POMS PRODUCTION AND OPERATIONS MANAGEMENT

PRODUCTION AND OPERATIONS MANAGEMENT

Vol. 29, No. 4, April 2020, pp. 955–972
ISSN 1059-1478 | EISSN 1937-5956 | 20 | 2904 | 0955
DOI 10.1111/poms.13143

Robust Vehicle Pre-Allocation with Uncertain Covariates

Zhaowei Hao
Institute of Supply Chain Analytics, Dongbei University of Finance and Economics, Dalian, 116025, Liaoning, China, haozhaowei@dufe.edu.cn

Long He, Zhenyu Hu*
NUS Business School and Institute of Operations Research and Analytics, National University of Singapore, 119245, Singapore, longhe@nus.edu.sg, bizhuz@nus.edu.sg

Jun Jiang
Institute of Data Science and NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, 119246, Singapore, jiangjun@u.nus.edu

Motivated by a leading taxi operator in Singapore, we consider the idle vehicle pre-allocation problem with uncertain demands and other uncertain covariate information such as weather. In this problem, the operator, upon observing its distribution of idle vehicles, proactively allocates the idle vehicles to serve future uncertain demands. With perfect information of demand distribution, the problem can be formulated as a stochastic transportation problem. Yet, the non-stationarity and spatial correlation of demands pose significant challenges in estimating its distribution accurately from historical data. We employ a novel distributionally robust optimization approach that can utilize covariate information as well as the moment information of demand to construct a scenario-wise ambiguity set. We further illustrate how the key parameters required by the new ambiguity set, such as the scenarios and their probabilities, can be estimated via multivariate regression tree. Although information about uncertain covariates provides no value when there is perfect knowledge of demand distribution, we show that it could alleviate the over-conservativeness of the robust solution. The resulting distributionally robust optimization problem can be exactly and tractably solved using linear decision rule technique. We further validate the performance of our solution via extensive numerical simulations, and a case study using trip and vehicle status data from our partner taxi operator, paired with the rainfall data from the Meteorological Service Singapore.

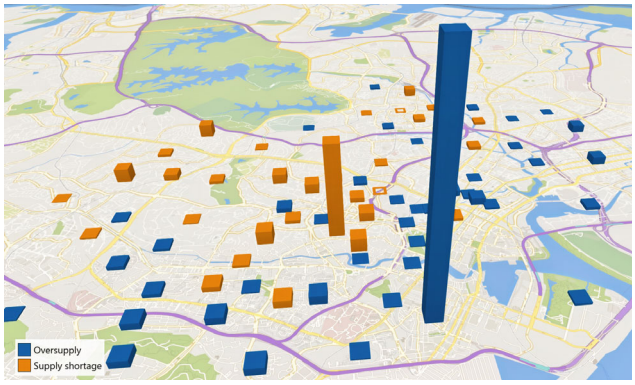
Key words: vehicle pre-allocation; distributionally robust optimization; covariate information; multivariate regression tree
History: Received: December 2018; Accepted: December 2019 by Qi (Annabelle) Feng after, 1 revision.

1. Introduction

In this paper, we study the single-period idle vehicle pre-allocation problem when there are uncertainties in both demands and other covariate information (also known as feature, attribute or predictor) that is correlated with demands. Our work stems from a partnership with a leading taxi operator in Singapore. One pressing issue faced by the operator is that a significant amount of its vehicles crammed into the downtown area during the morning rush hours each day and these vehicles, after dropping off their passengers, stayed idle in the area for a prolonged duration of time. As a result, the oversupply of vehicles exacerbates congestion in downtown and meanwhile, there are insufficient vehicles to meet demands in the surrounding areas. Figure 1 below illustrates a typical pattern of the vehicles' spatial distribution around the downtown area in Singapore during the morning rush hours. The highest blue bar is located in downtown, which indicates a large number of idle vehicles there and the orange bars scattered around the map are regions where passengers failed to find a vehicle via the mobile app. With this observation, their management team is mostly concerned with a one-time pre-allocation decision in the morning rush hour when the mismatch between supply and demand is prominent to direct idle vehicles from the downtown area to the surrounding regions. For exposition simplicity, in the following, we use allocation and pre-allocation interchangeably.

If the joint distribution of uncertain demands in all the regions is known, the vehicle allocation problem can be formulated as a classic stochastic transportation problem. Regions with idle vehicles are interpreted as supply nodes and all nearby regions with possible demands are viewed as demand nodes. The objective is to minimize the expected lost sales and transportation costs (or maximize the expected net profit) by sending flows from supply nodes to

Figure 1 Pattern of Vehicle's Spatial Distribution during 8:00 AM to 8:15 AM on 15 May 2017 [Color figure can be viewed at wileyonlinelibrary.com]



demand nodes. Even though in practice, the uncertain demands are correlated with other uncertain covariates such as weather, neither the knowledge of such dependence nor the distribution of covariates provides any value here since the objective only depends on the marginal distribution of demands, which is assumed to be perfectly known in stochastic transportation problem.

The main challenge in our practical problem, however, is that the joint distribution of uncertain demands is unknown. Obtaining an accurate estimate of the distribution from historical data can be very difficult due to the non-stationarity and spatial correlation of travel demands—a common feature in the urban transportation industry. While in many cases with observable covariate information, the predictability of demands can be greatly improved, one major covariate in our problem—the future weather information at the demand location—is uncertain at the point when allocation decisions are made. Facing such limited information, our goal is to utilize both historical demands as well as covariate information to deliver efficient and robust vehicle allocation decisions.

To address the issue of limited demand information, we employ the framework of distributionally robust optimization, where demand distribution is assumed to lie in a certain ambiguity set instead of being known perfectly and one chooses a decision that performs the best against the worst possible distribution within the ambiguity set. Classical distributionally robust optimization models that consider ambiguity sets based on the moments information (see, for example, Delage and Ye 2010, Scarf 1958, Wiesemann et al. 2014) usually ignore any covariate information that may be correlated with the uncertainty in the optimization models. The resulting solutions are sometimes criticized as being overly conservative.

We consider in this work a novel scenario-wise ambiguity set that is recently proposed in Chen et al. (2018). In this ambiguity set, the support of the random variables under consideration is first clustered into several scenarios, and the moments information is then specified for each of the scenarios. Chen et al. (2018) propose the scenario-wise ambiguity set as a unified format to specify both moments and statistical distance information, yet how the scenarios can be formed by utilizing covariate information is not discussed. In contrast, we propose to use a multivariate regression tree to classify demands into different scenarios based on the covariate information and we illustrate in our case study using weather data as covariate information to classify the travel demands. We further show that while the knowledge of the uncertain covariates provides no benefit when the marginal distribution of demands is known, it can lead to a less conservative solution when the marginal distribution of demands lies in the corresponding scenario-wise ambiguity set.

On the computational side, our distributionally robust optimization problem can be reformulated as a second-order cone program using standard techniques in robust optimization. However, due to the nonlinearity in the objective function, the number of constraints in the resulting reformulation grows exponentially with the number of demand regions. Instead, we linearize the objective using the linear decision rule technique. Our formulation can be solved very efficiently, and more interestingly we establish that the linearization step results in no loss of optimality. In our extensive numerical simulations, the model (abbreviated as SDR) is benchmarked against the one with only marginal moment information (MMM), the one with ambiguity set that specifies covariance information but without covariate information (SDP), and the sample average approximation method (SAA) that solves the stochastic transportation problem with the empirical distribution. The average out-of-sample performance of SDR is consistently better than both MMM and SDP and is close to that of SAA, and in terms of the standard deviation of the out-of-sample performance, SDR is always the lowest. Finally, in a case study using trip and vehicle status data from our partner taxi operator, paired with the rainfall data from the Meteorological Service Singapore, our SDR has a better average performance compared to SAA, SDP, and MMM models, and has much less variability in the out-of-sample profits as opposed to SAA.

2. Literature Review

The problem of vehicle allocation or more generally vehicle repositioning has been studied in both

ridesharing and vehicle sharing contexts. In ridesharing, vehicles (and drivers) are typically independent service providers and are matched to customers via on-demand service platforms such as Uber, Didi, and Grab. Due to the nature of self-scheduled service providers, the platform has less control over the vehicles and research on vehicle allocation problem in this context has been mainly focusing on how to effectively use economic tools such as price and wage to both create more vehicles and to incentivize the vehicles to flow to the desired places. For example, Cachon et al. (2017) compare different pricing schemes to examine the effectiveness of surge pricing. Taylor (2018) focuses on studying how delay sensitivity and driver independence impact the on-demand platform's price and wage decisions. Bai et al. (2018) extend Taylor (2018) by further analyzing the impact of demand rate, waiting time sensitivity, service rate, and the available driver capacity. Considering the spatially distributed demand, Bimpikis et al. (2019) examine how to dynamically adjust the prices at each location to rebalance the vehicles. While the imbalance of vehicles is a shared problem for both our partner taxi operator and the above ridesharing industry, the price for the taxi is usually fixed, but the taxi company has more control over its vehicles and hence it is easier to allocate vehicles via more direct methods (e.g., directly informing its drivers/employees). Based on this, our paper considers the firm fully controls the vehicles and proposes an optimization framework that is able to utilize the covariate information in fleet allocation decisions.

In the context of vehicle sharing, like taxi industry, the vehicles are owned by the vehicle sharing companies such as Car2go (for car) and Mobike (for bike) but the customers in this context are also drivers that move vehicles from one place to another. Similar vehicle allocation problems also arise due to unbalanced demand flows across different regions. Shu et al. (2013) study the problem of initial bicycle allocation in a stochastic network flow model. They assume demands arrive according to Poisson processes with known rates and the problem is approximated by a deterministic linear program. Lu et al. (2017) also focus on the problem of allocating the initial vehicle fleet. They approximate the multistage dynamic model as a two-stage stochastic program and in their case study, Gamma distribution is used to fit the empirical demand distributions. A dynamic inventory (e.g., vehicle) repositioning problem is considered in Benjaafar et al. (2018), where after initial allocation, one can further reallocate vehicles each time demands are realized. Given independent demands' distributions in each period, the problem is formulated as a multistage stochastic dynamic program, and an optimal repositioning policy is characterized. A similar dynamic problem is studied in He et al. (2020). They further formulate the distributionally robust counterpart of the problem and provide approximation algorithms. Our model may also be applied to the one-time overnight vehicle repositioning operations in these problems.

The vehicle allocation problem is also studied in the transportation literature in the context of taxi operators. In a series of papers, Miao et al. (2015, 2016) derive dynamic robust vehicle allocation decisions using only support information and Miao et al. (2017) utilize first two moments information of the demands to formulate a distributionally robust optimization model to arrive at static vehicle allocation decisions. Although we study a simpler single-period problem—which is the primal concern of the taxi operator we are working with, our focus is on how information regarding uncertain covariates can help improve the robust solution. In addition, by limiting ourselves in a one-period problem, we are able to provide exact solutions instead of resorting to approximation schemes that are commonly used in dynamic models.

While our problem is framed in a vehicle allocation context, it essentially originates from the stochastic transportation problem first studied by Williams (1963), which can also be applied in many other contexts. For example, one can interpret the supply nodes as warehouses and the demand nodes as retail stores. The problem is then to allocate inventories to the retail stores proactively to minimize expected lost sales and transportation costs. Many dynamic multiechelon inventory models (for instance, Federgruen and Zipkin 1984) are built upon this single-period problem. Computationally, including Williams (1963), various efficient algorithms are proposed to solve the stochastic transportation problem (see, for example, Holmberg and Joernsten 1984, Qi 1985). We remark that our distributionally robust counterpart can also be solved very efficiently via standard commercial solvers.

3. Vehicle Allocation Problem

We consider an operator providing service to an urban area that is further partitioned into well-defined regions. During the rush hours, e.g., 8:00 AM to 8:15 AM, the operator allocates idle vehicles (i.e., the oversupply) from N supply regions to M demand regions with supply shortage, which is defined as the extra number of vehicles needed to meet the demand in that region. Throughout the study, the terms "demand" and "supply shortage" are used interchangeably in the vehicle allocation context. The network structure of the problem is illustrated in Figure 2, where we use "supply nodes" to denote the regions

with an oversupply of idle vehicles and "demand nodes" to denote the regions with a supply shortage.

