdf/2015/05/pwc_etude_sharing_economy.pdf.

Rougès JF, Montreuil B (2014) Crowdsourcing delivery: New interconnected business models to reinvent delivery. *1st Internat. Physical Internet Conf., May 28–30, Quebec City, Canada*, 1–19.

Schaefer B, Konur D (2015) Economic and environmental considerations in a continuous review inventory control system with integrated transportation decisions. *Transportation Res. Part E: Logist. Transportation Rev.* 80(August):142–165.

Shu J, Chou MC, Liu Q, Teo CP, Wang IL (2013) Models for effective deployment and redistribution of bicycles within public bicycle-sharing systems. *Oper. Res.* 61(6):1346–1359.

Xie W, Ouyang Y (2015) Optimal layout of transshipment facility locations on an infinite homogeneous plane. *Transportation Res. Part B: Methodological* 75(May):74–88.

Yang H, Fung CS, Wong KI, Wong SC (2010) Nonlinear pricing of taxi services. *Transportation Res. Part A: Policy and Practice* 44(5): 337–348.

Yang H, Ye M, Tang WH, Wong SC (2005) Regulating taxi services in the presence of congestion externality. *Transportation Res. Part A: Policy and Practice* 39(1):17–40.