# Transportation asset acquisition under a newsvendor model with cutting-stock restrictions: Approximation and decomposition algorithms

| | |
|---|---|
| Authors | Wagenaar,J.C.; Fragkos,I.; Faro,W.L.C. |
| Published in | Transportation Science |
| DOI | 10.1287/trsc.2023.1201 |
| Publication Date | 2023-05 |
| Document Version | publishersversion |
| Link | https://research.tilburguniversity.edu/en/publications/97eddbd0-6e34-489c-b27d-993f72c2cb34 |
| Citation | Wagenaar, J C, Fragkos, I & Faro, W L C 2023, 'Transportation asset acquisition under a newsvendor model with cutting-stock restrictions: Approximation and decomposition algorithms', Transportation Science, vol. 57, no. 3, pp. 778-795. https://doi.org/10.1287/trsc.2023.1201 |
| Download Date | 2026-05-17 11:27:04 |
| Rights | |

This article was downloaded by: [178.85.59.86] On: 13 September 2023, At: 02:09
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA

Transportation Science

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

# Transportation Asset Acquisition under a Newsvendor Model with Cutting-Stock Restrictions: Approximation and Decomposition Algorithms

Joris Wagenaar, Ioannis Fragkos, W. L. C. Faro

To cite this article:
Joris Wagenaar, Ioannis Fragkos, W. L. C. Faro (2023) Transportation Asset Acquisition under a Newsvendor Model with Cutting-Stock Restrictions: Approximation and Decomposition Algorithms. Transportation Science 57(3):778-795. https://doi.org/10.1287/trsc.2023.1201

Full terms and conditions of use: https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2023, INFORMS

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

https://pubsonline.informs.org/journal/trsc

TRANSPORTATION SCIENCE
Vol. 57, No. 3, May–June 2023, pp. 778–795
ISSN 0041-1655 (print), ISSN 1526-5447 (online)

# Transportation Asset Acquisition under a Newsvendor Model with Cutting-Stock Restrictions: Approximation and Decomposition Algorithms

**Joris Wagenaar,**[a,*] **Ioannis Fragkos,**[b] **W. L. C. Faro**[c]

[a] Zero Hunger Laboratory, Department of Econometrics and Operations Research, Tilburg School of Economics and Management, Tilburg University, Tilburg 5000 LE, Netherlands; [b] Department of Technology and Operations Management, Rotterdam School of Management, Erasmus University, 3062 Rotterdam, Netherlands; [c] BlueRock TMS, 5223's-Hertogenbosch, Netherlands

*Corresponding author

**Contact:** j.c.wagenaar_1@uvt.nl, https://orcid.org/0000-0002-2867-3149 (JW); fragkos@rsm.nl, https://orcid.org/0000-0001-7654-2314 (IF); martinfaro@live.nl (WLCF)

**Received:** December 15, 2021
**Revised:** September 26, 2022; January 21, 2023
**Accepted:** February 4, 2023
**Published Online in Articles in Advance:** March 21, 2023

https://doi.org/10.1287/trsc.2023.1201

**Abstract.** Logistics service providers use transportation assets to offer services to their customers. To cope with demand variability, they may acquire additional assets on a one-off (spot) basis. The planner's problem is to determine the optimal level of assets acquired upfront, such that their cost is minimized, for a given planning horizon. Our formulation captures a nontrivial complication: Although ordering quantities are pertinent to asset acquisition, customer demand is in the form of service requests. Not only does each request have a stochastic duration, but also the total number of requests per customer is uncertain. We introduce a two-stage newsvendor model where demand for spot assets is derived through optimal cutting-stock patterns. Leveraging results from bin-packing, we propose polynomial algorithms that have worst-case guarantees for upper and lower bounds. Our method finds optimal solutions to instances intractable by commercial solvers. We investigate demand variability by means of a factorial experiment. We find that, whereas variability in the number of requests leads to higher costs, variability in each request's duration can reduce costs. Finally, we demonstrate the modularity of our approach with two extensions: asset routing and outsourcing. Our results provide a practical approach to transportation asset acquisition and offer insights on the differing impact of demand uncertainty on the total acquisition cost.

**Supplemental Material:** The online appendix is available at https://doi.org/10.1287/trsc.2023.1201.

**Keywords:** asset acquisition • newsvendor • column generation • cutting stock • bin packing

## 1. Introduction

Logistic service providers (LSPs) offer door-to-door transportation services to companies spanning multiple industries. In order to carry out such services with high operational efficiency, their capability to respond to fluctuating customer demand is paramount. Timely access to transportation infrastructure, such as train wagons and containers, requires the deployment of assets provisioned well in advance, so that resorting to reactive, last-minute alternatives is kept to a minimum. Consider, for instance, a provider of rail transportation services, who may commit to the acquisition of a certain number of train wagons on an annual basis, either by committing long-term leasing agreements or by commissioning the wagons ad hoc. If customer demand exhibits strong seasonal fluctuations and she needs additional wagons to cover it, she will have to use spot wagons, which are often priced at a premium. Therefore, committing to a small number of wagons for long-term usage exposes her to demand variability, leading to last-minute spot wagon orders and recurrent, one-off costs. Committing upfront, however, to an unnecessarily large number of wagons can lead to significant fixed costs, affect liquidity, and, ultimately, the prosperity of her operations. Such a position can lead to financial distress, especially if anticipated demand is not realized. Although such decisions have additional complexities, in the case of transportation assets, there is one aspect that makes them particularly challenging: Customer demand is in the form of transportation *services*, whereas the LSP needs to supply these services by managing *asset acquisitions*. A request for transportation services amounts to a customer requesting a certain number—or volume—of goods to be transferred to her location, every period. This, in turn, implies that the LSP should make a given number of trips to the customer's location. Because the volume of products the LSP needs to transfer to each customer is uncertain, the corresponding number of trips to the customer's location is uncertain. Assuming that each trip can carry out a given constant volume, customer demand in terms of volume of products can be readily translated to demand for a certain number of trips, hereafter called *customer requests*. Further, such trips generally suffer from delays of various, context-related causes. For the rail-based LSP case, cargo trains experience delays because they use the same network with passenger trains, which often have delays and take priority over cargo trains. Therefore, uncertainty stemming from travel times prolongs the duration of customer requests for transportation services. In this work, we consider customer requests for transportation services that have two dimensions of uncertainty: the number of visit requests per customer and the time spent to accommodate each request, which is the round-trip duration to the customer and the time spent in her premises to carry out the service. We proceed with a minimal example of the trade-offs we consider next.

**Example.** Following the example of the rail-based LSP, suppose that she knows, with perfect foresight, that each of her customers will require a certain number of trips for a specific week, as shown in Table 1. Here, $c_1$ and $c_2$ request 10 and 20 (single-wagon) trips to their destinations, respectively, to be carried out within this specific week. Note that $c_1$ and $c_2$ may also correspond to aggregated customer demand across two destinations, without loss of generality, in which case the total demand of all customers located in destination 1 is 10 trips. Return trips to $c_1$ and $c_2$ take 0.5 and 0.2 weeks, respectively, and the LSP needs to decide how many total wagons are required to satisfy these demands, given that one wagon can be transferred in each trip. Finally, the last column shows the allocation of wagons that satisfies these restrictions and minimizes the total number of wagons.

Although this example can be solved by inspection, in general, it is not clear how to translate customer requests (trips) to assets (wagons) in a multiperiod model that minimizes the costs of regular and spot assets in the presence of uncertainty. In effect, the single-period, deterministic version of the problem, as shown in the above example, requires the solution of a cutting-stock problem. The presence of multiple periods, uncertainty, and a strategic-level newsvendor-like structure make the overall problem nontrivial to solve.

In this work, we abstract away from the specifics of rail-based operations and focus on strategic asset-acquisition aspects pertinent to generic transportation environments. The asset-acquisition trade-off that we tackle is reminiscent of a multiperiod newsvendor problem (Chen et al. 2016). Concretely, we introduce a two-stage stochastic integer program with recourse, which is a two-stage newsvendor problem with demand generated by optimally solved cutting-stock patterns. We solve this problem using a custom bounding technique that utilizes column generation. The first-stage decision represents long-term commitments and corresponds to acquiring the number of assets that the LSP is going to utilize throughout the time horizon, hereafter called *regular* assets, and the second stage corresponds to the additional number of assets that the operator employs in each time period and scenario when demand cannot be covered by regular assets, hereafter called *spot* assets.

**Table 1.** Single-Period Example with Two Customers, Deterministic Demand (Trips, Duration), and Asset Allocation (Wagons)

| Customer | Trips | Duration/Trip (wk) | Wagons |
|---|---|---|---|
| $c_1$ | 10 | 0.5 | 5 |
| $c_2$ | 20 | 0.2 | 4 |

A straightforward formulation of the problem is computationally challenging to solve with regular mixed integer programming (MIP) solvers. To this end, we exploit an important structural property of the formulation, which makes the problem decomposable to period- and scenario-dependent cutting-stock subproblems, and a first-stage problem that can be solved in $\mathcal{O}(\log \max\{|T|, |S|\})$ time using binary search, where $T$ and $S$ are the number of periods and scenarios, respectively. We show that the absolute performance ratios for upper and lower bounds that are known for the cutting-stock problem can be used to derive such guarantees for our model and devise a computationally efficient algorithm that solves the cutting-stock problems using column generation.