Let S_i be the number of idle vehicles at supply node i, $i \in [N]$, where $[N]$ is the set of the running indices, that is, $\{1,\ldots,N\}$. The shortages are denoted by a random vector $\tilde{\mathbf{z}} = (\tilde{z}_j)$, where $\tilde{z}_j$ is the shortage at node j for $j \in [M]$. There are also uncertain covariates that can potentially correlate with the demand, denoted by an I-dimensional random covariate vector $\tilde{\mathbf{v}} \in \mathbb{R}^I$. We denote the joint demand and covariate information as $(\tilde{\mathbf{z}}, \tilde{\mathbf{v}}) \in \mathbb{R}^M \times \mathbb{R}^I$.

For each vehicle allocated to demand node j, the operator either earns r_j if the vehicle successfully picks up a passenger, or 0 otherwise. Meanwhile, it costs the operator w_{ij} to allocate a vehicle from supply node i to demand node j. In practice, r_j can be interpreted as the share of the profit that the operator receives from the customer trip, and w_{ij} can be the compensation fee that the operator provides to the allocated vehicles for their relocations.

Suppose the operator allocates x_{ij} vehicles from supply node i to demand node j, at the total allocation costs $\sum_{j\in[M]}\sum_{i\in[N]} w_{ij}x_{ij}$. After the random demand realizes, the operator collects $\sum_{j\in[M]} r_j(\tilde{z}_j \wedge \sum_{i\in[N]} x_{ij})$ revenue, where $a \wedge b$ denotes $\min(a,b)$. Let $\mathcal{P}_0(\mathbb{R}^M \times \mathbb{R}^I)$ be the set of all distributions of a random vector with dimension $M+I$. If the operator has perfect knowledge of the joint distribution of $(\tilde{\mathbf{z}}, \tilde{\mathbf{v}})$, say, $\mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^M \times \mathbb{R}^I)$, then we can formulate the vehicle allocation problem as the following stochastic transportation problem:

$$\begin{aligned} \max_{x_{ij}\geq 0} \; & -\sum_{j\in[M]}\sum_{i\in[N]} w_{ij}x_{ij} + \mathbb{E}_{\mathbb{Q}}\left[\sum_{j\in[M]} r_j(\tilde{z}_j \wedge \sum_{i\in[N]} x_{ij})\right] \\ \text{s.t.} \; & \sum_{j\in[M]} x_{ij} \leq S_i, \forall i \in [N]. \end{aligned} \tag{1}$$

Let $\mathbb{P} = \Pi_{\tilde{\mathbf{z}}}\mathbb{Q}$, where $\Pi_{\tilde{\mathbf{z}}}\mathbb{Q}$ denotes the marginal distribution of $\tilde{\mathbf{z}}$ under $\mathbb{Q}$. Note that the objective function in problem (1) is independent of the uncertain covariates $\tilde{\mathbf{v}}$. Therefore, problem (1) is equivalent to

$$\begin{aligned} \max_{x_{ij}\geq 0} \; & -\sum_{j\in[M]}\sum_{i\in[N]} w_{ij}x_{ij} + \mathbb{E}_{\mathbb{P}}\left[\sum_{j\in[M]} r_j(\tilde{z}_j \wedge \sum_{i\in[N]} x_{ij})\right] \\ \text{s.t.} \; & \sum_{j\in[M]} x_{ij} \leq S_i, \forall i \in [N], \end{aligned} \tag{2}$$

where the expectation is taken over $\mathbb{P}$, the marginal distribution of $\tilde{\mathbf{z}}$. The equivalence of problems (1) and (2) shows that the distributional information of uncertain covariates brings no benefits to the stochastic optimization formulation (1). However, this observation is no longer true if the marginal distribution $\mathbb{P}$ is not known perfectly as we will show in section 4.

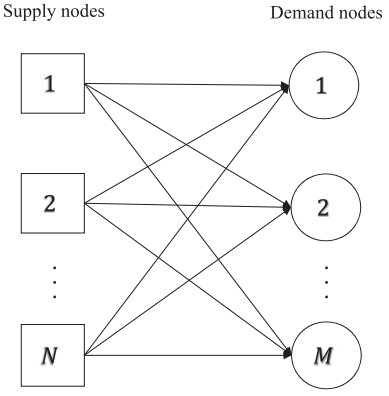
The stochastic optimization formulation (1) is the classic stochastic transportation problem first studied by Williams (1963). It considers the problem of sending a commodity from N supply nodes to M destinations, in face of uncertain demands. When $N = 1$, (1) reduces to the classic multi-item newsvendor problem with capacity constraint (Erlebacher 2000). The literature on stochastic transportation problem develops many efficient algorithms (see, for example, Holmberg and Joernsten 1984, Qi 1985) in solving problem (1) based on the assumption that $\mathbb{Q}$ or $\mathbb{P}$ is known perfectly. Yet, in practice, such information needs to be estimated from historical data. One commonly used approach is the SAA.

Suppose there are T samples of historical demand and covariate observations, that is, $\mathcal{T} = \{(\hat{\mathbf{z}}_1, \hat{\mathbf{v}}_1), (\hat{\mathbf{z}}_2, \hat{\mathbf{v}}_2), \ldots, (\hat{\mathbf{z}}_T, \hat{\mathbf{v}}_T)\}$. The SAA approach approximates the distribution $\mathbb{Q}$ using the empirical distribution where each sample $(\hat{\mathbf{z}}_t, \hat{\mathbf{v}}_t)$ has equal probability $\frac{1}{T}$. Correspondingly, problem (1) is approximated by

$$\begin{aligned} \Pi^{\text{SAA}} = \max_{x_{ij}\geq 0} \; & -\sum_{j\in[M]}\sum_{i\in[N]} w_{ij}x_{ij} + \frac{1}{T}\sum_{t\in[T]}\sum_{j\in[M]} r_j(\hat{z}_{jt} \wedge \sum_{i\in[N]} x_{ij}) \\ \text{s.t.} \; & \sum_{j\in[M]} x_{ij} \leq S_i, \forall i \in [N], \end{aligned} \tag{3}$$

where $\hat{z}_{jt}$ is the j-th component of $\hat{\mathbf{z}}_t$. The SAA formulation (3) can be solved by reformulating it into a linear program, whose detailed formulation is provided in Online Appendix A. Nevertheless, the empirical distribution may suffer from the issue of overfitting, especially when the underlying random vector is high dimensional and there is non-stationarity in the historical data. Existing computational evidence revealed that if the out-of-sample distribution deviates from the true distribution, the optimal

Figure 2 Illustration of the Vehicle Allocation Problem



solution from the SAA approach may perform poorly (see p.165 in Bertsimas and Thiele 2006). In contrast, "in many instances, only partial information about this distribution is available or reliable, such as means, variances, and covariances" (p. 98 in Popescu 2007). From the trip and vehicle status data sponsored by our partner taxi operator, we have a similar observation as the above literature—the mean is more stable and can be more accurately estimated than the distribution. Figure 3 shows the histograms of daily demand during morning rush hours in Robertson Quay, Singapore, for February and May in 2017 respectively. We note that the demand patterns (distribution shapes) for the same region in different months vary significantly.[1]

Consequently, from the operator's perspective, the true distribution is ambiguous and only belongs to a set of distributions that share consistent partial distributional information. To address the distributional ambiguity in our problem, we employ the framework of distributionally robust optimization and incorporate covariate information into an innovative scenario-wise ambiguity set, which is detailed in the following sections.

4. Robust Vehicle Allocation with Uncertain Covariates

Different from stochastic optimization, the framework of distributionally robust optimization assumes that the true distribution of $(\tilde{\mathbf{z}}, \tilde{\mathbf{v}})$: $\mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^M \times \mathbb{R}^I)$ lies in a certain ambiguity set $\mathbb{F} \subseteq \mathcal{P}_0(\mathbb{R}^M \times \mathbb{R}^I)$ that is characterized by the partial distributional information estimated from data rather than being known exactly. Similar to $\Pi_{\tilde{\mathbf{z}}}\mathbb{Q}$, the marginal distribution of $\tilde{\mathbf{z}}$ under $\mathbb{Q}$, we use $\Pi_{\tilde{\mathbf{z}}}\mathbb{F}$ to denote the set of all marginal distributions of $\tilde{\mathbf{z}}$ for all the distributions in $\mathbb{F}$, that is, $\Pi_{\tilde{\mathbf{z}}}\mathbb{F} = \cup_{\mathbb{Q}\in\mathbb{F}}\{\Pi_{\tilde{\mathbf{z}}}\mathbb{Q}\}$. Under this framework, the operator aims to maximize the worst-case expected profit over all possible distributions in $\mathbb{F}$. That is,

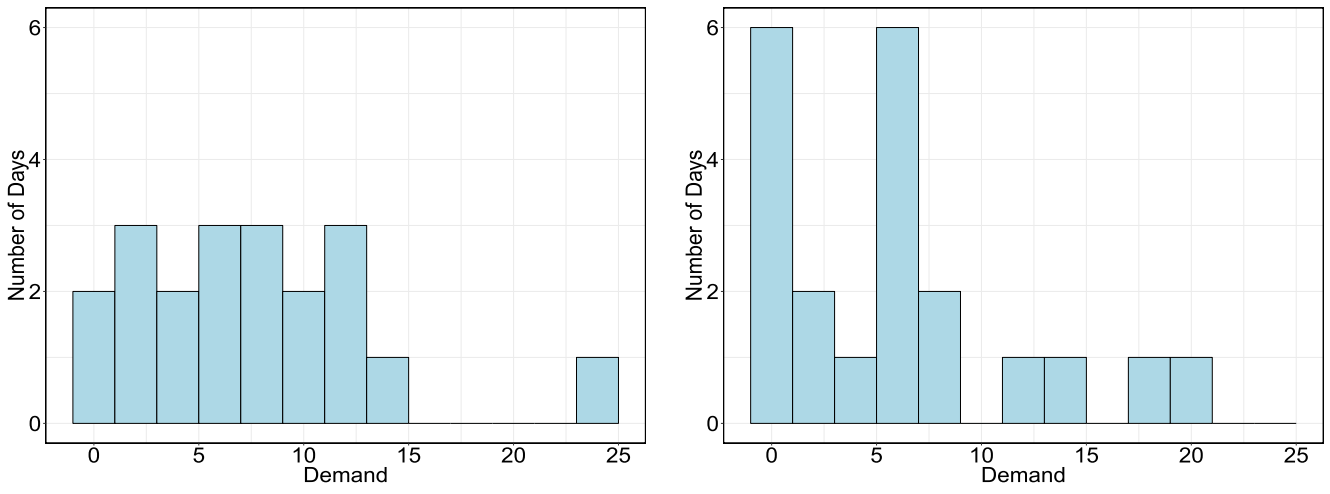
$$\begin{aligned} \max_{x_{ij}\geq 0} \; & -\sum_{j\in[M]}\sum_{i\in[N]} w_{ij}x_{ij} + \inf_{\mathbb{Q}\in\mathbb{F}} \mathbb{E}_{\mathbb{Q}}\left[\sum_{j\in[M]} r_j(\tilde{z}_j \wedge \sum_{i\in[N]} x_{ij})\right] \\ \text{s.t.} \; & \sum_{j\in[M]} x_{ij} \leq S_i, \forall i \in [N]. \end{aligned} \tag{4}$$

Both the solution of problem (4) and its out-of-sample performance depend on the ambiguity set $\mathbb{F}$. In the following, we first introduce the scenario-wise ambiguity set that incorporates the distributional information on both uncertain demands and uncertain covariates. We then illustrate the value of covariate information and how to estimate the scenario-wise ambiguity set based on historical data.

4.1. Scenario-wise Ambiguity Set with Covariate Information

We propose to jointly use covariate information and demand information to construct a scenario-wise ambiguity set. Specifically, to utilize the covariate information, we can partition all the possible realizations of $\tilde{\mathbf{v}}$, denoted by Ω, into L scenarios: Ω_l, $l \in [L]$ with $\Omega_l \cap \Omega_k = \emptyset$ for all $l,k \in [L], l \neq k$ and $\cup_{l\in[L]}\Omega_l = \Omega$. Let p_l denote the probability of scenario l happening.

