Last Mile Innovation: The Case of the Locker Alliance Network

Guodong Lyu[a,b], Chung-Piaw Teo[a,b]

[a]Institute of Operations Research and Analytics, National University of Singapore, Singapore 117602

[b]Department of Analytics and Operations, NUS Business School, National University of Singapore, Singapore 119245

Problem Definition: The Singapore government has recently proposed the concept of "Locker Alliance" (LA), an interoperable network of public lockers in residential areas and hot spots in community, to improve the efficiency of last mile parcel delivery operations. This is to complement the existing infrastructure, comprising mainly of proprietary lockers and collection points in commercial areas set up by large delivery companies. How do we determine the density and coverage of the LA network, to promote adoption of locker pickup in Singapore? What will be the impact on the delivery profile in the central business district, far from the residential areas?
Academic/Practical Relevance: We discuss the operational challenges associated with the problem of public locker installation in a city, following a new smart nation initiative in Singapore. We used data analytics to address the question: What are the chances that a customer will choose to pickup parcel from a locker, over home or office deliveries, based on walking distance (to lockers) and a variety of other features? Without knowing the transit routes of the customers, how do we design the LA network to ensure that the lockers will be well utilized?
Methodology: We use a set of locker usage data from a commercial courier company to calibrate a locker choice model, to determine the impact of walking distance on locker pickup intentions. We use the current (observed) parcel delivery profile to develop a facility location model for the LA network. We use this model to extrapolate and approximate the true adoption and new delivery profiles when the LA network is built.
Results: Contrary to conventional wisdom, our model does not always place lockers near areas with peak parcel volume (in pre-existing data), because the LA lockers provide another option for customers to pick up from lockers near residential areas. Furthermore, the model suggests that a coverage of 250 meters is appropriate for the LA network in Singapore.
Managerial Implications: Commercial parcel locker installation has traditionally focused on hot spots in the transit routes of the citizens in the city. The LA network is the first attempt in Singapore to allow public lockers in residential areas. This paper develops analytical method to determine network density and coverage based on a locker choice model, and argues how useful insights can be gleaned from the model, despite not having the full transit route information of all citizens in the city.

Key words: Smart Nation; Last Mile Innovation; Locker Network; Facility Location
History:

"Based on IDA's findings, the biggest problem faced by our local delivery companies is in making door-to-door deliveries and finding that no one is at home to receive the goods. Return visits add to costs and often inconvenience for customers."[1]

Singapore Deputy Prime Minister Tharman Shanmugaratnam, April 2016

[1] Locker system for parcel deliveries in residential areas to be implemented; retrieved from http://www.channelnewsasia.com/news/singapore/locker-system-for-parcel/2731536.html

1. Introduction

The rapid growth of business-to-consumer (B2C) e-commerce has boosted the sales of online retailers. According to a survey[2] reported by Ecommerce Foundation, global sales of B2C reached $2.3 trillion in year 2015, almost doubled the sales accrued in 2012. Intensive online transactions have given rise to the present challenges in urban logistics, especially in the last mile delivery of parcels from distribution center to consumers. This is a challenge that smart cities have to grapple with in their development into mega business centers to allow growing number of urban population to live and work in the cities.

E-commerce companies like Google, Amazon, eBay, and Uber etc. are operating and expanding services that allow shoppers to order goods online and have them delivered to their homes, preferably on the same day.[3] However, in this mode of operations, consumers often have to wait at home for their parcels. Existing innovations in this space focus on reducing the variability of the delivery time windows and/or using communication technologies to provide advanced notices prior to the arrival of the parcels. Despite much effort, unfortunately failed or missed deliveries still arise frequently, adding costs to last mile operations. In these instances, delivery companies have to either re-schedule another delivery or direct the customers to pick up the parcels at some designated places, including customer's pickup at workplace, retail stores, friends or family members, or locations pre-designated by customers in case of failure in the deliveries. A recent survey[4] conducted by UPS showed clearly that delivery to home (residential address) is stilled the preferred mode of delivery location choice (67% in 2015), although a growing number of customers surveyed have shifted their preferred choices to other non-traditional pickup locations.

A few delivery companies have experimented with the use of innovative technologies such as automated locker system as an alternate channel for parcel pickup by customers (Faugere and Montreuil 2017). This strategy effectively decouples the parcel delivery and customer's pickup processes within the last mile, reducing the coordination cost for a successful transaction, and also the needs for re-delivery. Song et al. (2009) demonstrated that lockers can indeed provide additional savings in operational cost and improve the delivery efficiency. In addition, it was also shown that the implementation of locker system can reduce the number of deliveries in city area and thus can reduce the liter consumption and CO_2 emission (e.g., Edwards et al. 2009, 2010, Iwan et al. 2016). DHL/Deutsche Post now operates around 2500 lockers in Germany, and Swedish PostNord supplied about 5000 pickup points for customers in Sweden, Norway, Finland, and Denmark (Morganti et al. 2014b).

[2] Global B2C E-Commerce report 2016; retrieved from `http://www.ecommercefoundation.org.`

[3] SAME-DAY DELIVERY: E-Commerce Giants Are Battling To Own The 'Last Mile'; retrieved from `http://www.businessinsider.sg/e-commerce-and-same-day-delivery-2014-9`

[4] 2015 UPS Pulse of the Online Shopper; retrieved from `https://www.pressroom.ups.com/assets/`

SingPost in Singapore runs a network of 140 lockers, strategically located to serve exclusively its customers. A few logistics players have also set up similar, albeit smaller scale parcel locker network system in Singapore. Most lockers can handle 3 or more parcel sizes, and pre-registration is normally required for proprietary locker systems.

To maximize utilization, some companies have installed lockers in hot spots near public places where consumers will congregate to. Amazon for instance puts their lockers in tube stations in London. Others have opted to put lockers in busy bus terminals/stops and commercial shopping areas. However, the upfront land and installation cost in these places can be prohibitive for small players. The location choices are also affected by regulations and policies imposed on the sector. Singapore government, for instance, prohibits the installation of commercial parcel lockers in train stations or crowded public areas, due to security concerns.

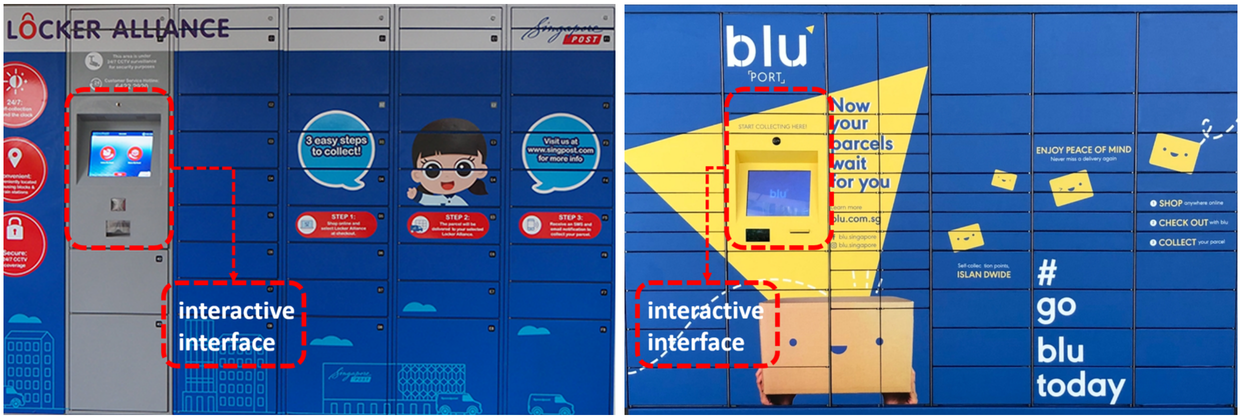
To level the playing field, and to shape the future of smart last mile operations, the Singapore government has recently announced its plan to roll out a Locker System (LA Network) as part of its smart nation vision.[5] The locker network allows parcel pickup and return to be performed without face-to-face contact with carriers. Instead, customers can directly interact with the locker system via a digital interface (cf. Figure 1). Through the automated infrastructure, the government aims to build a society of digitally receptive delivery companies and customers so as to provide more convenient, more cost-effective, and quicker parcel delivery services.[6] As a side effect, the government also targets to streamline the flows of parcels into crowded districts. The LA project will be open to all parcel delivery companies, making it the first country to do "large scale deployment of common parcel lockers" at a national level. This is a bold move, and hopefully will allow many logistics companies to break away from the lack of scale in their respective business models to offer seamless parcel delivery services to reduce the cost of delivery operations. The latter comes from the fact that delivery productivity can be expected to increase 3 to 4 folds with the LA Network, since many of the door-to-door deliveries in the same block can now be consolidated and delivered to a single locker station.

The LA Network is also expected to increase the volume of e-commerce retail transactions by making delivery more convenient for the consumers. Companies in several sectors can also re-engineer their operations to exploit the availability of the lockers. For instance, field service engineers no longer need to stock spares in their car trunks, but can get the needed spares delivered to the nearest lockers when the need arises. However, while these are indirect benefits that the LA Network can bring to

[5] Singapore's Smart Nation Initiative; retrieved from the Keynote talk of the "Smart Libraries for Tomorrow" Conference, `http://www.las.org.sg/wp/lft/files/0910-TKY-Smart-Libraries-for-Tomorrow-20-Sep-Final.pdf`

[6] Singapore Locker Alliance; retrieved from `https://www.lockeralliance.net`

Figure 1 Automated locker infrastructure deployed in the LA Network pilot.



(a) Singpost Locker (b) Blu Locker

Note. Two companies – Singpost and Blu – were involved in the LA Network pilot launched by Singapore government. Two towns (Punggol and Bukit Panjang) and eight MRT stations were selected to deploy 70 lockers.
The locker infrastructure figures were retrieved from `https://www.lockeralliance.net`.

the national economy, the long-term viability of the locker network nevertheless hinges on answers to the following key fundamental question:

What should be the right density and coverage of the LA network in the city? In particular, how near should we place a locker to a customer to make usage of the service appealing? The Singapore government aims to install a locker station within 250 meters of every public housing block in the city.[7] Is this coverage suitable? How will this affect utilization?

How many parcels in other part of the city will be shifted to self-pickup at lockers in residential areas? It has been argued that growth in parcel deliveries would contribute to slower city commutes and greater carbon emission. In the case of Singapore, shifting such deliveries away from the central business district (CBD) in the city will go a long way to curb the problems with traffic congestion in the CBD. How can a locker in residential areas, far away from the CBD, divert parcel volume away?

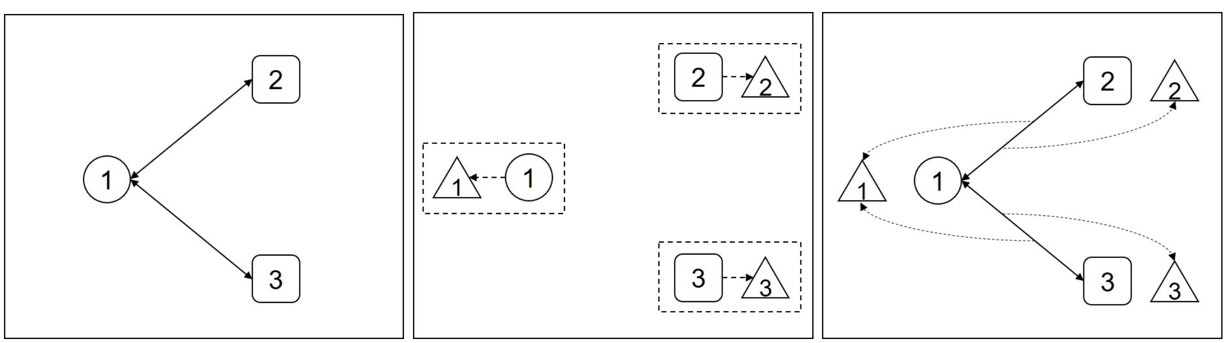
We propose a locker network design model to provide answers to these questions. To design the locker network, it is critical to address the issue of **locker choice** in the problem, using current parcel delivery dataset obtained in the case without the LA Network (i.e., when the customer does not have LA network as a choice). Note that we observe only the parcel volume at each delivery location, but do not know the choice processes underlying the decisions. To understand this, we provide an example to illustrate how this affects the design of locker networks.

[7] As comparison, the city has a public bus stop within 400 meters of every public housing block.

Example 1. Consider an environment with two groups of customers, all living in one home location $\mathcal{H} = \{1\}$, and working in one of the two office locations $\mathcal{W} = \{2, 3\}$ (cf. Figure 2(a)). The customers currently have their parcels (one parcel delivery for each customer) delivered to either home or office, leading to two different demand segments $(1, 2)$ and $(1, 3)$. Note that the demand segments here correspond to the transit routes taken by the customers. Let $D_{1,2} = D_{1,3} = 12$ denote the number of such customers in each segment. Furthermore, customers from each demand segment form a consideration set of parcel delivery options.

We assume that the customers' preferences towards parcel deliveries follow an attraction model (e.g., MNL choice model). For ease of exposition, we denote respectively the attraction of delivery-to-home and delivery-to-office by $\theta_{\mathcal{H}}$ and $\theta_{\mathcal{W}}$ for both groups of customers. Furthermore, we set $\theta_{\mathcal{H}} = 1, \theta_{\mathcal{W}} = 3$. Prior to the installation of lockers, customers from segment $(1, 2)$ can get their parcels delivered to either location 1 or location 2, while customers from segment $(1, 3)$ can do so at location 1 or location 3. Based on the MNL model, we observe that 6 customers have their parcels delivered to location 1 ($D_1^O = (D_{1,2} + D_{1,3}) \times \frac{\theta_{\mathcal{H}}}{\theta_{\mathcal{H}} + \theta_{\mathcal{W}}} = 6$), 9 customers have their parcels delivered to location 2 ($D_2^O = D_{1,2} \times \frac{\theta_{\mathcal{W}}}{\theta_{\mathcal{H}} + \theta_{\mathcal{W}}} = 9$), and the other 9 customers have their parcels delivered to location 3 ($D_3^O = D_{1,3} \times \frac{\theta_{\mathcal{W}}}{\theta_{\mathcal{H}} + \theta_{\mathcal{W}}} = 9$), where we use D_i^O to represent the observed parcel volume at each location $i \in \{1, 2, 3\}$.

Figure 2 Locker network design example.



(a) Demand Profile (b) Traditional Model (c) Model with Actual Demand $\boldsymbol{D}$

Note. The circle stands for home location and the squares for offices. Lockers are indicated as triangles. The lockers can be installed either next to home or offices.

There are three potential locations $\{1, 2, 3\}$ for parcel locker installation (cf. Figure 2(b)). Let $\theta_{\alpha,k}$ denote the attraction value of the locker installed at location $k \in \{1, 2, 3\}$ to the customer from demand segment $\alpha = (1, 2)$ or $(1, 3)$. We set $\theta_{(1,2),1} = \theta_{(1,2),2} = \theta_{(1,3),1} = \theta_{(1,3),3} = 2$, and $\theta_{\alpha,k} = 0$ otherwise. Suppose our objective is to install one locker so that the volume of delivery to the locker is maximized. Where should we locate the locker?

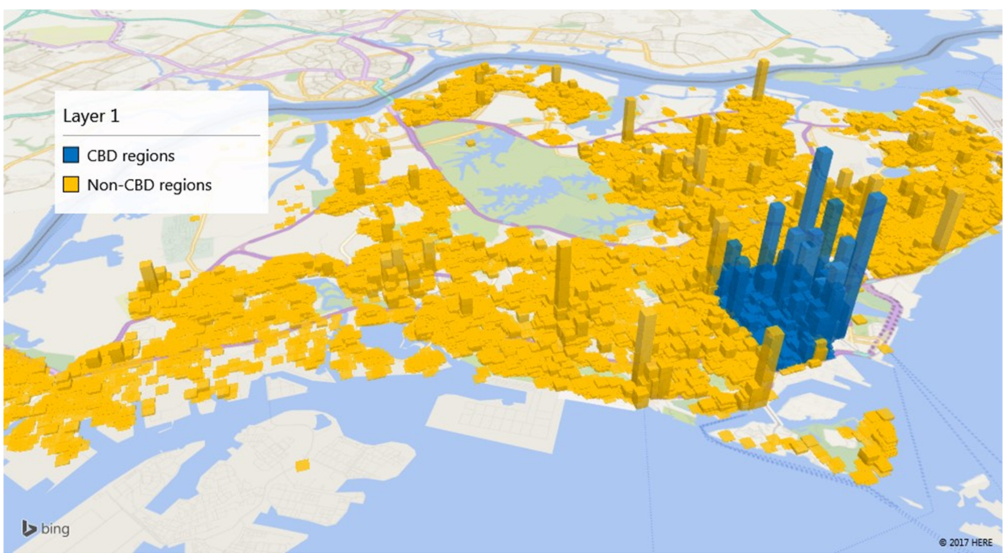
In the traditional facility location model, the volume of delivery to the locker is calculated based on the observed parcel volume $\boldsymbol{D}^O$, which shows 6 deliveries at location 1, 9 at location 2, and 9 at location 3. Putting locker at location 1 seems to be a bad choice. However, in our example, if a locker is installed at location 1, customers from both segments $(1,2)$ and $(1,3)$ will put locker 1 as an alternative delivery option into their considerations sets. As a result, the volume of parcels diverted to the locker is

$$\frac{\theta_{(1,2),1}}{\theta_{\mathcal{H}}+\theta_{\mathcal{W}}+\theta_{(1,2),1}}\times D_{1,2}+\frac{\theta_{(1,3),1}}{\theta_{\mathcal{H}}+\theta_{\mathcal{W}}+\theta_{(1,3),1}}\times D_{1,3}=\frac{2}{1+3+2}\times 12+\frac{2}{1+3+2}\times 12=8.$$

This turns out to be the optimal design to maximize locker utilization, as locker in location 2 or 3 will attract only 4 deliveries. Therefore, it is optimal to locate the parcel locker in area with low observed parcel volume! ■

This issue arises because locker at a location will be considered only if it lies on the transit route α of the customers. Unfortunately, obtaining the transit route of all customers in a city appears to be a daunting task, and we only have data on the current delivery volume to each location. In this paper, we develop a robust locker network design model to address this issue, and show how we can still glean useful insights on the potential impact of the LA network.

Figure 3 **Empirical parcel delivery to CBD and non-CBD regions in Singapore over a period of 3 months.**



Using real data on observed delivery volume to homes and offices provided by government agency and industry players, and a set of locker usage data from a commercial courier company, we examine the impact of the LA Network on the re-distribution of the volume of parcel deliveries into the central business district (CBD) in Singapore. We note that the delivery profile (cf. Figure 3) clearly indicates that the parcel density (more than 6000 parcels per day) in CBD is almost 3 times that of the average amount delivered to other areas. Given that the Government will allow LA lockers to be installed in residential blocks and community/transit hot spots like MRT stations outside of the CBD, it will be important to understand how many of these parcels will be switched to be picked up from lockers near their residential addresses? In this way, we can maintain high utilization of the lockers and yet reduce the number of parcels delivered to the CBD area at the same time.

Our key **contributions** in the paper are summarized as follows:

(1) **Customer Choice.** Using a set of proprietary locker usage data in a commercial locker network, we propose and calibrate a locker choice model to capture the appeal of the locker pickup option for the customers. This allows us to obtain a partial answer to the key question – how close do we need to locate a locker near the customer to ensure that it will have a high chance of being used?[8] The analysis suggests that the planning norm (250 meters) adopted by the Singapore government seems to be an appropriate trade-off for the design of the LA network.

(2) **Network Design.** If we know the transit route of each customer in the system, we could determine the appeal of each locker in the network for the customer, using the customer choice model developed earlier. However, in the absence of such information, and using only the volume of parcels delivered to each address in the country, we develop a facility location model to extrapolate and approximate the utilization of each locker installed in the LA network. We demonstrate that the optimality loss incurred by this approach is bounded above by a constant factor when the attractiveness of locker pickup is smaller than delivery to homes or offices.

(3) **Computational Efficiency.** Given the size of the LA Network, standard MIP approaches using the big-M formulation do not scale well. We propose a SOCP-MIP model for this large-scale network design problem, reducing drastically the size of the constraints and variables for the problem.

(4) **Managerial Insights.** With a planning norm of a locker within 250 meters of every public housing block, we estimated that the LA network (with 1500 lockers) has the ability to divert at least 7.5% of the parcel volume currently delivered into the CBD.

Outline of this paper. The rest of this paper is organized as follows. We review relevant literature in Section 2. In Section 3, we describe the dataset we used to study this problem. In Section 4, we calibrate a locker choice model that we will use in the network design model. In Section 5, we address the locker network design problem. We apply this model to study the impact of LA Network on the delivery profile in the CBD area in Section 6. Section 7 concludes the paper. We present a compact reformulation of the locker network design model in Appendix A. In Appendix B, we focus in the case when each transit route consists of only home-office pair, and estimate the volume of

[8] We assume that consumers enjoy one dollar discount by choosing to pick up at lockers, which is the operating model for this set of data. We ignore the pricing issue in this paper because such decisions are more operational in nature.

traffic on each pair, using a set of public transit data in Singapore. This allows us to estimate the hitherto unknown transit route for each customer and solve the locker network model with the transit route information incorporated. For completeness, we compare the market share maximization model (studied in this paper) and the customer welfare maximization model (a plausible alternative for the network designer) in Appendix C. More details on the Singapore LA Network case are provided in Appendix D. In Appendix E, we provide the technical proofs.

2. Literature Review

This work is motivated by the announcement of the plan by the Singapore Government to launch a nation wide LA Network, as part of the Singapore Smart Nation Initiatives. We restrict our focus to the literature related to locker operations. For other topics on last mile innovation, we refer the interested readers to Savelsbergh and Van Woensel (2016) and Ranieri et al. (2018).

Torres and Suggs (2015) provided a detailed description of the locker system. For each locker station, there are multiple compartments of different sizes, used for parcel delivery and retrieval. Some researches focused on the design and development of smart locker system (e.g., Fee 2015, Irwin et al. 2015) while some investigated the effect of locker system on city logistics. For example, Song et al. (2009) quantified the saving of delivery cost if failed deliveries can be diverted to lockers. Morganti et al. (2014a) described the last mile innovation by using locker networks in European countries, and Morganti et al. (2014b) demonstrated the effectiveness of locker network as an alternative to home delivery. Ranieri et al. (2018) introduced the locker system as an innovative strategy to reduce the externalities cost such as traffic congestion and pollution. In addition, the locker system was also demonstrated to be effective in reducing the number of deliveries in city area and thus reducing the liter consumption and CO_2 emission (cf. Edwards et al. 2009, 2010, Iwan et al. 2016). Notably, although it has been widely accepted that the locker can be used to improve the delivery efficiency and reduce the negative impact on environment, few researches have been done on the locker network design problem (Savelsbergh and Van Woensel 2016).

Locker network design is essentially a facility location problem, which is a long-standing topic in the operations management literature. Wu et al. (2015) used public transport data and historical delivery data to locate the locker stations. The key approach was to estimate transport mobility pattern of the population and they suggested that lockers should be installed at crowded places to serve more customers. Wang et al. (2017) developed a covering model to decide the optimal locker network so that the volume of captured demand could be maximized. Recently, Lin et al. (2020a) studied the locker location problem and used the multinomial logit (MNL) model to predict customer's choice towards different lockers. Moreover, Lin et al. (2020b) introduced a threshold luce model to capture the zero-probability choice in the locker location problem. Along this direction, we note that the

MNL model is a natural option to formulate customer choices in the facility location problem. For example, Zhang et al. (2012) studied the MNL model in a healthcare facility location problem. Aros-Vera et al. (2013) addressed the park-ride location problem under the MNL framework. In a similar vein, we adopt the MNL model to capture customer choices in the locker network design problem. Furthermore, we empirically calibrate the choice parameters and validate the performance of MNL model using a set of proprietary data.