We utilize our algorithm to investigate the impact of the average and variance of customer service requests and their duration on the optimal split between regular and spot assets and on the total cost. One of our key findings is that, although uncertainty in the number of customer requests leads to higher expected costs, uncertainty on the duration of requests can have a positive effect on expected cost. The intuition behind this result is that high variability in service requests allows more flexibility in scheduling assets efficiently, making it possible to "pack" long- and short-duration trips using the same asset. We also investigate the validity of our conclusions under different experimental conditions, such as when the number of service requests and their duration are correlated. We find that positive correlations diminish the benefit of highly variable request duration, but the effect is persistent and remains statistically significant. Our results remain qualitatively similar for instances with and without spot capacity—although the absolute costs and split between spot and regular resources are different.

In summary, our work offers the following contributions:

- We introduce the two-stage newsvendor model with cutting-stock-generated demand, which represents the asset-acquisition problem of LSPs carried out during strategic planning. The model captures the trade-off between the cost of regular and spot assets, by taking into account stochastic customer service requests and service duration. Despite its parsimony, the model is computationally challenging to solve, even for small instances.

- We take advantage of the structural properties of this model to identify optimal solutions when solving the problem using a decomposition scheme that divides the formulation in a first-stage problem, solvable in sublinear time, and in a series of second-stage cutting-stock subproblems. Further, we study how known theoretical guarantees of lower and upper bounds can be used to derive theoretical guarantees in our setting.
- We implement a column generation to solve each subproblem and integrate their solutions to the upper-level problem. Column generation allows us to derive excellent solutions very fast, which are then passed on as parameters to the upper-level problem. It is worth noting that the upper-level problem can be optimal, even if some second-stage subproblems are not solved to optimality. We show that our solution algorithm has superior computational performance relative to a default integer programming (IP) solver.
- We conduct a computational study to investigate how various configurations of alternative parameters influence the structure of optimal solutions and of the total cost. With respect to total cost, we report that a larger average number of customer requests and a longer average request duration both lead to higher total cost, all else being equal. Higher variability of customer requests also leads to higher total cost, but higher request-duration variability leads to lower total cost because of added flexibility. These results are persistent when we introduce correlations between duration lengths and number of requests and when we alter the ratio of spot to regular cost. When it comes to the split of total cost to regular and spot, we find that the proportion of spot cost increases with variability of customer requests and variability of service-request duration. Our main results are persistent for instances with capacitated spot resources.
- We show how to take advantage of the modular structure of our model and extend it to include fixed deployment costs, operational costs, and asset routing. For the case of routing, we conduct another study that investigates how duration variability influences acquisition costs. We find that, depending on the average duration of customer requests, duration variability could either help or hinder in forming feasible routes.

The rest of this paper is organized as follows. Section 2 gives an overview of existing literature on similar problems. Then, Section 3 describes the model formulation for the problem at hand. Section 4 studies the problem structure, and Section 5 describes a column-generation approach. Computational results are reported in Section 6, while Section 7 reports on the impact of the model parameters on solution structure and cost. Extensions of the basic problem are presented in Section 8, while Section 9 concludes the paper, by discussing future work.

## 2. Literature Review

One of the key strategic decisions for an LSP is how much capital to invest in its transportation infrastructure throughout the strategic planning phase (Yamada et al. 2009). In this paper, the main decision is how many transportation assets (such as vehicles, train cars, or containers) the LSP should acquire for the entire time horizon. The LSP further has the option to acquire assets from the spot market for a specific time period in case she faces a shortage throughout that time period (Wagenaar, Fragkos, and Zuidwijk 2021). The decision of acquiring assets for the entire time horizon has far-reaching cost implications (Loxton, Lin, and Teo 2012; Nourinejad and Roorda 2017).

Although the setting we consider is a multiperiod variant of the newsvendor model, to the best of our knowledge, prior studies have not incorporated nuances found in transportation settings. Hua, Wang, and Cheng (2012) consider a setting where a supplier offers free shipping to a retailer, who decides on her optimal ordering quantity. Here, the ordering quantity of the customer is endogenous, because it depends on the supplier's pricing structure (e.g., quantity discounts). Our model can be seen as a multiperiod extension of this setting, where the supplier decides on the level of assets based on the anticipated demand of the retailer. Other papers follow a similar line of work (Jiang, Shang, and Liu 2013; Noori-daryan, Taleizadeh, and Govindan 2018), but do not consider explicitly the fact that customer demand is expressed in services and LSP supply decisions are in terms of assets, thereby making their planning nontrivial.

Containership routing and scheduling (Meng et al. 2014) is another problem, the nature of which, when considered at the strategic level, is reminiscent of our formulation. For example, Meng and Wang (2011) mention that Maersk operated a mix of owned and chartered ships, which are rebalanced in every time period within a horizon. Purchasing ships in the long run seems to be a more profitable investment compared with chartering out ships, although the latter is cheaper than renting spot ships in the short run. Our model does not consider specific features of liner ship routing, but captures the essential trade-off between owning or renting assets under uncertainty. Another difference is that in their setting, there is operational flexibility in determining the number of lay-up days for each ship under each period and scenario, whereas in our model, the duration that an asset serves a customer is exogenous. Finally, their formulation attempts to integrate operational decisions, such as the number of containers transferred from one port to another, with strategic decisions, whereas we abstract away from problem-specific operational nuances, incorporating, however, that servicing a customer implies that an asset will be occupied for an uncertain amount of time.

Fleet composition problem (see for extensive literature surveys, e.g., Baldacci, Battarra, and Vigo 2008; Pantuso, Fagerholt, and Hvattum 2014; and Baykasoğlu et al. 2019) also face the problem of transportation asset allocation. Jabali, Gendreau, and Laporte (2012) consider a vehicle-fleet composition problem, where fleet size and mix are decision variables, and develop computationally efficient bounding procedures. The key differences between these models and ours is that (i) asset types in fleet composition can be heterogeneous in several dimensions, namely, capacity, cost, and duration of use; and (ii) they typically are modeled as single-period problems. We thus cover another dimension of fleet composition, where identical vehicles can have different ownership status and, therefore, different costs. It is worth noting that adding the routing part in our formulation will only influence the structure of the subproblems, leaving the key trade-off between regular and spot assets intact.

The paper of Klosterhalfen, Kallrath, and Fischer (2014) focuses on finding the optimal structure and size of a railcar fleet at a chemistry company. They combine mixed-integer linear programming models with stochastic inventory approaches to include uncertainty in both demand and travel times. One of the key differences is that in their work, demand is expressed in number of rail cars to be used, instead of in services required.

It is often the case that decisions taken at the tactical or operational level are constrained by strategic decisions, such as long-term resource-acquisition policies. This is the case in Wagenaar, Fragkos, and Zuidwijk (2021), where the authors consider the integrated tactical and operational planning of multimodal operations of transportation carriers. The models developed therein consider the explicit scenario-based circulation of rolling stock, given the option to acquire rolling stock on the spot market four weeks before the actual operations. Although this ensures cost-efficient operations, if the rolling stock determined during strategic planning is suboptimal, it is inevitable that tactical planning will be affected. However, integrating all levels of planning within a single modeling framework will result in a complicated model with several, hard-to-obtain inputs, spanning across different time frames and of different reliability—in addition to challenges related to solution methods and model maintenance. Similar trade-offs can be found in the scheduling of maritime operations, as described by Fragkos and De Reyck (2016). There, vessels arrive at a port randomly, but within a specific time window, and a series of (un)loading operations need to be scheduled for each vessel. The objective is to minimize the joint cost of penalties resulting from delayed operations and the cost of spot equipment. If the vessel-arrival time windows, which are negotiated at the beginning of each year, have significant overlaps, there is limited space to achieve operational efficiency. In this case as well, the operational complexities of the problem and the uncertainty on vessel arrival when viewed from a strategic perspective make it futile to integrate all stages in a single model. In this paper, our model attempts to strike a balance between including the nuances of operational models and representing decisions that are relevant from a strategic perspective.

Finally, there is an important stream of literature in variants of bin-packing problems that share similarities with our model. The work of Crainic, Fomeni, and Rei (2021) studies a generalized deterministic bin-packing problem with item-to-bin costs and intertemporal constraints. When it comes to stochastic bin-packing problems, much of the literature has focused on online versions of bin packing, and only a few papers, such as Crainic et al. (2014, 2016), focus on stochastic variants used for strategic and tactical capacity-planning problems. Crainic et al. (2016), in particular, generalize the model of Crainic et al. (2014), which is a two-stage stochastic program with recourse that considers spot and regular assets of varying types. The problem is challenging to solve for a regular MIP solver, and the authors develop a progressive hedging meta-heuristic that delivers high-quality, robust solutions. The key difference of our approach with these models is that the introduction of variables that represent regular- and spot-asset acquisitions creates a modular model structure that disentangles the asset-acquisition part from the operational part of allocating assets to customer requests. As such, it allows us to extend the base model with additional features, such as routing, and take advantage of its convexity, which leads to efficient computation algorithms and approximation guarantees.