Figure 3 Demand Pattern of Robertson Quay during Morning Rush hour in Feb. and May 2017 [Color figure can be viewed at wileyonlinelibrary.com]



(a) Demand histogram in Feb. 2017

(b) Demand histogram in May 2017

By definition, we have $\mathbb{Q}(\tilde{\mathbf{v}} \in \Omega_l) = p_l$ and $\sum_{l\in[L]} p_l = 1$. Combining all the scenarios with the marginal moment information of demand, we construct the following scenario-wise ambiguity set:

$$\bar{\mathbb{F}} = \left\{ \mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^M \times \mathbb{R}^I) \middle| \begin{array}{ll} (\tilde{\mathbf{z}}, \tilde{\mathbf{v}}) \sim \mathbb{Q} & \\ \mathbb{E}_{\mathbb{Q}}(\tilde{\mathbf{z}} \mid \tilde{\mathbf{v}} \in \Omega_l) = \boldsymbol{\mu}_l, & \forall l \in [L] \\ \mathbb{E}_{\mathbb{Q}}((\tilde{z}_{jl} - \mu_{jl})^2 \mid \tilde{\mathbf{v}} \in \Omega_l) \le \sigma_{jl}^2, & \forall l \in [L], j \in [M] \\ \mathbb{Q}(\tilde{\mathbf{v}} \in \Omega_l) = p_l, & \forall l \in [L] \\ \mathbb{Q}(\tilde{\mathbf{z}} \in \mathcal{Z}_l \mid \tilde{\mathbf{v}} \in \Omega_l) = 1, & \forall l \in [L] \end{array} \right\},$$

where $\mathcal{Z}_l = [\underline{\mathbf{z}}_l, \bar{\mathbf{z}}_l]$.[2]

In $\bar{\mathbb{F}}$, the first set of equality constraints specify the first marginal moment information (or mean) of demand conditional on $\tilde{\mathbf{v}}$ belonging to each scenario l. The second set of inequality constraints provide upper bounds on the variance of demand conditional on $\tilde{\mathbf{v}}$ belonging to each scenario l. The third set of equality constraints specify the probability of each scenario l, and the last set of equality constraints specify the maximum and minimum values $\tilde{z}_{jl}$ can take in each scenario l.

The scenario-wise ambiguity set $\bar{\mathbb{F}}$ is in the same format as the one proposed in Chen et al. (2018). However, in Chen et al. (2018), the random variable $\tilde{\mathbf{v}}$ is not interpreted as covariate information but rather as generic random scenarios. They then discuss how $\bar{\mathbb{F}}$ can be used to model the mixture distribution ambiguity set (where $\tilde{\mathbf{v}}$ represents the index of the distribution in the mixture), K-means clustering ambiguity set (where $\tilde{\mathbf{v}}$ represents the index of the cluster) and the Wasserstein ambiguity set (where $\tilde{\mathbf{v}}$ represents the support of the reference distribution). Hence, we add a new interpretation to the generic ambiguity set proposed in Chen et al. (2018).[3] More importantly, in section 4.3 below, we address the issue of given L, how one may utilize the historical data on demands and covariates to estimate the parameters in the scenario-wise ambiguity set $\bar{\mathbb{F}}$, including $\boldsymbol{\Omega}_l$ and p_l. Finally, we theoretically establish the value of utilizing covariate information (see Proposition 1 below) and perform extensive numerical comparisons with the ambiguity set that ignores covariate information (see section 6) as well.

We highlight here the flexibility of the scenario-wise ambiguity set $\bar{\mathbb{F}}$ in terms of modeling. Given T samples of demand and covariate: $\{(\hat{\mathbf{z}}_1, \hat{\mathbf{v}}_1), (\hat{\mathbf{z}}_2, \hat{\mathbf{v}}_2), \ldots, (\hat{\mathbf{z}}_T, \hat{\mathbf{v}}_T)\}$, by enforcing $L = T$, we have $\hat{\boldsymbol{\Omega}}_l = \{\hat{\mathbf{v}}_l\}$, $\hat{\boldsymbol{\mu}}_l = \hat{\mathbf{z}}_l$, $\hat{\sigma}_{jl}^2 = 0$ for all $j \in [M]$, $l \in [L]$ and $\hat{p}_l = \frac{1}{T}$, $l \in [L]$. That is, $\Pi_{\tilde{\mathbf{z}}}\bar{\mathbb{F}}$ contains only the empirical distribution of $\tilde{\mathbf{z}}$ and the robust problem (4) reduces to the SAA formulation in problem (3). By enforcing $L = 1$, it is easy to see that the ambiguity set $\bar{\mathbb{F}}$ reduces to the classical marginal moment ambiguity set that ignores the covariate information, which we denote by $\hat{\mathbb{F}}$, that is,

$$\hat{\mathbb{F}} = \left\{ \mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^M \times \mathbb{R}^I) \middle| \begin{array}{l} (\tilde{\mathbf{z}}, \tilde{\mathbf{v}}) \sim \mathbb{Q} \\ \mathbb{E}_{\mathbb{Q}}(\tilde{\mathbf{z}}) = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{Q}}(\tilde{z}_j - \mu_j)^2 \le \sigma_j^2, \forall j \in [M] \\ \mathbb{Q}(\tilde{\mathbf{z}} \in \mathcal{Z}) = 1 \end{array} \right\},$$

where $\mathcal{Z} = [\underline{\mathbf{z}}, \bar{\mathbf{z}}]$. The first set of equality constraints specify the first marginal moment information; the second set of inequality constraints provide upper bounds on the variance of the demand, and the last set of equality constraints specify the maximum and minimum values $\tilde{z}_j$ can take. The marginal moment ambiguity set $\hat{\mathbb{F}}$ is first proposed in the seminal paper of Scarf (1958), after which it has been widely adopted in the literature (e.g., Mak et al. 2014 for appointment scheduling problem and Wang and Zhang 2015 for process flexibility problem) due to its simplicity and computational tractability. Note that $\hat{\mathbb{F}}$ does not utilize any information on the uncertain covariates $\tilde{\mathbf{v}}$. While this is not an issue if we know the marginal distribution $\mathbb{P}$ perfectly (or equivalently $\Pi_{\tilde{\mathbf{z}}}\mathbb{F} = \{\mathbb{P}\}$), we illustrate how covariate information can bring value when the distribution is not known exactly below.

4.2. Value of Covariates

We first use a simple example to illustrate the benefit of $\bar{\mathbb{F}}$ that utilizes covariate information compared to $\hat{\mathbb{F}}$ that ignores it.

Example 1. Consider the case when both $\tilde{\mathbf{z}}$ and $\tilde{\mathbf{v}}$ are one dimensional, denoted as $(\tilde{z}, \tilde{v})$. Suppose the underlying true model is as follows: $\tilde{v}$ follows a Bernoulli distribution with $\mathbb{Q}(\tilde{v} = 1) = 1/2$. Conditioning on $\tilde{v}$, we have $\mathbb{Q}(\tilde{z} = z_1 | \tilde{v} = 1) = 1$, $\mathbb{Q}(\tilde{z} = z_0 | \tilde{v} = 0) = 1$, where $z_0 \le z_1$. One can interpret, for example, $\tilde{v} = 1$ as the event that the weather is rainy. Correspondingly, $\mathbb{Q}(\tilde{z} = z_1 | \tilde{v} = 1) = 1$ implies that we will see a high demand z_1 on a rainy day. Suppose further that we have access to the distribution of $\tilde{v}$ but not that of $\tilde{z}$, and given a realization of $\tilde{v}$ we know the exact mean and variance of $\tilde{z}$. Without utilizing covariate information, we can write the ambiguity set $\hat{\mathbb{F}}$ using the mean and variance of $\tilde{z}$ as:

$$\hat{\mathbb{F}} = \left\{ \mathbb{Q} \in \mathcal{P}_0(\mathbb{R} \times \mathbb{R}) \middle| \begin{array}{l} (\tilde{z}, \tilde{v}) \sim \mathbb{Q} \\ \mathbb{E}_{\mathbb{Q}}(\tilde{z}) = \frac{z_1 + z_0}{2} \\ \mathbb{E}_{\mathbb{Q}}(\tilde{z} - \frac{z_1 + z_0}{2})^2 \le (\frac{z_1 - z_0}{2})^2 \\ \mathbb{Q}(\tilde{z} \in \mathbb{R}) = 1 \end{array} \right\},$$

which contains many possible distributions. However, once conditioning on the values of $\tilde{v}$, we can formulate the following ambiguity set:

$$\bar{\mathbb{F}} = \left\{ \mathbb{Q} \in \mathcal{P}_0(\mathbb{R}\times\mathbb{R}) \;\middle|\; \begin{array}{l} (\tilde{z},\tilde{v}) \sim \mathbb{Q} \\ \mathbb{E}_{\mathbb{Q}}(\tilde{z} \mid \tilde{v} = i) = z_i, \text{ for } i = 0,1 \\ \mathbb{E}_{\mathbb{Q}}((\tilde{z}-z_i)^2 \mid \tilde{v} = i) \le 0, \text{ for } i = 0,1 \\ \mathbb{Q}(\tilde{v} = i) = \frac{1}{2}, \text{ for } i = 0,1 \\ \mathbb{Q}(\tilde{z} \in \mathbb{R} \mid \tilde{v} = i) = 1, \text{ for } i = 0,1 \end{array} \right\}.$$

Note that, in $\bar{\mathbb{F}}$, conditioning on $\tilde{v} = i$, for $i = 0,1$, the mean and variance of $\tilde{z}$ are given by $\mathbb{E}_{\mathbb{Q}}(\tilde{z}|\tilde{v} = i) = z_i$ and $\mathbb{E}_{\mathbb{Q}}((\tilde{z} - z_i)^2|\tilde{v} = i) \le 0$ due to $\mathbb{Q}(\tilde{z} = z_i|\tilde{v} = 1) = 1$. Consequently, the ambiguity set $\bar{\mathbb{F}}$ contains only a single distribution, that is, $\mathbb{Q}(\tilde{z} = z_0, \tilde{v} = 0) = \frac{1}{2}$, $\mathbb{Q}(\tilde{z} = z_1, \tilde{v} = 1) = \frac{1}{2}$, which is the true distribution.

The above example implies the covariates can help us obtain a less conservative ambiguity set. Let MMM and SDR denote the distributionally robust optimization model with $\mathbb{F} = \hat{\mathbb{F}}$ and $\mathbb{F} = \bar{\mathbb{F}}$, and let Π^{MMM} and Π^{SDR} denote the corresponding optimal objective values, respectively. We demonstrate the benefit of considering covariate information in Proposition 1.

Proposition 1. *Given ambiguity sets $\hat{\mathbb{F}}$ and $\bar{\mathbb{F}}$, suppose $\boldsymbol{\mu} = \sum_{l\in[L]} \boldsymbol{\mu}_l p_l$, $\sigma_j^2 + \mu_j^2 = \sum_{l\in[L]}(\sigma_{jl}^2 + \mu_{jl}^2)p_l$, $\forall j \in [M]$, and $\cup_{l\in[L]}\mathcal{Z}_l \subseteq \mathcal{Z}$, we then have $\Pi^{SDR} \ge \Pi^{MMM}$.*

Proof. Please see Appendix B.1 in the Online Supplement.

The conditions in Proposition 1, that is, $\boldsymbol{\mu} = \sum_{l\in[L]} \boldsymbol{\mu}_l p_l$, $\sigma_j^2 + \mu_j^2 = \sum_{l\in[L]}(\sigma_{jl}^2 + \mu_{jl}^2)p_l$, $\forall j \in [M]$ and $\cup_{l\in[L]}\mathcal{Z}_l \subseteq \mathcal{Z}$, essentially requires the consistency in moment specification for both ambiguity sets $\hat{\mathbb{F}}$ and $\bar{\mathbb{F}}$. For instance, the ambiguity sets $\hat{\mathbb{F}}$ and $\bar{\mathbb{F}}$ in Example 1 are consistent with each other since it is easy to verify that $\mu = \frac{z_0+z_1}{2} = \frac{1}{2}(\mu_0 + \mu_1)$ and $\sigma^2 + \mu^2 = \frac{z_0^2+z_1^2}{2} = \frac{1}{2}(\mu_0^2 + \sigma_0^2) + \frac{1}{2}(\mu_1^2 + \sigma_1^2)$. In fact, those conditions are very mild in that they naturally hold when the mean, variance, and scenarios are estimated from the same dataset (see more discussion in section 4.3 below). With the consistency guaranteed, Proposition 1 shows that with covariate information, the ambiguity set $\bar{\mathbb{F}}$ results in a less conservative solution as compared to $\hat{\mathbb{F}}$.