Based on the data analytics, we observe that substitution effect exists among different lockers. Glaeser et al. (2019) empirically studied the substitution effect among different pickup points operated by one online retailer. They showed that the pickup points negatively affect the demands of their nearby points within 0.5 miles. To capture this effect in their location problem, they developed a mixed quadratic and integer programming model. A similar location model was also developed in Kung and Liao (2017). However, Kung and Liao (2017) claimed that a facility would be more attractive with more nearby facilities. In the present work, we also develop an econometric model to study the substitution effects across nearby lockers.

In the end, we highlight that we develop a robust location model to address the locker network design problem since the demand information is not directly observable. Notably, robust optimization techniques have been used to solve different variants of facility location problems under uncertain environments. For example, Baron et al. (2011) studied a multi-period facility location problem with demand uncertainty. Chan et al. (2017) developed a robust model to decide the locations for automated external defibrillator while the uncertainty comes from the spatial distribution of potential demands. Wang et al. (2020) studied a robust hub location problem in the face of uncertain commodity demand and cost information. Under our robust model, we derive a compact formulation to address the locker network problem.

3. Description of Dataset

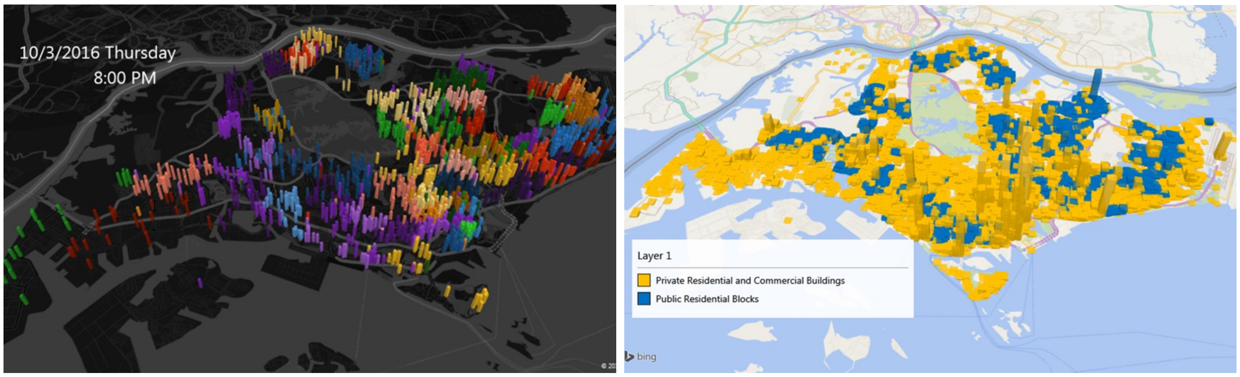
Singapore is a heavily urbanized city-state in Southeast Asia and has a total land area of 719.1 square kilometers.[9] Its central business district (CBD) occupies 17.84 square kilometers,[10] accounting for 2.5% of the total land area. However, among the average daily volume of parcels delivered, 14.39% of them are delivered to the CBD area. These estimates are obtained from a set of proprietary delivery data provided by a commercial operator in Singapore. The average distance of the last mile operation in Singapore (measured from the distribution center of the operator to the addresses of the delivery operations) is around 10.16 km for this company. The delivery activities are organized based on a zonal system. Figure 4(a) shows the delivery activities of different workers on a typical day in the

[9] Geography of Singapore; retrieved from `https://en.wikipedia.org/wiki/Geography_of_Singapore`

[10] Central Area, Singapore; retrieved from `https://en.wikipedia.org/wiki/Central_Area,_Singapore`

country. The deliveries start usually near the central area, and fan out gradually to the outskirts near the end of the day. The workers are also organized in the zonal system, and each worker delivers only to a small zone in the country.

Figure 4 Last mile delivery in Singapore.



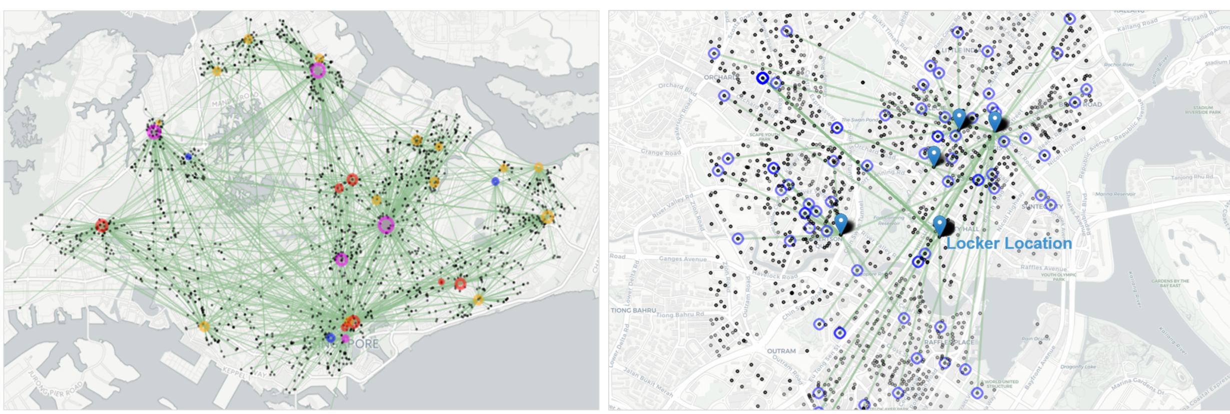
(a) Parcel Delivery Activity (b) Demand Distribution over the Island

Note. (a) Each delivery worker, identified by a unique color, is responsible for a specific zone in the country at the end of a typical day. (b) The yellow/blue (light/dark) bars represent the commercial buildings/residential blocks, respectively. The height visualizes the volume of delivery at the location.

Our dataset contains three months of parcel delivery records of a parcel delivery company in Singapore, from January 2016 to March 2016, with deliveries to $28,953$ locations over the island. For each parcel delivery record, we obtain the (unique) delivery ID, delivery time, delivery destination (address, longitude, and latitude), courier ID, and delivery status (delivered or failed). The estimated volume of daily average deliveries is 42,650. According to our collected delivery records, more than 20% of parcels could not be physically delivered to their destinations. Among the failed deliveries, we observe that almost 40% were due to customers not showing up at the pre-arranged time windows. We also collected the delivery records between January 2017 and February 2017 to examine the predictive performance of our proposed customer choice model towards the locker usage. The geographical distributions of delivery to public residential and private/commercial blocks are plotted in Figure 4(b). According to the spatial distribution of these buildings, public residential deliveries are mainly along the outskirt of the island, while most commercial deliveries are clustered at the downtown areas (the central part of the island). As Singapore is a densely populated country, we cluster the public residential and private/commercial blocks (including shopping malls and public transit stations) into population centers. We use K-Centroids approach to cluster the dataset into 1000 public residential centers (represented by set $\mathcal{H}$) and 2000 private/commercial centers (represented by set $\mathcal{W}$), based on geographic proximity. We estimate the daily parcel volume (i.e., market share) to each center, using the parcel delivery dataset.

Prior to the announcement of the LA Network, our industry partner has experimented with locker pickup operations in 29 selected locations, in collaboration with the convenience store 7-11 in Singapore. As an incentive, customers were given a dollar discount off their delivery charges if they opt for locker pickup. This dataset provides a unique glimpse into the way customer chooses to pick up parcels from the lockers. For each parcel delivery (to locker) record, we obtain the delivery ID, parcel delivery time to locker, parcel collection time from locker, receiver information (customer address), locker station, locker size (larger, medium, and small), and delivery status (delivered or failed). Note that some of the reported addresses are in private/commercial estates, indicating that the customers were picking up parcels at lockers near their work areas. As estimated from the dataset, we observe that most of the customers collected their parcels from the lockers within one day. The median of parcel retrieval times is 7.49, 6.36 and 6.97 hours, in January, February and March, respectively. Around 3/4 customers collected their parcels within 24 hours, while the remaining customers picked up their parcels much later for various reasons.

Figure 5 Locker usage patterns.



(a) Parcel Collection Pattern (Singapore) (b) Parcel Collection Pattern (CBD Region)

Note. (a) The dots represent the customer addresses while the circles represent the locker locations. The line between the circle and the dot visualizes the parcel collection pattern of a locker user. The size of each circle corresponds to parcel volume to the locker. (b) The markers indicate the installed lockers in the CBD region. The blue circles represent the locations of these locker users while the black dots represent the locations of those customers choosing outside options. The line between the circle and the marker visualizes the parcel collection pattern of a locker user.

The reported customer addresses and the locker locations selected for pickup are shown in Figure 5. We collected 1680 locker usage observations and plot the customer pickup behavior. The 29 lockers spread all over the island (determined by the location and suitability of 7-11 outlets), and hence the dataset is sufficiently representative to characterize the locker usage behaviors of the entire population. We observe that some lockers are at major shopping centers or close to MRT (subway) stations, whereas some are at convenience stores in residential areas. Unfortunately, with 29 locker locations, the volume of parcels picked up at lockers are around 1% of the total parcel volume delivered by the company. Many of the lockers were not well utilized at this scale. In addition, we observe that 69.3% customers chose the lockers within 3 kilometers of their addresses, while some customers clearly prefer to pick up parcels near the MRT stations that may be far away. Notably, these lockers located at both shopping malls and MRT stations attract more customers to use. To capture the effect of locker location type on customer choice, we introduce an index *LockerType* to indicate whether the locker is installed near both MRT and Shopping mall or not. Furthermore, we observe that customers from public residential blocks are more likely to use lockers, compared to the customers from commercial blocks. Motivated by this observation, we calibrate the choice models for customers from residential blocks and commercial districts separately.

4. Customer Choice Model

Previous research focuses on exploring the incentives of customers to use lockers, but ignores their choice behavior in selecting different lockers available to them. Collins (2015) used a survey in Australia to investigate customer choice between conventional delivery and lockers. They showed that the distance to lockers plays a significant role in affecting customer choice. Weltevreden (2008) empirically studied the utilization of lockers in Netherlands and they showed that customers will obtain higher utility to use lockers when there are more lockers in the vicinity of their home. Based on these studies, we hypothesize that, *the distance from customer reported address to the selected locker location* has a significant impact on customer locker choice behavior. To be more concrete, we hypothesize that customer would be more sensitive to distance if the distance is within a certain range. Their utilities towards lockers decrease drastically with the increase of distance to lockers. But once the distance exceeds a threshold, customers would be indifferent to all locker options. Therefore, we can use a concave nondecreasing function to re-scale the distance measurement.

We also notice that the substitution happens across different lockers, i.e., the delivery to one locker would be affected by the presence of other lockers in the vicinity. Glaeser et al. (2019) demonstrated that this effect exists in a retail location setting and introduced a regression model to calibrate this effect (which is called spatial cannibalization effect in their paper). The substitution effect is widely studied in the assortment literature (e.g., Talluri and Van Ryzin 2004) and the MNL choice model is commonly used to capture this effect across different products. Motivated by the results in Glaeser et al. (2019), we use a similar econometric model to demonstrate that the substitution effect exists in the locker network and then use the MNL model to calibrate customer choice towards the lockers.

Last but not least, we collect an out-of-sample dataset (i.e., the records between January and February 2017) to test the model, and compare the predicted market share by the econometric model and our MNL model. It turns out that the MNL model provides better prediction accuracy. Hence, we calibrate the customer choice using the MNL model in the locker network design study.

4.1. Substitution Effect in the Locker Network

We combine two sources of delivery datasets from January 2016 to March 2016. The first source contains the traditional delivery to home/office records, and the second dataset contains the delivery to locker records. Following Glaeser et al. (2019), we aggregate the monthly average demand to each location in the following regression model:

$$\text{(Model 1)} \quad ParcelsToLocker_{k,m} = \beta_0 + \beta_1 LockerType_k + \beta_2 NearbyLockers_k + \beta_3 TotalDelivery_m(log) + \beta_4 LockerLocation_k + \epsilon_{k,m}.$$

The subscript $k \in \mathcal{S}^0 = \{1, 2, \ldots, 29\}$ refers to each locker station, and $m = \{1, 2, 3\}$ represents the index of month. The response variable $ParcelsToLocker_{k,m}$ represents the total volume of parcels delivered to locker k during month m. The explanatory variable $LockerType_k$ indicates whether the locker k is located near both MRT and Shopping mall or not. $NearbyLockers_k$ indicates if there are other lockers within the vicinity (i.e., within stipulated distance). $TotalDelivery_m$ represents the total volume of parcels during month m. We do not use the volume of parcels at the vicinity of each locker as additional feature, since the variable $LockerLocation_k$ has already been used to control for the locational effect, i.e., the zonal information of each locker location. We assume the error term $\epsilon_{k,m}$ captures other factors that are not specified in our model.

Table 1 Substitution effect across neighbor lockers.

Variables	Model 1	Model 2	Model 3
Nearby Lockers within 1 km	-7.322^{**}	-7.319^{*}	-7.319^{*}
	(2.208)	(2.824)	(3.134)
Nearby Lockers from 1 km to 3 km	2.044	−2.383	−2.383
	(4.274)	(5.409)	(6.003)
Nearby Lockers from 3 km to 5 km	3.302	2.901	2.901
	(3.885)	(4.969)	(5.515)
Locker Type	21.122^{***}	–	–
	(2.969)	–	–
Volume of Total Deliveries (log)	19.643^{***}	19.643^{***}	–
	(3.495)	(4.470)	–
(Intercept)	-198.225^{***}	-189.325^{***}	18.188^{**}
	(37.138)	(47.476)	(5.415)
Location Control	True	True	True
Number of Observations	78	78	78
R-squared	0.555	0.263	0.080

Signif. codes: $***p < 0.001$, $**p < 0.01$, $*p < 0.05$

Robust standard errors are provided in parentheses.

The regression results are summarized in Table 1. We observe that the coefficient of $Nearby\ lockers\ within\ 1\ km$ is significantly negative, which implies the existence of substitution effect among neighbor lockers. In other words, nearby lockers within 1 km will decrease demand by 7 on average. However, additional lockers farther than 1 km do not have significant effect. To check the robustness of this effect, we remove the variables $LockerType_k$ and $TotalDelivery_m$ from Model 1 and develop two simplified models, say Model 2 and 3, respectively. Notably, a consistent substitution effect can be observed in all models. We also find that these lockers located near both MRT and Shopping mall are significantly more attractive than the others.

In fact, it is natural to apply this class of econometric models to predict the market share captured by each locker station, but it is hard to directly capture the customer's utility. Next, we use the MNL model to calibrate customer choice towards the lockers.

4.2. Calibration of MNL Choice Model

We use the locker usage dataset to calibrate the MNL choice model for locker operation, to understand the impact of "distance of the locker from home/office", and the locational feature like whether the locker is located near a major shopping mall and/or train station. Besides the option of choosing lockers for parcel delivery, the outside options (i.e., delivery-to-home and delivery-to-office) are also considered in the discrete choice model. This forms the basis for our LA Network design model. Furthermore, to control for wealth effect, we split the whole delivery dataset into residential block location set (denoted by $\mathcal{H}^0$) and commercial district location set (denoted by $\mathcal{W}^0$), based on the addresses reported by the customers, and estimate the choices for two groups of customers separately. We assume that customers are homogeneous within each group, and they are utility maximizer in using lockers. Due to space constraint, we describe only the case of customers with reported addresses at residential blocks to illustrate the model calibration process.

We have 29 lockers (denoted by $\mathcal{S}^0$) installed at different locations (cf. Figure 5(a)). We assume that each customer $i \in \mathcal{H}^0$ can opt to use any locker $k \in \mathcal{S}^0$ for parcel delivery. Let $\theta_{i,k}$ denote the utility obtained from locker k for customer i. Besides, she can choose to receive the parcel at the reported address (outside option), represented by $\{0^{\mathcal{H}}\}$. Let $\theta_{\mathcal{H}}$ denote the utility obtained from the outside option. In this way, the consideration set for customer $i \in \mathcal{H}^0$ can be represented by $\{0^{\mathcal{H}}\} \cup \mathcal{S}^0$. Here we assume that all the customers share the same consideration set in the locker experiments. Under the MNL choice model (Talluri and Van Ryzin 2004), the probability that customer i selects locker k for parcel delivery is given by:

$$
\begin{aligned}
\mathbb{P}_{i,k} &= \mathbb{P}(\text{customer } i \text{ selects locker } k)\\
&= \mathbb{P}(\text{customer } i \text{ selects locker } k | \text{customer chooses to pickup from locker})\\
&\qquad \times \mathbb{P}(\text{customer chooses to pick up from locker})\\
&= \frac{\theta_{i,k}}{\sum_{l \in \mathcal{S}^0} \theta_{i,l}} \times \frac{\sum_{l \in \mathcal{S}^0} \theta_{i,l}}{\theta_{\mathcal{H}} + \sum_{l \in \mathcal{S}^0} \theta_{i,l}},
\end{aligned}
$$

where $\sum_{l \in \mathcal{S}^0} \theta_{i,l}$ represents the total utility gained from the set of lockers against outside option.

Indeed, it is standard to jointly estimate the utility from outside option $\theta_{\mathcal{H}}$ together with the utility of using lockers $\theta_{i,k}$. However, our locker experiment dataset is highly imbalanced, with very few customers using the delivery-to-locker option (cf. Figure 5(b)). To deal with this challenge, we introduce a two-stage approach to calibrate $\theta_{i,k}$ and $\theta_{\mathcal{H}}$, respectively. At stage 1, we implement the R package *mlogit*[11] to estimate the customer utility towards different lockers, given that customers have chosen lockers. At stage 2, we apply the *maximum likelihood estimation* method to estimate the utility of outside option.

Stage 1: Customer Choices over Different Lockers

We model the utility $\theta_{i,k}$ obtained from locker k by customer i as

$$\log(\theta_{i,k}) = \beta_1(Distance_{i,k})^{\gamma} + \beta_2 LockerType_k$$

where the explanatory variable $Distance_{i,k}$ represents the distance from customer location i to the selected locker k. We let $\gamma \in (0,1)$ model the diminishing effect of distance on locker choice. The best fit obtained from our dataset is $\gamma = 1/3$. The other (binary) explanatory variable $LockerType = 1$ if the locker k is installed at a location near both MRT and Shopping Mall; $LockerType = 0$ otherwise. $\{\beta_1, \beta_2\}$ are the coefficients to be estimated.

The model calibrations for customers from public residential blocks and commercial blocks are provided in Table 2. The coefficients for ($Distance$) are negative in both choice models, and hence the utility of using locker decreases with the distance from the locker to customer address. Furthermore, the lockers located at both MRT and Shopping Mall are more attractive to customers. This observation is consistent with the results from Model 1. We also find that the MRT and Shopping Mall lockers are more attractive to the customers from residential blocks, compared to those from commercial blocks.

Table 2 Logit of customers from public residential blocks.

Variables	Customers from Public Residential Blocks	Customers from Commercial Blocks
$(Distance_{i,k})^{\frac{1}{3}}$	-4.59^{***}	-4.47^{***}
	(0.14)	(0.19)
$LockerType$	1.50^{***}	0.52^{***}
	(0.11)	(0.15)
Number of Observations	1106	574
Log-Likelihood	-946.8	-587.4

Signif. codes: $***p < 0.001$, $**p < 0.01$, $*p < 0.05$

Robust standard errors are provided in parentheses.

[11] Please see `https://cran.r-project.org/web/packages/mlogit/mlogit.pdf`

Stage 2: Estimation of Outside Option

We estimate next the utility of outside option $\theta_{\mathcal{H}}$. Let a_i denote the population of customers opting to having the parcels delivered to their home addresses (i.e. outside option), and b_i be the population who opted to picking up parcels from the lockers. We can represent the probability of choosing outside option for customer i as $\frac{\theta_{\mathcal{H}}}{\theta_{\mathcal{H}}+\sum_{l\in\mathcal{S}^0}\theta_{i,l}}$, which is treated as the likelihood function. Similarly, we can represent the probability of choosing the existing lockers for delivery as $\frac{\sum_{l\in L}\theta_{i,l}}{\theta_{\mathcal{H}}+\sum_{l\in\mathcal{S}^0}\theta_{i,l}}$. We target to find the optimal $\theta_{\mathcal{H}}$ such that the total likelihood function is maximized by the following model:

$$\max \sum_{i\in\mathcal{H}} \left[a_i \log\left(\frac{\theta_{\mathcal{H}}}{\theta_{\mathcal{H}}+\sum_{l\in\mathcal{S}^0}\theta_{i,l}}\right) + b_i \log\left(\frac{\sum_{l\in\mathcal{S}^0}\theta_{i,l}}{\theta_{\mathcal{H}}+\sum_{l\in\mathcal{S}^0}\theta_{i,l}}\right)\right]$$

$$\text{s.t.}\quad \theta_{\mathcal{H}} \geq 0$$

Recall that we let $\theta_{\mathcal{H}}$ denote the utility of outside option for the customers from public residential blocks, and $\theta_{\mathcal{W}}$ the utility of outside option for the customers from private/commercial blocks. Using this two-stage calibration approach, we obtain that $\theta_{\mathcal{H}} = 7.06$ and $\theta_{\mathcal{W}} = 17.79$. The difference between these two utilities implies that customers from public residential blocks are more likely to use lockers, compared to those from private/commercial blocks.

To summarize, the calibrated MNL choice models for the two groups of customers are given by:

$$\begin{cases} \log(\theta_{i,k}) = -4.59 \times (Distance_{i,k})^{\frac{1}{3}} + 1.50 \times LockerType_k,\ \theta_{\mathcal{H}} = 7.06, \quad (i \in \mathcal{H}^0) \\ \log(\theta_{j,k}) = -4.47 \times (Distance_{j,k})^{\frac{1}{3}} + 0.52 \times LockerType_k,\ \theta_{\mathcal{W}} = 17.79,\ (j \in \mathcal{W}^0) \end{cases} \tag{1}$$

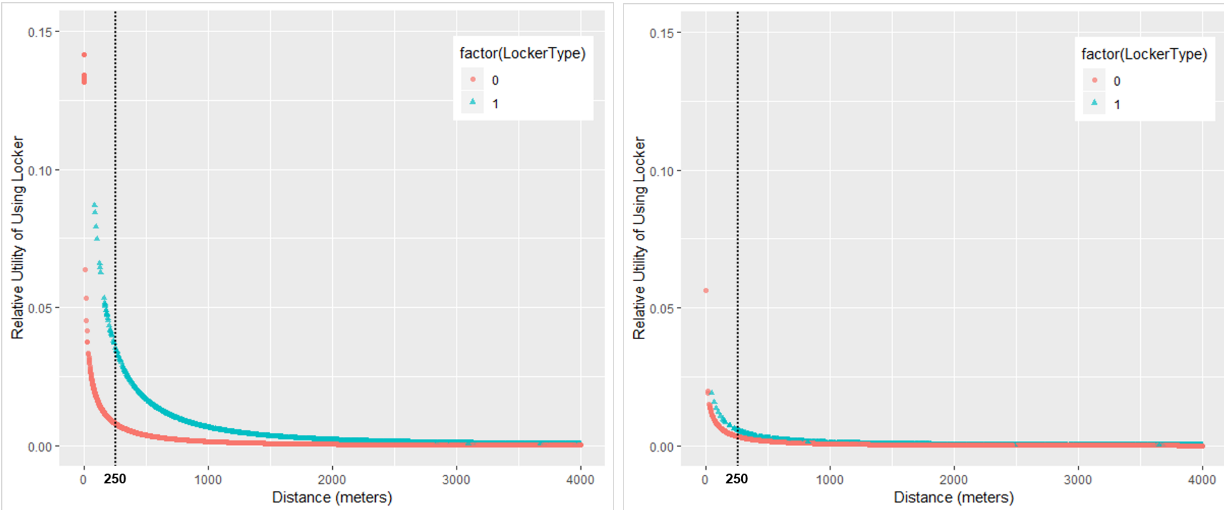
We acknowledge that the time-variant attributes (e.g., weekday effect) and the price were not controlled in the locker experiments. The MNL model can be extended to incorporate the pricing issue if this information is available. Next, we show the impact of parcel pickup distance and locker type on the relative utility of using lockers (i.e., the value of $\theta_{i,k}/\theta_{\mathcal{H}}$ for $i \in \mathcal{H}^0$, and $\theta_{j,k}/\theta_{\mathcal{W}}$ for $j \in \mathcal{W}^0$) in Figure 6. It shows that: (1) the utility of using lockers decreases in the distance from the customer address to the locker. In particular, the utility drops fast when the distance exceeds 250 meters; (2) customers become indifferent to the lockers far away from them (e.g., over 2000 meters away from them); (3) Lockers, especially those located near both major shopping mall and train station, are more attractive to the customers from residential blocks.