## 3. The Model

In this section, we discuss the formulation for the *Two-stage Newsvendor Model with Cutting-Stock Generated Demand*, hereafter called NVCSD. To this end, we make the following assumptions:

- A tour made by an asset needs to be fully completed within a period. That means that an asset reaches the customer and returns to the LSP's base before the end of each period.
- The operating cost of using an asset to serve a customer is not taken into account. This is because when an asset is operated by the LSP, the operating cost is the same, regardless of how the asset was acquired. This assumption is lifted later when we consider the case that a third-party operates the asset, and, therefore, the operational costs can become relevant.

- To simplify the notation, we assume that the capacity of each asset is identical and discuss how variable asset prices and capacities can be included in our formulation without further analysis.
- Regular assets are procured for the entire planning horizon. Apart from the regular assets, there are no operational constraints that span several time periods.
- Assets cannot be routed. In our motivating application, asset routing is not common because customers request full-capacity rail trips. This is also the case for full-truckload services. We lift this assumption later when we study a routing extension.

Taking these assumptions into account, the following notation is introduced.

### 3.1. Sets

$C = \{1, \ldots, |C|\}$, customers, indexed by $c$.
$T = \{1, \ldots, |T|\}$, time periods, indexed by $t$.
$N = \{1, \ldots, |N|\}$, assets, indexed by $n$.
$S = \{1, \ldots, |S|\}$, scenarios, indexed by $s$.

Note that a single asset might be able to satisfy multiple service requests from, possibly different, customers. We incorporate uncertainty both on the demand in service requests of a customer, $\theta_{cts}$, and on the time it takes an asset to perform a service request, $\tau_{cts}$, respectively. Finally, note that, without loss of generality, customer demand can be thought of as being the aggregated demand of specific locations. For example, when two actual customers are in the same location, they can be represented as one customer object in our model.

### 3.2. Parameters

$\rho$, maximum duration an asset can be in service per time period.

$\theta_{cts}$, service request demand of customer $c$ in time period $t$ under scenario $s$ (number of trips required)

$\tau_{cts}$, duration an asset is deployed to perform one service request of customer $c$ in time period $t$ under scenario $s$

$K_t$, number of spot assets available during time period $t$ in the spot market.

$p_{ts}$, probability that in time period $t$ scenario $s$ is realized.

$\Lambda$, regular asset acquisition cost.

$\Gamma$, spot asset acquisition cost.

### 3.3. Decision Variables

$\omega$, number of assets acquired at the beginning of the planning horizon (regular assets).

$\zeta_{ts}$, number of additional assets rented from the spot market, during time period $t$ under scenario $s$ (spot assets).

$y_{nts}$, $=1$, if asset $n$ is utilized in time period $t$ under scenario $s$ 0 otherwise.

$a_{cnts}$, number of service requests of customer $c$ satisfied by asset $n$, during time period $t$ under scenario $s$.

### 3.4. IP Formulation

Using the sets, parameters, and variables, the NVCSD is defined as follows.

$$\min \quad \Lambda\omega + \Gamma\sum_{t\in T}\sum_{s\in S} p_{ts}\zeta_{ts}, \tag{1}$$

$$\text{s.t.} \quad \omega + \zeta_{ts} \geq \sum_{n\in N} y_{nts}, \qquad \forall t \in T, \quad \forall s \in S, \tag{2}$$

$$\sum_{n\in N} a_{cnts} = \theta_{cts}, \qquad \forall c \in C, \ \forall t \in T, \quad \forall s \in S, \tag{3}$$

$$\sum_{c\in C} \tau_{cts} a_{cnts} \leq \rho y_{nts}, \qquad \forall n \in N, \ \forall t \in T, \quad \forall s \in S, \tag{4}$$

$$0 \leq \zeta_{ts} \leq K_t; \zeta_{ts} \text{ integer} \qquad \forall t \in T, \quad \forall s \in S, \tag{5}$$

$$y_{nts} \in \{0, 1\}, \qquad \forall n \in N, \ \forall t \in T, \quad \forall s \in S, \tag{6}$$

$$a_{cnts} \geq 0; \quad a_{cnts} \text{ integer} \qquad \forall c \in C, \ \forall n \in N, \ \forall t \in T, \quad \forall s \in S, \tag{7}$$

$$\omega \geq 0; \quad \omega \text{ integer}. \tag{8}$$

The objective function minimizes the total expected cost, which consists of the cost of the regular assets and the expected cost of spot assets. Constraints (2) ensure that the assets required should be lower than or equal to the total number of available assets, spot and regular, for every time period and scenario. Constraints (3) require that service requests from customer $c$, $\theta_{cts}$, must be satisfied by asset deployments in a scenario/time-period combination. Constraints (4) impose that the total time an asset is in use during a time period under a scenario should be no higher than its available operating time in every scenario/time period combination. Finally, Constraints (5)–(8) enforce the bounding and integrality restrictions of the decision variables.

Note that cases, such as those of a rail-based LSP, with a variable number of wagons in each trip to a customer can be captured with a small modification to our model. Concretely, if an asset $n$ has capacity equivalent to $cap_n$ assets per trip, then writing (3) as $\sum_{n\in N} cap_n a_{cnts} = \theta_{cts}$ and (2) as $\omega + \zeta_{ts} \geq \sum_{n\in N} cap_n y_{nts}$ allows us to tackle the case where assets can have variable, nonunit capacities per trip. In what follows, we focus on the unit-capacity case and leave the additional complication of variable capacity for future work. Model NVCSD is a two-stage stochastic integer program with the number of regular assets $\omega$ being the first-stage decisions, the spot assets $\zeta_{ts}$, the amount of demand satisfied by a specific asset $a_{cnts}$, and asset-usage indicators $y_{nts}$ being the recourse variables.

Note that the inclusion of two distinct sources of uncertainty has different operational and financial implications. Demand uncertainty directly increases the number of visits that an asset has to pay to a customer, and this increase can be absorbed if enough assets have been allocated. An increase of the time needed to perform one service request of a customer, on the other hand, although it can also be accommodated if there are enough assets, does not affect the number of times an asset needs to visit a customer. Finally, note that, implicitly, the model assumes that the set of scenarios accurately represents the support of the underlying distributions. In particular, the worst-case realizations of $\tau_{cts}$ and $\theta_{cts}$ are included as one of the scenarios $s \in S$, so that the model remains feasible in the presence of capacity constraints.

## 4. Model Structure and Decomposition

In order to solve (1)–(8), one needs to have an estimate of $N$. A way to obtain such an estimate is to assume that each customer's demand is served by dedicated assets in each time period and scenario, leading to a scenario- and period-specific upper bound, $|N_{ts}| = \sum_{c\in C}\lceil(\tau_{cts}\theta_{cts})/\rho\rceil$. This cardinality might be unnecessary large, because a single asset can serve multiple customers, and, therefore, fewer assets than $|N_{ts}|$ are required. In what follows, we exploit the structure of (1)–(8) and circumvent this difficulty.

A high-level schematic overview of our solution approach is given in Figure 1. The original problem (1)–(8) gives rise to a series of independent cutting-stock problems for each period/scenario combination. The optimal objective values of these problems are then used in a master problem formulation, which can be solved using single-variable convex minimization methods (e.g., golden section search).

### 4.1. Decomposing the Two-Stage Newsvendor Model with Cutting-Stock-Generated Demand

A key property of (1)–(8) is that the variables that are in the objective function appear only in Constraints (2), lower bounded by $\sum_{n\in N} y_{nts}$, the number of required assets in each period and scenario. Therefore, a simpler version of (1)–(8) can be formulated, assuming that the number of required assets in each time period and scenario, $\delta_{ts}$, is given exogenously. We use $\boldsymbol{\delta}$ to denote the vector of $\delta_{ts}$, hereafter assumed to be integer without loss of generality. Then, Model (9)–(12) describes this reformulation.

$$z(\boldsymbol{\delta}) = \min \quad \Lambda\omega + \Gamma\sum_{t\in T}\sum_{s\in S} p_{ts}\zeta_{ts}, \tag{9}$$

$$\text{s.t.} \quad \omega + \zeta_{ts} \geq \delta_{ts}, \qquad \forall t \in T,\ \ \forall s \in S, \tag{10}$$

$$\omega \geq 0, \qquad \omega \text{ integer}, \tag{11}$$

$$\zeta_{ts} \in \{0, 1, \ldots, K_t\}, \qquad \forall t \in T,\ \ \forall s \in S. \tag{12}$$