4.3. Estimation of Ambiguity Set $\bar{\mathbb{F}}$

For a given L,[4] the key in estimating the parameters in the ambiguity set $\bar{\mathbb{F}}$ is to construct the partition $\Omega_l, l \in [L]$ in a way such that the resulting $\tilde{\mathbf{z}}$ conditioning on $\tilde{\mathbf{v}} \in \Omega_l$ has minimum variance (recall Example 1 at the beginning of section 4.1). For this purpose, we use the regression tree, an intuitive and widely used non-parametric method, to arrive at a partition. The univariate regression tree (URT) is popular in handling a one-dimensional-dependent variable. However, in our vehicle allocation problem, the dependent variable $\tilde{\mathbf{z}}$ is a M-dimensional vector. Therefore, we adopt the multivariate regression tree (MRT), which is a natural extension of URT in predicting the multivariate dependent variable. In the following, we briefly introduce URT and MRT (for more details, please refer to Breiman et al. 1984, De'Ath and Fabricius 2000 for URT, De'Ath 2002 for MRT).

Tree models are popular in supervised learning such as classification and regression. In these models, the tree structure is specified by a set of splitting rules (hyperplanes) that lead to a partition of the covariate space. More specifically, a tree is formed by recursively splitting Ω and its subsets into two further subsets. For a resulting tree with L leaf nodes, each leaf node represents a subset $\Omega_l, l \in [L]$, which together form a partition of Ω. The dataset $\mathcal{T} = \{(\hat{z}_1, \hat{\mathbf{v}}_1), (\hat{z}_2, \hat{\mathbf{v}}_2), \ldots, (\hat{z}_T, \hat{\mathbf{v}}_T)\}$ then falls into L labeled clusters according to the tree, where each cluster l can be denoted as $\mathcal{K}_l = \{t|\hat{\mathbf{v}}_t \in \Omega_l\}$, that is, the index set of the data points whose covariates belong to Ω_l.

In URT, the impurity of a partition, denoted as E^{URT}, is then defined as the sum of squared errors of the univariate dependent variable, that is, $E^{\text{URT}} = \sum_{l\in[L]}\sum_{t\in\mathcal{K}_l}(\hat{z}_t - \mu_l)^2$, where $\mu_l = \frac{1}{|\mathcal{K}_l|}\sum_{t\in\mathcal{K}_l}\hat{z}_t$ and $|\mathcal{K}_l|$ is the cardinality of $\mathcal{K}_l$. The URT then aims to minimize E^{URT} by choosing a tree with L leaf nodes:

$$\min_{\text{Tree with L Leaf Nodes}} \quad \sum_{l\in[L]}\sum_{t\in\mathcal{K}_l}(\hat{z}_t - \mu_l)^2 \tag{5}$$

MRT simply extends the impurity function in URT to the total sum of squared errors of the multivariate dependent variable. That is, with the samples $(\hat{\mathbf{z}}_1, \hat{\mathbf{v}}_1)$, $(\hat{\mathbf{z}}_2, \hat{\mathbf{v}}_2), \ldots, (\hat{\mathbf{z}}_T, \hat{\mathbf{v}}_T)$, the impurity function becomes $E^{\text{MRT}} = \sum_{l\in[L]}\sum_{t\in\mathcal{K}_l}\|\hat{\mathbf{z}}_t - \boldsymbol{\mu}_l\|_2^2$, where $\|\cdot\|_2$ denotes the l_2-norm and $\boldsymbol{\mu}_l = \frac{1}{|\mathcal{K}_l|}\sum_{t\in\mathcal{K}_l}\hat{\mathbf{z}}_t$.

Many existing software packages including scikit-learn for Python (which we use in our case study) apply the top-down induction method as a heuristic to solve problem (5). Starting from the root set Ω, in each split, the heuristic myopically generates a hyperplane to minimize the sum of squared errors of the two resulting subsets, but ignores the possible impact of future splits in the tree. The heuristic terminates when it obtains L subsets.[5]

With MRT, the scenario probability p_l for scenario l can be estimated from the size of the leaf node as

$$\hat{p}_l = \frac{|\mathcal{K}_l|}{T},$$

Furthermore, the mean $\boldsymbol{\mu}_l$ and variance σ_{jl}^2 of demand for scenario l can be estimated as

$$\hat{\boldsymbol{\mu}}_l = \frac{1}{|\mathcal{K}_l|}\sum_{t\in\mathcal{K}_l}\hat{\mathbf{z}}_t,$$

and

$$\hat{\sigma}_{jl}^2 = \frac{1}{|\mathcal{K}_l|}\sum_{t\in\mathcal{K}_l}(\hat{z}_{jt}-\hat{\mu}_{jl})^2.$$

The set $\mathcal{Z}_l$ for scenario l can also be set as $\bar{z}_{jl} = \max_{t\in\mathcal{K}_l}\hat{z}_{jt}$, and $\underline{z}_{jl} = \min_{t\in\mathcal{K}_l}\hat{z}_{jt}$, $\forall j \in [M]$. Note that, with the above estimations, the conditions in Proposition 1 hold since $\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t\in[T]}\hat{\mathbf{z}}_t = \sum_{l\in[L]}\frac{|\mathcal{K}_l|}{T}\frac{1}{|\mathcal{K}_l|}\sum_{t\in\mathcal{K}_l}\hat{\mathbf{z}}_t = \sum_{l\in[L]}\hat{p}_l\hat{\boldsymbol{\mu}}_l$, $\hat{\sigma}_j^2+\hat{\mu}_j^2 = \frac{1}{T}\sum_{t\in[T]}\hat{z}_{jt}^2 = \sum_{l\in[L]}\hat{p}_l\frac{1}{|\mathcal{K}_l|}\sum_{t\in\mathcal{K}_l}\hat{z}_{jt}^2 = \sum_{l\in[L]}\hat{p}_l(\hat{\sigma}_{jl}^2+\hat{\mu}_{jl}^2)$. Also, $\cup_{l\in[L]}\mathcal{Z}_l \subseteq \mathcal{Z}$ since $\bar{z}_j \geq \bar{z}_{jl}$ and $\underline{z}_j \leq \underline{z}_{jl}$ for $l \in [L]$.

Finally, we remark that the regression tree is not the only method to obtain a partition. There are other unsupervised learning methods, e.g., *K*-means clustering, which can also be used to obtain the partition. However, we note that those unsupervised learning methods do not utilize the covariate information. We show in the following example that in some cases, with covariate information, the regression tree can provide a more accurate scenario classification compared to the unsupervised clustering methods that only cluster based on the dependent variables.

Example 2. Consider the one dimensional case where $(\tilde{z}, \tilde{v})$ follows the distribution: $\mathbb{Q}(\tilde{v}=1)=1/2$, $\mathbb{Q}(\tilde{v}=2)=1/2$, and conditioning on $\tilde{v}$, the probability of $\tilde{z}$ is $\mathbb{Q}(\tilde{z}=1|\tilde{v}=1)=\frac{2}{3}$, $\mathbb{Q}(\tilde{z}=3|\tilde{v}=1)=\frac{1}{3}$, $\mathbb{Q}(\tilde{z}=2|\tilde{v}=2)=\frac{1}{3}$, $\mathbb{Q}(\tilde{z}=4|\tilde{v}=2)=\frac{2}{3}$. The conditional distribution is summarized in Table 1 below.

Table 1 Description of Conditional Probability

$\tilde{z}$	$\tilde{v}$	$\mathbb{Q}(\tilde{z}\|\tilde{v})$
1	1	2/3
3		1/3
2	2	1/3
4		2/3

Suppose we obtain 6 samples of $(\tilde{z}, \tilde{v})$ from this distribution: (1,1), (1,1), (3,1), (2,2), (4,2), and (4,2). For $L = 2$, with regression tree, {(1,1), (1,1), (3,1)} is identified as scenario 1 with probability 0.5 and {(2,2), (4,2), (4,2)} is identified as scenario 2 with probability 0.5. In contrast, with *K*-means clustering that ignores the covariate information, one only utilizes the data points of $\tilde{z}$, that is, 1, 1, 2, 3, 4, and 4. Consequently, under *K*-means clustering with $K = 2$, {1, 1, 2} is clustered as scenario 1 with probability 0.5, and {3, 4, 4} is clustered as scenario 2 with probability 0.5.

Example 2 shows that the regression tree captures the accurate scenarios and the statistics of $\tilde{z}$ for each scenario while the *K*-means clustering method misspecifies both as it only relies on the distances between the values of $\tilde{z}$. The same issue exists for other popular unsupervised clustering methods, e.g., *K*-medoids clustering and hierarchical clustering.

Regression tree method also has the advantage of good interpretability, that is, it shows how the partition is connected to the covariates. Such interpretability also grants the scenario-wise ambiguity set the flexibility of incorporating forecast information, e.g., the probability of the covariates falling into scenario l, $\hat{p}_l$, can be obtained from expert forecast instead of using the historical frequency.

5. Solution Approach

In this section, we discuss how one can solve problem (4) with $\mathbb{F} = \bar{\mathbb{F}}$ in an efficient way. We first reformulate the ambiguity set $\bar{\mathbb{F}}$ into a standard form proposed in Wiesemann et al. (2014) by introducing auxiliary random variables $\tilde{\mathbf{u}}_l$ for $l \in [L]$:

$$\mathbb{G} = \left\{ \mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^M \times \mathbb{R}^M \times \mathbb{R}^I) \;\middle|\; \begin{array}{ll} (\tilde{\mathbf{z}}, \tilde{\mathbf{u}}, \tilde{\mathbf{v}}) \sim \mathbb{Q} & \\ \mathbb{E}_{\mathbb{Q}}(\tilde{\mathbf{z}} \mid \tilde{\mathbf{v}} \in \Omega_l) = \boldsymbol{\mu}_l, & \forall l \in [L] \\ \mathbb{E}_{\mathbb{Q}}(\tilde{u}_{jl} \mid \tilde{\mathbf{v}} \in \Omega_l) \leq \sigma_{jl}^2, & \forall l \in [L], j \in [M] \\ \mathbb{Q}(\tilde{\mathbf{v}} \in \Omega_l) = p_l, & \forall l \in [L] \\ \mathbb{Q}((\tilde{\mathbf{z}}, \tilde{\mathbf{u}}) \in \mathcal{W}_l \mid \tilde{\mathbf{v}} \in \Omega_l) = 1, & \forall l \in [L] \end{array} \right\},$$

where $\mathcal{W}_l$ is the "lifted support set" defined as

$$\mathcal{W}_l = \left\{ (\mathbf{z}, \mathbf{u}) \in \mathbb{R}^M \times \mathbb{R}^M | \mathbf{z} \in \mathcal{Z}_l, (z_{jl} - \mu_{jl})^2 \leq u_{jl}, \forall j \in [M] \right\}.$$

Compared to $\bar{\mathbb{F}}$, the terms inside the expectation constraints in $\mathbb{G}$ are all linear functions of the random vectors and the nonlinearities are encoded in the set $\mathcal{W}_l$. Using standard duality argument, one can reformulate problem (4) with $\mathbb{F} = \bar{\mathbb{F}}$ into the following semi-infinite program in Lemma 1, which can then be reformulated as a second-order cone program (SOCP).