4.3. Out-of-Sample Prediction Performance

We use the delivery records collected between January 2017 and February 2017 for out-of-sample performance test. Note that the choice model is calibrated based on the delivery data from January to March 2016, given the locker pickup operations in 29 selected locations.

To predict the volume of delivery-to-locker, one straightforward approach is to use the historical volume. However, this approach is unable to estimate the volume of delivery to new lockers. For

Figure 6 **Relative utility of using lockers.**



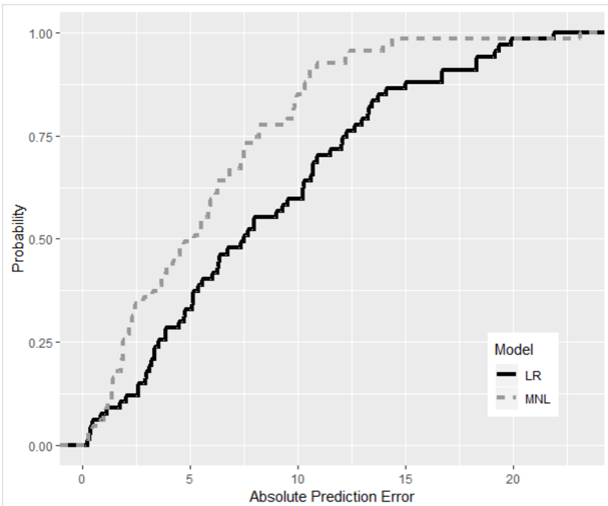
(a) Customers from Public Residential Blocks (b) Customers from Commercial Blocks

example, the company expanded the locker network to 34 lockers in January 2017. 7 lockers were removed from previous 7-11 stores and 12 lockers were installed to another stores. Only 22 lockers were still installed at the same locations. In February 2017, the company removed one more locker from the network. In fact, the company conducted a trial-and-error experiment to roll out the locker network. The lockers with low utilization were removed from the network and some new locations were explored.

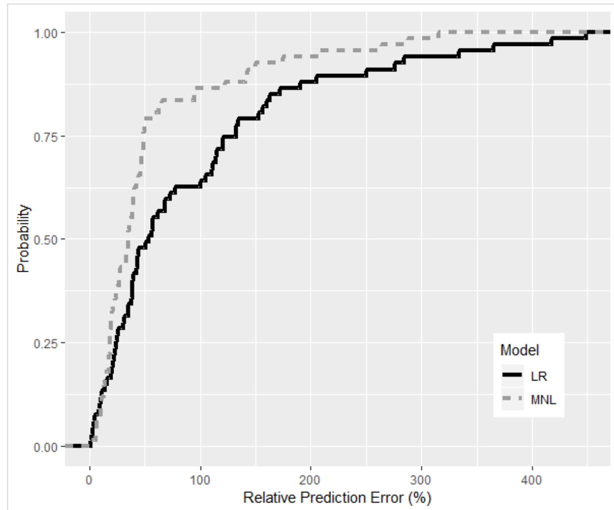
Given the observed parcel volumes in January/February 2017 and the expanded locker network, we predict the monthly volume of delivery to locker using the MNL choice model. The total volume of delivery-to-locker is 989 from January to February 2017, while the predicted volume is 984. Clearly, the MNL model performs well in predicting the total market share captured by the locker network. Furthermore, we calculate the monthly volume of delivery to each locker as well as the predicted volume to each locker. It shows that the correlation coefficient between the set of actual volumes and predicted volumes is 0.77. This result demonstrates the good prediction performance of our MNL choice model in terms of the individual locker market share.

To make the discussion clearer, we provide the density plot of prediction errors under the MNL model in Figure 7. The absolute prediction error is defined as the absolute difference between the (monthly) actual volume and predicted volume; and the relative prediction error is defined as the absolute prediction error divided by the actual volume ($\times 100\%$). We also apply the linear regression (LR) model (i.e., the Model 1 in Table 1) to predict the volume of delivery-to-locker. As shown in Figure 7, the LR model suffers from larger prediction errors, compared with the MNL model. The average absolute prediction errors under the LR model and MNL model are 8.33 and 5.63, respectively. The average relative prediction errors under the LR model and MNL model are 93.02% and 54.40%, respectively. This result indicates better prediction performance of the MNL choice model in our locker network design problem.

Figure 7 **Density plot of prediction errors.**



(a) Absolute Prediction Error



(b) Relative Prediction Error

5. Network Design Model

Let D_α denote the number of parcel deliveries for customers in segment α. For instance, customers who live in home i and work at office j, and will consider picking up parcels at either location i or j are clustered into the same demand segment $\alpha = (i, j)$.

Let the binary decision variable x_k denote the decision whether to locate a locker station in location k or not. We use $g_\alpha(\boldsymbol{x})$ to denote the (convex) disutility of the network design solution $\boldsymbol{x}$ for this segment, i.e., the proportion of deliveries in segment α that will not be picked up at the LA network. In general, D_α is unknown to the planner.

We consider two disjoint sets of locations – the set of public residential housing blocks $\mathcal{H}$ and the set of private/commercial buildings $\mathcal{W}$. The demand segments can be separated into the following three classes:

- **Class I:** Customers who will consider home delivery option, for some $i \in \mathcal{H}$, and pickup from lockers, if there is one installed in the neighborhood of location i (denoted N_i). This segment of demand is denoted by $\alpha = (i)$ for $i \in \mathcal{H}$. Let $D_{i,i}$ denote the parcel volume of such customers in the population, and $\theta_{i,k}$ denote the attraction of locker station k to customers in segment i. Let $\theta_{\mathcal{H}}$ denote the attraction of delivery-to-home for this class of customers, identical for all segment $i \in \mathcal{H}$.

The disutility function, representing the proportion of parcels that will not be using the lockers by these customers, is given by:

$$g_{i,i}(\boldsymbol{x}) = \left\{ \frac{\theta_{\mathcal{H}}}{\theta_{\mathcal{H}} + \sum_{k \in N_i} \theta_{i,k} x_k} \right\};$$

and the proportion of pickup in locker station $k \in N_i$ is given by:

$$g_{i,k}(\boldsymbol{x}) = \left\{ \frac{\theta_{i,k} x_k}{\theta_{\mathcal{H}} + \sum_{k \in N_i} \theta_{i,k} x_k} \right\};$$

- **Class II:** Customers who will consider delivery to office, for $j \in \mathcal{W}$, and pickup from lockers if one is installed in the neighbor set N_j. This segment of demand is denoted by $\alpha = (j)$ for each $j \in \mathcal{W}$. Let $D_{j,j}$ denote the parcel volume of such customers in the population, and $\theta_{j,k}$ denote the attraction of locker station k to customers in segment j. Let $\theta_{\mathcal{W}}$ denote the attraction of delivery-to-office for this class of customers. The proportion of customers who continue to request for office delivery is given by:

$$g_{j,j}(\boldsymbol{x}) = \left\{ \frac{\theta_{\mathcal{W}}}{\theta_{\mathcal{W}} + \sum_{k \in N_j} \theta_{j,k} x_k} \right\};$$

and the proportion of pickup in locker station $k \in N_j$ is given by:

$$g_{j,k}(\boldsymbol{x}) = \left\{ \frac{\theta_{j,k} x_k}{\theta_{\mathcal{W}} + \sum_{k \in N_j} \theta_{j,k} x_k} \right\};$$

- **Class III:** Customers who live in home block $i \in \mathcal{H}$ and work at office building $j \in \mathcal{W}$, and will consider pickup at locker stations near either home i (denoted by N_i) or office j (denoted by N_j). This segment of demand is denoted by $\alpha = (i, j)$ for each $i \in \mathcal{H}$, $j \in \mathcal{W}$. Let $D_{i,j}$ denote the volume of such customers in the population, and $\theta_{(i,j),k}$ denote the attraction of locker station k to customers in segment (i, j). The proportion of customers in this class who continue to request for home or office delivery is given by:

$$g_{i,j}(\boldsymbol{x}) = \left\{ \frac{\theta_{\mathcal{H}} + \theta_{\mathcal{W}}}{\theta_{\mathcal{H}} + \theta_{\mathcal{W}} + \sum_{k \in N_i \cup N_j} \theta_{(i,j),k} x_k} \right\},$$

and

$$\theta_{(i,j),k} = \begin{cases} \theta_{i,k} & k \in N_i \setminus N_i \cap N_j, \\ \theta_{j,k} & k \in N_j \setminus N_i \cap N_j, \\ \theta_{i,k} + \theta_{j,k} & k \in N_i \cap N_j. \end{cases}$$

This gives rise to

$$g_{i,j}(\boldsymbol{x}) = \left\{ \frac{\theta_{\mathcal{H}} + \theta_{\mathcal{W}}}{\theta_{\mathcal{H}} + \theta_{\mathcal{W}} + \sum_{k \in N_i} \theta_{i,k} x_k + \sum_{k \in N_j} \theta_{j,k} x_k} \right\}.$$

For these three classes of customers, there are $|\mathcal{H}| + |\mathcal{W}| + |\mathcal{H}| \times |\mathcal{W}|$ customer segments in our demand model, with $g_\alpha(\cdot)$ represented by $g_{i,i}(\cdot)$, $g_{j,j}(\cdot)$, and $g_{i,j}(\cdot)$, respectively.

Note that there could be other type of demand segments, corresponding to different transit routes (with multiple stops in transit hotspots besides home and office addresses). These customers may consider lockers installed near their transit routes, other than homes or offices. This results in a wider set of lockers in the consideration sets of these customers. For ease of exposition, we focus only on the three classes of customers identified above in this paper. Our results can be extended directly to the more general case.

The empirical delivery profile is obtained in the case $\boldsymbol{x} = \mathbf{0}$. i.e., prior to the installation of the LA Network. Hence we observe only D_i^O and D_j^O, the original number of parcels delivered to each location $i \in \mathcal{H}$ and $j \in \mathcal{W}$ in our dataset, when $\boldsymbol{x} = \mathbf{0}$. Furthermore, let $D_{i,j}^O$ denote possibly lower bound on the volume in the (i,j) segment.[12] Based on the definition of the three classes of customers, we have

$$D_i^O := D_{i,i} + \sum_{j \in \mathcal{W}} \frac{\theta_{\mathcal{H}}}{\theta_{\mathcal{H}} + \theta_{\mathcal{W}}} D_{i,j}, \ i \in \mathcal{H}, \tag{2}$$

and

$$D_j^O := D_{j,j} + \sum_{i \in \mathcal{H}} \frac{\theta_{\mathcal{W}}}{\theta_{\mathcal{H}} + \theta_{\mathcal{W}}} D_{i,j}, \quad j \in \mathcal{W}, \tag{3}$$

where $D_{i,i}$, $D_{j,j}$, and $D_{i,j}$ represent the actual demand at each customer segment (i), (j), and (i,j), respectively. However, the data $D_{i,i}$, $D_{j,j}$ and $D_{i,j}$ are not directly observable, since the home-office pair information of each customer in the population might not be available. In what follows, we present two locker network design models – one deterministic model with actual demand profile $\boldsymbol{D}$ and one robust model with unknown demand profile. By comparing the two models, we explicitly characterize the performance loss if the demand profile $\boldsymbol{D}$ is not available.

Locker Network Design with Locker Choice

We formulate the *LA Network Design* model with demand profile $\boldsymbol{D}$ as follows:

$$\begin{aligned}
(\mathrm{P}) \qquad \min_{\boldsymbol{x}} \ & V(\boldsymbol{x}, \boldsymbol{D}) := \left\{ \sum_{i \in \mathcal{H}} D_{i,i} g_{i,i}(\boldsymbol{x}) + \sum_{i \in \mathcal{W}} D_{j,j} g_{j,j}(\boldsymbol{x}) + \sum_{i \in \mathcal{H}, j \in \mathcal{W}} D_{i,j} g_{i,j}(\boldsymbol{x}) \right\} \\
\text{s.t.} \ & \sum_{k \in \mathcal{S}} x_k \le C \\
& \sum_{k \in M_i} x_k \ge 1, \ \forall i \in \mathcal{H} \\
& x_k \in \{0,1\}, \ \forall k \in \mathcal{S}
\end{aligned}$$

where the objective minimizes the proportion of parcels not delivered to the lockers among all the demand segments $\{D_{i,i}, D_{j,j}, D_{i,j}\}$. In the Singapore LA Network case, the desired goal is to ensure that there is always a locker station within the vicinity of every public residential block (e.g., 250 meters). Therefore, the second set of "egalitarian" constraints forces that $\sum_{k \in M_i} x_k \ge 1, \ \forall i \in \mathcal{H}$,

[12] In Appendix B, we provide a systematic way to estimate the volume of $D_{i,j}^O$ in each segment (i,j), using a set of public transit data in Singapore.

where M_i denotes the locker sets within the vicinity of the residential block i. We numerically examine the impact of egalitarian constraint on the design of locker network in Appendix D.1.

We need the demand profile $\{D_{i,i}, D_{j,j}, D_{i,j}\}$ to solve problem (P). In the Singapore LA Network case, we observe that the volume of parcel delivery is (weakly) positively correlated to the volume of public transit records on origin-destination trips. Therefore, we can use a set of public transit data to obtain a rough estimate of the value of $\{D_{i,j}\}$, to solve problem (P). The numerical study is detailed in Appendix B. We also develop an exact second-order cone programming & mixed integer programming (SOCP-MIP) reformulation for this problem in Appendix A. This reformulation allows us to use available solver to construct a locker network solution in a reasonable computational time, provided $\boldsymbol{D}$ can be reasonably estimated.

Robust Locker Network Design with Locker Choice

Recall that the data $D_{i,i}$, $D_{j,j}$ and $D_{i,j}$ are not directly observable, since the home-office pair information of each customer in the population is not available. To overcome this information gap and issues associated with unknown demand profile, we introduce a robust model $(\mathcal{P})$ to address the *LA Network Design* problem:

$$
\begin{aligned}
(\mathcal{P}) \quad \min_{\boldsymbol{x}} \max_{\boldsymbol{E}\in\mathcal{D}} \; & V(\boldsymbol{x},\boldsymbol{E}) = \left\{ \sum_{i\in\mathcal{H}} E_{i,i} g_{i,i}(\boldsymbol{x}) + \sum_{i\in\mathcal{W}} E_{j,j} g_{j,j}(\boldsymbol{x}) + \sum_{i\in\mathcal{H},\; j\in\mathcal{W}} E_{i,j} g_{i,j}(\boldsymbol{x}) \right\} \\
\text{s.t.} \; & \sum_{k\in\mathcal{S}} x_k \le C \\
& \sum_{k\in M_i} x_k \ge 1, \;\forall i\in\mathcal{H} \\
& x_k \in \{0,1\}, \;\forall k\in\mathcal{S}
\end{aligned}
$$

where the demand uncertainty set (assume $\mathcal{D}\neq\emptyset$) is represented as:

$$
\mathcal{D} := \left\{ \boldsymbol{E}\in\mathbb{R}_+^{|\mathcal{H}|+|\mathcal{W}|+|\mathcal{H}|\times|\mathcal{W}|} \; : \; \begin{array}{l}
E_{i,i} + \sum\limits_{j\in\mathcal{W}} \dfrac{\theta_{\mathcal{H}}}{\theta_{\mathcal{H}}+\theta_{\mathcal{W}}} E_{i,j} = D_i^O, \;\forall i\in\mathcal{H}, \\
E_{j,j} + \sum\limits_{i\in\mathcal{H}} \dfrac{\theta_{\mathcal{W}}}{\theta_{\mathcal{H}}+\theta_{\mathcal{W}}} E_{i,j} = D_j^O, \;\forall j\in\mathcal{W}, \\
E_{i,i}\ge 0, \; E_{j,j}\ge 0, \; E_{i,j}\ge D_{i,j}^O, \;\forall i\in\mathcal{H}, j\in\mathcal{W}.
\end{array} \right\}.
$$

We note that this model is more general than model (P) as we consider more flexible demand profiles. Interestingly, the worst case scenario in the robust model is independent of the solution $\boldsymbol{x}$. The result is formally stated in Proposition 1.

PROPOSITION 1. *In the worst case solution to model ($\mathcal{P}$), assuming $\mathcal{D} \neq \emptyset$, we have closed-form demand profile:*

$$\left\{ \boldsymbol{E} \in \mathbb{R}_+^{|\mathcal{H}|+|\mathcal{W}|+|\mathcal{H}|\times|\mathcal{W}|} \ : \ \begin{array}{ll} E_{i,i} = D_i^O - \sum_{j\in\mathcal{W}} \frac{\theta_{\mathcal{H}}}{\theta_{\mathcal{H}}+\theta_{\mathcal{W}}} D_{i,j}^O, \ \forall\ i \in \mathcal{H}, & \\ E_{j,j} = D_j^O - \sum_{i\in\mathcal{H}} \frac{\theta_{\mathcal{W}}}{\theta_{\mathcal{H}}+\theta_{\mathcal{W}}} D_{i,j}^O, \ \forall j \in \mathcal{W}, & \\ E_{i,j} = D_{i,j}^O, & \forall i \in \mathcal{H}, j \in \mathcal{W}. \end{array} \right\} \tag{4}$$

In the case when $D_{ij}^O = 0$ for $\forall i \in \mathcal{H}$ and $j \in \mathcal{W}$, the worst case solution to problem ($\mathcal{P}$) reduces to $E_{i,i} = D_i^O, E_{j,j} = D_j^O$, and $E_{i,j} = 0$ for $\forall i \in \mathcal{H}$ and $j \in \mathcal{W}$. We denote this by the demand profile $\boldsymbol{E}^O$. This is indeed the observed delivery profile to each location. Proposition 1 implies that model (P) with the worst case demand profile is equivalent to the robust model ($\mathcal{P}$). Furthermore, in the case of $D_{i,j}^O = 0$ for $\forall i \in \mathcal{H},\ j \in \mathcal{W}$, we represent $\mathcal{D}$ as $\mathcal{D}^{\mathbf{0}}$, and model ($\mathcal{P}$) as ($\mathcal{P}^{\mathbf{0}}$) to avoid confusion. Similarly, it is straightforward to show that model ($\mathrm{P}^{\mathbf{0}}$) is equivalent to the robust model ($\mathcal{P}^{\mathbf{0}}$). Note that for any demand profile $\boldsymbol{E} \in \mathcal{D}$, we must have $\boldsymbol{E} \in \mathcal{D}^{\mathbf{0}}$. Therefore, we claim that $\mathcal{D} \subseteq \mathcal{D}^{\mathbf{0}}$.

With a slight abuse of notation, we denote respectively the objective values for model ($\mathrm{P}^{\mathbf{0}}$) and (P) as $V^{\mathbf{0}}(\boldsymbol{x}) := V(\boldsymbol{x}, \boldsymbol{E}^O)$ and $V(\boldsymbol{x}) := V(\boldsymbol{x}, \boldsymbol{D})$, given a feasible locker network solution $\boldsymbol{x}$. It is straightforward to show that the performance gap between $V^{\mathbf{0}}(\boldsymbol{x})$ and $V(\boldsymbol{x})$ is always nonnegative.

THEOREM 1. $V^{\mathbf{0}}(\boldsymbol{x}) - V(\boldsymbol{x}) \geq 0$ *holds for any feasible locker network solution* $\boldsymbol{x}$*.*

Theorem 1 implies that *the market share of the volume of parcels delivered to the lockers, obtained using the demand profile* $\boldsymbol{E}^O$*, is a* **lower bound** *to the locker network design problem under the unknown demand profile* $\boldsymbol{D}$, since the disutility of the former is higher. Next, we formally characterize the performance gap between model ($\mathrm{P}^{\mathbf{0}}$) and (P).

Given a feasible locker network solution $\boldsymbol{x}$, we define the gap between $V^{\mathbf{0}}(\boldsymbol{x})$ and $V(\boldsymbol{x})$ as:

$$\mathrm{Gap}(\boldsymbol{x}) := \frac{V^{\mathbf{0}}(\boldsymbol{x}) - V(\boldsymbol{x})}{\sum_{i\in\mathcal{H}} D_i^O + \sum_{j\in\mathcal{W}} D_j^O}, \tag{5}$$

where $\sum_{i\in\mathcal{H}} D_i^O + \sum_{j\in\mathcal{W}} D_j^O$ refers to the total delivery volume observed from data sets.

For the sake of notational simplicity, let $\delta_{\mathcal{H}}^i(\boldsymbol{x}) := \sum_{k\in N_i} \theta_{i,k} x_k$ denote the total utility of using lockers under the solution $\boldsymbol{x}$ for customers living at residential block $i \in \mathcal{H}$, and $\delta_{\mathcal{W}}^j(\boldsymbol{x}) := \sum_{k\in N_j} \theta_{j,k} x_k$ denote the total utility for customers working at office building $j \in \mathcal{W}$. Furthermore,

$$\chi_{i,j}(\boldsymbol{x}) := \frac{\left[\theta_{\mathcal{H}}\delta_{\mathcal{W}}^j(\boldsymbol{x}) - \theta_{\mathcal{W}}\delta_{\mathcal{H}}^i(\boldsymbol{x})\right]^2}{\left[\theta_{\mathcal{H}} + \delta_{\mathcal{H}}^i(\boldsymbol{x})\right]\left[\theta_{\mathcal{W}} + \delta_{\mathcal{W}}^j(\boldsymbol{x})\right]\left[\theta_{\mathcal{H}} + \theta_{\mathcal{W}}\right]\left[\theta_{\mathcal{H}} + \theta_{\mathcal{W}} + \delta_{\mathcal{H}}^i(\boldsymbol{x}) + \delta_{\mathcal{W}}^j(\boldsymbol{x})\right]}$$

measures the discrepancy between $\delta_{\mathcal{H}}^i(\boldsymbol{x})$ and $\delta_{\mathcal{W}}^j(\boldsymbol{x})$ for each pair (i,j). It is easy to see that $\chi_{i,j}(\boldsymbol{x}) \in [0,1]$ for all (i,j)'s and $\boldsymbol{x}$'s. After some algebra, we can obtain the following results to characterize $\mathrm{Gap}(\boldsymbol{x})$ under the locker network $\boldsymbol{x}$.

Proposition 2. *For any feasible solution $\boldsymbol{x}$, we have $V^{\mathbf{0}}(\boldsymbol{x}) - V(\boldsymbol{x}) = \sum_{i\in\mathcal{H},j\in\mathcal{W}} \chi_{i,j}(\boldsymbol{x}) D_{i,j}$.*

Let $\rho^i_{\mathcal{H}}(\boldsymbol{x}) = \delta^i_{\mathcal{H}}(\boldsymbol{x})/\theta_{\mathcal{H}}$ and $\rho^j_{\mathcal{W}}(\boldsymbol{x}) = \delta^j_{\mathcal{W}}(\boldsymbol{x})/\theta_{\mathcal{W}}$ denote respectively the relative attractiveness of locker pickup at home i and office j for the network $\boldsymbol{x}$. Next, we formally show that the performance gap between model (P$^{\mathbf{0}}$) and (P) can be upper bounded in Theorem 2.