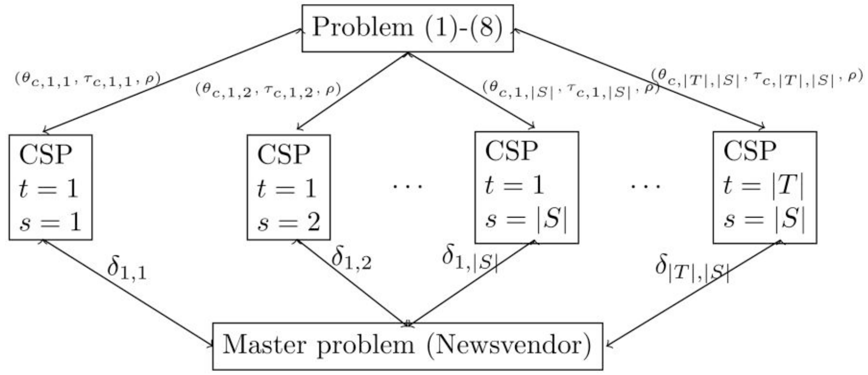
Note that the Integrality Restrictions (11) and (12) can be dropped for integer vectors $\boldsymbol{\delta}$. Here, the objective, (9), is identical to the original objective of the NVCSD, and Constraints (10) state that there need to be at least as many assets available in each time-period/scenario combination as there is demand for. Function $z(\boldsymbol{\delta})$ is a value function of the corresponding linear program, and, as such, it is jointly concave on $\boldsymbol{\delta}$. The solution of (9)–(12) depends on the choice of $\delta_{ts}$, and each $\delta_{ts}$ has to be selected such that the corresponding Constraints (3) and (4) are feasible. In order to determine each $\delta_{ts}$, only Constraints (3) and (4) indexed by $(t, s)$ are relevant. In fact, selecting each $\delta_{ts} = \sum_{n\in N} y_{nts}$ such that $\sum_{n\in N} y_{nts}$ is minimized, subject to the corresponding Constraints (3) and (4) being feasible, guarantees that (9)–(12) has an optimal objective value equal to the original formulations, (1)–(8). Moreover, higher or lower $\delta_{ts}$ for all $(t, s)$ pairs will make (9)–(12) generate upper or lower bounds on the objective of the Original Model (1)–(8), respectively. This is stated formally in the following remark.

**Remark 1.** Define

$$\begin{aligned}
\delta^*_{ts} := \min \quad & \sum_{n\in N} y_{nts} \\
\text{s.t.} \quad & \sum_{n\in N} a_{cnts} = \theta_{cts}, \ \ \forall c \in C \\
& \sum_{c\in C} \tau_{cts} a_{cnts} \leq \rho y_{nts}, \ \ \forall n \in N \\
& (6) - (7).
\end{aligned} \tag{13}$$

Then,

Figure 1. Schematic Overview of the Solution Approach



*Note.* For each scenario and period, a cutting-stock problem (CSP) is solved.

- If $\delta_{ts} = \delta^*_{ts}$ for all $t \in T, s \in S$, (9)–(12) and (1)–(8) have the same optimal objective value.
- If $\delta_{ts} \geq \delta^*_{ts}$ for all $t \in T, s \in S$, then the optimal objective value of (9)–(12) is an upper bound to the optimal objective of (1)–(8).
- If $\delta_{ts} \leq \delta^*_{ts}$ for all $t \in T, s \in S$ with at least one inequality strict, then the optimal objective value of (9)–(12) is a lower bound to the objective solution of (1)–(8). □

The two last bullets of Remark 1 follow because $z(\boldsymbol{\delta})$ is the value function of the corresponding integer program and because of the monotonicity properties of value functions (see, for example, Wolsey and Nemhauser 1999, proposition 2.1). Further, Remark 1 can be used to devise an algorithm to solve (1)–(8): Calculate $\delta^*_{ts}$ by solving (13) for each $t$, $s$, and then use them to solve (9)–(12). Before implementing this algorithm, we take a closer look at Subproblem (13) and discuss existing polynomial-time algorithms with worst-case performance guarantees.

The structure of the NVCSD Model (1)–(8) makes some lower and upper bounds for $\delta^*_{ts}$ readily available. Concretely, the quantity $\sum_{c \in C} \lceil (\tau_{cts}\theta_{cts})/\rho \rceil$ is an upper bound, as it represents the total number of used assets if each customer is allocated $\lceil (\tau_{cts}\theta_{cts})/\rho \rceil$ dedicated assets. This corresponds to scheduling each customer individually. For example, if a customer requires seven trips ($\theta_{cts} = 7$) of duration 0.6 weeks each ($\tau_{cts}/\rho = 0.6$), we would schedule five trains ($\lceil 4.2 \rceil$). This is clearly an upper bound, because the fifth train is underutilized. For the lower bound, note that summing (4) over $n \in N$ and substituting $\sum_{n \in N} a_{cnts}$ from (3) gives $\sum_{c \in C}((\tau_{cts}\theta_{cts})/\rho) \leq \sum_{n \in N} y_{nts}$, which can be rounded up because the number of assets is integer. Thus, we have established the following fact.

**Remark 2.** $\lceil \sum_{c \in C}((\tau_{cts}\theta_{cts})/\rho) \rceil \leq \delta^*_{ts} \leq \sum_{c \in C} \lceil (\tau_{cts}\theta_{cts})/\rho \rceil \leq \sum_{c \in C} \lfloor (\tau_{cts}\theta_{cts})/\rho \rfloor + |C|$, $\forall t \in T, s \in S$.

Although the bounds provided in Remark 2 are generally loose, they tend to be tight when durations $\tau_{cts}$ are large and when customers have a large number of requests, in the sense that their relative gap tends to zero. In the case of an LSP that utilizes a railway network, customers can represent distinct locations, and customer requests $\theta_{cts}$ is the aggregate request demand of location $c$. If the LSP serves multiple customers in specific locations, then the relative bound difference is small.

The upper and lower bounds on $\delta^*_{ts}$ can be strengthened further if we observe that Problem (13) is a one-dimensional cutting-stock problem for each pair $(t, s)$: Each customer $c$ poses a demand $\theta_c$, for services of duration $\tau_c$, which should be satisfied ($\sum_n a_{cn} = \theta_c$) using the minimum number of assets ($\min \sum_n y_n$), so that each used asset's time utilization does not exceed its capacity, ($\sum_c \tau_c a_{cn} \leq \rho y_n$). It is possible to utilize results from the cutting-stock literature to show that existing algorithms have performance guarantees for our model.

First, note that, without loss of generality, one can create $\theta_{cts}$ customers with unit demand for each customer $c \in C$. Then, one has to fit each unit-demand customer to a minimum number of assets, so that their capacity is not exceeded, which is a one-dimensional bin-packing problem. Upper bounds can then be obtained by applying the *First-Fit Decreasing* (FFD) or *Best-Fit Decreasing* (BFD) algorithms. An outline of algorithm FFD can be found in Online Appendix A. Both FFD and BFD have an absolute ratio of 1.5; this is the best possible ratio in polynomial time, unless $\mathcal{P} = \mathcal{NP}$ (Simchi-Levi 1994). Therefore, if $\overline{\delta}_{ts}$ is the solution obtained by applying FFD or BFD and $\delta^*_{ts}$, the optimal solution of (13), then $\overline{\delta}_{ts} \leq 1.5\delta^*_{ts}$, for all $t$ and $s$.

Similar assertions can be made for lower bounds on $\delta^*_{ts}$. There are several lower bounds for cutting-stock and bin-packing problems, but no such bound computable in polynomial time can have a worst-case performance ratio of 2/3 (see Martello and Toth 1990 for computations of such bounds). Therefore, if $\underline{\delta}_{ts}$ is such a lower bound, then $\underline{\delta}_{ts} \geq 2/3\delta^*_{ts}$, for all $t$ and $s$.

We can show that by injecting the upper bounds $\overline{\delta}_{ts}$ in the Reduced Problem (9)–(12), we can obtain a solution that is at most 1.5 times higher than the optimal solution, whereas by injecting the lower bounds $\underline{\delta}_{ts}$, we can obtain a lower bound that is no lower than 2/3 of the optimal solution. This is formally stated in the following proposition, whose proof can be found in Online Appendix B.

**Proposition 1.** *Let* $z^*$ *be the optimal objective value of* (1)–(8), $\delta^*_{ts}$ *defined by* (13) *and* $\boldsymbol{\delta}^*$ *the vector of* $\delta^*_{ts}$. *For* $2/3\boldsymbol{\delta}^* \leq \underline{\boldsymbol{\delta}} \leq \boldsymbol{\delta}^* \leq \overline{\boldsymbol{\delta}}_{\boldsymbol{ts}} \leq 3/2\boldsymbol{\delta}^*$, *let* $\underline{z} = z(\underline{\boldsymbol{\delta}})$ *and* $\overline{z} = z(\overline{\boldsymbol{\delta}})$, *defined by* (9)–(12). *Then*, $2/3z^* \leq \underline{z} \leq z^* \leq \overline{z} \leq 3/2z^*$.

The proof of Proposition 1 relies on the fact that the Objective Function (9) is nondecreasing in $\boldsymbol{\delta}$ and has a multiplicative scaling behavior, which allows the scaling of $\delta$s to be transferable to the scaling of the objective of our problem. Computing an upper bound $\overline{\delta}_{ts}$ with FFD or BFD has time complexity $\mathcal{O}(|C|\log|C|)$ and, therefore, overall $\mathcal{O}(|C||T||S|\log|C|)$. Once a solution of (13) is available, one should inject it in Formulation (9)–(12) and solve for $\omega$ and $\zeta_{ts}$ to obtain a solution of the original model. We show that (9)–(12) can be solved efficiently next.