Lemma 1. *The SDR model, that is, problem* (4) *with* $\mathbb{F} = \bar{\mathbb{F}}$, *is equivalent to the following optimization problem with the lifted ambiguity set* $\mathbb{G}$ *and lifted support set* $\mathcal{W}_l$*:*

$$\Pi^{\text{SDR}} = \max_{\substack{x_{ij},\gamma_l \\ \boldsymbol{\alpha},\boldsymbol{\delta}_l}} \left\{ \sum_{j\in[M]} \sum_{i\in[N]} (r_j - w_{ij}) x_{ij} - \sum_{l\in[L]} \alpha_l - \sum_{l\in[L]} \left(\boldsymbol{\delta}_l' \boldsymbol{\mu}_l + \sum_{j\in[M]} \gamma_{jl} \sigma_{jl}^2 \right) \right\} \tag{6}$$

$$\text{s.t.} \quad \sum_{j\in[M]} x_{ij} \le S_i, \forall i \in [N] \tag{7}$$

$$\begin{aligned} &\alpha_l + \boldsymbol{\delta}_l' \mathbf{z} + \boldsymbol{\gamma}_l' \mathbf{u} \ge p_l \sum_{j\in[M]} r_j \Big(\sum_{i\in[N]} x_{ij} - z_{jl} \Big)^+, \\ &\forall (\mathbf{z},\mathbf{u}) \in \mathcal{W}_l, l \in [L] \\ &x_{ij} \in \mathbb{R}_+, \boldsymbol{\alpha} \in \mathbb{R}^L, \boldsymbol{\delta}_l \in \mathbb{R}^M, \boldsymbol{\gamma}_l \in \mathbb{R}_+^M, \forall l \in [L] \end{aligned} \tag{8}$$

Proof. *Please see Appendix B.2 in the Online Supplement.*

The formulation in problem (6) is not yet directly solvable, due to the infinite number of constraints as well as the nonlinear terms $(\sum_{i\in[N]} x_{ij} - z_{jl})^+$ in constraints (8). We note that the nonlinear constraints can be replaced by the following sets of linear constraints:

$$\begin{aligned} \alpha_l + \boldsymbol{\delta}_l' \mathbf{z} + \boldsymbol{\gamma}_l' \mathbf{u} \ge p_l \sum_{j\in\mathcal{M}} r_j \Big(\sum_{i\in[N]} x_{ij} - z_{jl} \Big), \forall (\mathbf{z},\mathbf{u}) \\ \in \mathcal{W}_l, l \in [L], \mathcal{M} \in \mathcal{P}(M), \end{aligned} \tag{9}$$

where $\mathcal{P}(M)$ is the power set of $[M]$. The infinite number of constraints, that is, the inequality in constraints (9) is required to hold for any $(\mathbf{z},\mathbf{u}) \in \mathcal{W}_l$, can also be transformed into a finite number of second-order cone constraints via standard techniques in the literature. We remark that even though the number of the corresponding second-order cone constraints is finite, it grows exponentially with the number of demand regions due to the fact that the number of constraints in (9) is exponential in M. Hence, the standard SOCP formulation cannot be tractably solved for a problem of practical size (for the specific problem that our industry partner is interested in, the number of demand regions can be more than 10). In the following subsection, we develop an alternative reformulation that gets around this difficulty.

5.1. Linear Decision Rule Approximation

The main source of computational complexity in our formulation above comes from the step in representing the nonlinear constraint (8) as an exponential number of linear constraints (9). Instead, our idea here is to linearize the nonlinear term $(\sum_{i\in[N]} x_{ij} - z_{jl})^+$. Specifically, for all $l \in [L]$, $j \in [M]$, we replace $(\sum_{i\in[N]} x_{ij} - z_{jl})^+$ in constraint (8) by $y_j^l(\mathbf{z},\mathbf{u})$ and impose the constraint $y_j^l(\mathbf{z},\mathbf{u}) \ge (\sum_{i\in[N]} x_{ij} - z_{jl})^+$, where $y_j^l(\mathbf{z},\mathbf{u}) = y_j^{0l} + \sum_{k\in[M]} y_{jk}^{1l} z_{kl} + \sum_{k\in[M]} y_{jk}^{2l} u_{kl}$, with the coefficients y_j^{0l}, y_{jk}^{1l} and y_{jk}^{2l} being our new decision variables. More compactly, we can write this linearization step as $\mathbf{y}(\cdot) = (y_j^l(\cdot)) \in \mathcal{L}$, where

$$\mathcal{L} = \left\{ \mathbf{y}(\cdot) : \mathbb{R}^M \times \mathbb{R}^M \mapsto \mathbb{R}^{M\times L} \;\middle|\; \begin{array}{l} \text{for all } j \in [M], l \in [L]: \\ y_j^l(\mathbf{z},\mathbf{u}) = y_j^{0l} + \sum_{k\in[M]} y_{jk}^{1l} z_{kl} + \sum_{k\in[M]} y_{jk}^{2l} u_{kl} \\ \text{for some } y_j^{0l}, y_{jk}^{1l}, y_{jk}^{2l} \in \mathbb{R}, j \in [M], l \in [L] \end{array} \right\}.$$

In other words, the constraint $\mathbf{y}(\cdot) \in \mathcal{L}$ restricts the mapping $\mathbf{y}(\cdot)$ to be affine. The semi-infinite nonlinear program (6) can then be approximated by the following infinite linear program:

$$\begin{aligned} \hat{\Pi} = \max_{\substack{x_{ij},\gamma_l \\ \boldsymbol{\alpha},\boldsymbol{\delta}_l, y_j^l(\cdot)}} & \left\{ \sum_{j\in[M]} \sum_{i\in[N]} (r_j - w_{ij}) x_{ij} - \sum_{l\in[L]} \alpha_l - \sum_{l\in[L]} \left(\boldsymbol{\delta}_l' \boldsymbol{\mu}_l + \sum_{j\in[M]} \gamma_{jl} \sigma_{jl}^2 \right) \right\} \\ \text{s.t.} \quad & \sum_{j\in[M]} x_{ij} \le S_i, \forall i \in [N] \\ & \alpha_l + \boldsymbol{\delta}_l' \mathbf{z} + \boldsymbol{\gamma}_l' \mathbf{u} \ge p_l \sum_{j\in[M]} r_j y_j^l(\mathbf{z},\mathbf{u}), \forall (\mathbf{z},\mathbf{u}) \in \mathcal{W}_l, l \in [L] \\ & y_j^l(\mathbf{z},\mathbf{u}) \ge \sum_{i\in[N]} x_{ij} - z_{jl}, \forall (\mathbf{z},\mathbf{u}) \in \mathcal{W}_l, j \in [M], l \in [L] \\ & y_j^l(\mathbf{z},\mathbf{u}) \ge 0, \forall (\mathbf{z},\mathbf{u}) \in \mathcal{W}_l, j \in [M], l \in [L] \\ & \mathbf{y}(\cdot) \in \mathcal{L} \\ & x_{ij} \in \mathbb{R}_+, \boldsymbol{\alpha} \in \mathbb{R}^L, \boldsymbol{\delta}_l \in \mathbb{R}^M, \boldsymbol{\gamma}_l \in \mathbb{R}_+^M, \forall l \in [L] \end{aligned} \tag{10}$$

At the cost of introducing more decision variables (the linear approximation step adds $O(M^2L)$ more decision variables), the constraints in problem (10) are now all linear and for a fixed $(\mathbf{z},\mathbf{u}) \in \mathcal{W}_l$, the total number of constraints is now $O(M(L + N))$, which is polynomial in M. Applying the same standard approach we use in dealing with the infinite constraints in problem (6), the infinite linear

program (10) can be transformed into an SOCP with $O(M^2L + MN)$ number of decision variables and $O(M^2L + N)$ number of constraints, which can be solved very efficiently.

Interestingly, even though we have imposed restrictive constraints $\mathbf{y}(\cdot) \in \mathcal{L}$ in problem (10), it results in no loss of optimality as the following proposition shows.

PROPOSITION 2. *The linear decision rule approximation results in no loss of optimality, that is,* $\hat{\Pi} = \Pi^{SDR}$.

PROOF. *Please see Appendix B.3 in the Online Supplement.*

There are relatively scarce results on the optimality of linear decision rule approximation in the (distributionally) robust literature. With the ambiguity set only specifying support information, Bertsimas et al. (2010) and Iancu et al. (2013) show the optimality of linear decision rule under certain conditions. When the ambiguity set contains additional distributional information, e.g., mean and variance, Bertsimas et al. (2018) prove the optimality of linear decision rule for a two-stage problem with one-dimensional second stage decision, and He et al. (2020) further provide the sufficient conditions for its optimality for one particular two-stage problem with multidimensional recourse decisions. Note that in the case when $L = 1$, we can directly apply Proposition 4 of He et al. (2020) to show the optimality of the linear decision rule approximation for our problem. Proposition 2 shows that linear decision rule can be optimal for multidimensional recourse decisions with a scenarios-wise ambiguity set when $L > 1$.

Finally, note that in practice, one needs to obtain an integer solution to implement the allocation decision. One approach is to impose integer constraints on x_{ij} in problem (10), which results in a mixed-integer SOCP (MISOCP). Although the MISOCP can also be solved using existing commercial solvers, e.g., Gurobi, the drawback is that the computational complexity is much higher than that of the SOCP formulation, especially when the number of demand regions is high. Instead, we can directly round x_{ij} solved from the SOCP to the nearest integer to obtain the integer solution. We compare the performance of the heuristic rounding solution to the MISOCP solution in the case study of section 6.2. The results show that the optimality gap between them is negligible. In addition, the insights from the performance comparison among different models with continuous solutions continue to hold for the models with heuristic rounding solutions and the models with optimal integer solutions (please see Figure 1 and 2 in Online Appendix D for the detailed comparison).[6] Hence, we directly compare the numerical performance of different models with continuous allocation solutions below.

6. Numerical Studies

Proposition 1 in section 4.2 shows that the ambiguity set that incorporates covariate information $\bar{\mathbb{F}}$ (SDR) results in higher worst-case expected profit compared to $\hat{\mathbb{F}}$ which ignores such information (MMM). Note that the worst-case distribution that achieves Π^{SDR} is in general different from that achieves Π^{MMM}. It is unclear whether the same conclusion from Proposition 1 would hold when their respective solutions are evaluated in out-of-sample tests. It would also be interesting to compare the scenario-wise ambiguity set with the ambiguity set that contains more moment information, such as the one from Delage and Ye (2010) that specifies covariance information of demand (but still ignores covariate information). We denote the ambiguity set that contains both marginal moment and covariance information by $\mathbb{F}_\Sigma$ and use SDP to denote the robust vehicle allocation model with $\mathbb{F}_\Sigma$ (see Online Appendix C for the construction of $\mathbb{F}_\Sigma$ and the detailed formulation of the SDP model).[7] In addition, since SAA is also a widely popular approach in addressing problems with uncertainty, it is important to benchmark the performance of our SDR solution to SAA in both simulated environments and those cases generated by real data. We evaluate their model performances in this section by first conducting an extensive simulation study. We then use the trip and vehicle status data obtained from our partner taxi operator and the rainfall data collected from the Meteorological Service Singapore to test our solution in a case study.

In accordance to the practice of the taxi operator, we follow the parameter settings discussed below in both the simulation experiment and the case study. Suppose the average trip revenue from region j is $\hat{r}_j$ for each order. We assume the operator collects a fixed share of $\theta\hat{r}_j$, where $0 \le \theta \le 1$. For the allocation cost, we assume the operator pays b_j for each driver who is allocated to region j but fails to pick up a customer. In practice, we can set b_j as the online booking fee. When there is a successful pick-up, the booking fee will be paid by the passenger. In addition, the operator also pays an inconvenience cost $\hat{w}_{ij}$ for a driver allocated from region i to region j. In summary, the per-order average revenue and the allocation cost are set as $r_j = \theta\hat{r}_j + b_j$ and $w_{ij} = \hat{w}_{ij} + b_j$ respectively.

6.1. Simulation Experiment

Since our main motivation is to resolve the oversupply issue in the downtown area of Singapore (recall Figure 1), we focus on the case when there is only one supply region, and we consider the problem of allocating the supply to five demand regions. For notational convenience, we simplify S_i, x_{1j} and w_{1j} as S, x_j and w_j, respectively. We set $\hat{w}_{ij} = 0$ and enumerate θ from 0.01 to 0.1 with step 0.01. The per-order average revenue and booking fee are set as $\hat{r}_j = 12.5 - 0.5j, b_j = 3, j \in [5]$.

We prepare two demand datasets to examine the performance of the four models: the (in-sample) training data to derive the allocation decisions x_j, and the (out-of-sample) test data which is used to evaluate the out-of-sample performance of the decisions x_j. In both datasets, we set a one-dimensional covariate $\tilde{v} = 1, 2, 3, 4$ with equal probabilities and condition on $\tilde{v} = i$, demand samples are generated from truncated normal distributions with different means. In our implementation, instead of uniformly generating $\tilde{v}$ first and then sample demand conditioning on the value of $\tilde{v}$, we simply generate the samples under 4 scenarios with equal sample size. For the training data, when $\tilde{v} = 1$, the expected demand across all regions are $\mu_{11} = 150$, $\mu_{21} = 140$, $\mu_{31} = 130$, $\mu_{41} = 120$ and $\mu_{51} = 110$. For $\tilde{v} = 2, 3, 4$, we set $\mu_{jl} = \mu_{j1} - 20(l-1)$, $j \in [5]$, $l = 2,3,4$. We also vary the standard deviation of demand in each scenario as $\sigma_{jl} = q\mu_{il}$, where $q = 0.1,0.2,\ldots,0.5$. The demand pattern of the four scenarios used in the training data is illustrated in Figure 4.