Theorem 2. *$Gap(\boldsymbol{x}) \leq \max_{i\in\mathcal{H},j\in\mathcal{W}} \left\{ \frac{\left[\rho^j_{\mathcal{W}}(\boldsymbol{x}) - \rho^i_{\mathcal{H}}(\boldsymbol{x})\right]^2}{\left[1+\rho^i_{\mathcal{H}}(\boldsymbol{x})\right]\left[1+\rho^j_{\mathcal{W}}(\boldsymbol{x})\right]} \right\} \left\{ \frac{\theta_{\mathcal{H}}\theta_{\mathcal{W}}}{(\theta_{\mathcal{H}}+\theta_{\mathcal{W}})^2} \right\}$ holds for any feasible locker network solution $\boldsymbol{x}$.*

We observe that in the case of locker delivery services, the attractiveness of locker pickup is usually smaller than delivery to homes or offices, regardless of the density of the locker network. In this case, we have $\rho^i_{\mathcal{H}}(\boldsymbol{x}) \leq 1$, and $\rho^j_{\mathcal{W}}(\boldsymbol{x}) \leq 1$. Therefore, our result guarantees that for any feasible solution $\boldsymbol{x}$, the gap between model (P$^{\mathbf{0}}$) and (P) is bounded above by $\frac{1}{8}$, since

$$\frac{\left[\rho^j_{\mathcal{W}}(\boldsymbol{x}) - \rho^i_{\mathcal{H}}(\boldsymbol{x})\right]^2}{\left[1+\rho^i_{\mathcal{H}}(\boldsymbol{x})\right]\left[1+\rho^j_{\mathcal{W}}(\boldsymbol{x})\right]} \leq \frac{1}{2}, \text{ whenever } \rho^i_{\mathcal{H}}(\boldsymbol{x}) \leq 1,\ \rho^j_{\mathcal{W}}(\boldsymbol{x}) \leq 1,$$

and

$$\left\{ \frac{\theta_{\mathcal{H}}\theta_{\mathcal{W}}}{(\theta_{\mathcal{H}}+\theta_{\mathcal{W}})^2} \right\} \leq \frac{1}{4}, \text{ for } \theta_{\mathcal{H}} \geq 0,\ \theta_{\mathcal{W}} \geq 0.$$

Corollary 1. *When $\rho^i_{\mathcal{H}}(\boldsymbol{x}) \leq 1$, $\rho^j_{\mathcal{W}}(\boldsymbol{x}) \leq 1$ for $\forall i \in \mathcal{H}$ and $j \in \mathcal{W}$, $Gap(\boldsymbol{x}) \leq \frac{1}{8}$ holds for any feasible locker network solution $\boldsymbol{x}$.*

5.1. Performance of the Optimal Solution

The previous result shows that planning based on $\boldsymbol{E}^O$, the observed delivery profile before the installation of the LA network, can be used to estimate the amount of demand lost to home and office deliveries, under the unknown demand profile $\boldsymbol{D}$, when the network $\boldsymbol{x}(\boldsymbol{E}^O)$ is built. Here we denote $\boldsymbol{x}(\boldsymbol{E}^O)$ as the optimal solution to problem (P$^{\mathbf{0}}$). By optimizing the network over the demand profile $\boldsymbol{E}^O$, we can obtain a good estimation of the amount of parcels shifted in the network. In the rest of this section, we analyze the "regret", using the solution $\boldsymbol{x}(\boldsymbol{E}^O)$, under the true demand $\boldsymbol{D}$. This is formally defined as follows.

Definition 1. Let $\boldsymbol{x}(\boldsymbol{D})$ denote the optimal solution obtained in problem (P) when the demand profile $\boldsymbol{D}$ is known. For any feasible network $\boldsymbol{x}$, $\text{Reg}(\boldsymbol{x}, \boldsymbol{D}) := V(\boldsymbol{x}, \boldsymbol{D}) - V(\boldsymbol{x}(\boldsymbol{D}), \boldsymbol{D})$ denotes the performance loss from solution $\boldsymbol{x}$, compared with the optimal solution $\boldsymbol{x}(\boldsymbol{D})$.

Next, we analyze $\max_{\boldsymbol{D}\in\mathcal{D}} \text{Reg}(x(\boldsymbol{E}^O), \boldsymbol{D})$ in the rest of this section.

According to Proposition 1, we have

$$V(\boldsymbol{x}(\boldsymbol{D}), \boldsymbol{E}^O) - V(\boldsymbol{x}(\boldsymbol{D}), \boldsymbol{D}) = \sum_{i\in\mathcal{H},j\in\mathcal{W}} \chi_{i,j}(\boldsymbol{x}(\boldsymbol{D})) D_{i,j}.$$

Note that $V(\boldsymbol{x}(\boldsymbol{D}),\boldsymbol{E}^O) \geq V(\boldsymbol{x}(\boldsymbol{E}^O),\boldsymbol{E}^O)$ since $\boldsymbol{x}(\boldsymbol{D})$ is also feasible to problem ($\mathrm{P}^{\mathbf{0}}$). Therefore,

$$\begin{aligned}\sum_{i\in\mathcal{H},j\in\mathcal{W}} \chi_{i,j}(\boldsymbol{x}(\boldsymbol{D}))D_{i,j} &\geq V(\boldsymbol{x}(\boldsymbol{E}^O),\boldsymbol{E}^O) - V(\boldsymbol{x}(\boldsymbol{D}),\boldsymbol{D}) \\ &= V(\boldsymbol{x}(\boldsymbol{E}^O),\boldsymbol{D}) + \sum_{i\in\mathcal{H},j\in\mathcal{W}} \chi_{i,j}(\boldsymbol{x}(\boldsymbol{E}^O))D_{i,j} - V(\boldsymbol{x}(\boldsymbol{D}),\boldsymbol{D}).\end{aligned}$$

Altogether, it is straightforward to see

$$\mathrm{Reg}(\boldsymbol{x}(\boldsymbol{E}^O),\boldsymbol{D}) = V(\boldsymbol{x}(\boldsymbol{E}^O),\boldsymbol{D}) - V(\boldsymbol{x}(\boldsymbol{D}),\boldsymbol{D}) \leq \sum_{i\in\mathcal{H},j\in\mathcal{W}} \Big(\chi_{i,j}(\boldsymbol{x}(\boldsymbol{D})) - \chi_{i,j}(\boldsymbol{x}(\boldsymbol{E}^O))\Big) D_{i,j}.$$

By the previous results, we have:

Corollary 2. *When $\rho^j_{\mathcal{W}}(\boldsymbol{x}) \leq 1$, $\rho^i_{\mathcal{H}}(\boldsymbol{x}) \leq 1$ for $\forall i \in \mathcal{H}$ and $j \in \mathcal{W}$,*

$$\max_{\boldsymbol{D}\in\mathcal{D}} Reg(x(\boldsymbol{E}^O),\boldsymbol{D}) \leq \frac{1}{8}\left(\sum_{i\in\mathcal{H}} D_i^O + \sum_{j\in\mathcal{W}} D_j^O\right).$$

5.2. Extension to More General Demand Profile

Note that we have only considered two disjoint classes of locations (i.e., home set $\mathcal{H}$ and office $\mathcal{W}$) in the model, and assume that only lockers near these locations will appeal to the users. In practice, the users may consider lockers in hot spots near their transit routes, and hence the demand segments in our problem may consist of more than the three classes as described in the earlier section. We generalize next the locker network design model to $\{1,2,\ldots,\Gamma\}$ disjoint sets of choices in the consideration set, to allow for pickup near hot spots[13] other than home or office. There are in total up to $|\Phi| = \binom{\Gamma}{1} + \binom{\Gamma}{2} + \ldots + \binom{\Gamma}{\Gamma}$ classes of customers. We let $\phi \in \Phi$ ($\phi \neq \emptyset$) denote each class (subset) of customers and Φ denote the collections of all non-empty subsets. With a slight abuse of notation, we let $\Phi(\gamma)$ denote the collections of all the subsets of customers who at least use location γ (and may also use some other locations) for pick up or delivery of parcels. For example, the pair $(\gamma,\psi) \in \Phi(\gamma)$ represents the class of customers who use both location γ and some other locations in set ψ.

Note that we can only observe the empirical demand volume D^O_γ to each location γ when the facilities have not been installed, as we cannot directly observe the volume of different classes of customers $D_{\gamma,\psi}$, $\forall(\gamma,\psi)\in\Phi(\gamma)$ (e.g., the home-office pair information is not observable). Hence, we have

$$D^O_\gamma := D_{\gamma,\gamma} + \sum_{(\gamma,\psi)\in\Phi(\gamma)} \frac{\theta_\gamma}{\theta_\gamma + \sum_{j\in\psi}\theta_j} D_{\gamma,\psi}, \quad \forall \gamma = 1,2,\ldots,\Gamma,$$

[13] In the dataset, some customers pickup their parcels near hot spots such as 7-11 stores that are located at/nearby public transit stations. Beside the consideration sets $\mathcal{H}$ and $\mathcal{W}$, we introduce one more set $\mathcal{K}$ to represent the delivery option at hot spots $\mathcal{K}$, and we also use D^O_k to represent the observed delivery volume at location $k \in \mathcal{K}$. In this setting, there are 7 classes of customers who may opt for 1, 2, or 3 consideration sets, respectively.

where $D_{\gamma,\gamma}$ denotes the demand volume to location γ from the class of customers who only use the location γ, while $D_{\gamma,\psi}$ denotes the demand volume to location γ from the class of customers who use both the location γ and some other locations in set ψ. The parameter θ_j represents the attraction of delivery to location j for this class of customers.

Given the customer at location γ and nearby facility $\{x_k,\ k \in N_\gamma\}$ if it is installed, the proportion of customers in class γ who reject to use the facilities $\boldsymbol{x}$ is given by:

$$g_{\gamma,\gamma}(\boldsymbol{x}) = \left\{ \frac{\theta_\gamma}{\theta_\gamma + \sum_{k \in N_\gamma} \theta_{\gamma,k} x_k} \right\}.$$

To avoid confusion, we remark that the subscript (γ,γ) is used to highlight the customer segments who consider only delivery to location γ, and pickup from lockers if there is one nearby γ. This is consistent with the notion (i,i) and (j,j) in the home-office setting. For these customers who come from the demand segment $(\gamma,\psi) \in \Phi(\gamma)$, we denote $\psi(j)$ as the location belong to class $j \in \psi$ and $N_{\psi(j)}$ as the neighbor set of location $\psi(j)$. The proportion of customers in class (γ,ψ) who reject to use the facilities $\boldsymbol{x}$ is given by:

$$g_{\gamma,\psi}(\boldsymbol{x}) = \left\{ \frac{\theta_\gamma + \sum_{j \in \psi} \theta_j}{\theta_\gamma + \sum_{j \in \psi} \theta_j + \sum_{k \in N_\gamma} \theta_{\gamma,k} x_k + \sum_{j \in \psi} \sum_{k \in N_{\psi(j)}} \theta_{j,k} x_k} \right\}.$$

Similar to the home-office setting, we represent the demand profile $\boldsymbol{D}$ as:

$$\left\{ \boldsymbol{D} \in \mathbb{R}_+^{|\Phi|} \ : \ D_{\gamma,\gamma} + \sum_{(\gamma,\psi) \in \Phi(\gamma)} \frac{\theta_\gamma}{\theta_\gamma + \sum_{j \in \psi} \theta_j} D_{\gamma,\psi} = D_\gamma^O,\ \forall \gamma = 1,2,\ldots,\Gamma \right\}. \tag{6}$$

We develop the following generic framework to address the class of locker network design problems with actual demand information $\boldsymbol{D}$:

$$\begin{aligned}
(\mathrm{G}) \quad & \min_{\boldsymbol{x}} U(\boldsymbol{x},\boldsymbol{D}) := \left\{ \sum_{\gamma=1}^{\Gamma} \left[g_{\gamma,\gamma}(\boldsymbol{x}) D_{\gamma,\gamma} + \sum_{(\gamma,\psi) \in \Phi(\gamma)} \frac{1}{|\{\gamma,\psi\}|} g_{\gamma,\psi}(\boldsymbol{x}) D_{\gamma,\psi} \right] \right\} \\
& \text{s.t.} \ \sum_{k \in \mathcal{S}} x_k \le C \\
& \qquad \sum_{k \in M_i} x_k \ge 1,\ \forall i \in \mathcal{H} \\
& \qquad x_k \in \{0,1\},\ \forall k \in \mathcal{S}
\end{aligned}$$

where $|\{\gamma,\psi\}|$ denotes the number of elements in set $\{\gamma,\psi\}$. In the objective function, we divide the second term by $|\{\gamma,\psi\}|$ to account for "double-counting".

In the case when $D_{\gamma,\psi} = 0$ for all pair (γ,ψ), we can represent $E_{\gamma,\gamma} = D_\gamma^O$, and $E_{\gamma,\psi} = 0$ for $\forall(\gamma,\psi) \in \Phi(\gamma)$, $\gamma = 1,2,\ldots,\Gamma$. We denote this by the demand profile $\boldsymbol{E}^O$. Next, we formulate the locker network design problem based on the observed demand profile $\boldsymbol{E}^O$ as:

$$(\mathrm{G}^{\mathbf{0}}) \quad \min_{\boldsymbol{x}} U(\boldsymbol{x},\boldsymbol{E}^O) := \left\{ \sum_{\gamma=1}^{\Gamma} \left[g_{\gamma,\gamma}(\boldsymbol{x}) D_\gamma^O \right] \right\}$$

$$
\begin{aligned}
\text{s.t.}\ & \sum_{k \in \mathcal{S}} x_k \leq C \\
& \sum_{k \in M_i} x_k \geq 1,\ \forall i \in \mathcal{H} \\
& x_k \in \{0, 1\},\ \forall k \in \mathcal{S}
\end{aligned}
$$

With a slight abuse of notation, we denote respectively the objective values for model ($\mathrm{G}^{\mathbf{0}}$) and (G) as $U^{\mathbf{0}}(\boldsymbol{x}) := U(\boldsymbol{x}, \boldsymbol{E}^O)$ and $U(\boldsymbol{x}) := U(\boldsymbol{x}, \boldsymbol{D})$, given a feasible network solution $\boldsymbol{x}$. We show that Theorem 1 can be extended to this generic case, i.e., the volume of demands covered by the lockers, obtained using the observed demand profile $\boldsymbol{E}^O$, is a lower bound to the locker network design problem under the unknown demand profile $\boldsymbol{D}$.

THEOREM 3. *$U^{\mathbf{0}}(\boldsymbol{x}) - U(\boldsymbol{x}) \geq 0$ holds for any feasible locker network solution $\boldsymbol{x}$.*

6. Impact of Singapore LA Network

In the Singapore LA Network case, we cluster the country into 3000 delivery points, including 1000 residential locations ($|\mathcal{H}| = 1000$) and 2000 commercial locations ($|\mathcal{W}| = 2000$). Since the access to public residential blocks can be granted by the Authority, the 1000 public residential centers can be used as locker locations. In addition, as suggested by the senior manager of the delivery company, convenience stores such as 7-11 outlets and DBS ATM locations are also feasible locations to install the lockers. We manually collected 980 such locations. Some of these are located in busy shopping malls, or near train stations. In total, the whole locker set contains 1980 locations ($|\mathcal{S}| = 1980$). Therefore, there would be 1980 binary decision variables in our problem. We allow all the customers to choose any of the available lockers, i.e., we let the neighbor set $N_i = N_j = \mathcal{S}$ for each $i \in \mathcal{H}$ and $j \in \mathcal{W}$ in the case study. However, the utility obtained from those lockers far away from them would be negligible, based on Equation (1). The parameter M_i in the egalitarian constraint is defined as the collection of locker stations within 250 meters to block i. In the numerical experiments, we formulate the locker network design models using Java language and solve the optimization problems using Gurobi on a 2.70 GHz i7-6820HQ CPU Windows PC with 16GB RAM.

We first evaluate the value of $\mathrm{Gap}(\boldsymbol{x})$ in the Singapore LA Network case. By varying the budget C from 400 to 1900, we solve problem ($\mathrm{P}^{\mathbf{0}}$) to obtain a set of locker network solutions. Our numerical results show that the performance gaps are consistently less than 0.07 under different budgets in this range. The result holds in our case due to the small attractiveness of locker pickup (i.e., small values of $\rho^i_{\mathcal{H}}(\boldsymbol{x})$ and $\rho^j_{\mathcal{W}}(\boldsymbol{x})$). It implies that the observed delivery profile $\boldsymbol{E}^O$ can be appropriately used for locker network design, without scarifying too much optimality. Therefore, we solve problem ($\mathrm{P}^{\mathbf{0}}$) based on the observed profile $\boldsymbol{E}^O$ in the rest of this paper, unless otherwise stated.

We proceed next to evaluate the impact of locker networks on the volume of parcels delivered to the CBD area. Note that the parcel density in CBD is almost 3 times that of the average amount delivered to other areas (cf. Figure 3), it is natural to look for ways to reduce the volume of parcels delivered to the CBD area. We show that with a well-chosen LA Network, we can hope to reduce the volume of parcels delivered to the CBD by at least 7.5%. However, if the commercial goal dominates and more lockers are installed in the CBD area because of the high volume of parcel deliveries there, the volume of parcels into the CBD area may actually increase in that case! This illustrates the importance of a well crafted locker network strategy in the Singapore LA Network.

We compare different locker network strategies and study the impact on the delivery to CBD:

- No Locker: the case when no locker is installed;
- Without CBD Lockers: lockers are not allowed to be installed at CBD. We rule out these CBD lockers from the potential locker set;
- With CBD Lockers: lockers are allowed to be installed at CBD. When we plot the volume of parcels to CBD, we consider two cases. The first case excludes the volume of delivery to the CBD lockers (i.e., only the deliveries to CBD residential blocks and commercial buildings are counted), whereas the second one includes the volume of delivery to the CBD lockers.

The current volume of parcels delivered to the CBD (i.e. without lockers) is used as the benchmark. We define the *Delivery Change (%)* as:

$$\text{Delivery Change} := \frac{\text{Vol. delivered to CBD (Lockers) } - \text{Vol. delivered to CBD (w/o Lockers)}}{\text{Volume delivered to CBD (w/o Lockers)}} \times 100\%.$$

If the change is negative, then fewer parcels are delivered to the CBD areas, which may go a long way in reducing traffic congestion there. We also compare the total volume of parcels covered by the LA Network, which is defined as:

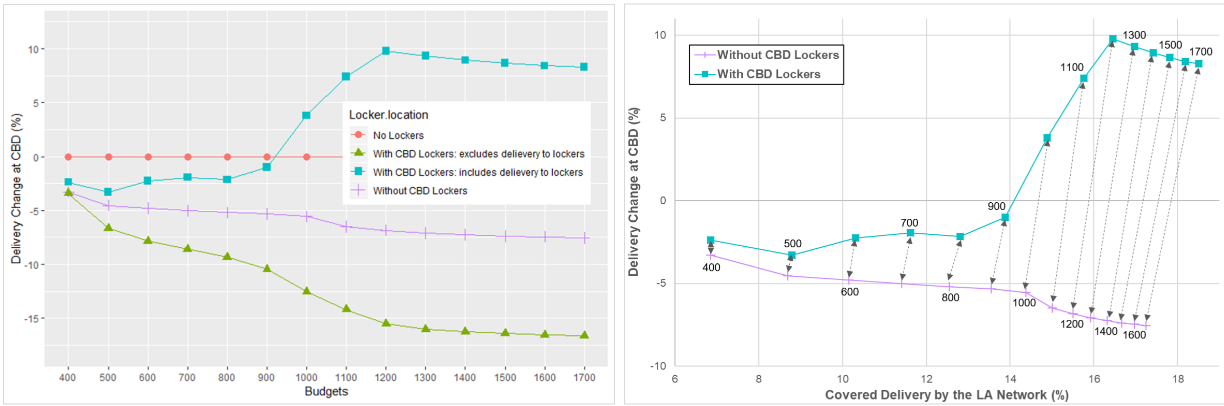
$$\text{Covered Delivery} := \frac{\text{Total Delivery Volume} - V^{\mathbf{0}}(\boldsymbol{x})}{\text{Total Delivery Volume}} \times 100\%.$$

We plot the delivery change and the total volume of parcels covered by different LA Networks in Figure 8(a) and (b), respectively.

The findings are summarized as follows:

- If lockers are not allowed to be installed at CBD, the deliveries to CBD decrease with the expansion of locker network. For example, when the budget is around 1500, at least 7.5% of the deliveries to CBD will be diverted out (cf. the +-dotted curve).
- If lockers are allowed to be installed at CBD, the delivery of parcels to CBD may also be shifted to the lockers inside. As a result, with the expansion of network, the total delivery to CBD (including delivery to lockers and home/office in CBD) actually increases in the optimal network!
- It turns out that the locker network will divert some traffic to lockers in the CBD zone, reducing the parcel traffic to the public or commercial buildings in the CBD. When only the parcel deliveries to residential blocks and commercial buildings in CBD are counted (cf. the triangle-dotted curve), the reduction in the CBD parcel volume is almost 17.5%, with 1500 lockers.

Figure 8 Effects of different LA networks on the deliveries to CBD.



(a) Delivery Change at CBD (b) Covered Delivery by the LA Network

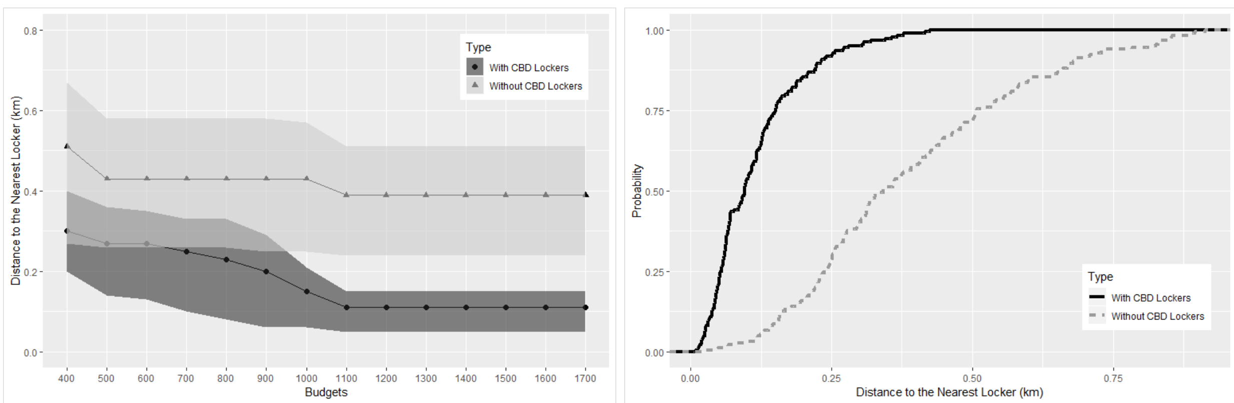
Note. There are 184 lockers located in the CBD area. For ease of comparison, we vary the budget C from 400 to 1700.

- The LA Network without CBD lockers maintains similar utilization (i.e., total volume of covered deliveries) as the one with CBD lockers. As shown in Figure 8(b), the performance gaps in terms of covered deliveries between these two sets of locker solutions are consistently below 1.5%.