### 4.2. Solving the Asset-Acquisition Model (9)–(12)

Model (9)–(12) can be solved to optimality efficiently for given $\delta_{ts}$. Concretely, for $\omega$ fixed, note that each $\zeta_{ts}$ will be positive and equal to $\delta_{ts} - \omega$ only when $\omega < \delta_{ts}$—that is, $\zeta_{ts} = \max\{\delta_{ts} - \omega, 0\}$—provided that there is enough asset capacity in each period, $K_t$. The model can then be solved by substituting $\zeta_{ts} = \max\{\delta_{ts} - \omega, 0\}$ and using a one-dimensional search over $\omega$. Proposition 2 characterizes formally the structure of optimal solutions of

Model (9)–(12), for the general case with bounds on spot assets.

**Proposition 2.** *The objective function of* (9)–(12) *is convex and piecewise linear in* $\omega$. *Further, let* $\omega_{min} = \max_{t\in T}\{\max_{s\in S}\{\delta_{ts}\} - K_t\}$ *and* $P = \{(t,s) : \delta_{ts} \geq \omega_{min}\}$. *Then, the optimal number of regular assets* $\omega^*$ *is in the set* $\Omega^* := \{\omega_{min}\} \cup \{\delta_{ts},\ \forall (t,s) \in P\}$, *and the optimal number of spot assets follows from* $\zeta^*_{ts} = \max(0, \delta_{ts} - \omega^*)$.

**Proof.** First, note that $\omega \geq \omega_{min}$ is necessary to ensure feasibility. In addition, because (9) is a minimization problem, Constraints (10) are binding at an optimal solution, and, therefore, $\zeta_{ts} = \min(K_t, \max(0, \delta_{ts} - \omega))$, which can be reduced to $\zeta_{ts} = \max(0, \delta_{ts} - \omega)$ because $\omega \geq \omega_{min}$. By substituting this expression for $\zeta_{ts}$ at the objective function, Problem (9)–(12) can be recast to a single-variable unconstrained optimization problem, as follows:

$$\min f(\omega) = \Lambda\omega + \Gamma\sum_{t\in T}\sum_{s\in S} p_{ts}\max(0, \delta_{ts} - \omega),$$

for $\omega \in \Omega := [\omega_{min}, \max\{\delta_{ts}\}]$. Function $f(\omega)$ is bounded and convex, as the sum of convex functions. Therefore, it attains a minimum value. To see that $f$ has linear segments, we fix $\omega$ to an arbitrary $\overline{\omega} \in \Omega\backslash P$ and denote $\overline{P} := \{(t,s) \in P : \delta_{ts} - \omega > 0\}$. Then, $f(\overline{\omega}) = \Lambda\,\overline{\omega} + \Gamma\sum_{(t,s)\in\overline{P}} p_{ts}(\delta_{ts} - \overline{\omega})$, and for a small $\epsilon > 0$, $\Delta f = f(\overline{\omega} + \epsilon) - f(\overline{\omega}) = \epsilon(\Lambda - \Gamma\sum_{(t,s)\in\overline{P}} p_{ts})$. Thus, the incremental difference of $f$ is linear in $\omega$ when $\omega$ does not take value at any of the points in $P$. Generalizing, if we reorder the terms $\delta_{ts}$ for $(t,s) \in P$ from smallest to largest, $\delta^1_{ts}, \ldots, \delta^l_{ts}$, we can see that $f$ is piecewise-linear in the intervals $[\delta^1_{ts}, \delta^2_{ts}], \ldots, [\delta^{l-1}_{ts}, \delta^l_{ts}]$. When $\omega$ increases beyond some $\delta^k_{ts}$, the term $\max(0, \delta^k_{ts} - \omega)$ vanishes from the objective, and the rate of change changes, correspondingly. Thus, because $f$ is convex and piecewise-linear, there exists a minimum that is attained at one of the breakpoints $\{\delta_{ts},\ \forall (t,s) \in P\}$ or at the minimum possible number of regular assets that ensures feasibility, $\omega_{min}$. □

An immediate consequence of convexity on $\omega$ is that the complexity of calculating the optimal value is $\mathcal{O}(\log(|T| \cdot |S|))$, because we can solve the problem using binary search.

In addition, the fact that function $f$ is piecewise-linear implies that we can characterize further the optimal solution if we have some information on the distributions of asset-requests demand and request duration. The following result assumes that the source uncertainties (demand and request durations) are independent, identically distributed (i.i.d.) and uses marginal analysis to characterize $\omega^*$ explicitly.

**Proposition 3.** *If* $\omega^*$ *is the optimal solution of* (9)–(12), *then* $\omega^* = \max\{\omega_{min}, \omega_{NB}\}$, *where* $\omega_{NB}$ *is the minimum value for which* $\mathbb{P}(\delta \leq \omega) > 1 - (\Lambda/(|T|\Gamma))$.

**Proof.** First, note that because $f$ is convex and its domain is bounded from below, the optimal solution will be either at the boundary $\omega_{min}$, if the global minimizer is infeasible, or $\omega_{NB}$, otherwise. In order to calculate $\omega_{NB}$, we can use an incremental analysis argument to see that the unit cost of one additional regular asset beyond $\omega_{NB}$ should be higher than the expected cost of spot assets. Concretely, $\Lambda - \Gamma\sum_t\sum_s p_{ts}I(\delta_{ts} > \omega_{NB}) > 0$, where $I(x) = 1$ if $x > 0$, and 0 otherwise. The expression $\sum_s p_{ts}I(\delta_{ts} > \omega_{NB})$ is the probability that we will need spot assets in period $t$—that is, $\mathbb{P}(\delta_t > \omega_{NB})$. Assuming i.i.d. uncertainties in request duration and number of requests across periods, the random variables $\delta$ are also i.i.d. Therefore, we can write the incremental condition as $\Lambda - \Gamma\sum_t\mathbb{P}(\delta > \omega_{NB}) = \Lambda - \Gamma|T| + \Gamma|T|\mathbb{P}(\delta \leq \omega_{NB}) > 0$, and, thus, $\mathbb{P}(\delta \leq \omega_{NB}) > 1 - (\Lambda/(\Gamma|T|))$. □

An alternative proof of Proposition 3 that sets the model up in a continuous domain is presented in Online Appendix C. The critical ratio $1 - (\Lambda/(|T|\Gamma))$ implies that the optimal number of regular assets $\omega^*$ is decreasing with the regular asset's price $\Lambda$ and increasing with the spot asset's price $\Gamma$ and the length of the horizon, $|T|$. Further characterizations require knowledge of the distribution of $\delta_{ts}$, which, however, is only possible to derive computationally, from the distributions of $\theta_{cts}$ and $\tau_{cts}$, as we show in our computational experiments. In what follows, we focus on solving the period/scenario-specific subproblems using column generation.

## 5. Using Column Generation to Solve the Subproblems

We utilize the column-generation formulation first formulated by Kantorovich in his seminal paper (Kantorovich 1960) to find lower and upper bounds for the subproblems in a computationally efficient manner. In the original cutting-stock problem, the goal is to cut paper rolls in a number of pieces of a certain length (not necessary identical lengths), while minimizing the total number of paper rolls that are needed. In our setting, the problem is to select *schedules* defining the number of service requests that an asset will satisfy of different customers during time period $t \in T$ under scenario $s \in S$—that is, a cutting pattern represents the allocation of an asset to customers. The set of all possible asset schedules is denoted by $Q$. Then, the service requests satisfied by an asset to customer $c \in C$ in asset schedule $q \in Q$ is signified by $v_{qc}$. For example, suppose we have three customers with the amount of time an asset is deployed to perform one service request equal to 20, 30, and 40 hours, respectively, and one asset available that can be in service for 100 hours per time period. Examples of feasible asset schedules are given by $[5\,0\,0]^T$, $[0\,3\,0]^T$, $[0\,0\,2]^T$, and $[0\,2\,1]^T$. Here, the $c-$th entry in asset schedule $q$ is represented by $v_{qc}$.

In what follows, we fix and drop the indexes $t$ and $s$ to avoid notation clutter. First, we define the integer variable $\lambda_q$, which represents the number of assets following schedule $q$. Then, Subproblem (13) can be formulated as follows.

$$\min \quad \sum_{q \in Q} \lambda_q, \tag{14}$$

$$\text{s.t.} \quad \sum_{q \in Q} \lambda_q v_{qc} = \theta_c, \qquad \forall c \in C, \tag{15}$$

$$\lambda_q \geq 0; \quad \lambda_q \in \mathbb{Z} \qquad \forall q \in Q. \tag{16}$$

The new model is presented in Equations (14)–(16). It is evident that the new formulation is no longer dependent on $|N|$. Here, the Objective (14) equals the total number of assets used in a specific period/scenario combination, which we want to minimize. Furthermore, Constraints (15) enforce service-demand requests to be fulfilled by a precise number of asset schedules, and (16) are domain constraints.