In order to capture the possibility that the out-of-sample distribution may deviate from the in-sample distribution, we add a perturbation to the mean demand of the out-of-sample data: $\mu_{jl}^{\text{out}} = (1+\Delta)\mu_{jl}$ with $\sigma_{jl}^{\text{out}} = q\mu_{jl}^{\text{out}}$. We enumerate Δ, which denotes the deviation of out-of-sample distribution from the in-sample distribution, from -0.2 to 0.2 with step 0.04. We consider supply $S \in \{100,400,800\}$, to represent the low-, medium- and high-supply cases.

The combination of different values of θ, Δ, q and S results in a total of $10\times11\times5\times3 = 1650$ experimental instances. In each instance, we generate 80 training samples—20 demand samples for each $\tilde{v} = 1, 2, 3, 4$—which we use to determine the allocation decisions x_j under SDR, MMM, SDP, SAA, respectively. Meanwhile, we generate 20 test samples from the out-of-sample distribution according to the value of Δ, for the evaluation in each instance. We then calculate the average and standard deviation of the 20 out-of-sample profits for x_j from each method.

Note that even though the demand distribution is unknown to all the four models we are training, by using a discrete covariate $\tilde{v}$, we are implicitly assuming L and $\Omega_l, l \in [L]$ are known to our SDR model. That is, $L = 4$, $\Omega_l = \{l\}$, $p_l = \frac{1}{4}$ and we do not need to use multivariate regression tree to estimate these parameters for the simulation.

We report both the average and standard deviation of out-of-sample profits in Figure 5a and b (please refer to Table 1 in Online Appendix E.2 for a detailed summary statistics of a subset of experimental instances). We also provide Figure 5c to integrate the mean and standard deviation to further elaborate on the trade-offs of the four models.

First, note that on average SDP slightly outperforms MMM with 0.4%–2.5% higher mean of out-of-sample profits, while the difference between standard deviation is not significant (within 0.3%). The result shows that compared to the marginal moment information, incorporating covariance information via SDP can only slightly improve the performance over MMM. In addition, one drawback of SDP is that the computational complexity is much higher than that of all other three models because the number of the semi-definite constraints increases exponentially in the number of demand regions (see the SDP formulation in Online Appendix C).

Second, we can observe that on average SDR consistently outperforms MMM and SDP irrespective of the magnitude of perturbation in the out-of-sample distribution. This confirms that our observation in Proposition 1 is robust—covariate information is indeed helpful in arriving at a less conservative solution. To our surprise, the covariate information can also help in reducing the variance of the out-of-sample profits—the SDR consistently has the lowest variability in performance. The results also demonstrate that when the impact of covariates on demand is significant, the ambiguity set $\bar{\mathbb{F}}$ that is able to capture the nonlinear

Figure 4 Scenario Pattern of the Simulated Data

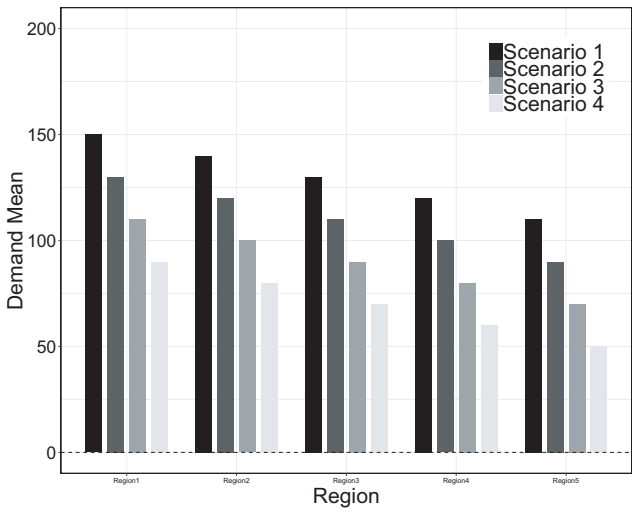
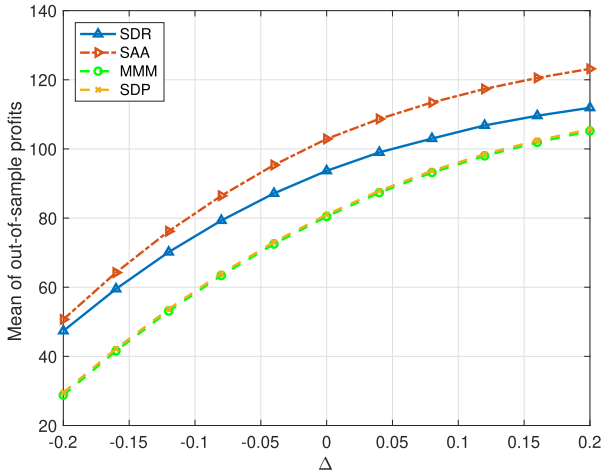
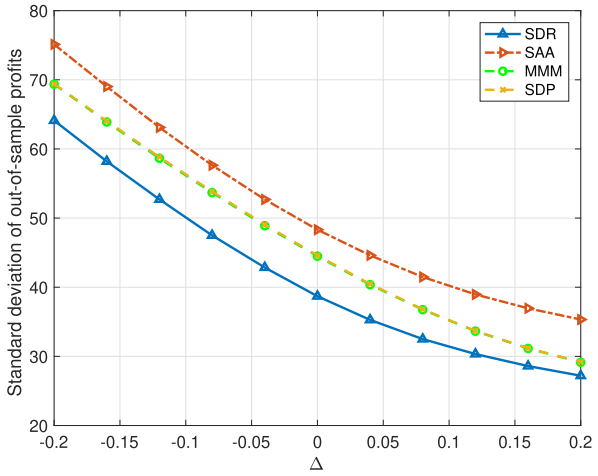


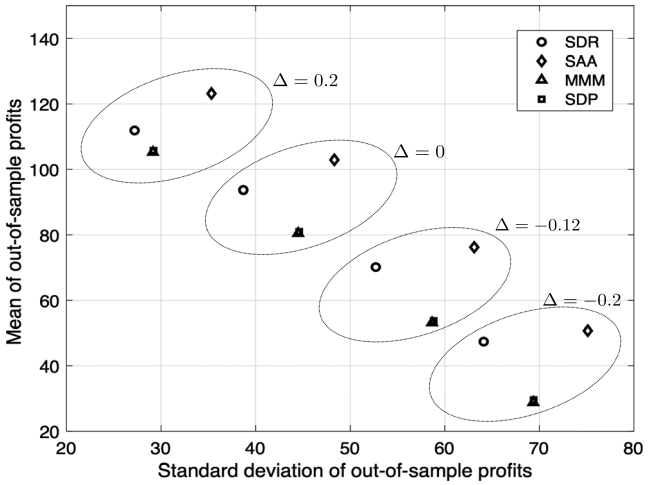
Figure 5 Out-of-Sample Performance of the Scenario-Wise Distributionally Robust, Sample Average Approximation, SDP and Marginal Moment Information Models for the Simulation Experiment [Color figure can be viewed at wileyonlinelibrary.com]



(a) Mean of out-of-sample profits



(b) Standard deviation of out-of-sample profits



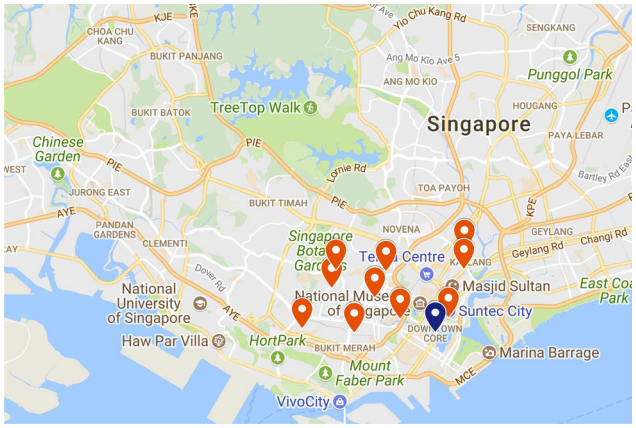
(c) Mean against standard deviation

dependence of demand is more effective than $\mathbb{F}_\Sigma$ which only captures the linear dependence of demand.

On the other hand, although the average performance of SDR cannot beat that of SAA in our tests, it tends to close the gap when Δ is small, that is, the out-of-sample distribution has much smaller mean and variance compared with the in-sample distribution. We believe the main reason that SAA still performs very well even when the mean and variance of the out-of-sample distribution differ is that SAA still captures the right shape of the distribution (in our case the normal distribution). The worst-case distributions for SDR, SDP or MMM, in comparison, are commonly quite extreme discrete distributions, and hence the solutions tend to be more conservative.

Figure 5c illustrates the trade-offs between the mean and standard deviation of the out-of-sample profits of the four models. For a given Δ, we group the results from the four models by a dotted ellipse. Since a higher mean and a lower standard deviation of out-of-sample profits are more desirable, we note that a model performs better when its result is closer to the upper left corner. We can observe that for a given Δ, SDR dominates both MMM and SDP by achieving a higher mean and a lower standard deviation of out-of-sample profits. It also achieves a more balanced performance than SAA with a similar mean and a lower standard deviation.

Figure 6 Illustration of Downtown and Selected Demand Regions [Color figure can be viewed at wileyonlinelibrary.com]

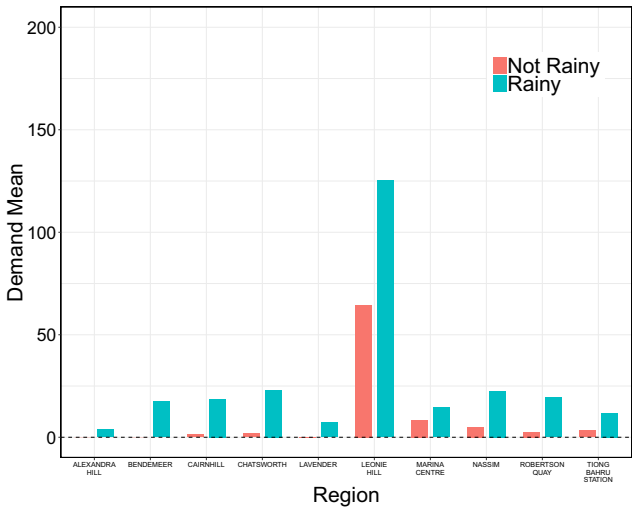


6.2. Case Study

In this section, we conduct a case study using real-world operational data. Our taxi operational dataset consists of booking trips, street-hail trips, and vehicle locations, ranging from January 1st, 2017 to May 31st, 2017, among which 98 working days are selected. As we have motivated in the introduction (see also Figure 1), we select the downtown region as the only supply node and the surrounding regions as the demand nodes. Figure 6 below shows the selected 10 demand regions (marked as red) which are within a 15-minute commuting distance from downtown (marked as blue).

For each working day, we select trips recorded between 8:00 AM and 8:15 AM. We use the real-time observed number of idle vehicles as supply S in the downtown region. We then estimate the corresponding demand $\hat{\mathbf{z}}$ and average trip revenue $\hat{\mathbf{r}}$ for the demand regions. Consistent with the simulation experiments, we set $\hat{w}_j = 0$ and enumerate θ from 0.04 to 0.12 with step 0.02. The allocation cost is set as the same as the average booking fee, that is, $\mathbf{w} = \hat{\mathbf{b}}$. We provide the detailed values of $\hat{\mathbf{r}}$ and $\hat{\mathbf{b}}$ in Table 3 in Online Appendix E.4.