To summarize, we numerically show that the delivery to CBD is significantly affected by the LA Network. If lockers are allowed to be installed at CBD, some deliveries, originally to residential blocks and commercial buildings, would partially shift to lockers at CBD, but this in turn increases the total parcel volume into CBD. A natural way to reduce the CBD deliveries is to exclude the CBD lockers from the potential locker set and attract customers to pick up their parcels at these lockers near their homes outside the CBD. In this way, our numerical experiment shows that 7.5% of deliveries to CBD can be reduced if the budget is around 1500, without scarifying too much utilization of the LA Network. We note that the locker network solution $\boldsymbol{x}(\boldsymbol{E}^{\mathbf{0}})$ above is obtained from model ($\mathrm{P}^{\mathbf{0}}$), and the *Delivery Change* is estimated based on the observed delivery profile $\boldsymbol{E}^{\mathbf{0}}$. Indeed, this is different from the actual *Delivery Change* under the true demand $\boldsymbol{D}$. We show that the estimated CBD parcel volume reduction with $\boldsymbol{E}^{\mathbf{0}}$ is a lower bound on the actual one with $\boldsymbol{D}$ in the Singapore LA Network case (under mild conditions). In other words, the actual parcel volume to CBD could be reduced for more than 7.5% given a well-chosen LA Network with budget 1500. Therefore, we can extrapolate the reduced parcel volume to CBD based on solution $\boldsymbol{x}(\boldsymbol{E}^{\mathbf{0}})$. More detailed discussions on this issue are provided in Appendix D.2.

In the end, we highlight that if the lockers are not allowed to be installed at CBD, customers from CBD in turn may need to travel a longer distance for parcel pickup from lockers; otherwise they will continue to use delivery-to-office at the CBD area. As a result, the customer welfare, measured by the parcel pickup distance from the nearest locker, will decrease in the LA Network without CBD lockers. To see this, we calculate the average parcel pickup distance (from the nearest locker) across different customer segments in the CBD area. As shown in Figure 9(a), the average parcel pickup distance decreases with the number of lockers in both LA Networks with/without CBD lockers. This is intuitive since customers have more locker options in a larger scale network. However, these customers have slightly longer pickup distance in the network without CBD locker. Taking the case of 1500 lockers for illustration (cf. Figure 9(b)), the parcel pickup distance from the nearest locker is longer than 250 meters for more than 75% of CBD customers if lockers are not allowed to be installed at CBD. In such a case, the utility of using locker would be very low, and delivery-to-office is still the better option.

Figure 9 Parcel pickup distance from the nearest locker for CBD customers.



(a) Average Parcel Pickup Distance at CBD (b) Density Plot for the Case of 1500 Lockers

Note. The shaded region in the Figure (a) indicates the distance interval from 25 percentile to 75 percentile.

7. Concluding Remarks

Last mile innovation has piqued a surge of interests in the smart nation initiatives. To improve the efficiency of last mile operations, the automated parcel lockers are already widely used in the US, Europe, and China etc. However, relatively few studies have been done to address the optimal locker network design challenge. Inspired by the proposal of Singapore LA Network, we study the locker network design problem to maximize the utilization as well as to serve the public residents.

We develop a robust framework to solve the class of facility location problems in the absence of the transit route information of all customers. Interestingly, we show that the volume of parcels delivered to the LA Network, obtained under the observed delivery profile, provides a lower bound to the case when we know the actual demand information. Furthermore, we explicitly characterize the optimality loss if the LA network is built based on the observed delivery profile. We also numerically show that we can hope to reduce the volume of parcels delivered to the CBD by at least 7.5% with a well-chosen LA Network. In fact, if the utility of using lockers increases 3 times in the Singapore LA

Network case, more than 19% deliveries to CBD would be attracted to lockers given a well-chosen LA Network with budget 1500. Therefore, the planning agencies need to incentivize customers to use lockers, for example, by compensating the delivery fees.

Last but not least, we highlight that some lockers suffer from low utilization issue even in the case when we attempt to maximize the total utilization. In the LA network case, we observe that some lockers can cover around 20 parcels per day, whereas some only witness 2 parcels. The median utilization is between 6 and 8 parcels. Furthermore, with the increase of network scale, the median utilization decreases gradually because of the competition effect in the locker network. To boost the utilization of LA Network, commercial operators can use the LA Network as a storage option to avoid another visit, to serve customers whose parcels were not delivered due to various reasons, for example, those estates that are far away from the distribution hub are known to have higher incidences of failed deliveries. This is a particularly appealing option for the LA Network, since there is already a station within the vicinity of every public residential block. The utilization of lockers will be boosted if failed deliveries to these blocks can be channeled to the lockers.

In this paper, we have focused solely on network design from the consumer end. With the LA Network, we expect to see an increasing trend in the volume of e-commerce retail transactions. It will be interesting to examine the impact of the open locker system on the entire e-commerce value chain, and how to incentivize e-commerce vendors to consider the usage of lockers in their delivery operations. At the national scale, it will be interesting to understand how this solution concept can help to streamline traffic flows into congested zones in the city, and the associated impact on the environment. We leave these and other issues to our future research.

Acknowledgments

The authors sincerely thank Prof. Christopher TANG, the associate editor, and the anonymous reviewers for their constructive suggestions. The authors also would like to thank the colleagues from *IMDA* and *Yamato Asia Pte. Ltd.* Their valuable feedbacks have helped the authors to clarify many practical issues pertaining to the Singapore Locker Alliance Network. This study was supported in part by the Ministry of Education, Singapore, under its 2019 Academic Research Fund Tier 3 [Grant MOE-2019-T3-1-010]; and the Ministry of Education, Singapore, under it Social Science Research Thematic Grant [Grant MOE2016-SSRTG-059].

References

Alizadeh, F., Goldfarb, D. 2003. Second-order cone programming. *Mathematical programming*, **95(1)**, 3–51.

Aros-Vera, F., Marianov, V., Mitchell, J. E. 2013. p-Hub approach for the optimal park-and-ride facility location problem. *European Journal of Operational Research*, **226(2)**, 277–285.

Baron, O., Milner, J., Naseraldin, H. 2011. Facility location: A robust optimization approach. *Production and Operations Management*, **20(5)**, 772–785.

Berman, O., Krass, D. 2002. The generalized maximal covering location problem. *Computers & Operations Research*, **29(6)**, 563–581.

Berman, O., Krass, D., Drezner, Z. 2003. The gradual covering decay location problem on a network. *European Journal of Operational Research*, **151(3)**, 474–480.

Chan, T. C., Shen, Z. J. M., Siddiq, A. 2017. Robust Defibrillator Deployment Under Cardiac Arrest Location Uncertainty via Row-and-Column Generation. *Operations Research*, **66(2)**, 358–379.

Collins, A. 2015. Travel behaviour in the context of parcel pickups. *Working paper.*

Dasci, A., Laporte, G. 2005. A continuous model for multistore competitive location. *Operations Research*, **53(2)**, 263–280.

Derakhshan, M., Golrezaei, N., Manshadi, V., Mirrokni, V. 2018. Product ranking on online platforms. *Working Paper.*

Edwards, J., McKinnon, A., Cherrett, T., McLeod, F., Song, L. 2009. The impact of failed home deliveries on carbon emissions: Are collection/delivery points environmentally-friendly alternatives? *Proceedings of the Logistics Research Network.*

Edwards, J., McKinnon, A., Cherrett, T., McLeod, F., Song, L. 2010. Carbon dioxide benefits of using collection-delivery points for failed home deliveries in the United Kingdom. *Transportation Research Record: Journal of the Transportation Research Board*, **(2191)**, 136–143.

Faugere, L., Montreuil, B. 2017. Hyperconnected pickup & delivery locker networks. In *Proceedings of the 4th International Physical Internet Conference.*

Fee, K. A. 2015. U.S. Patent Application No. 14/063,080.

Glaeser, C., Fisher, M., Su, X. 2019. Optimal Retail Location: Empirical Methodology and Application to Practice. *Manufacturing & Service Operations Management*, **21(1)**, 86–102.

Goyal, V., Levi, R., Segev, D. 2016. Near-optimal algorithms for the assortment planning problem under dynamic substitution and stochastic demand. *Operations Research*, **64(1)**, 219–235.

Irwin, D. E., Yessin, G. M., Shenhar, J., Dewnandan, J. L., Spears, M. L., Tartal, W. A. 2015. U.S. Patent Application No. 14/743,260.

Iwan, S., Kijewska, K., Lemke, J. 2016. Analysis of parcel lockers' efficiency as the last mile delivery solution– the results of the research in Poland. *Transportation Research Procedia*, **12**, 644–655.

Kung, L. C., Liao, W. H. 2017. An approximation algorithm for a competitive facility location problem with network effects. *European Journal of Operational Research*, **267(1)**, 176–186.

Lin, Y. H., Wang, Y., He, D., Lee, L. H. 2020. Last-mile delivery: Optimal locker location under multinomial logit choice model. *Transportation Research Part E: Logistics and Transportation Review*, **142**, 102059.

Lin, Y. H., Wang, Y., Lee, L. H. 2020. Parcel Locker Location Problem under Threshold Luce Model. *arXiv preprint arXiv:2002.10810.*

Méndez-Díaz, I., Miranda-Bront, J. J., Vulcano, G., Zabala, P. 2014. A branch-and-cut algorithm for the latent-class logit assortment problem. *Discrete Applied Mathematics*, **164**, 246–263.

Morganti, E., Dablanc, L., Fortin, F. 2014. Final deliveries for online shopping: The deployment of pickup point networks in urban and suburban areas. *Research in Transportation Business & Management*, **11**, 23–31.

Morganti, E., Seidel, S., Blanquart, C., Dablanc, L., Lenz, B. 2014. The impact of e-commerce on final deliveries: Alternative parcel delivery services in France and Germany. *Transportation Research Procedia*, **4**, 178–190.

Ranieri, L., Digiesi, S., Silvestri, B., Roccotelli, M. 2018. A Review of Last Mile Logistics Innovations in an Externalities Cost Reduction Vision. *Sustainability*, **10(3)**, 782.

Savelsbergh, M., Van Woensel, T. 2016. 50th anniversary invited article—city logistics: Challenges and opportunities. *Transportation Science*, **50(2)**, 579–590.

Şen A., Atamtürk A., Kaminsky P. 2018. A conic integer optimization approach to the constrained assortment problem under the mixed multinomial logit model. *Operations Research*, **66(4)**, 994–1003.

Song, L., Cherrett, T., McLeod, F., Guan, W. 2009. Addressing the last mile problem: transport impacts of collection and delivery points. *Transportation Research Record: Journal of the Transportation Research Board*, **(2097)**, 9–18.

Talluri, K., Van Ryzin, G. 2004. Revenue management under a general discrete choice model of consumer behavior. *Management Science*, **50(1)**, 15–33.

Torres, L., Suggs, B. 2015. U.S. Patent Application No. 14/582,088.

Train, K. E. 2009. Discrete choice methods with simulation. *Cambridge university press.*

Wang, Y., Ong, T., Lee, L. H., Chew, E. P. 2017. Capacitated competitive facility location problem of self-collection lockers by using public big data. In *Automation Science and Engineering (CASE), 2017 13th IEEE Conference on*, 1344–1344.

Wang, S., Chen, Z., Liu, T. 2020. Distributionally robust hub location. *Transportation Science*, **54(5)**, 1189–1210.

Weltevreden, J. W. 2008. B2c e-commerce logistics: the rise of collection-and-delivery points in The Netherlands. *International Journal of Retail & Distribution Management*, **36(8)**, 638–660.

Wu, H., Shao, D., Ng, W. S. 2015. Locating self-collection points for last-mile logistics using public transport data. *In Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer International Publishing.* 498–510.

Zhang, Y., Berman, O., Verter, V. 2012. The impact of client choice on preventive healthcare facility network design. *OR spectrum*, **34(2)**, 349–370.

Appendix

Organization of the Appendix. In Appendix A, we present a compact reformulation of the locker network design model to solve large-scale problem. In Appendix B, we use a set of Singapore public transit data to estimate the volume of traffic on each home-office pair. This allows us to estimate the hitherto unknown transit route for each demand segment and solve the locker network model with the transit route information incorporated. For completeness, we compare the market share maximization model (studied in this paper) and the customer welfare maximization model (a plausible alternative) in Appendix C. More details on the Singapore LA Network case are provided in Appendix D. In Appendix E, we provide the technical proofs.

A. Computational Approach for Large-Scale Problems

To facilitate the analysis, we first simplify the "heavy" notations in model (P). Let $\mathcal{B}$ denote the whole demand segment set ($\mathcal{H} \cup \mathcal{W} \cup \mathcal{H} \times \mathcal{W}$) and $\mathcal{S}$ the set of possible locker locations. We normalize the utility of outside option to be 1 and denote $\mu_{i,k}$ as the utility of using locker k for customer from segment i. We allow the locker consideration set N_i for block i to be the entire set $\mathcal{S}$, and remove the egalitarian constraints. Next, we replace the volume of parcel delivery by $\hat{d}_i$. Therefore, we can re-formulate model (P) as the following *0-1* fractional programming problem:

$$
\begin{aligned}
\text{(FB)} \qquad \min \; & \sum_{i \in \mathcal{B}} \hat{d}_i \left(\frac{1}{1 + \sum_{k \in \mathcal{S}} \mu_{i,k} x_k} \right) \\
\text{s.t.} \; & \sum_{k \in \mathcal{S}} x_k \le C \\
& x_k \in \{0,1\}, \; \forall k \in \mathcal{S}
\end{aligned}
$$

It is straightforward to see that this fractional minimization problem is equivalent to the cardinality constrained assortment problem with multiple customer segments problem, i.e., the classic mixed MNL (MMNL) assortment problem:

$$
\text{(MMNL)} \qquad \max_{\boldsymbol{x} \in \{0,1\}^{|\mathcal{S}|}} \left\{ \sum_{i \in \mathcal{B}} \left[\hat{d}_i \left(\frac{\sum_{k \in \mathcal{S}} \mu_{i,k} x_k}{1 + \sum_{k \in \mathcal{S}} \mu_{i,k} x_k} \right) \right] \middle| \sum_{k \in \mathcal{S}} x_k \le C \right\}.
$$

Therefore, we can relate our locker network design problem to the inventory planning problem considered in Goyal et al. (2016), and show that this class of locker network design problem is NP-hard. We use a reduction from vertex cover problem to show the hardness of our problem.

THEOREM 4. *The locker network design problem (P) is NP-hard.*

Furthermore, we note that the objective in model (FB) is non-linear in the decision variable $\boldsymbol{x}$ due to the convex disutility function. A standard way to solve this convex programming problem is to linearize the fractional objective so that it can be transformed to a mixed integer programming problem (MIP). However, this reformulation technique does not scale well in our LA Network. In this paper, we develop an exact second-order cone programming & mixed integer programming (SOCP-MIP) reformulation for this problem and we show that this technique can reduce drastically the size of decision variables and constraints. Furthermore, this SOCP-MIP reformulation allows us to use available solver to construct a locker network solution in a reasonable computational time.

A.1. MIP Formulation

We first consider a common MIP formulation to Model (FB). Zhang et al. (2012) introduced this linearization technique to solve a healthcare facility network design problem by incorporating the client choice. Méndez-Díaz et al. (2014) also did this MIP reformulation for the assortment problem. We briefly summarize the technique as follows.

Introduce two sets of auxiliary decision variables

$$w_i := \frac{1}{1+\sum_{k\in\mathcal{S}}\mu_{i,k}x_k}, \text{ and } \quad t_{i,k} := w_i x_k,$$

then the fractional term can be replaced by

$$\begin{aligned}
& w_i + \sum_{k\in\mathcal{S}} \mu_{i,k} t_{i,k} = 1, \\
& t_{i,k} \le x_k, \quad t_{i,k} \le w_i, \\
& t_{i,k} \ge (x_k - 1) + w_i, \quad t_{i,k} \ge 0.
\end{aligned}$$

Notice that $w_i \in [0,1]$. Therefore, we can derive the following MIP reformulation:

$$\begin{aligned}
\text{(MIP)} \quad & \min \sum_{i\in\mathcal{B}} \hat{d}_i w_i \\
& \text{s.t. } w_i + \sum_{k\in\mathcal{S}} \mu_{i,k} t_{i,k} = 1, \ \forall\, i \in \mathcal{B} \\
& \quad t_{i,k} \le x_k, \ \forall\, i\in\mathcal{B}, \ k\in\mathcal{S} \\
& \quad t_{i,k} \le w_i, \ \forall\, i\in\mathcal{B}, \ k\in\mathcal{S} \\
& \quad t_{i,k} \ge (x_k - 1) + w_i, \ \forall\, i\in\mathcal{B}, \ k\in\mathcal{S} \\
& \quad \sum_{k\in\mathcal{S}} x_k \le C \\
& \quad t_{i,k} \ge 0, \ \forall\, i\in\mathcal{B}, \ k\in\mathcal{S} \\
& \quad 0 \le w_i \le 1, \ \forall\, i\in\mathcal{B} \\
& \quad x_k \in \{0,1\}, \ \forall\, k\in\mathcal{S}
\end{aligned}$$

In this model, there are $|\mathcal{S}|$ binary decision variables x_k, and $|\mathcal{B}|(1+|\mathcal{S}|)$ continuous variables, w_i and $t_{i,k}$. The number of constraints is $|\mathcal{B}| + 3|\mathcal{S}||\mathcal{B}| + 1$ in total. Nevertheless, if the scale of delivery location and locker set are large, e.g., in thousands level, then the product term $|\mathcal{B}||\mathcal{S}|$ would be incredibly gigantic. As a result, this MIP approach does not scale well. We show next a SOCP-MIP reformulation for this problem can reduce drastically the size of decision variables and constraints.

A.2. SOCP-MIP Formulation

Before we proceed to the SOCP-MIP reformation, we briefly describe the second-order cone programming (SOCP) problem. For more details about this type of optimization model, we refer readers to Alizadeh and Goldfarb (2003).

Definition 1 (Alizadeh and Goldfarb 2003) The standard formulation of SOCP problem can be represented:

$$\begin{aligned} &\min \ \boldsymbol{c}^T\boldsymbol{x} \\ &\text{s.t.} \ \ A\boldsymbol{x} = \boldsymbol{b} \\ &\qquad \boldsymbol{x} \succcurlyeq_Q \boldsymbol{0} \end{aligned}$$

where $\boldsymbol{c} \in \mathbb{R}^n$, $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{b} \in \mathbb{R}^m$, and $A \in \mathbb{R}^{m\times n}$. We let $\boldsymbol{x} = (x_0, \bar{\boldsymbol{x}}) \in \mathbb{R}^n$, and the second-order cone is defined as

$$Q := \{\boldsymbol{x} = (x_0, \bar{\boldsymbol{x}}) \in \mathbb{R}^n : x_0 \geq ||\bar{\boldsymbol{x}}||\},$$

where $||\cdot||$ refers to the Euclidean norm. In addition, the inequality $\boldsymbol{x} \succcurlyeq_Q \boldsymbol{0}$ represents the second-order cone inequality. The SOCP problem can be solved efficiently. Next, we show that a second order cone and mixed integer programming representable formulation can reduce drastically the size of the locker network design problem. Similar SOCP reformulation techniques were also used to solve other applications (e.g., Alizadeh and Goldfarb 2003, Şen et al. 2018). More concretely, we introduce two more auxiliary decision variables:

$$z_i := 1 + \sum_{k\in\mathcal{S}} \mu_{i,k} x_k, \text{ and } \ w_i := \frac{1}{z_i}.$$

We can re-formulate the model (FB) as

$$\begin{aligned} \text{(SOCP-MIP)} \quad &\min \sum_{i\in\mathcal{B}} \hat{d}_i w_i \\ &\text{s.t.} \ \ 1 + \sum_{k\in\mathcal{S}} \mu_{i,k} x_k \geq z_i, \ \forall \ i \in \mathcal{B} \\ &\qquad \begin{pmatrix} w_i & 1 \\ 1 & z_i \end{pmatrix} \succeq 0, \ \forall \ i \in \mathcal{B} \\ &\qquad \sum_{k\in\mathcal{S}} x_k \leq C \\ &\qquad z_i \geq 0, \quad 0 \leq w_i \leq 1, \ \forall \ i \in \mathcal{B} \\ &\qquad x_k \in \{0,1\}, \ \forall \ k \in \mathcal{S} \end{aligned}$$

where the second set of constraints, as expressed in rotated cone form, implies that $w_i z_i \geq 1, \forall i \in \mathcal{B}$ (Şen et al. 2018). In this SOCP-MIP model, there are $|\mathcal{S}|$ binary decision variables x_k, and $2|\mathcal{B}|$ continuous variables, z_i and w_i. The number of constraints is $2|\mathcal{B}| + 1$. Compared with the MIP formulation, this SOCP-MIP formulation involves much less decision variables and constraints!

We show the equivalence of model (FB) and (SOCP-MIP) in the following Proposition.

Proposition 3. *The SOCP-MIP formulation is equivalent to the FB formulation.*

Next, we compare the computational performances of the SOCP-MIP and the MIP reformulations. We formulate the problems using Java language and Gurobi solver on a 2.70 GHz i7-6820HQ CPU Windows PC with 16GB RAM. Recall that in the MIP formulation, there are $(|\mathcal{B}| + 3|\mathcal{S}||\mathcal{B}| + 1)$ constraints while there are only $(2|\mathcal{B}| + 1)$ constraints in the SOCP-MIP formulation. This comparison suggests that commercially available software (e.g., Gurobi, Cplex) can be applied to solve a relatively large size problem with the SOCP-MIP formulation. In the LA Network case, we compare the computation time of two formulations on problem ($\mathrm{P}^{\mathbf{0}}$), with 3000 demand segments and 1980 potential locker locations. The budget C is varied from 400 to 1900 with step size 100, and we calculate the average computational time. For the SOCP-MIP formulation, it takes around $2,543$ seconds to solve for each budget case. However, for the MIP formulation, we cannot solve the problem with such size using the same Windows PC. Therefore, we are limited to compare the performances of two formulations for smaller size problems. We re-cluster the demand segments and potential locker locations to generate the simulation environment. As shown in Table 3, we let the number of demand segment $|\mathcal{B}| = \{4, 8, 16, 32, 64, 128, 256\}$, and the number of potential candidates $|\mathcal{S}| = \{2, 4, 8, 16, 32, 64, 128\}$. The budget C is set to be $|\mathcal{S}|/2$ for all cases. It turns out that the MIP approach is as efficient as the SOCP-MIP approach when the number of demand segments is less than 100. For a larger size problem (e.g., $|\mathcal{B}| = 256$), the SOCP-MIP approach outperforms the MIP approach significantly.

Table 3 Computation time comparison for problem ($\mathbf{P}^{\mathbf{0}}$).

CPU Time (s) \ Segments	4	8	16	32	64	128	256
SOCP-MIP	0.010	0.011	0.024	0.067	0.151	0 .348	7.274
MIP	0.007	0.012	0.033	0.200	2.470	3.828	66.882

We also consider a synthetic example of problem (P). More concretely, we let the number of home/office locations $|\mathcal{H}| = |\mathcal{W}| = N$, where $N = \{4, 8, 16, 32, 64, 128, 256\}$. For each pair of demand segment (i, j), $\forall i \in \mathcal{H}$, $j \in \mathcal{W}$, we uniformly generate the volume of $D_{i,j}$ between 0 and 1. In each case, there are N^2 demand segments. The number of potential candidates is set to be $|\mathcal{S}| = \{2, 4, 8, 16, 32, 64, 128\}$, and the budget C is $|\mathcal{S}|/2$ in each case. As shown in Table 4, the traditional MIP approach requires much longer time to solve the large-scale cases, while the SOCP-MIP approach scales well. In particular, the case of $N = 256$ cannot be solved by the MIP approach due to the "out of memory" problem returned by the Gurobi solver.

Table 4 Computation time comparison for problem (P).

CPU Time (s) \ Segments	4^2	8^2	16^2	32^2	64^2	128^2	256^2
SOCP-MIP	0.019	0.032	0.227	1.221	8.291	64.254	11563.637
MIP	0.026	0.031	0.319	3.202	221.908	8449.005	N.A.