Enumerating all possible asset schedules that belong to the set $Q$ is computationally intractable, and, therefore, we use a column-generation scheme that starts with a limited subset of schedules $Q' \subset Q$ and adds schedules dynamically based on a pricing subproblem. The set $Q'$ is initialized using a schedule $q'$, in which assets serve the maximum number of requests to each customer—that is, $v_{q'c} = \lfloor \rho / \tau_c \rfloor$—for each $c \in C$. Then, after solving the linear programming (LP) relaxation of (14)–(16) using $Q'$ instead of $Q$, we retrieve the dual prices of (15), $\pi_c$, and solve the following pricing subproblem:

$$\alpha = \min \left\{ 1 - \sum_{c \in C} \pi_c q_c : \sum_{c \in C} \tau_c q_c \leq \rho;\ q_c \in \mathbb{N}_0,\ \ \forall c \in C \right\}. \tag{17}$$

As long as a column with negative reduced cost $\alpha < 0$ is found, it is added to the Master Problem (14)–(16), and the process repeats. Several algorithmic refinements are possible, such as reusing columns generated from (17) for a fixed $(t, s)$ pair to Master Problems (14)–(16) of other $(t, s)$ subproblems that share the same $\tau_{cts}$, but have different demand, or to stop the column generation prematurely in case the current best integer solution is smaller than the current lower bound in $\omega$. For a detailed account of these refinements, we refer the reader to Faro (2018).

The LP relaxation of (14)–(16) has a very strong lower bound in practice (Berge and Johnson 1977). It has been conjectured—and disproved, however—that rounding up the optimal objective value of (14)–(16) leads to an optimal solution of the integer problem: Marcotte (1985) showed that this is true only for certain classes of knapsack polyhedra. A weaker conjecture by Scheithauer and Terno (1995) claims that the optimal integer solution cannot be more than one unit higher than the rounded LP relaxation value. This has been proven true for $|C| \leq 6$ by Nitsche, Scheithauer, and Terno (1998) and remains open for other cases. In our computational experiments, we solve the linear programming relaxation of (14)–(16) and the corresponding integer program using the columns found during column generation. If the two values are less than one unit apart, the integer solution is optimal. Otherwise, we solve the original Formulation (13), pruning the tree using the best integer solution found during column generation, using a limit of 2,000 nodes.

## 6. Computational Experiments

A set of computational experiments are used to demonstrate the performance of the column-generation approach in combination with the Asset-Allocation Model (9)–(12) in comparison with a benchmark solution method by solving the NVCSD using CPLEX in terms of solution quality and computation times. All computational experiments are executed using CPLEX 12.8.0 with a maximum computation time of 7,200 seconds.

### 6.1. Case Environment

The computational experiments represent a specific problem of a Logistic Service Provider that is able to deploy trains leased at the beginning of the strategic planning phase or trains rented from the spot market throughout individual time periods. All problem instances can be defined by five features; (i) $|T|$ (time periods), (ii) $|C|$ (Customers), (iii) $|S|$ (Scenarios), (iv) cost ratio between spot and regular trains ($\Gamma|T|/\Lambda$), and (v) the load factor describing the maximum proportion of demand that can be served using spot trains only once per time period ($\sum_{c \in C} \theta_{cts} / K_t$). Table 2 shows the values of the features we consider; all combinations are possible, leading to 192 different instances. Regarding the number of chosen scenarios, we report further computational experiments on the variability of the obtained solutions in Online Appendix F. Note that our approach scales well with the size of the scenario set because each scenario requires the solution of an independent cutting-stock problem.

The 192 instances are divided into 54 easy, 90 medium, and 48 difficult instances, based on the average computation time of the column-generation (CG) approach. The detailed classification of input combinations to Easy, Medium, and Difficult can be found in Table 6 in Online Appendix D. Reflecting upon the hardness

**Table 2.** Overview of the Values of the Features in the Problem Instances

| Parameter | $\lvert T\rvert$ | $\lvert C\rvert$ | $\lvert S\rvert$ | $\Gamma\lvert T\rvert/\Lambda$ | $\sum_{c \in C} \theta_{cts}/k_t$ |
|---|---|---|---|---|---|
| Values | {26, 52} | {2, 4, 8, 16} | {5, 10, 20, 50} | {1.3, 2.4} | {0.6, 1} |

classification reveals that the key dimensions that influence computational performance are the number of customers and the number of scenarios. Indeed, instances with eight or 16 customers are either Medium or Difficult, whereas instances with 16 customers (the maximum considered) are Difficult, unless they have only five scenarios. Instances with 50 scenarios (the maximum considered) are distributed across Easy, Medium, and Difficult, whereas the number of Time Periods seems to be the less influential element. These findings make intuitive sense because the key difficulty of the formulation lies in the combinatorial nature of the cutting-stock subproblems, which become harder with a larger number of customers and more in number with a larger number of scenarios. The number of scenarios is more influential than the number of periods, for our instances, because increasing the problem by a period generates $|S|$ more cutting-stock-like "subproblem" structures, whereas increasing it by one scenario creates $|T|$ ones, and $|T| > |S|$.

### 6.2. Computational Benchmarks

Table 3 gives a summary of the performance of CPLEX and CG. All results are averages over the instances related to the specific problem size. Note that only the averages are taken over the instances for which CPLEX was able to find a solution, meaning that the results of 42 instances are not included in the first nine columns. The results for these 42 instances using the CG approach are presented in Table 7 in Online Appendix E. We consider two major performance indicators: solution quality and computation time. The solution quality is demonstrated by the values for lower bound (LB), upper bound (UB), gap sizes, and whether a feasible solution has been found. Regarding the easy instances, the lower bound, upper bound, gap sizes, and number of feasible solutions found are identical. In all instances, both approaches were able to find the optimal solution. The only difference is with respect to the computation times, which are larger for CG, but still small in absolute terms. The medium and difficult instances demonstrate a different perspective. Both lower and upper bounds found by CG are better than CPLEX. This is further reflected by the gap size, which signifies that the solutions found by CG are closer to optimality. In addition, CPLEX was not able to find a feasible solution in 42 (seven medium and 35 difficult) instances. Finally, the CG approach also outperforms CPLEX in terms of CPU time. In summary, our approach shows superior computational performance, which hinges on decomposing the original formulation to subproblems that can be solved efficiently and on using these solutions to find an optimal solution of the remaining problem, which reduces to a single-variable convex minimization problem. Finally, note that increasing the number of periods or scenarios would result in a linear increase in computational time for our approach, which is not the case for the original formulation used by CPLEX.

## 7. Impact of Uncertainty on Asset Allocation and Performance

In this part, we investigate how uncertainty affects the structural characteristics of optimal solutions. Concretely, we seek to understand how the magnitude and variability of customer requests and their duration influence the overall expected cost and the mix of spot and regular assets. Both uncertainty dimensions—namely, the number of customer requests and their duration—influence the total number of assets required in each period and scenario. Although it is reasonable to hypothesize that longer durations and more customer requests will result in more assets, it is not trivial to quantify each individual impact on the overall cost and on the optimal proportion of spot resources. To answer this question, one should characterize the dependency of the distribution of required assets on the durations and customer requests. This is challenging (if at all possible) to do analytically because the mapping between source uncertainty distributions—namely, requests and request duration—to the distribution of interest (minimum number of assets) requires the solution of a combinatorial optimization problem for each sampled realization. We tackle this question numerically, by designing a set of instances that vary in requests and their durations in a controlled randomized fashion.

### 7.1. Instance Generation

To this end, we have designed a new set of instances with four customers, 10 periods, and 10 scenarios per period. This is a problem size that we could always solve to optimality, as the optimal solution is at most one unit above the lower bound obtained by column

**Table 3.** Performance Indicators for CPLEX and Column Generation (CG)

| | LB | | UB | | Gap (%) | | CPU (s) | | No solution | |
|---|---|---|---|---|---|---|---|---|---|---|
| Problem class | CPLEX | CG | CPLEX | CG | CPLEX | CG | CPLEX | CG | CPLEX | CG |
| Easy | 134.72 | 134.72 | 134.72 | 134.72 | 0 | 0 | 1.23 | 15.86 | 0 | 0 |
| Medium | 200.85 | 202.58 | 205.24 | 202.77 | 1.26 | 0.03 | 3,978 | 255 | 7 | 0 |
| Difficult | 298.33 | 305.89 | 322.35 | 307.31 | 7.11 | 0.01 | 7,200 (TL) | 1,728 | 35 | 0 |

*Note.* TL denotes time limit.

generation (Nitsche, Scheithauer, and Terno 1998). The probability of each scenario in each period is drawn from a uniform distribution and normalized so that all scenario probabilities per period add up to 100%. We then vary the following parameters, with the intention to cover a broad range of configurations.

**7.1.1. Duration of Customer Requests.** We normalize durations between zero and one, reflecting the proportion of time that a customer request—that is, a return trip to a customer—may take and sample them from truncated logNormal distributions. Travel times, in general, have a known minimum value, which is the fastest possible time in which a given distance can be covered. In case of small, frequent delays, travel times deviate upward, creating high probability concentrations. Larger delays tend to be more infrequent and are captured by the long tail of logNormal distributions. Such distributions have been used extensively in the literature to model travel-time duration (Strathman and Hopper 1993, Hollander and Liu 2008, Guessous et al. 2014, Bertsimas et al. 2019, Chen and Fan 2020). Average duration has three levels—namely, Low, Medium, and High, corresponding to 0.1, 0.4, and 0.7 of total time capacity, respectively. Variability also has Low, Medium, and High levels, selected so that the average coefficient of variance of each level is 8%, 31%, and 53%, respectively. Figure 2 shows the frequencies of two instances with Low/High and High/Medium average and variability, respectively.