Figure 7 Demand Pattern with Different Weather Condition [Color figure can be viewed at wileyonlinelibrary.com]



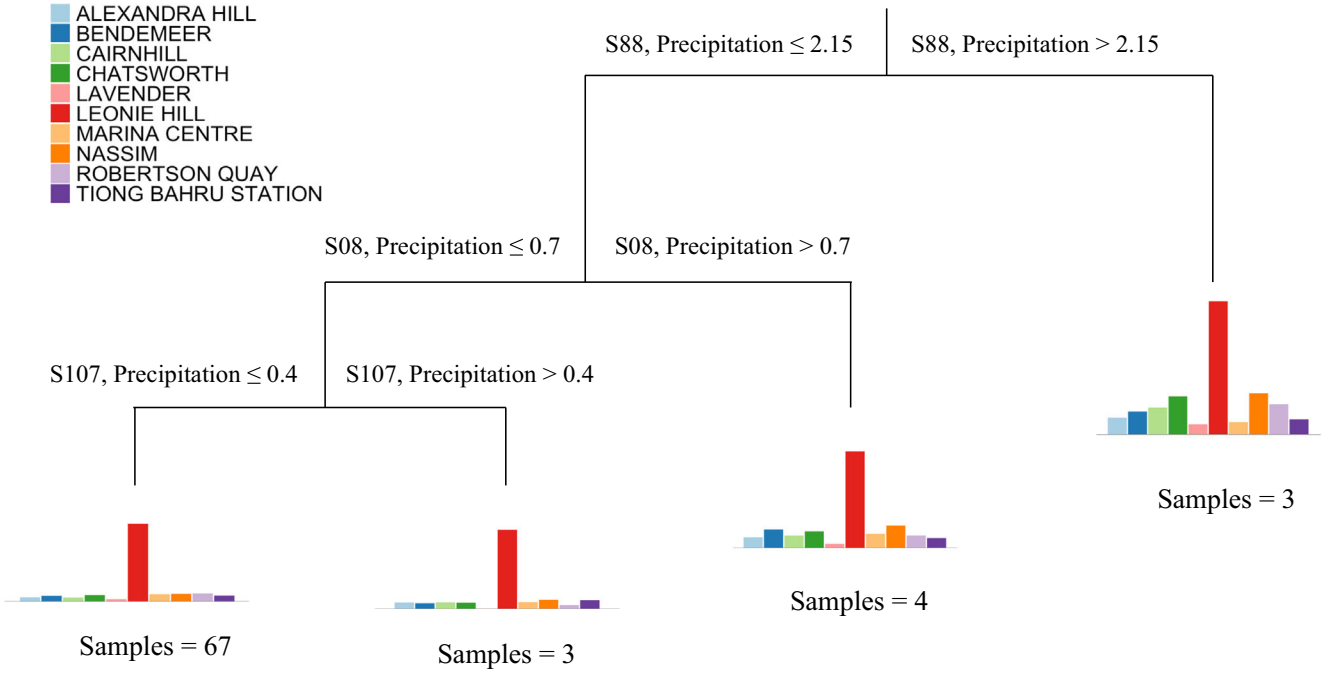
A key covariate that affects the taxi demand in Singapore is the weather. Therefore, we also acquire the rainfall data which records precipitation (mm/hour) in each hour and the duration of rainfall for the study period from January 1st, 2017 to April 30th, 2017. Figure 7 provides an example in comparing the average demand of each region conditional on the weather being rainy or not. We can see from Figure 7 that rainfall has a significant positive impact on passenger demand. To associate the weather information with the demand, records from 6 weather stations (denoted as S08, S107, S108, S78, S79, S88) distributed near the demand regions are selected, with each station recording both the precipitation and raining duration between 7:00 AM and 9:00 AM. Therefore, we set the covariates $\tilde{\mathbf{v}}$ as a 12-dimensional vector of weather information.

We select the demand samples in the 77 working days from January to April in 2017 as the training data, leaving the samples in May 2017 as the test data for which we do not have the weather information. In particular, for SDR, we apply MRT to classify demand into L (L = 2,4,6,8) scenarios based on the covariates $\tilde{\mathbf{v}}$. Figure 8 provides the resulting regression tree from MRT for $L = 4$, where the four-leaf tree is formed by three splits. The first split is formed based on the precipitation at station S88 (≤ 2.15 and > 2.15). For the low levels of precipitation, a second split is formed according to the precipitation of the station S08 with a different threshold at 0.7. Lastly, for the leaf with low levels of precipitation, the third split is formed based on the precipitation of station S107 with the splitting point of 0.4. Each leaf in the MRT is counted as a scenario, with the number of data points in each scenario indicated below each leaf. The probability of each scenario, the mean, and variance of demand in each region for each scenario are calculated according to the equations in section 4.3 (the estimation results are provided in Tables 4–8 in Online Appendix E.4).

As we have mentioned in section 4.3, in a real implementation, each scenario probability of the regression tree in Figure 8 can be a predicted value instead of the one estimated from historical data. For instance, if the weather forecast predicts there is a more than 90% chance of precipitation above 2.15, and the operator is quite confident about the forecast, then the operator can simply let the probability of this scenario be 0.9 instead of 3/77 estimated from historical data.

For MMM and SDP, we calculate the mean and variance of the demand for each demand region and

Figure 8 Regression Tree with 4-Scenarios for Demands at 10 Demand Regions [Color figure can be viewed at wileyonlinelibrary.com]

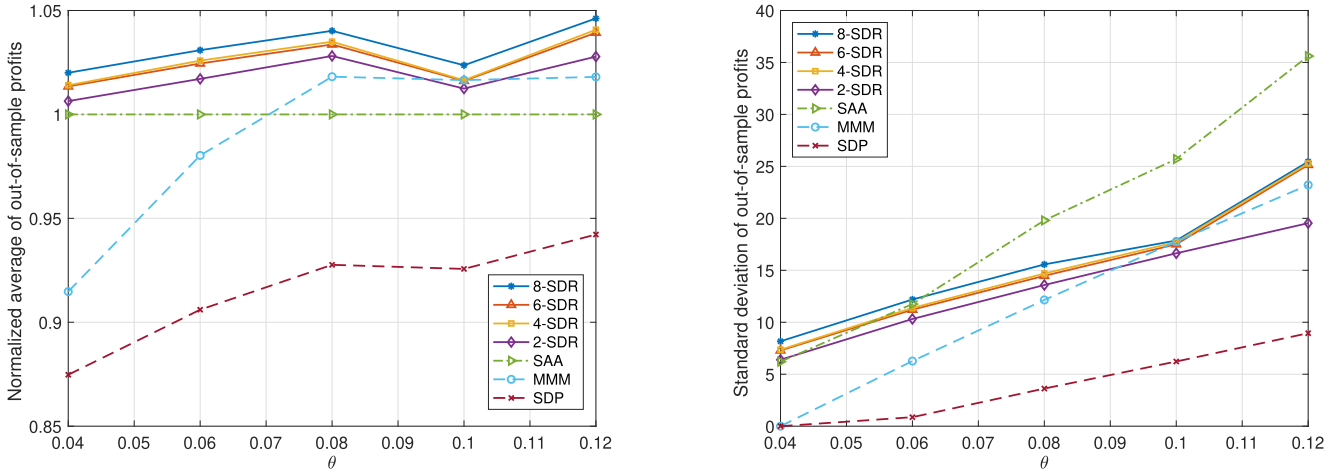


the covariance matrix using the entire training data as a single scenario. Similar to the simulation experiments, we solve SAA as a linear program. The performance of the four models is evaluated using the test data of demand samples from the 21 working days in May 2017.

The performance result is summarized in Figure 9 below. For the ease of exposition, we use 8-SDR, 6-SDR, 4-SDR, 2-SDR to denote the SDR models with 8 scenarios, 6 scenarios, 4 scenarios, and 2 scenarios respectively.

Figure 9a sets the mean of the out-of-sample profits from SAA as the benchmark and the average out-of-sample profits of all the other models are normalized by dividing that of SAA. In sharp contrast to our simulation results in Figure 5, the mean of the out-of-sample profits from all SDR models are consistently higher than that of SAA, and are higher than that of MMM and SDP in most of the cases. Recall from Figure 3 that the in-sample and out-of-sample distributions of demand in our dataset can be very different. Therefore, we believe the main reason that SAA

Figure 9 Out-of-Sample Performance of the Scenario-Wise Distributionally Robust, Sample Average Approximation, SDP and Marginal Moment Information Models for the Case Study [Color figure can be viewed at wileyonlinelibrary.com]



(a) Normalized average of out-of-sample profits

(b) Standard deviation of out-of-sample profits

performs poorly in this case could be that it is over-fitting the in-sample data.

MMM performs much worse than SAA and the SDR models when θ is relatively small due to the over-conservativeness—MMM allocates much fewer vehicles in this case. Surprisingly, different from the numerical simulation, the SDP model performs even worse than the MMM model in terms of out-of-sample performance for all the values of θ. One possible reason could be that the covariance matrix emphasizes primarily the linear dependence of demand while the true dependence relationship of the demand across different regions may be highly nonlinear due to the impact of covariates. As a result, incorporating additional covariance information may not necessarily be beneficial to the vehicle allocation decisions.

Figure 9b shows the stability of the SDR models. In terms of the standard deviation of the out-of-sample profits, 2-SDR, for example, is close to SAA when θ is small and significantly outperforms SAA as well as MMM, when θ is large. In summary, compared to SAA, the SDR outperforms it with 2%–4% higher average out-of-sample profits and up to 35% lower in the standard deviation of the out-of-sample profits. Compared to MMM and SDP, the SDR achieves a better balance of average performance and risk—it obtains a higher mean of the out-of-sample profits without bringing significantly greater standard deviation. The results demonstrate the efficiency of the SDR model in vehicle allocation and highlight the importance of covariate information in alleviating the conservativeness of the MMM and SDP framework in real-world settings.

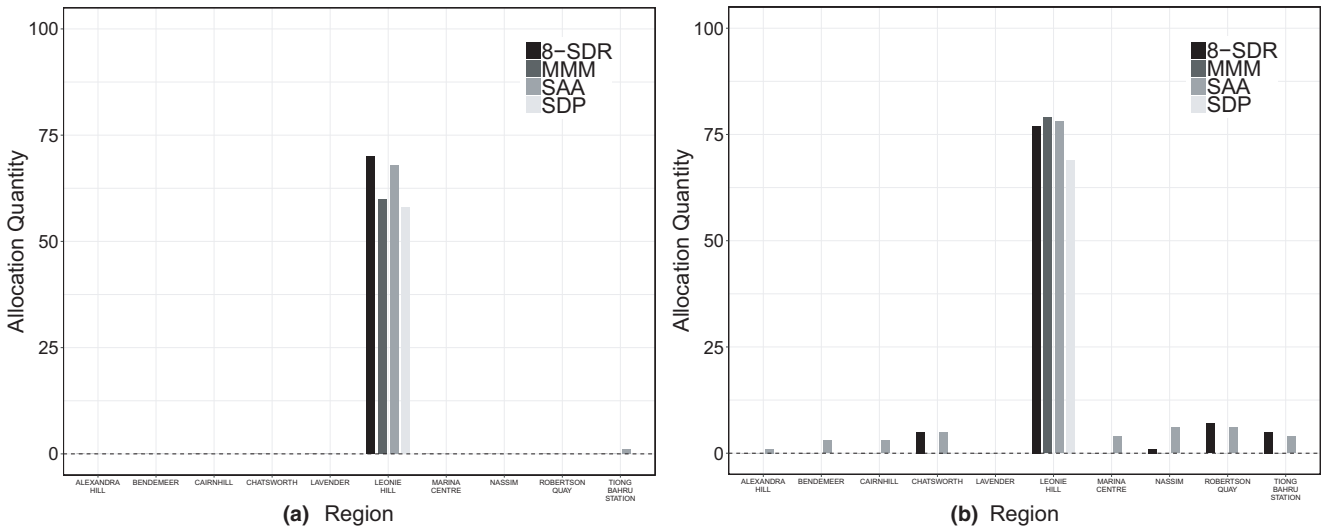
We further compare the allocation decision in Figure 10. First, we can observe in Figure 10a that when θ is low, almost all the four models only send vehicles to Leoniehill, which has the highest mean demand.[8] We also note that the number of the allocated vehicles of SDR and SAA are close to each other, while the MMM and SDP models tend to be more conservative in allocating vehicles to Leoniehill.

Interestingly, as shown by Figure 10b, when θ is high, both SDR and SAA tend to send more vehicles to more regions in addition to Leoniehill, while SDP and MMM still only send vehicles to Leoniehill. The allocation decision of SAA is more aggressive than that of SDR. One can observe both SAA and SDR send vehicles to Chatsworth, Robertson Quay, and Tiong Bahru Station, while SAA sends vehicles to additional regions including Marina Center, Cairnhill, Bendemeer, etc. In summary, from the perspective of allocation decision, the advantage of SDR is that it delivers the most balanced allocation solution among all the four models in terms of both the number of allocated vehicles and the number of target regions.

7. Conclusion

In this paper, we study the vehicle allocation problem with uncertain demand and other uncertain covariates. We show it can be formulated as a stochastic transportation problem when information is perfect. When information is incomplete, we employ the distributionally robust optimization framework, and we utilize the covariate information by adopting a novel scenario-wise ambiguity set. The construction and estimation of the ambiguity set using covariate information can be achieved via multivariate regression tree. We further show that the resulting distributionally robust optimization

Figure 10 Solution Comparison of the Scenario-Wise Distributionally Robust, Sample Average Approximation, SDP and Marginal Moment Information Models for the Case Study with $S = 218$



problem can be exactly and tractably solved using linear decision rule technique.