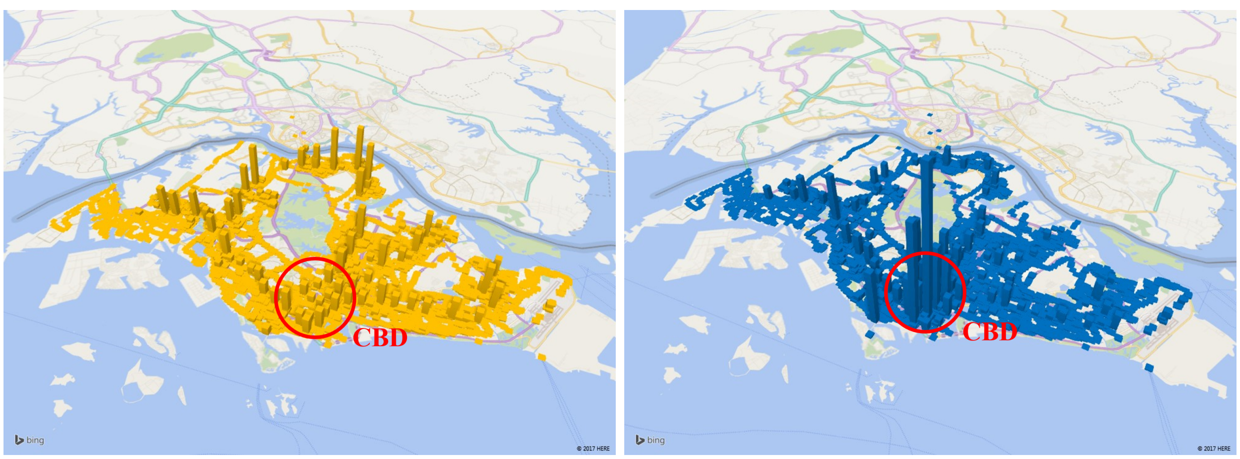
B. Estimation of the Demand on Each Home-Office Pair

In this section, we formally address the estimation challenge of each home-office demand $\{D_{i,j}\}$ so that the model (P) can be solved. To facilitate the estimation, we first describe a set of Singapore public transport data, and then show that the public transit patterns are positively correlated to the parcel delivery volume. This observation leads us to estimate the demand of $D_{i,j}$ between home location $i \in \mathcal{H}$ and office location $j \in \mathcal{W}$, using the public transit flow between location i and j.

B.1. Singapore Public Transport System

Singapore has constructed cost-efficient and convenient public transport systems in the world.[14] According to Wikipedia,[15] Singapore public transport systems include both bus rides and mass rapid transit (MRT) system, and more than half of Singapore residents go to work by taking public transport. We were provided with a set of public transport data provided by the Singapore Land Transport Authority (LTA). The dataset contains one weekday's worth of public transit information, with around 4 million (daily) transactional records, and 4530 stations/stops over the island. In addition, the average traveling time for each trip is about 19.74 minutes. There are around 2 million unique passenger ID in the dataset, and 72.50% of them generated at least 2 records per day. This implies that a large proportion of passengers took round trips by the public transportation system. Note that it would be crowed to visualize the passenger transit patterns across 4530 stations, we provide the zonal transit patterns instead according to the Singapore postal district map.[16] As shown in Figure 10, we observe that many residents board from non-CBD regions before noon, and these residents alight at CBD regions before noon.

Figure 10 Singapore public transit patterns.



(a) Boarding Records before 12:00 pm (b) Alighting Records before 12:00 pm

Note. (a) The height of bar represents the volume of passenger boarding records at a specific station/stop. (b) The height of bar represents the volume of passenger alighting records at a specific station/stop.

We match the bus/MRT station (stop) to the nearest residential block i or commercial block j, and display the volume of parcel delivery (i.e., D_i^O and D_j^O) at each location. Next, we aggregate the transportation records and delivery volume at the district level. Figure 11(a) compares the total transportation records with the delivery volume in each district (28 districts in total). It is straightforward to observe that the

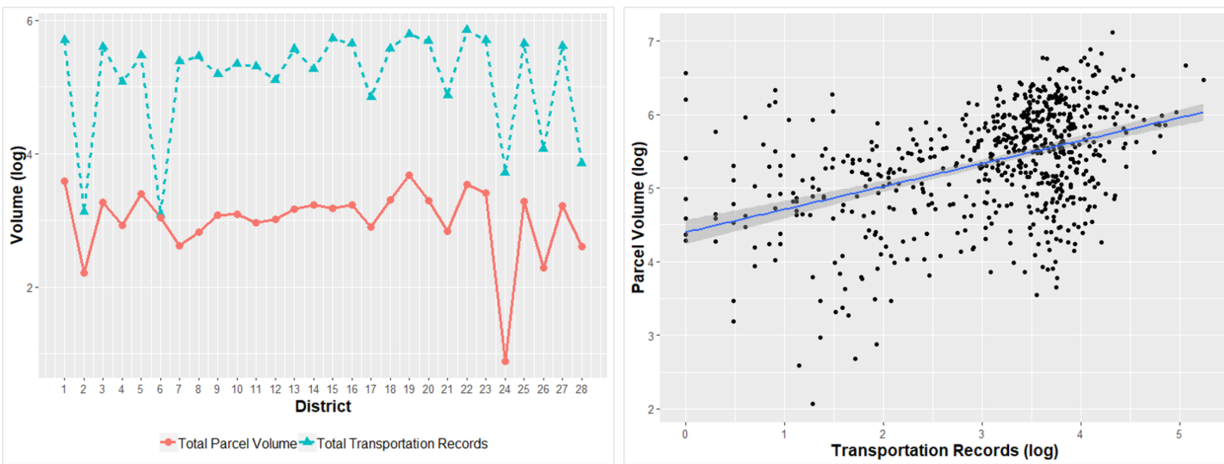
[14] Singapore's public transport system one of world's most efficient; retrieved from `http://www.straitstimes.com/singapore/transport/study-singapores-public-transport-system-one-of-worlds-most-efficient`

[15] Transport in Singapore; retrieved from `https://en.wikipedia.org/wiki/Transport_in_Singapore\#cite_note-Facts-2`

[16] Singapore is divided into 28 typical zones (districts). The district information is retrieved from `https://en.wikipedia.org/wiki/Postal_codes_in_Singapore`

transportation records and parcel volume are highly correlated. Furthermore, as the target is to estimate the transit flow for the population who lives in location i and work at location j, we pick up these transport records boarding from district i and alighting at district j as the transit flow $t_{i,j}$ between these two districts. Note that the majority of residents go to work in the morning and go back home in the afternoon, we select the records before 12:00 p.m. to estimate the transit flow. Since we cannot observe the volume of demand $\{D_{i,j}\}$ directly, we use the product term $D_i^O \times D_j^O$ to indicate that this term is related to the parcel volumes to both district i and j. Figure 11(b) visualizes the correlation between $D_i^O \times D_j^O$ and $t_{i,j}$, with a positive coefficient 0.43. It also shows clearly that heavier transit flow is more likely to be accompanied with larger parcel volume. Therefore, we can estimate the demand $D_{i,j}$ between home location $i \in \mathcal{H}$ and office location $j \in \mathcal{W}$ based on the volume of public transit flow between location i and j.

Figure 11 Correlation between public transit records and parcel deliveries.



(a) Aggregated Volume at Each District

(b) Aggregated Flow between Districts

B.2. Estimation of Parcel Volume

For ease of exposition, we apply K-Centroids approach to cluster 280 public residential blocks (centers) and 280 private/commercial districts (centers). We match the station/stop to the nearest public/commercial center so that we can estimate the public transit records between each pair of residential center and commercial center. We acknowledge that it is challenging to estimate the delivery volume on each home-office pair because the home-office pair information cannot be obtained accurately. To overcome this information gap, we assume that the volume of demand $D_{i,j}$ is linearly correlated with the transit flow $t_{i,j}$ since the parcel volume is positively correlated with the public transit volume (as stated in Figure 11). Let $|D|$ denote the total parcel volume and $|P|$ denote the total population in Singapore. We let

$$D_{i,j} = \frac{|D|}{|P|} t_{i,j} + \tau_{i,j},$$

where $\tau_{i,j}$'s represent the "noise" terms and need to be determined such that the demand $D_{i,j}$'s are feasible to the delivery profile. In fact, according to the expression of $\boldsymbol{D}$, the delivery volume $D_{i,j}$'s have to satisfy some side constraints. For example, if $D_i^O = 0$, then $D_{i,j} = 0$ for $\forall j$. However, $t_{i,j}$ may not be 0. Therefore, we introduce the auxiliary term $\tau_{i,j}$ to make the estimation of delivery profile $\boldsymbol{D}$ feasible. Note that we are trying to provide a set of delivery flow that is close to the estimator $\frac{|D|}{|P|}t_{i,j}$, and hence we formulate the following model to determine $\{\tau_{i,j}\}$:

$$\min_{\boldsymbol{\tau}} \sum_{i\in\mathcal{H},\ j\in\mathcal{W}} (\tau_{i,j})^2$$

$$\text{s.t. } D_i^O \geq \frac{\theta_{\mathcal{H}}}{\theta_{\mathcal{H}}+\theta_{\mathcal{W}}} \sum_{j\in\mathcal{W}} \left\{\frac{|D|}{|P|}t_{i,j}+\tau_{i,j}\right\}, \forall i\in\mathcal{H}$$

$$D_j^O \geq \frac{\theta_{\mathcal{W}}}{\theta_{\mathcal{H}}+\theta_{\mathcal{W}}} \sum_{i\in\mathcal{H}} \left\{\frac{|D|}{|P|}t_{i,j}+\tau_{i,j}\right\}, \forall j\in\mathcal{W}$$

The first and second set of constraints force the estimated delivery volume $\{D_{i,j}\}$ to satisfy the delivery profile $\boldsymbol{D}$. The result shows that the estimated volume of $(\sum_{i\in\mathcal{H},j\in\mathcal{W}} D_{i,j})$ accounts for 54.43% of the total observed delivery volume $(\sum_{i\in\mathcal{H}} D_i^O + \sum_{j\in\mathcal{W}} D_j^O)$ in the Singapore LA case. We remark that we find a feasible lower bound for the 'actual' demand $D_{i,j}$ since we rule out the private transit records from our analysis. A more accurate estimation may require a nationwide survey to elicit this information. Given the information of $\{D_{i,j}\}$, we can solve the model (P) based on the SOCP-MIP reformulation, as shown in Section A.2.

B.3. Numerical Comparison between Model ($\mathbf{P^0}$) and (P)

For ease of exposition, we consider a relatively small-scale network to numerically compare model ($\mathrm{P^0}$) and (P) in the Singapore LA case. We cluster the delivery locations into 280 residential block centers and 280 commercial centers, i.e, $|\mathcal{H}| = |\mathcal{W}| = 280$. We choose a subset of potential locker locations with $|\mathcal{S}| = 409$. This subset includes the residential block centers and some 7-11 stores. We also allow all the customers to choose any of the available lockers, i.e., we let the neighbor set $N_i = N_j = \mathcal{S}$ for each $i \in \mathcal{H}$ and $j \in \mathcal{W}$ in the choice model. Furthermore, the parameter M_i in the egalitarian constraint is revised to represent the collection of locker stations within 1 kilometer to block i. In this case, there are $(280 + 280 + 280 \times 280 = 78960)$ demand segments, and (409) binary variables. The delivery profile $\boldsymbol{D}$ is estimated based on the transit data.

Table 5 Comparison between model ($\mathbf{P^0}$) and (P).

Budget	Market Share		RB Coverage		CPU Time (s)	
	$\mathrm{P^0}$	P	$\mathrm{P^0}$	P	$\mathrm{P^0}$	P
50	410.79	410.79	280	280	2.20	1014.52
100	799.29	799.32	280	280	2.15	1292.06
150	1068.09	1068.09	280	280	1.63	1388.73
200	1293.74	1294.22	280	280	1.54	1167.88
250	1495.96	1496.08	280	280	1.62	845.60
300	1674.06	1674.54	280	280	1.72	898.15
350	1828.01	1828.73	280	280	1.98	668.76
400	1934.56	1934.56	280	280	2.39	638.68

Note. The *Market Share* captures the total volume of parcels delivered to the Lockers based on the actual demand profile $\boldsymbol{D}$; The *RB Coverage* counts the number of residential block (RB) centers that can be served by at least one locker within 1 kilometer; The *CPU Time* reports the computational time in second.

We vary the budget C from 50 to 400, and numerically compare the difference between model ($\mathrm{P^0}$) and (P). As shown in Table 5, the difference in market share obtained from the two models appears to be negligible.

This implies that planning based on the observed delivery profile by solving model ($\mathrm{P}^{\mathbf{0}}$) can be used to estimate the amount of demand lost to home and office deliveries under the unknown demand $\boldsymbol{D}$ in the Singapore LA Network. In fact, this result holds in our case due to the small utility obtained from using lockers. We remark that this numerical result does mean that the issues associated with unknown demand are not important in the facility location problems. In the case with large utility towards lockers, the optimal locker network solution would be affected significantly if we do not use the actual demand information for planning (see Example 1 in Section 1 for illustration). In addition, since the feasible regions of model ($\mathrm{P}^{\mathbf{0}}$) and (P) are identical, all the residential block (RB) centers are covered by the locker networks (i.e., there is at least one locker installed within 1 kilometer of each residential block) under both models. Notably, compared with model (P), model ($\mathrm{P}^{\mathbf{0}}$) reveals great computational advantages under different budgets in this range.

C. Comparison of Different Locker Network Design Models

There are two foremost objectives – maximizing the market share covered by the locker network and maximizing the customer welfare obtained from using lockers – in the locker network design problem. In this section, we numerically compare the performance of (1) market share maximization model (Q^{Ms}), and (2) customer welfare maximization model ($\mathrm{Q}^{\mathrm{Wel}}$).

Under our MNL choice model, the probability that a consumer from demand segment i shifts to locker network $\boldsymbol{x}$ for parcel delivery is given by $\frac{\sum_{k\in\mathcal{S}}\theta_{i,k}x_k}{\theta_{i,0}+\sum_{k\in\mathcal{S}}\theta_{i,k}x_k}$ (Talluri and Van Ryzin 2004), where the utility of outside option $\theta_{i,0}=\theta_{\mathcal{H}}$ if $i\in\mathcal{H}$ and $\theta_{i,0}=\theta_{\mathcal{W}}$ if $i\in\mathcal{W}$. Given the demand volume D_i^O at segment i, the market share obtained from demand segment i can be represented by $D_i^O\left(\frac{\sum_{k\in\mathcal{S}}\theta_{i,k}x_k}{\theta_{i,0}+\sum_{k\in\mathcal{S}}\theta_{i,k}x_k}\right)$. In this way, the market share maximization model (Q^{Ms}) can be formulated as follows:

$$
\begin{aligned}
(\mathrm{Q}^{\mathrm{Ms}}) \qquad & \max_{\boldsymbol{x}} \sum_{i\in\mathcal{H}} D_i^O\left(\frac{\sum_{k\in\mathcal{S}}\theta_{i,k}x_k}{\theta_{\mathcal{H}}+\sum_{k\in\mathcal{S}}\theta_{i,k}x_k}\right)+\sum_{i\in\mathcal{W}} D_i^O\left(\frac{\sum_{k\in\mathcal{S}}\theta_{i,k}x_k}{\theta_{\mathcal{W}}+\sum_{k\in\mathcal{S}}\theta_{i,k}x_k}\right)\\
& \text{s.t. } \sum_{k\in\mathcal{S}} x_k \le C\\
& \qquad x_k\in\{0,1\},\ \forall k\in\mathcal{S}
\end{aligned}
$$

where the first constraint limits the number of lockers installed to a budget C, and the second set of constraints indicates the decision variable to be binary. For ease of exposition, here we remove the egalitarian constraints. It is directly to implement the SOCP-MIP formulation introduced in Appendix A.2 to solve this market share maximization problem.

In the customer welfare maximization problem, Train (2009) and Derakhshan et al. (2018) demonstrated that the welfare obtained for a customer from demand segment i is given by $\log\left(\theta_{i,0}+\sum_{k\in\mathcal{S}}\theta_{i,k}x_k\right)$ under the locker network solution $\boldsymbol{x}$, where the utility of outside option $\theta_{i,0}=\theta_{\mathcal{H}}$ if $i\in\mathcal{H}$ and $\theta_{i,0}=\theta_{\mathcal{W}}$ if $i\in\mathcal{W}$. Therefore, we can formulate the welfare maximization model as follows:

$$
\begin{aligned}
(\mathrm{Q}^{\mathrm{Wel}}) \qquad & \max_{\boldsymbol{x}} \sum_{i\in\mathcal{H}} D_i^O\log\left(\theta_{\mathcal{H}}+\sum_{k\in\mathcal{S}}\theta_{i,k}x_k\right)+\sum_{i\in\mathcal{W}} D_i^O\log\left(\theta_{\mathcal{W}}+\sum_{k\in\mathcal{S}}\theta_{i,k}x_k\right)\\
& \text{s.t. } \sum_{k\in\mathcal{S}} x_k \le C\\
& \qquad x_k\in\{0,1\},\ \forall k\in\mathcal{S}
\end{aligned}
$$

Note that the logarithm function $\log(\cdot)$ is concave but non-linear. In the numerical experiments, we introduce a sequence of piece-wise linear functions to approximate the objective function in model ($\mathrm{Q}^{\mathrm{Wel}}$) so that we can use standard optimization solvers (e.g., Gurobi, Cplex) to solve the optimization problem.

Notably, the aforementioned locker network design models assume that the demand is not fully covered even if there is one locker installed within certain distance from the demand segment. Instead, the demand can only be partially covered based on the utility functions. In this regard, our problem is also related to the class of partial/gradual covering models in the facility location literature. We refer interested readers to the partial covering model (e.g., Berman and Krass 2002) and gradual covering model (e.g., Berman et al. 2003) for more modeling choices in the locker network design problem.

Next, we consider the similar numerical setting as Appendix B.3, except that all the calculations below are based on the observed delivery profile $\boldsymbol{E}^{O}$. Table 6 shows that the numerical performances of model (Q^{Ms}) and ($\mathrm{Q}^{\mathrm{Wel}}$) appear to be similar in the Singapore LA Network case, even though they are clearly different in terms of the objective functions. This observation holds in our case due to the small utility obtained from using lockers. Notably, the number of residential blocks covered by model (Q^{Ms}) is slightly larger than that by model ($\mathrm{Q}^{\mathrm{Wel}}$), but the performance gap shrinks quickly with the increase of budget. We also observe that model (Q^{Ms}) is more computationally efficient than model ($\mathrm{Q}^{\mathrm{Wel}}$). This motivates us to focus on the market share maximization objective for large-scale network design.

Table 6 Comparison of different locker network design models.

Budget	Market Share (%)		Customer Welfare (%)		RB Coverage		CPU Time (s)	
	Q^{Ms}	Q^{Wel}	Q^{Ms}	Q^{Wel}	Q^{Ms}	Q^{Wel}	Q^{Ms}	Q^{Wel}
50	1.28	1.28	0.53	0.53	140	140	1.49	43.73
100	1.97	1.97	0.82	0.82	186	181	0.86	43.19
150	2.54	2.54	1.06	1.06	232	229	0.93	37.80
200	3.05	3.05	1.27	1.27	245	244	0.86	50.16
250	3.51	3.51	1.46	1.46	256	256	0.99	42.75
300	3.92	3.92	1.64	1.64	271	271	0.91	36.13
350	4.27	4.27	1.78	1.79	278	277	0.88	32.48
400	4.52	4.52	1.89	1.89	280	280	1.15	22.09

Note. The *Market Share* is calculated by (Volume of Parcels Delivered to the Lockers)/(Total Parcel Volume) $\times 100\%$ based on the observed demand profile $\boldsymbol{E}^{O}$; The *Customer Welfare* is obtained by (Total Welfare $-$ Welfare from Outside Option)/(Welfare from Outside Option) $\times 100\%$; The *RB Coverage* counts the number of residential block (RB) centers that can be served by at least one locker within 1 kilometer; The *CPU Time* reports the computational time in second.

To make the discussion clearer, we also consider the case with larger utility of using lockers. For ease of exposition, we manually increase the utility of using lockers by 10 times (i.e., $\theta_{i,k} \leftarrow \theta_{i,k} \times 10,\ \forall i \in \mathcal{H} \cup \mathcal{W},\ k \in \mathcal{S}$), and re-examine the differences between the three models. Table 7 shows that the performance gaps between model (Q^{Ms}) and ($\mathrm{Q}^{\mathrm{Wel}}$) become more apparent. It is straightforward to see that model (Q^{Ms}) attains higher market share while model ($\mathrm{Q}^{\mathrm{Wel}}$) achieves higher customer welfare. Furthermore, the number of residential blocks covered by model (Q^{Ms}) is also larger than that by model ($\mathrm{Q}^{\mathrm{Wel}}$), which implies that the lockers installed under model (Q^{Ms}) are more "evenly" distributed at the public residential areas so that there is at least one locker station within the vicinity of every public residential block. Therefore, model (Q^{Ms}) implicitly favors the egalitarian concern. In addition, model (Q^{Ms}) becomes much more computationally efficient than model ($\mathrm{Q}^{\mathrm{Wel}}$), and hence the market share maximization objective is still the recommended option for large-scale network design in this case.

Table 7 Comparison of different locker network design models with synthetic data.

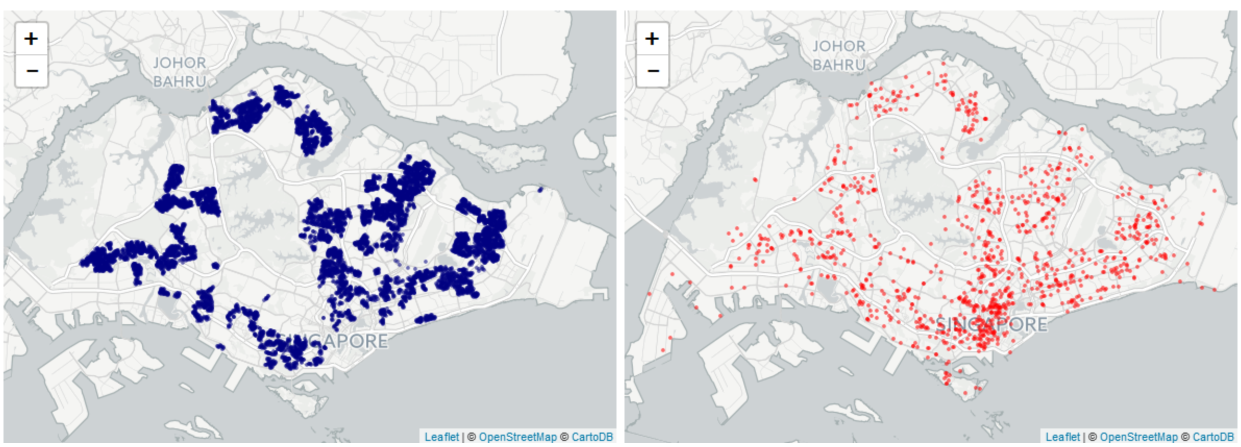
Budget	Market Share (%)		Customer Welfare (%)		RB Coverage		CPU Time (s)	
	Q^{Ms}	Q^{Wel}	Q^{Ms}	Q^{Wel}	Q^{Ms}	Q^{Wel}	Q^{Ms}	Q^{Wel}
50	10.14	10.09	4.59	4.62	166	157	3.83	178.12
100	14.70	14.57	6.86	6.90	240	212	2.10	160.08
150	18.27	18.14	8.69	8.76	259	247	13.50	145.56
200	21.31	21.20	10.34	10.40	269	256	2.72	106.42
250	23.95	23.85	11.79	11.85	271	270	8.76	55.12
300	26.25	26.16	13.09	13.13	277	271	4.67	52.46
350	28.22	28.19	14.23	14.25	278	278	2.29	41.17
400	29.60	29.60	15.02	15.02	280	280	1.52	28.77

D. Elaborations on Singapore LA Network

D.1. Visualization of Locker Network Solutions

In the Singapore LA Network case, we consider 3000 delivery points, including 1000 residential blocks and 2000 commercial buildings. We proceed next to describe the potential locker set $\mathcal{S}$ in the LA Network.

Figure 12 Potential locker locations.