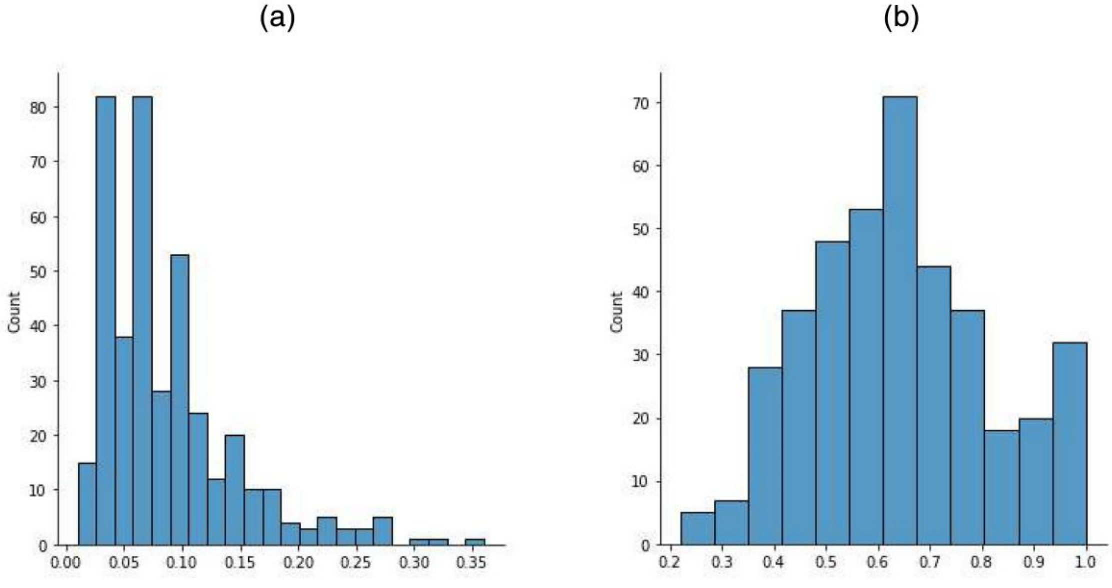
**7.1.2. Customer Requests.** We sample customer requests from discrete uniform distributions having three average levels, Low, Medium, and High, corresponding to 10, 20, and 30 requests, respectively, and, similarly, three variability levels, reflecting business cycles of varying market uncertainty. To this end, Figure 3 shows two instances with Low/Low and High/High average and variability in customer requests.

The different configurations capture different demand scenarios. Preliminary experiments with higher averages and variability level showed qualitatively similar results.

**7.1.3. Cost Ratio.** Another factor influencing the solution is the cost ratio between spot and regular resources. Without loss of generality, we normalize $\Gamma = 1$ and define the cost ratio as $cr = \Gamma|T|/\Lambda$. Observe that $cr > 1$; otherwise, it would be optimal to use spot resources only. We thus design a High and a Low condition with $cr \in \{4, 1.3\}$, respectively. Note that high $cr$ values make the option to acquire spot assets unattractive. Because of the application-agnostic nature of our study, we selected two $cr$ values that represent different extremes in terms of attractiveness of spot assets. In specific applications (such as rail-based LSPs), $cr$ will have a more well-defined range and is likely to be scenario- and period-specific, as the output of a spot-price forecasting model (Boin et al. 2020).

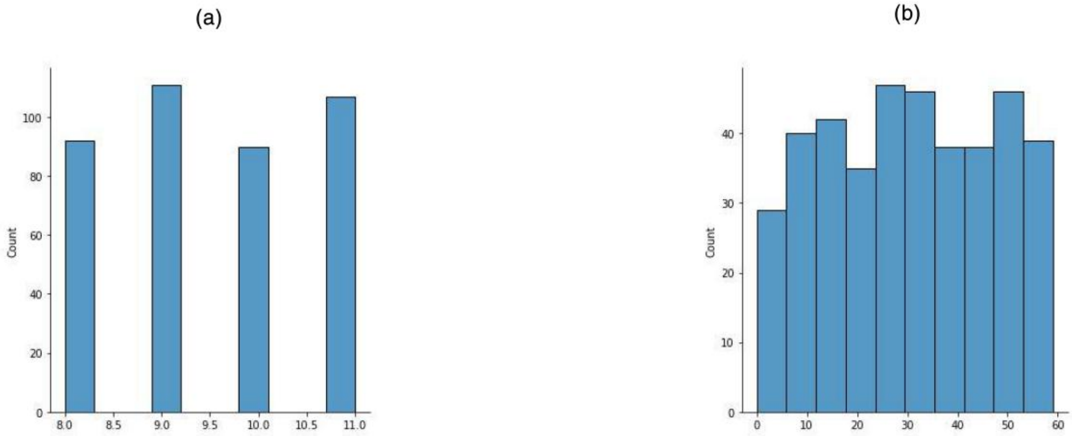
**7.1.4. Load Factor.** Furthermore, we vary the capacity of spot resources so that we obtain uncapacitated instances and instances with High and Low load factors, where low load factors indicate harder spot capacities. We focus on uncapacitated instances and report detailed results for capacitated instances in the online appendix because capacity constraints may overshadow the influence of

**Figure 2.** (Color online) Customer Request Duration Distribution Illustrations



*Notes.* (a) Low average, high variability (CV = 70%). (b) High average, medium variability (CV = 30%; truncated).

**Figure 3.** (Color online) Distribution of Customer Requests



*Notes.* (a) Low average, low variability. (b) High average, high variability.

other factors, as they have a direct impact on the balance between regular and spot assets.

**7.1.5. Correlations.** Finally, we also experimented with correlations between travel times and customer requests. In particular, we constructed instances where travel times and customer requests are uncorrelated, or positively (0.80) or negatively (−0.80) correlated. We report these results in the online appendix, as the conclusions were similar to the conclusion of previous results.

An overview of the instance classification is given in Table 4. Our factorial design entails 972 capacitated instances, of which half have a high and half have a low load factor, and 972 uncapacitated instances, for a total of 1,944 instances.

Note that our design allows us to investigate how each factor influences total costs and asset allocation because it is fully balanced across all dimensions. For example, Table 5 shows that partitioning the instances with respect to their average service-duration category gives subsets with identical average and standard deviation of requests, as well as correlations. The standard deviation of service duration is proportional to the average duration, having a near-constant coefficient of variation. This is the case when slicing other parameters as well.

The distributions we sample from are identical for each time period and for each customer. As such, they can be seen as time-stationary, with homogeneous customers. Note that, despite the symmetry between periods and scenarios in the formulation, having a problem with $|T|$ periods and $|S|$ scenarios per period is not equivalent to having a single period problem with $|T| \cdot |S|$ scenarios because the probability weights $p_{ts}$ in the objective function are different, because $\sum_s p_{ts} = 1$, for each period $t$.

We next report on the influence of each of the parameters on the overall cost and optimal asset allocation, the first-order effects for the 972 uncapacitated instances.

## 7.2. First-Order Effects

We first report on the impact of varying average customer requests and request duration. To this end, Figure 4 shows the optimal cost and proportion of spot cost averaged over each level of expected customer requests and customer request duration.

Figure 4 implies that increasing either the average number of service requests or the average duration of each request results in higher costs for the LSP. This is expected, because in both cases, more assets are required to accommodate a growing customer demand. The proportion of spot cost appears to remain largely constant across levels for service requests and request duration. For request duration, the spike in the middle level is statistically significant compared with the other two levels (using paired t-tests at the 5% level). This might be related to the asymmetric effect of the truncating logNormal duration of medium and high levels of request duration. Specifically, although variability for

**Table 4.** Overview of Qualitative Experiments

| Service duration | | Customer requests | | | | |
|---|---|---|---|---|---|---|
| Mean | Variability | Mean | Variability | Cost ratio | Load factor | Correlation |
| H, M, L | H, M, L | H, M, L | H, M, L | H, L | H, L | 0, −0.80, 0.80 |

*Note.* H, high; M, medium; L, low.