Our SDR model is benchmarked with the SAA, SDP, and MMM models in an extensive numerical study. Using simulated data, we show that by incorporating the covariate information, one can significantly improve the performance of the solution as compared to the SDP and MMM models. Our case study using the real taxi operational data combined with rainfall data demonstrates that the SDR model yields better average performance compared to the SAA, SDP, and MMM models, with small variability.

Our current model focuses on solving the single-period vehicle allocation problem. Therefore, one future research direction is to address the multi-period vehicle allocation problem. There are two potential challenges as we elaborate below. On one hand, inter-temporal demand dependence is commonly observed in the urban transportation industry (see He et al. 2020). With uncertain covariates, it is plausible that the covariates at different time periods would be correlated as well. It is not yet clear how to incorporate the dependence in both demand and covariate in a dynamic setting, which calls for future investigation. On the other hand, the dynamic allocation problem—when modeled as a dynamic program—suffers from the well-known "curse of dimensionality" issue since the state, which is the number of idle vehicles in each region, grows exponentially with the number of regions. While our linear decision rule approximation can still be applied to multi-period problems, its performance is not yet theoretically proven. One heuristic in providing dynamic allocation decisions is to implement our SDR model in a rolling-horizon fashion. That is, the operator can split a time horizon into several time periods. In each time period, the operator can make an (myopic) allocation decision by solving our single-period model based on updated demand history, covariate information, and real-time supply information. It would be an interesting future research direction to study how to split the time horizon or how frequently the operator should reoptimize the system and its impact on performance.

To obtain an integer allocation decision, instead of using the heuristic rounding we propose in the study, one can also explore other efficient algorithms, e.g., cutting-plane approaches, to solve the MISOCP (see, for example, Atamtürk and Narayanan 2008, Bhardwaj 2015, Zhang et al. 2018). The SDR model can also be directly extended to the case where the operator is risk-averse and uses the Conditional Value-at-Risk (CVaR) to quantify the risk (see Chan et al. 2017, for a recent application in healthcare device allocation problem). In contrast to the risk-neutral stochastic transportation problem, the risk-averse counterpart with CVaR can be shown to be #P-hard (see, for example, Hanasusanto et al. 2016). Yet, our SDR model can still be solved very efficiently in such a case.

Instead of using covariate information to alleviate the over-conservativeness of the robust solution, many works in the literature have managed to achieve this by specifying more distributional information, such as shape and tail-behavior in the ambiguity set (see Das et al. 2018, Natarajan et al. 2017). It would be a promising future research direction to incorporate the covariate information into the more sophisticated ambiguity sets considered in those works.

Acknowledgment

The authors gratefully acknowledge the departmental editor, the senior editor, and three anonymous referees for constructive comments. The authors thank Melvyn Sim, Zhi Chen, and Peng Xiong for valuable discussion, and our partner taxi company in Singapore for providing the data and sharing their insights. This research was partially supported by the Singapore Ministry of Education Social Science Research Thematic Grant, MOE2016- SSRTG-059, SPIRE; National University of Singapore, Institute of Operations Research and Analytics (IORA) Grant R-726-000-007-646.

Notes

[1]We apply Kolmogorov–Smirnov test (KS test), one common method to test whether two underlying probability distributions statistically differ (Chakravarty et al. 1967). The *p*-value of the test result is 0.026, which rejects the hypothesis that the underlying two shortage probability distributions are statistically equal. Meanwhile, we apply a non-parametric test, the Wilcoxon Rank Sum Test (Hollander et al. 2013) to test whether their means are statistically equal. The resulting *p*-value is 0.102, which is not significant enough to reject the hypothesis that the underlying two means are statistically equal.

[2]Here, we use $[\underline{\mathbf{z}}_l, \bar{\mathbf{z}}_l]$ as a shorthand notation for hyperrectangle, that is, $[\underline{\mathbf{z}}, \bar{\mathbf{z}}] = [\underline{z}_{1l}, \bar{z}_{1l}] \times \ldots \times [\underline{z}_{Ml}, \bar{z}_{Ml}]$.

[3]In their latest version of the paper Chen et al. (2019), they have included and cited our interpretation as well.

[4]In our case study, we compare the solution performance of the SDR model with different values of L. In practice, the best L can be calibrated using cross-validation.

[5]Recently, there is also a mixed-integer optimization model proposed in Bertsimas and Dunn (2017) to obtain a global optimal regression tree. However, it requires more computational time and is not yet available in the commonly used software packages.

[6]We also obtain the optimal integer solutions and the heuristic rounding solutions of SAA and MMM. Since the insights are the same, we only provide the results of the models with heuristic rounding solution in Online Appendix D.

[7]We also compare the integer solution and the heuristic rounding solution of the SDP model in the case study. The result shows that the profit gap is negligible. Please see Figure 1 and 2 in Online Appendix D.
[8]We have confirmed this result with the management team of our partner taxi company, who also revealed that both the residential density and the average income of the residents are relatively high in this region.

References

Atamtürk, A., V. Narayanan. 2008. Polymatroids and mean-risk minimization in discrete optimization. *Oper. Res. Lett.* **36**(5): 618–622.

Bai, J., K. C. So, C. S. Tang, X. Chen, H. Wang. 2018. Coordinating supply and demand on an ondemand service platform with impatient customers. *Manuf. Serv. Oper. Manag.* **21**(3): 556–570.

Benjaafar, S., D. Jiang, X. Li, X. Li. 2018. Inventory repositioning in on-demand product rental networks. Working paper, Available at SSRN 2942921.

Bertsimas, D., J. Dunn. 2017. Optimal classification trees. *Mach. Learn.* **106**(7): 1039–1082.

Bertsimas, D., A. Thiele. 2006. A robust optimization approach to inventory theory. *Oper. Res.* **54**(1): 150–168.

Bertsimas, D., D. A. Iancu, P. A. Parrilo. 2010. Optimality of affine policies in multistage robust optimization. *Math. Oper. Res.* **35** (2): 363–394.

Bertsimas, D., M. Sim, M. Zhang. 2018. Adaptive distributionally robust optimization. *Management Sci.* **65**(2): 604–618.

Bhardwaj, A. 2015. Binary conic quadratic knapsacks. PhD thesis, UC Berkeley.

Bimpikis, K., O. Candogan, D. Saban. 2019. Spatial pricing in ride-sharing networks. *Oper. Res.* **67**(3): 744–769.

Breiman, L., J. H. Friedman, R. A. Olshen, C. J. Stone. 1984. *Classification and Regression Trees*, The Wadsworth Statistics and Probability Series, Wadsworth International Group, Belmont California (pp. 356).

Cachon, G. P., K. M. Daniels, R. Lobel. 2017. The role of surge pricing on a service platform with self-scheduling capacity. *Manuf. Serv. Oper. Manag.* **19**(3): 368–384.

Chakravarty, I. M., J. Roy, R. G. Laha. 1967. *Handbook of Methods of Applied Statistics*. Wiley, Hoboken.

Chan, T. C., Z.-J. M. Shen, A. Siddiq. 2017. Robust defibrillator deployment under cardiac arrest location uncertainty via row-and-column generation. *Oper. Res.* **66**(2): 358–379.

Chen, Z., M. Sim, P. Xiong. 2018. Adaptive robust optimization with scenario-wise ambiguity sets. Working paper, Available at Optimization Online.

Chen, Z., M. Sim, P. Xiong. 2019. Robust stochastic optimization made easy with rsome. Working paper, Available at Optimization Online.

Das, B., A. Dhara, K. Natarjan. 2018. On the heavy-tail behavior of the distributionally robust newsvendor. Working paper, arXiv preprint arXiv:1806.05379.

De'Ath, G. 2002. Multivariate regression trees: A new technique for modeling species–environment relationships. *Ecology* **83** (4): 1105–1117.

De'Ath, G., K. E. Fabricius. 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology* **81**(11): 3178–3192.

Delage, E., Y. Ye. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* **58**(3): 595–612.

Erlebacher, S. J. 2000. Optimal and heuristic solutions for the multi-item newsvendor problem with a single capacity constraint. *Prod. Oper. Manag.* **9**(3): 303–318.

Federgruen, A., P. Zipkin. 1984. Approximations of dynamic, multilocation production and inventory problems. *Manage. Sci.* **30**(1): 69–84.

Hanasusanto, G. A., D. Kuhn, W. Wiesemann. 2016. A comment on "computational complexity of stochastic programming problems". *Math. Program.* **159**(1-2): 557–569.

He, L., Z. Hu, M. Zhang. 2020. Robust repositioning for vehicle sharing. *Manuf. Serv. Oper. Manag.* **22**(2): 223–428. https://doi.org/10.1287/msom.2018.0734.

Hollander, M., D. A. Wolfe, E. Chicken. 2013. *Nonparametric Statistical Methods*, Volume 751. John Wiley & Sons, Hoboken, NJ.

Holmberg, K., K. O. Joernsten. 1984. Cross decomposition applied to the stochastic transportation problem. *Eur. J. Oper. Res.* **17** (3): 361–368.

Iancu, D. A., M. Sharma, M. Sviridenko. 2013. Supermodularity and affine policies in dynamic robust optimization. *Oper. Res.* **61**(4): 941–956.

Lu, M., Z. Chen, S. Shen. 2017. Optimizing the profitability and quality of service in carshare systems under demand uncertainty. *Manuf. Serv. Oper. Manag.* **20**(2): 162–180.

Mak, H.-Y., Y. Rong, J. Zhang. 2014. Appointment scheduling with limited distributional information. *Manage. Sci.* **61**(2): 316–334.

Miao, F., S. Han, S. Lin, G. J. Pappas. 2015. Robust taxi dispatch under model uncertainties. Decision and Control (CDC), 2015 IEEE 54th Annual Conference on, IEEE, pp. 2816–2821.

Miao, F., S. Han, S. Lin, J. A. Stankovic, D. Zhang, S. Munir, H. Huang, T. He, G. J. Pappas. 2016. Taxi dispatch with real-time sensing data in metropolitan areas: A receding horizon control approach. *IEEE Trans. Autom. Sci. Eng.* **13**(2): 463–478.

Miao, F., S. Han, A. M. Hendawi, M. E. Khalefa, J. A. Stankovic, G. J. Pappas. 2017. Data-driven distributionally robust vehicle balancing using dynamic region partitions. Proceedings of the 8th International Conference on Cyber-Physical Systems, ACM, pp. 261–271.

Natarajan, K., M. Sim, J. Uichanco. 2017. Asymmetry and ambiguity in newsvendor models. *Management Sci.* **64**(7): 3146–3167.

Popescu, I. 2007. Robust mean-covariance solutions for stochastic optimization. *Oper. Res.* **55**(1): 98–112.

Qi, L. 1985. Forest iteration method for stochastic transportation problem. R. W. Cottle ed. *Mathematical Programming Essays in Honor of George B. Dantzig Part II*. Springer, Berlin, 142–163.

Scarf, H. 1958. A min-max solution of an inventory problem. K. J. Arrow, S. Karlin, H. Scarf, eds. *Studies in the Mathematical Theory of Inventory and Production*. Stanford University Press, Stanford, CA, 201–209.

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

Taylor, T. A. 2018. On-demand service platforms. *Manuf. Serv. Oper. Manag.* **20**(4): 704–720.

Wang, X., J. Zhang. 2015. Process flexibility: A distribution-free bound on the performance of k-chain. *Oper. Res.* **63**(3): 555–571.

Wiesemann, W., D. Kuhn, M. Sim. 2014. Distributionally robust convex optimization. *Oper. Res.* **62**(6): 1358–1376.

Williams, A. 1963. A stochastic transportation problem. *Oper. Res.* **11**(5): 759–770.

Zhang, Y., R. Jiang, S. Shen. 2018. Ambiguous chance-constrained binary programs under meancovariance information. *SIAM J. Optim.* **28**(4): 2922–2944.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix A: Linear Program Formulation of SAA.
Appendix B: Proofs.
Appendix C: Formulation of Robust Vehicle Allocation Problem with Covariance information.
Appendix D: Performance Comparison between the Integer Solution and the Heuristic Rounding Solution of Case Study.
Appendix E: Data and Computational Results of the Numerical Study.