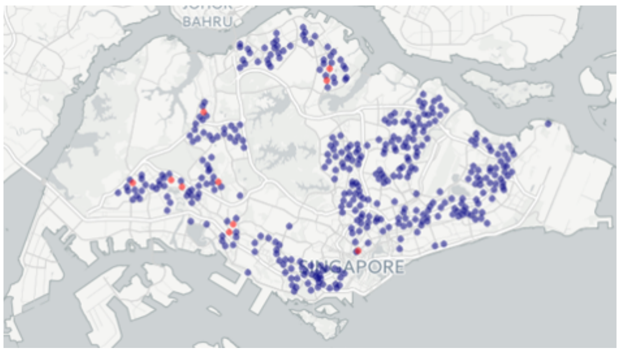
(a) Locker Locations at Residential Blocks (b) Locker Locations at Commercial Buildings

Note. (a) The dot represents the potential locker location at residential blocks. (b) The dot represents the potential locker location at commercial buildings.

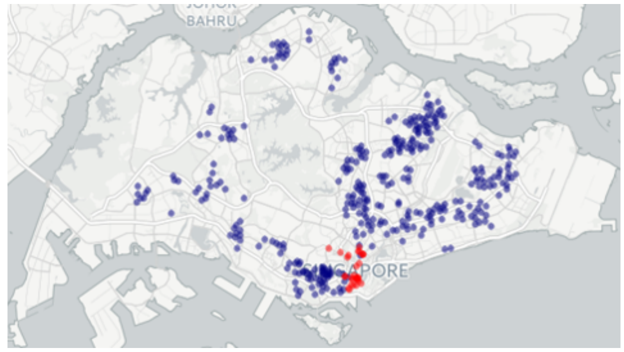
Since the access to public residential blocks can be granted by the Authority, the 1000 public residential centers can be used as locker locations. We plot these centers in Figure 12(a). In addition, as suggested by the senior manager of the delivery company, convenience stores such as 7-11 outlets and DBS ATM locations are also feasible locations to install the lockers. We manually collected 980 such locations and plot them in Figure 12(b). Some of these are located in busy shopping malls, or near train stations.

Next, we examine the impact of egalitarian constraints on the design of locker networks. To see this, we consider two types of models: (1) model ($\mathrm{P}^{\mathbf{0}}$) with egalitarian constraints, which is considered by the Singapore government; and (2) model ($\mathrm{P}^{\mathbf{0}}$) without egalitarian constraints, which is commonly considered by the commercial operators to maximize the locker usage. For ease of illustration, we solve the two models with budget $C = \{400, 1000\}$, respectively. We apply the R Package "leaflet" to plot the locker networks.[17]

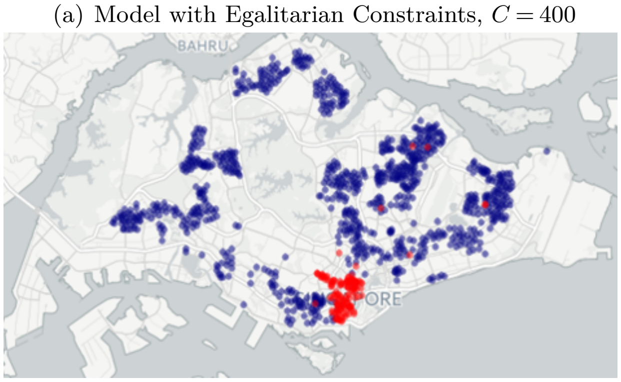
Figure 13 LA Network under different considerations.



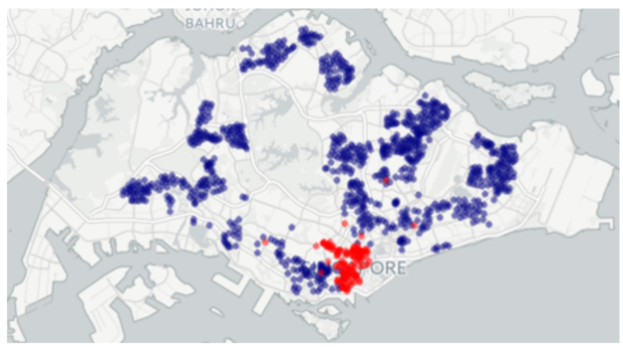
(a) Model with Egalitarian Constraints, $C = 400$



(b) Model without Egalitarian Constraints, $C = 400$



(c) Model with Egalitarian Constraints, $C = 1000$



(d) Model without Egalitarian Constraints, $C = 1000$

Figure 13(a) and (b) compare the networks obtained with and without the egalitarian constraints when budget $C = 400$. To satisfy the egalitarian constraints in model ($\mathrm{P}^{\mathbf{0}}$), the lockers are "evenly" distributed at the public residential areas so that there is at least one locker station within 250 meters of all public residential blocks. It turns out that only 10 commercial lockers are selected in 13(a) in this case. Figure 13(b) ignores the egalitarian constraints so that the network is designed to maximize the total utilization. In this case, more lockers are installed at the commercial areas where more parcel deliveries are gathered, and 30 private/commercial locations are selected. The egalitarian constraint reveals significant impact on the locker network design under this small budget case. Nevertheless, when the budget increases to 1000, there is essentially no difference between two networks (cf. Figure 13(c) and (d)). 139 commercial locations are selected in 13(c), and 144 such locations are selected in 13(d). Therefore, we would expect that an

[17] Please see `https://CRAN.R-project.org/package=leaflet`

appropriately scaled network, with the right notion of egalitarian consideration, can mitigate the gaps in the performance targets between the government and commercial operators.

We acknowledge that some lockers suffer from the problem of low utilization (e.g., the daily parcel volume is below 10) if we only have factored in the first-attempt delivery. To boost the utilization of LA Network, commercial operators can use the LA Network as a storage option to avoid another visit, to serve customers whose parcels were not delivered due to various reasons, for example, those estates that are far away from the distribution hub are known to have higher incidences of failed deliveries. This is a particularly appealing option for the LA Network, since there is already a station within the vicinity of every public residential blocks. The utilization of lockers will be boosted if failed deliveries to these blocks can be channeled to the lockers. Alternatively, since the locker is composed of different compartments and the size of each locker could be very flexible, right-sizing the LA Network appears to be vital for this new concept in the last mile delivery domain.

D.2. Impact of LA Network on the Parcel Delivery to CBD

In Section 6, recall that the locker network problem is solved by model ($\mathrm{P}^{\mathbf{0}}$), and the *Delivery Change* is calculated based on the observed parcel volume to CBD (i.e., D_j^O for $j \in \mathcal{W}_{\mathrm{CBD}}$, where $\mathcal{W}_{\mathrm{CBD}}$ represents the set of office buildings located at CBD). Given the locker network solution $\boldsymbol{x}$, we estimate the parcel volume delivered to CBD as $V_{\mathrm{CBD}}^{\mathbf{0}}(\boldsymbol{x}) = \sum_{j \in \mathcal{W}_{\mathrm{CBD}}} D_j^O g_{j,j}(\boldsymbol{x})$. With a slight abuse of notation, we also let $\delta_{\mathcal{H}}^i(\boldsymbol{x}) = \sum_{k \in N_i} \theta_{i,k} x_k$ denote the utility of using lockers under the solution $\boldsymbol{x}$ for customers living at public residential block $i \in \mathcal{H}$, and $\delta_{\mathcal{W}}^j(\boldsymbol{x}) = \sum_{k \in N_j} \theta_{j,k} x_k$ the utility for customers working at CBD block $j \in \mathcal{W}_{\mathrm{CBD}}$. According to Equation (3) and the definition of $g_{j,j}(\boldsymbol{x})$, we have

$$V_{\mathrm{CBD}}^{\mathbf{0}}(\boldsymbol{x}) = \sum_{j \in \mathcal{W}_{\mathrm{CBD}}} \left\{ D_{j,j} \left(\frac{\theta_{\mathcal{W}}}{\theta_{\mathcal{W}} + \delta_{\mathcal{W}}^j(\boldsymbol{x})} \right) + \sum_{i \in \mathcal{H}} \left(\frac{\theta_{\mathcal{W}}}{\theta_{\mathcal{H}} + \theta_{\mathcal{W}}} \right) \left(\frac{\theta_{\mathcal{W}}}{\theta_{\mathcal{W}} + \delta_{\mathcal{W}}^j(\boldsymbol{x})} \right) D_{i,j} \right\}, \tag{7}$$

where $\{D_{j,j}, D_{i,j}\}$ represent respectively the actual demands from Class-II and Class-III customers (as defined in Section 5) who are working at CBD.

In fact, if the demand $\{D_{j,j}, D_{i,j}\}$ can be exactly obtained, the "actual" parcel volume delivered to CBD should be calculated by

$$V_{\mathrm{CBD}}(\boldsymbol{x}) = \sum_{j \in \mathcal{W}_{\mathrm{CBD}}} \left\{ D_{j,j} \left(\frac{\theta_{\mathcal{W}}}{\theta_{\mathcal{W}} + \delta_{\mathcal{W}}^j(\boldsymbol{x})} \right) + \sum_{i \in \mathcal{H}} \left(\frac{\theta_{\mathcal{W}}}{\theta_{\mathcal{H}} + \theta_{\mathcal{W}} + \delta_{\mathcal{H}}^i(\boldsymbol{x}) + \delta_{\mathcal{W}}^j(\boldsymbol{x})} \right) D_{i,j} \right\}. \tag{8}$$

Notably, if $V_{\mathrm{CBD}}^{\mathbf{0}}(\boldsymbol{x}) - V_{\mathrm{CBD}}(\boldsymbol{x}) \geq 0$ holds for any given $\boldsymbol{x}$, the parcel volume $V_{\mathrm{CBD}}^{\mathbf{0}}(\boldsymbol{x})$ serves as an upper bound on the actual parcel volume $V_{\mathrm{CBD}}(\boldsymbol{x})$. In other words, the delivery reduction plotted in Figure 8 would be a lower bound to the actual case when we know the actual demand profile. For example, the actual parcel volume to CBD could be more than 7.5% given a well-chosen LA Network with budget 1500.

Given the locker network $\boldsymbol{x}$, the gap between the estimated parcel volume and actual parcel volume to CBD can be expressed as

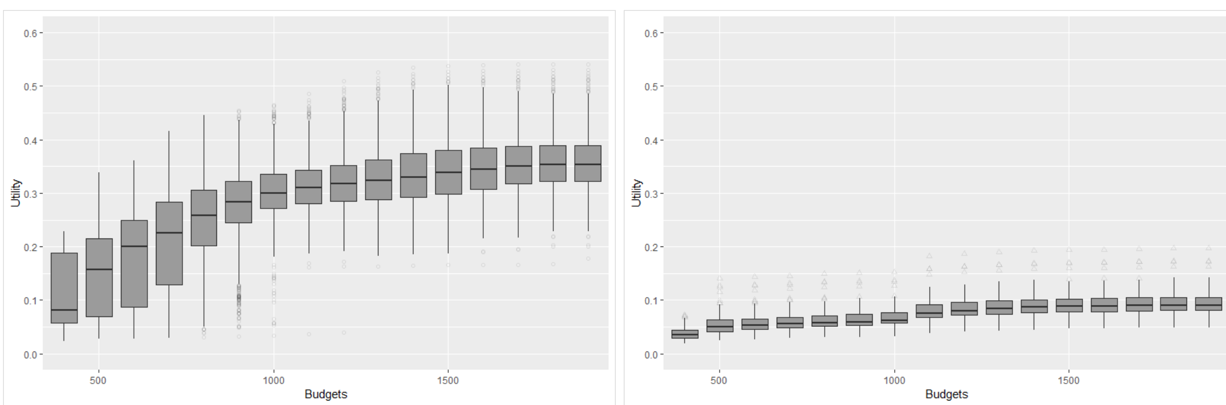
$$V_{\mathrm{CBD}}^{\mathbf{0}}(\boldsymbol{x}) - V_{\mathrm{CBD}}(\boldsymbol{x}) = \sum_{j \in \mathcal{W}_{\mathrm{CBD}}} \sum_{i \in \mathcal{H}} \left\{ \frac{\theta_{\mathcal{W}} \left(\theta_{\mathcal{W}} \delta_{\mathcal{H}}^i(\boldsymbol{x}) - \theta_{\mathcal{H}} \delta_{\mathcal{W}}^j(\boldsymbol{x}) \right)}{\left(\theta_{\mathcal{H}} + \theta_{\mathcal{W}} + \delta_{\mathcal{H}}^i(\boldsymbol{x}) + \delta_{\mathcal{W}}^j(\boldsymbol{x}) \right) \left(\theta_{\mathcal{H}} + \theta_{\mathcal{W}} \right) \left(\theta_{\mathcal{W}} + \delta_{\mathcal{W}}^j(\boldsymbol{x}) \right)} D_{i,j} \right\}.$$

It is straightforward to see that $V_{\mathrm{CBD}}^{\mathbf{0}}(\boldsymbol{x}) - V_{\mathrm{CBD}}(\boldsymbol{x}) \geq 0$ holds under some mild conditions, as stated in the following Proposition 4.

Proposition 4. *Given any home-office pair (i,j) for $i\in\mathcal{H},\ j\in\mathcal{W}_{CBD}$, if $\delta^j_{\mathcal{W}}(\boldsymbol{x})\le\frac{\theta_{\mathcal{W}}}{\theta_{\mathcal{H}}}\delta^i_{\mathcal{H}}(\boldsymbol{x})$ holds for a feasible locker network solution $\boldsymbol{x}$, then $V^{\mathbf{0}}_{CBD}(\boldsymbol{x})$, the estimated parcel volume to CBD under the observed demand profile $\boldsymbol{E}^O$, is an upper bound on the actual volume $V_{CBD}(\boldsymbol{x})$ when we know the actual demand $\boldsymbol{D}$.*

Next, we numerically show that the condition $\delta^j_{\mathcal{W}}(\boldsymbol{x})\le\frac{\theta_{\mathcal{W}}}{\theta_{\mathcal{H}}}\delta^i_{\mathcal{H}}(\boldsymbol{x})$ is satisfied for almost all the home-office pairs in our LA Network case when lockers are not allowed to be installed at CBD. Note that the attraction of locker k to customers in block j decreases with the distance between location k and j. Whenever there is no lockers installed nearby CBD blocks, the utility term $\delta^j_{\mathcal{W}}(\boldsymbol{x})=\sum_{k\in N_j}\theta_{j,k}x_k$ for $j\in\mathcal{W}_{\text{CBD}}$ would be a small positive number. As shown in Figure 14, we vary the budget from 400 to 1900, and obtain the optimal locker network solutions by solving model ($\text{P}^{\mathbf{0}}$). With the increase of locker budget, more lockers are installed around public residential blocks. As a result, the utility term $\delta^i_{\mathcal{H}}(\boldsymbol{x})$ significantly increases, whereas the term $\delta^j_{\mathcal{W}}(\boldsymbol{x})$ slightly increases a little bit. In addition, recall that we have $\theta_{\mathcal{W}}>\theta_{\mathcal{H}}$ in the LA Network case. Therefore, we would expect $V^{\mathbf{0}}_{\text{CBD}}(\boldsymbol{x})-V_{\text{CBD}}(\boldsymbol{x})>0$, and the estimated parcel volume reduction in CBD based on Equation (7) is a lower bound on the actual volume reduction.

Figure 14 Comparison between the utilities $\delta^i_{\mathcal{H}}(\boldsymbol{x})$ and $\delta^j_{\mathcal{W}}(\boldsymbol{x})$ for $i\in\mathcal{H},\ j\in\mathcal{W}_{\text{CBD}}$.



(a) Residential Blocks (b) CBD Blocks

E. Proof of Main Results

E.1. Proof of Proposition 1

Proposition 1. *In the worst case solution to model $(\mathcal{P})$, assuming $\mathcal{D}\neq\emptyset$, we have closed-form demand profile:*

$$\left\{\boldsymbol{E}\in\mathbb{R}_+^{|\mathcal{H}|+|\mathcal{W}|+|\mathcal{H}|\times|\mathcal{W}|}\ :\ \begin{array}{ll} E_{i,i}=D^O_i-\sum_{j\in\mathcal{W}}\dfrac{\theta_{\mathcal{H}}}{\theta_{\mathcal{H}}+\theta_{\mathcal{W}}}D^O_{i,j}, & \forall\ i\in\mathcal{H},\\ E_{j,j}=D^O_j-\sum_{i\in\mathcal{H}}\dfrac{\theta_{\mathcal{W}}}{\theta_{\mathcal{H}}+\theta_{\mathcal{W}}}D^O_{i,j}, & \forall j\in\mathcal{W},\\ E_{i,j}=D^O_{i,j}, & \forall i\in\mathcal{H},j\in\mathcal{W}.\end{array}\right\} \tag{4}$$

Proof: Given any feasible solution $\boldsymbol{x}$ to problem $(\mathcal{P})$, we denote the objective of the inner maximization problem as:

$$
\begin{aligned}
&\max_{\boldsymbol{E}} \left\{ \sum_{i\in\mathcal{H}} g_{i,i}(\boldsymbol{x}) E_{i,i} + \sum_{i\in\mathcal{W}} g_{j,j}(\boldsymbol{x}) E_{j,j} + \sum_{i\in\mathcal{H},\ j\in\mathcal{W}} g_{i,j}(\boldsymbol{x}) E_{i,j} \right\} \\
&\text{s.t. } \ E_{i,i} + \sum_{j\in\mathcal{W}} \frac{\theta_{\mathcal{H}}}{\theta_{\mathcal{H}}+\theta_{\mathcal{W}}} E_{i,j} = D_i^O,\ \forall i\in\mathcal{H}, \\
&\qquad E_{j,j} + \sum_{i\in\mathcal{H}} \frac{\theta_{\mathcal{W}}}{\theta_{\mathcal{H}}+\theta_{\mathcal{W}}} E_{i,j} = D_j^O,\ \forall j\in\mathcal{W}, \\
&\qquad E_{i,i}\ge 0,\ E_{j,j}\ge 0,\ E_{i,j}\ge D_{i,j}^O,\ \forall i\in\mathcal{H}, j\in\mathcal{W}.
\end{aligned}
$$

Suppose the solution $\tilde{\boldsymbol{E}}$ is the optimal solution to the problem above. If $\tilde{E}_{i,j} = D_{i,j}^O$ for $\forall i\in\mathcal{H},\ j\in\mathcal{W}$, then we are done. Therefore, without loss of generality, we assume that there exists at least one component $\tilde{E}_{i_0,j_0} = D_{i_0,j_0}^O + \epsilon\ (\epsilon > 0)$, for a home-office pair (i_0, j_0). Then we construct another feasible solution $\hat{\boldsymbol{E}} \neq \tilde{\boldsymbol{E}}$ but $\hat{E}_{i_0,j_0} = D_{i_0,j_0}^O$ and $\hat{E}_{i,j} = \tilde{E}_{i,j}$ for other (i,j)'s. $\hat{E}_{i_0,i_0} = \tilde{E}_{i_0,i_0} + \frac{\theta_{\mathcal{H}}}{\theta_{\mathcal{H}}+\theta_{\mathcal{W}}}\epsilon$, $\hat{E}_{j_0,j_0} = \tilde{E}_{j_0,j_0} + \frac{\theta_{\mathcal{W}}}{\theta_{\mathcal{H}}+\theta_{\mathcal{W}}}\epsilon$, and $\hat{E}_{i,i} = \tilde{E}_{i,i}$, $\hat{E}_{j,j} = \tilde{E}_{j,j}$ for all other i and j. It is straightforward to check the feasibility of solution $\hat{\boldsymbol{E}}$. Denote the objective value corresponding to solution $\hat{\boldsymbol{E}}$ and $\tilde{\boldsymbol{E}}$ as $Z(\hat{\boldsymbol{E}})$ and $Z(\tilde{\boldsymbol{E}})$, respectively. Next, we show that $Z(\hat{\boldsymbol{E}}) \ge Z(\tilde{\boldsymbol{E}})$.

Given the solution $\boldsymbol{x}$, for ease of expression, let $\mu_{\mathcal{H}} = \sum_{k\in N_i} \theta_{i,k} x_k$ and $\mu_{\mathcal{W}} = \sum_{k\in N_j} \theta_{j,k} x_k$. Then we can re-express

$$
g_{i,i}(\boldsymbol{x}) = \frac{\theta_{\mathcal{H}}}{\theta_{\mathcal{H}}+\mu_{\mathcal{H}}},\ g_{j,j}(\boldsymbol{x}) = \frac{\theta_{\mathcal{W}}}{\theta_{\mathcal{W}}+\mu_{\mathcal{W}}},\ \text{and } g_{i,j}(\boldsymbol{x}) = \frac{\theta_{\mathcal{W}}+\theta_{\mathcal{H}}}{\theta_{\mathcal{W}}+\theta_{\mathcal{H}}+\mu_{\mathcal{W}}+\mu_{\mathcal{H}}}.
$$

With some algebra, we have

$$
\begin{aligned}
&Z(\hat{\boldsymbol{E}}) - Z(\tilde{\boldsymbol{E}}) \\
&= \left\{\frac{\theta_{\mathcal{H}}}{\theta_{\mathcal{H}}+\mu_{\mathcal{H}}}\right\}\left\{\frac{\theta_{\mathcal{H}}}{\theta_{\mathcal{H}}+\theta_{\mathcal{W}}}\right\}\epsilon + \left\{\frac{\theta_{\mathcal{W}}}{\theta_{\mathcal{W}}+\mu_{\mathcal{W}}}\right\}\left\{\frac{\theta_{\mathcal{W}}}{\theta_{\mathcal{H}}+\theta_{\mathcal{W}}}\right\}\epsilon - \left\{\frac{\theta_{\mathcal{W}}+\theta_{\mathcal{H}}}{\theta_{\mathcal{W}}+\theta_{\mathcal{H}}+\mu_{\mathcal{W}}+\mu_{\mathcal{H}}}\right\}\epsilon \\
&= \left\{\frac{(\theta_{\mathcal{H}}\mu_{\mathcal{W}} - \theta_{\mathcal{W}}\mu_{\mathcal{H}})^2}{(\theta_{\mathcal{H}}+\mu_{\mathcal{H}})(\theta_{\mathcal{W}}+\mu_{\mathcal{W}})(\theta_{\mathcal{H}}+\theta_{\mathcal{W}})(\theta_{\mathcal{H}}+\theta_{\mathcal{W}}+\mu_{\mathcal{H}}+\mu_{\mathcal{W}})}\right\}\epsilon \\
&\ge 0.
\end{aligned}
$$

Therefore, we get the contradiction $Z(\hat{\boldsymbol{E}}) \ge Z(\tilde{\boldsymbol{E}})$, which implies that $\tilde{\boldsymbol{E}}$ is not optimal.

Repeating the procedure, we can conclude that for the optimal solution $\boldsymbol{E}$, we have $E_{i,j} = D_{i,j}^O, E_{i,i} = D_i^O - \sum_{j\in\mathcal{W}} \frac{\theta_{\mathcal{H}}}{\theta_{\mathcal{H}}+\theta_{\mathcal{W}}} D_{i,j}^O$, and $E_{j,j} = D_j^O - \sum_{i\in\mathcal{H}} \frac{\theta_{\mathcal{W}}}{\theta_{\mathcal{H}}+\theta_{\mathcal{W}}} D_{i,j}^O,\ \forall\ i\in\mathcal{H},\ j\in\mathcal{W}$. ■

E.2. Proof of Theorem 2

THEOREM 2. *Gap*$(\boldsymbol{x}) \le \max_{i\in\mathcal{H}, j\in\mathcal{W}} \left\{ \frac{\left[\rho_{\mathcal{W}}^j(\boldsymbol{x}) - \rho_{\mathcal{H}}^i(\boldsymbol{x})\right]^2}{\left[1+\rho_{\mathcal{H}}^i(\boldsymbol{x})\right]\left[1+\rho_{\mathcal{W}}^j(\boldsymbol{x})\right]} \right\} \left\{ \frac{\theta_{\mathcal{H}}\theta_{\mathcal{W}}}{(\theta_{\mathcal{H}}+\theta_{\mathcal{W}})^2} \right\}$ *holds for any feasible locker network solution* $\boldsymbol{x}$.

Proof: Based on the definition of $\chi_{i,j}(\boldsymbol{x})$, we can re-write

$$
\chi_{i,j}(\boldsymbol{x}) D_{i,j} := \frac{\left[\rho_{\mathcal{W}}^j(\boldsymbol{x}) - \rho_{\mathcal{H}}^i(\boldsymbol{x})\right]^2 D_{i,j}\theta_{\mathcal{H}}\theta_{\mathcal{W}}}{\left[1+\rho_{\mathcal{H}}^i(\boldsymbol{x})\right]\left[1+\rho_{\mathcal{W}}^j(\boldsymbol{x})\right]\left[\theta_{\mathcal{H}}+\theta_{\mathcal{W}}\right]\left[\theta_{\mathcal{H}}(1+\rho_{\mathcal{H}}^i(\boldsymbol{x})) + \theta_{\mathcal{W}}(1+\rho_{\mathcal{W}}^j(\boldsymbol{x}))\right]}.
$$

Note that

$$\sum_{j\in\mathcal{W}} D_{i,j} \leq \frac{\theta_{\mathcal{H}}+\theta_{\mathcal{W}}}{\theta_{\mathcal{H}}} D_i^O \text{ and } \sum_{i\in\mathcal{H}} D_{i,j} \leq \frac{\theta_{\mathcal{H}}+\theta_{\mathcal{W}}}{\theta_{\mathcal{W}}} D_j^O.$$

Hence, we have

$$\begin{aligned}\sum_{j\in\mathcal{W}} \chi_{i,j}(\boldsymbol{x}) D_{i,j} &\leq \max_{j\in\mathcal{W}} \left\{ \frac{\left[\rho_{\mathcal{W}}^j(\boldsymbol{x})-\rho_{\mathcal{H}}^i(\boldsymbol{x})\right]^2}{\left[1+\rho_{\mathcal{H}}^i(\boldsymbol{x})\right]\left[1+\rho_{\mathcal{W}}^j(\boldsymbol{x})\right]} \right\} \sum_{j\in\mathcal{W}} \left\{ \frac{D_{i,j}\theta_{\mathcal{H}}\theta_{\mathcal{W}}}{\left[\theta_{\mathcal{H}}+\theta_{\mathcal{W}}\right]\left[\theta_{\mathcal{H}}(1+\rho_{\mathcal{H}}^i(\boldsymbol{x}))+\theta_{\mathcal{W}}(1+\rho_{\mathcal{W}}^j(\boldsymbol{x}))\right]} \right\} \\ &\leq \max_{j\in\mathcal{W}} \left\{ \frac{\left[\rho_{\mathcal{W}}^j(\boldsymbol{x})-\rho_{\mathcal{H}}^i(\boldsymbol{x})\right]^2}{\left[1+\rho_{\mathcal{H}}^i(\boldsymbol{x})\right]\left[1+\rho_{\mathcal{W}}^j(\boldsymbol{x})\right]} \right\} \left\{ \frac{\theta_{\mathcal{W}}}{\theta_{\mathcal{H}}+\theta_{\mathcal{W}}} \right\} D_i^O.\end{aligned}$$

It is straightforward to derive the following upper bound

$$\sum_{i\in\mathcal{H},j\in\mathcal{W}} \chi_{i,j}(\boldsymbol{x}) D_{i,j} \leq \max_{i\in\mathcal{H},j\in\mathcal{W}} \left\{ \frac{\left[\rho_{\mathcal{W}}^j(\boldsymbol{x})-\rho_{\mathcal{H}}^i(\boldsymbol{x})\right]^2}{\left[1+\rho_{\mathcal{H}}^i(\boldsymbol{x})\right]\left[1+\rho_{\mathcal{W}}^j(\boldsymbol{x})\right]} \right\} \left\{ \frac{\theta_{\mathcal{W}}}{\theta_{\mathcal{H}}+\theta_{\mathcal{W}}} \right\} \sum_{i\in\mathcal{H}} D_i^O.$$

Similarly, we have

$$\sum_{i\in\mathcal{H},j\in\mathcal{W}} \chi_{i,j}(\boldsymbol{x}) D_{i,j} \leq \max_{i\in\mathcal{H},j\in\mathcal{W}} \left\{ \frac{\left[\rho_{\mathcal{W}}^j(\boldsymbol{x})-\rho_{\mathcal{H}}^i(\boldsymbol{x})\right]^2}{\left[1+\rho_{\mathcal{H}}^i(\boldsymbol{x})\right]\left[1+\rho_{\mathcal{W}}^j(\boldsymbol{x})\right]} \right\} \left\{ \frac{\theta_{\mathcal{H}}}{\theta_{\mathcal{H}}+\theta_{\mathcal{W}}} \right\} \sum_{j\in\mathcal{W}} D_j^O.$$

Taking a convex combination of the above, we can derive

$$\sum_{i\in\mathcal{H},j\in\mathcal{W}} \chi_{i,j}(\boldsymbol{x}) D_{i,j} \leq \max_{i\in\mathcal{H},j\in\mathcal{W}} \left\{ \frac{\left[\rho_{\mathcal{W}}^j(\boldsymbol{x})-\rho_{\mathcal{H}}^i(\boldsymbol{x})\right]^2}{\left[1+\rho_{\mathcal{H}}^i(\boldsymbol{x})\right]\left[1+\rho_{\mathcal{W}}^j(\boldsymbol{x})\right]} \right\} \left\{ \frac{\theta_{\mathcal{H}}\theta_{\mathcal{W}}}{(\theta_{\mathcal{H}}+\theta_{\mathcal{W}})^2} \right\} \left(\sum_{i\in\mathcal{H}} D_i^O + \sum_{j\in\mathcal{W}} D_j^O \right).$$

Recall that

$$\mathrm{Gap}(\boldsymbol{x}) = \frac{V^{\mathbf{0}}(\boldsymbol{x})-V(\boldsymbol{x})}{\sum_{i\in\mathcal{H}} D_i^O + \sum_{j\in\mathcal{W}} D_j^O}, \text{ and } V^{\mathbf{0}}(\boldsymbol{x})-V(\boldsymbol{x}) = \sum_{i\in\mathcal{H},j\in\mathcal{W}} \chi_{i,j}(\boldsymbol{x}) D_{i,j}.$$

It is straightforward to upper bound the $\mathrm{Gap}(\boldsymbol{x})$ as follows:

$$\mathrm{Gap}(\boldsymbol{x}) \leq \max_{i\in\mathcal{H},j\in\mathcal{W}} \left\{ \frac{\left[\rho_{\mathcal{W}}^j(\boldsymbol{x})-\rho_{\mathcal{H}}^i(\boldsymbol{x})\right]^2}{\left[1+\rho_{\mathcal{H}}^i(\boldsymbol{x})\right]\left[1+\rho_{\mathcal{W}}^j(\boldsymbol{x})\right]} \right\} \left\{ \frac{\theta_{\mathcal{H}}\theta_{\mathcal{W}}}{(\theta_{\mathcal{H}}+\theta_{\mathcal{W}})^2} \right\}.$$

■

E.3. Proof of Theorem 3

Following the proof of Theorem 1, we also develop a robust model to address this generic locker network design problem. In this setting, we represent the uncertainty set (assume $\mathcal{D} \neq \emptyset$) as:

$$\mathcal{D} := \left\{ \boldsymbol{E} \in \mathbb{R}^{|\Phi|} : \begin{array}{ll} E_{\gamma,\gamma} + \sum_{(\gamma,\psi)\in\Phi(\gamma)} \frac{\theta_\gamma}{\theta_\gamma + \sum_{j\in\psi}\theta_j} E_{\gamma,\psi} = D_\gamma^O, & \forall \gamma = 1,2,\ldots,\Gamma, \\ E_{\gamma,\gamma} \geq 0, & \forall \gamma = 1,2,\ldots,\Gamma, \\ E_{\gamma,\psi} \geq D_{\gamma,\psi}, & \forall (\gamma,\psi)\in\Phi(\gamma),\ \gamma = 1,2,\ldots,\Gamma. \end{array} \right\}$$

The generic robust location framework can be formulated as:

$$\begin{aligned} (\mathcal{G}) \quad \min_{\boldsymbol{x}} \max_{\boldsymbol{E}\in\mathcal{D}} & \left\{ \sum_{\gamma=1}^{\Gamma} \left[g_{\gamma,\gamma}(\boldsymbol{x}) E_{\gamma,\gamma} + \sum_{(\gamma,\psi)\in\Phi(\gamma)} \frac{1}{|\{\gamma,\psi\}|} g_{\gamma,\psi}(\boldsymbol{x}) E_{\gamma,\psi} \right] \right\} \\ \text{s.t. } & \sum_{k\in\mathcal{S}} x_k \leq C \\ & x_k \in \{0,1\},\ \forall k \in \mathcal{S} \end{aligned}$$

where $|\{\gamma,\psi\}|$ denotes the number of elements in set $\{\gamma,\psi\}$. In the objective function, we divide the second term by $|\{\gamma,\psi\}|$ to avoid "double-counting".

Next, we demonstrate that the demand profile specified by Equation (6) is optimal to problem $\mathcal{G}$ in the worst case scenario, as stated in Proposition 5.

PROPOSITION 5. *In the worst case solution to model* ($\mathcal{G}$)*, we have closed-form demand profile:*

$$\left\{ \boldsymbol{E} \in \mathbb{R}_+^{|\Phi|} \;:\; \begin{array}{ll} E_{\gamma,\gamma} + \sum_{(\gamma,\psi)\in\Phi(\gamma)} \dfrac{\theta_\gamma}{\theta_\gamma + \sum_{j\in\psi}\theta_j} D_{\gamma,\psi} = D_\gamma^O,\ \forall\gamma = 1,2,\ldots,\Gamma, & \\ E_{\gamma,\psi} = D_{\gamma,\psi}, & \forall(\gamma,\psi)\in\Phi(\gamma),\ \gamma = 1,2,\ldots,\Gamma. \end{array} \right\}$$

Proof: Given any feasible solution $\boldsymbol{x}$ to problem ($\mathcal{G}$), we denote the objective of the inner maximization problem as:

$$\begin{aligned}
&\max_{\boldsymbol{E}} \left\{ \sum_{\gamma=1}^{\Gamma} \left[g_{\gamma,\gamma}(\boldsymbol{x}) E_{\gamma,\gamma} + \sum_{(\gamma,\psi)\in\Phi(\gamma)} \frac{1}{|\{\gamma,\psi\}|} g_{\gamma,\psi}(\boldsymbol{x}) E_{\gamma,\psi} \right] \right\} \\
&\text{s.t. } E_{\gamma,\gamma} + \sum_{(\gamma,\psi)\in\Phi(\gamma)} \frac{\theta_\gamma}{\theta_\gamma + \sum_{j\in\psi}\theta_j} E_{\gamma,\psi} = D_\gamma^O,\ \forall\gamma = 1,2,\ldots,\Gamma, \\
&\quad E_{\gamma,\gamma} \ge 0,\ \forall\gamma = 1,2,\ldots,\Gamma, \\
&\quad E_{\gamma,\psi} \ge D_{\gamma,\psi},\ \forall(\gamma,\psi)\in\Phi(\gamma),\ \gamma = 1,2,\ldots,\Gamma.
\end{aligned}$$

Suppose the solution $\tilde{\boldsymbol{E}}$ is the optimal solution to the problem above. If $\tilde{E}_{\gamma,\psi} = D_{\gamma,\psi}$ for $\forall(\gamma,\psi)\in\Phi(\gamma)$, then we are done. Therefore, without loss of generality, we assume that there exists at least one component $\tilde{E}_{\gamma_1,\gamma_2,\ldots,\gamma_n} = D_{\gamma_1,\gamma_2,\ldots,\gamma_n} + \epsilon$ ($\epsilon > 0$), for a specific pair $(\gamma_1,\gamma_2,\ldots,\gamma_n)$. Then we construct another feasible solution $\hat{\boldsymbol{E}} \ne \tilde{\boldsymbol{E}}$ with $\hat{E}_{\gamma_1,\gamma_2,\ldots,\gamma_n} = D_{\gamma_1,\gamma_2,\ldots,\gamma_n}$, $\hat{E}_{\gamma_i,\gamma_i} = \tilde{E}_{\gamma_i,\gamma_i} + \frac{\theta_{\gamma_i}}{\sum_{i=1}^n \theta_{\gamma_i}}\epsilon$ for $i = 1,2,\ldots,n$, and $\hat{E}_{\gamma,\psi} = \tilde{E}_{\gamma,\psi}$ for other (γ,ψ)'s. It is straightforward to check the feasibility of solution $\hat{\boldsymbol{E}}$. Denote the objective value corresponding to solution $\hat{\boldsymbol{E}}$ and $\tilde{\boldsymbol{E}}$ as $Z(\hat{\boldsymbol{E}})$ and $Z(\tilde{\boldsymbol{E}})$, respectively. Next, we show that $Z(\hat{\boldsymbol{E}}) \ge Z(\tilde{\boldsymbol{E}})$.

Given the solution $\boldsymbol{x}$, for ease of expression, let $\mu_{\gamma_i} = \sum_{k\in N_{\gamma_i}} \theta_{\gamma_i,k} x_k$ for $i = 1,2,\ldots,n$. Then we can re-express the choice model as:

$$g_{\gamma_i,\gamma_i}(\boldsymbol{x}) = \frac{\theta_{\gamma_i}}{\theta_{\gamma_i} + \mu_{\gamma_i}},\ \forall i = 1,2,\ldots,n, \text{ and } g_{\gamma_1,\gamma_2,\ldots,\gamma_n}(\boldsymbol{x}) = \frac{\sum_{i=1}^n \theta_{\gamma_i}}{\sum_{i=1}^n \theta_{\gamma_i} + \sum_{i=1}^n \mu_{\gamma_i}}.$$

With some algebra, we have

$$\begin{aligned}
Z(\hat{\boldsymbol{E}}) - Z(\tilde{\boldsymbol{E}}) &= \sum_{i=1}^n \left\{ \frac{\theta_{\gamma_i}}{\theta_{\gamma_i} + \mu_{\gamma_i}} \right\} \left\{ \frac{\theta_{\gamma_i}}{\sum_{j=1}^n \theta_{\gamma_j}} \right\} \epsilon - \left\{ \frac{\sum_{i=1}^n \theta_{\gamma_i}}{\sum_{i=1}^n \theta_{\gamma_i} + \sum_{i=1}^n \mu_{\gamma_i}} \right\} \epsilon \\
&= \left\{ \sum_{i=1}^n \left[\frac{\theta_{\gamma_i}^2}{(\sum_{j=1}^n \theta_{\gamma_j})(\theta_{\gamma_i} + \mu_{\gamma_i})} \right] - \frac{[\sum_{i=1}^n \theta_{\gamma_i}]^2}{\sum_{i=1}^n \left[(\sum_{j=1}^n \theta_{\gamma_j})(\theta_{\gamma_i} + \mu_{\gamma_i}) \right]} \right\} \epsilon.
\end{aligned}$$

Let $\alpha_i = \frac{\theta_{\gamma_i}}{\sqrt{(\sum_{j=1}^n \theta_{\gamma_j})(\theta_{\gamma_i} + \mu_{\gamma_i})}}$, $\beta_i = \sqrt{(\sum_{j=1}^n \theta_{\gamma_j})(\theta_{\gamma_i} + \mu_{\gamma_i})}$, we have

$$Z(\hat{\boldsymbol{E}}) - Z(\tilde{\boldsymbol{E}}) = \left\{ \sum_{i=1}^n \alpha_i^2 - \frac{(\sum_{i=1}^n \alpha_i \beta_i)^2}{\sum_{i=1}^n \beta_i^2} \right\} \epsilon \ge 0 \text{ (by Cauchy–Schwarz inequality).}$$

Therefore, we get the contradiction $Z(\hat{\boldsymbol{E}}) \geq Z(\tilde{\boldsymbol{E}})$, which implies that $\tilde{\boldsymbol{E}}$ is not optimal. Repeating the procedure, we can conclude that for the optimal solution $\boldsymbol{E}$, we have $E_{\gamma,\psi} = D_{\gamma,\psi}$, $E_{\gamma,\gamma} = D_\gamma^O - \sum_{(\gamma,\psi)\in\Phi(\gamma)} \frac{1}{|\{\gamma,\psi\}|} g_{\gamma,\psi}(\boldsymbol{x}) D_{\gamma,\psi}$, $\forall(\gamma,\psi)\in\Phi(\gamma)$, $\gamma = 1,2,\ldots,\Gamma$. ■

Next, we are ready to prove Theorem 3.

Theorem 3. *$U^{\mathbf{0}}(\boldsymbol{x}) - U(\boldsymbol{x}) \geq 0$ holds for any feasible locker network solution $\boldsymbol{x}$.*

Proof: Proposition 5 implies that model (G) is equivalent to the robust model ($\mathcal{G}$). In the case of $D_{\gamma,\psi} = 0$ for $\forall(\gamma,\psi)\in\Phi(\gamma)$, $\gamma = 1,2,\ldots,\Gamma$, we represent $\mathcal{D}$ as $\mathcal{D}^{\mathbf{0}}$, and model ($\mathcal{G}$) as ($\mathcal{G}^{\mathbf{0}}$) to avoid confusion. It is straightforward to show that model ($\mathrm{G}^{\mathbf{0}}$) is equivalent to the robust model ($\mathcal{G}^{\mathbf{0}}$). Clearly, we also have $\mathcal{D}\subseteq\mathcal{D}^{\mathbf{0}}$ in this generic setting. We denote respectively the objective values for problem ($\mathcal{G}$) and ($\mathcal{G}^{\mathbf{0}}$) as $U(\boldsymbol{x})$ and $U^{\mathbf{0}}(\boldsymbol{x})$, given a feasible solution $\boldsymbol{x}$. Note that $U(\boldsymbol{x})$ and $U^{\mathbf{0}}(\boldsymbol{x})$ also represent the objective values for problem (G) and ($\mathrm{G}^{\mathbf{0}}$), respectively. Since the inner optimization problem in ($\mathcal{G}$) is a maximization problem over the uncertainty demand set $\mathcal{D}$, we prove that $U^{\mathbf{0}}(\boldsymbol{x}) - U(\boldsymbol{x}) \geq 0$. ■

E.4. Proof of Theorem 4

Theorem 4. *The locker network design problem (P) is NP-hard.*

Proof: To show the NP-hardness of the locker network design problem (P), we show the (MMNL) assortment problem is NP-hard even when every segment has unit demand and there are only two locker preference types.

Let the demand in each segment be 1 and total budget on locker network is C. We consider a simplified setting in which every segment has exactly two preferences towards different lockers with each utility being a sufficiently large number M, i.e., the unit demand will be satisfied immediately once at least one locker in its preference list has been installed to serve the segment.

Next, we use a reduction from the vertex cover problem to prove the hardness of our problem. Consider an undirected graph $G = (V,E)$, vertex cover is to check whether there exists a subset of vertices $V' \subset V$ of cardinality at most C, such that every edge $e \in E$ is incident to at least one vertex in V'. Let I be an instance of vertex cover. We will construct an instance I' of assortment problem corresponding to I. Let the number of locker locations be $N = |V|$, each of them corresponding to each vertex in graph G, and let the set of locker preference list be $\{(i,j):(v_i,v_j)\in E\}$, each of them corresponding to each customer's choice. We show that the maximal coverage of I' is $|E|$ if and only if there is a vertex cover of size at most k in instance I.

Suppose there exists a vertex cover $V' \subset V$ such that $|V'| \leq C$. In the locker setup solution $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$, where $x_i = 1$ if $v_i \in V'$ and $x_i = 0$ otherwise. Therefore, the total demands covered by the locker can be expressed as

$$R(\boldsymbol{x}) = \left\{ \sum_{(v_i,v_j)\in E} \max\{x_i, x_j\} \right\}.$$

Since V' is a vertex cover, at least one of $\{v_i, v_j\}$ is in V'. Therefore, $\max\{x_i, x_j\} = 1$ for $\forall\ (v_i, v_j) \in E$, and the total coverage is $|E|$.

Conversely, consider the locker solution $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$ containing at most C non-zero components and providing total coverage $|E|$, we argue that $V' = \{v_i : x_i = 1\}$ is a vertex cover. Notice that the total coverage can be rewritten as

$$R(\boldsymbol{x}) = \left\{ \sum_{(v_i, v_j) \in E} \min(\max\{x_i, x_j\}, 1) \right\},$$

and thus $\max\{x_i, x_j\} \geq 1$ for $\forall\ (v_i, v_j) \in E$, implying that at least one of $\{v_i, v_j\}$ is in set V', Therefore, set V' is a vertex cover. ■

E.5. Proof of Proposition 3

Proposition 3. *The SOCP-MIP formulation is equivalent to the FB formulation.*

Proof: Denote the optimal solution to Model (FB) as $\boldsymbol{x}^1$ and the optimal solution to Model (SOCP-MIP) as $(\boldsymbol{x}^2, \boldsymbol{z}^*, \boldsymbol{w}^*)$. Denote the objective of Model (FB) and (SOCP-MIP) as $f_1(\cdot)$ and $f_2(\cdot)$, respectively. We show that $f_1(\boldsymbol{x}^1) = f_2(\boldsymbol{x}^2, \boldsymbol{z}^*, \boldsymbol{w}^*)$.

(1) Given the optimal solution $\boldsymbol{x}^1$ to Model (FB), we construct a feasible solution $(\boldsymbol{x}^1, \boldsymbol{z}, \boldsymbol{w})$ to Model (SOCP-MIP) where $w_i = 1/z_i = 1/(\sum_{k \in \mathcal{S}} \mu_{i,k} x_i^1)$, $\forall i \in \mathcal{B}$. The feasibility of solution $(\boldsymbol{x}^1, \boldsymbol{z}, \boldsymbol{w})$ can be easily checked. Therefore, we have

$$f_1(\boldsymbol{x}^1) = f_2(\boldsymbol{x}^1, \boldsymbol{z}, \boldsymbol{w}) \geq f_2(\boldsymbol{x}^2, \boldsymbol{z}^*, \boldsymbol{w}^*).$$

The first inequality comes from the construction of solution $(\boldsymbol{x}^1, \boldsymbol{z}, \boldsymbol{w})$ and the second inequality is due to the optimality of the solution $(\boldsymbol{x}^2, \boldsymbol{z}^*, \boldsymbol{w}^*)$.

(2) Given the optimal solution $(\boldsymbol{x}^2, \boldsymbol{z}^*, \boldsymbol{w}^*)$ to Model (SOCP-MIP). The feasibility of this solution demonstrates that $w_i^* \geq 1/(\sum_{k \in \mathcal{S}} \mu_{i,k} x_i^2)$, $\forall i \in \mathcal{B}$ and thus $f_2(\boldsymbol{x}^2, \boldsymbol{z}^*, \boldsymbol{w}^*) \geq f_1(\boldsymbol{x}^2)$. Therefore, we have

$$f_2(\boldsymbol{x}^2, \boldsymbol{z}^*, \boldsymbol{w}^*) \geq f_1(\boldsymbol{x}^2) \geq f_1(\boldsymbol{x}^1).$$

The first inequality comes from the feasibility of solution $(\boldsymbol{x}^2, \boldsymbol{z}^*, \boldsymbol{w}^*)$ and the second inequality is due to the optimality of solution $\boldsymbol{x}^1$.

Hence, $f_1(\boldsymbol{x}^1) = f_2(\boldsymbol{x}^2, \boldsymbol{z}^*, \boldsymbol{w}^*)$ demonstrates the equivalence of (SOCP-MIP) formulation to the (FB) formulation. ■