**Table 5.** Slices with Respect to Service Duration (SD)

| Slice | SD avg | SD std | Requests avg | Requests std | Correlation avg |
|---|---|---|---|---|---|
| H | 0.62 | 0.17 | 19.51 | 6.77 | 0.0 |
| M | 0.37 | 0.12 | 19.52 | 6.77 | 0.0 |
| L | 0.09 | 0.03 | 19.49 | 6.80 | 0.0 |

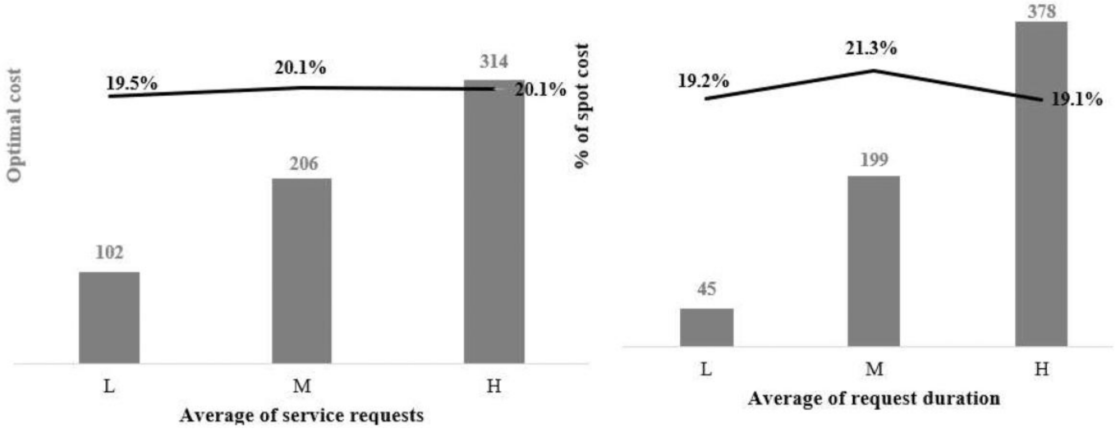
*Note.* Standard deviation (std) of service duration is proportional to average (avg) service duration, having a near-constant coefficient of variation.

low average duration can be accommodated by packing, and for high average duration, it may be less in magnitude, due to truncating, this is not the case for the medium duration—although the effect size is small. What is worth noticing is that the optimal cost appears to increase linearly with the number of requests and superlinearly with the request duration. Specifically, an increase from 10 to 20 requests (low to medium level) results in an average cost increase of 104 units, whereas a further increase to 30 requests (high level) increases cost by 108 units. Alternatively, every percent increase in the average number of requests is followed by a similar percent increase in average costs. For request duration, an increase from 0.1 to 0.4 (low to medium level) results in a more than four-fold increase in costs, whereas a further increase from 0.4 to 0.7 (that is, 75%) increases cost by about 80%. Comparing the overall increase in cost from low to high average request duration, we see that a six-fold increase in average request duration causes a more than seven-fold increase in costs. A possible explanation of this association is that low service duration results in more flexibility when it comes to allocating assets to requests. In other words, the direct mapping between assets and requests is weakened because each asset can accommodate many requests. This effect is diminished when the duration of each request approaches its maximum, where in the limit every request duration is so large that a dedicated asset is required. This is also the reason why the overall range of optimal cost when varying duration is wider than when varying the number of requests. When duration varies from near zero to 100% of time capacity, the corresponding number of assets varies from one to the number of requests, whereas the effect of varying the number of requests is more limited.

When it comes to the split between regular and spot assets, we observe that varying the average number of requests or their average duration does not seem to exert a systematic influence. The key takeaway, therefore, is that the average customer-request duration seems to have an overall stronger effect on the optimal cost compared with the average number of requests. This is manifested by a wider range of optimal costs, which appear to follow duration increases at near-constant rate.

The behavior of our model has some similarities with a standard newsvendor model. Concretely, if we assume that assets follow a Normal Distribution $\mathcal{N}(\mu,\sigma)$ in a single period, the optimal regular assets can be written as $\omega^* = \mu + \kappa(CR)\sigma$, where $\kappa(CR)$ is the safety factor, determined by the critical ratio $CR$. Because the critical ratio depends on the overage ($\Lambda$) and underage ($\Gamma|T| - \Lambda$) costs, it remains constant, and we can write $\omega^* = \mu(1 + \kappa(CR)CV)$, which shows that an increase in $\mu$ that preserves the coefficient of variation results in a linear increase in $\omega^*$. The resulting increase in costs will be convex because total cost is a convex function depending on $\omega^*$. Furthermore, in such a model, the fraction of spot costs would remain intact under a $CV$-preserving transformation. This is because such a transformation implies $\Delta \rightarrow \Delta(1+\lambda)$ for some given $\lambda$ (to see this, note that $\mathbb{E}[\Delta] \rightarrow (1+\lambda)\mathbb{E}[\Delta]$ and $\sigma_\Delta \rightarrow (1+\lambda)\sigma_\Delta$). The proportion of spot cost, in turn, depends on the ratio $\mathbb{E}[(\Delta - \omega^*)^+]/\omega^*$, which remains constant under such a transformation. Although such insights are in line with the behavior observed in Figure 4, the combinatorial

**Figure 4.** Impact of Customer Requests (Number and Duration) over Cost and Asset Allocation



*Note.* The black line shows the proportion of spot cost, and the bar chart shows the average optimal cost of each slice.

dependency of assets on service requests and request duration prohibits analytical derivations that relate changes in average service requests and request duration directly to the cost or proportion of spot assets.

### 7.3. Second-Order Effects

We next investigate how the optimal cost and the allocation to spot or regular assets change when we change the variability of the number of requests and of their duration. Results are reported in Figure 5.

When it comes to the optimal cost values, we see that higher levels of service-request variability increase costs, whereas higher variability in request duration leads to lower costs. Regarding optimal cost allocation, the proportion of spot cost increases with the variability of either service requests or request duration. To shed light on these findings, we examine the cumulative distribution functions (cdfs) of the corresponding $\delta^*$.

To this end, Figure 6 shows the $\delta^*$ cdf for each level of service-request variability. We see that, although overall, the cdfs that correspond to different variability levels are similar, high variability in $\theta_{cts}$ leads to a long tail in the cdf of $\delta^*$. In practice, this means that there are a number of limited scenarios that require significantly more assets, which are then accommodated by more expensive spot acquisitions. This increases not only the overall cost, but also the proportion of spot cost over the overall cost.

Next, Figure 7 shows the cdf of $\delta^*$ for each level of request duration.

**Figure 6.** (Color online) $\delta^*$ for Varying Levels of Customer-Request Variability (D_STD)

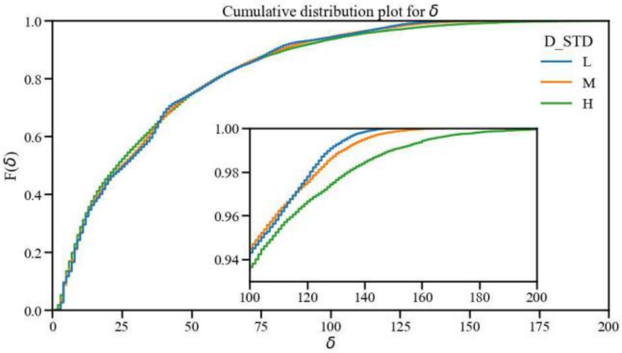


Figure 7 reveals that, although the support of all three cdfs is the same, for a given level of $\delta^*$, we can cover *fewer* scenarios under low request-duration variability. Given that request-duration variability follows a logNormal distribution (which is common for travel times), the high-variability condition implies that the bulk of duration density is lower than average, although some extreme scenarios do exist. However, a limited number of high-duration instances can likely be accommodated by a given number of assets because they are packed with shorter services. In contrast, the lower-variability condition implies that the duration distribution is more symmetric (for the same average), which results in a limited capacity to accommodate requests of high duration with the same number of assets. In Online Appendix G, we report additional computations using uniformly distributed request-duration times and also find that higher variability leads to lower costs, albeit only for requests that, on average, exceed 0.5—namely, half the available time in a period. Finally, the fact that for any given $\delta$, there are more scenarios that exceed it in the low-variability condition makes regular assets more attractive. As variability increases, it is less likely to exceed any given level of resources, and, therefore, the proportion of spot cost increases accordingly.

We report on additional analysis in Online Appendix I, including varying levels of correlation between requests and duration and spot-resource capacity levels. It is worth noting that when request duration has low variability, correlations do not have a significant effect on cost. For medium and high levels of duration variability, however, negative correlations have a higher effect of diminishing the overall cost than positive correlations. This is intuitive because with positive correlations, there

**Figure 5.** Impact of Customer Request Variability (Number and Duration) over Cost and Asset Allocation

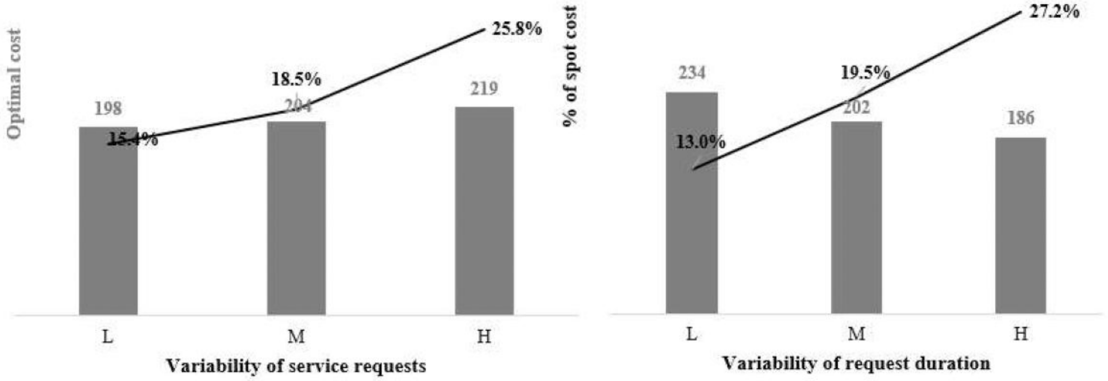
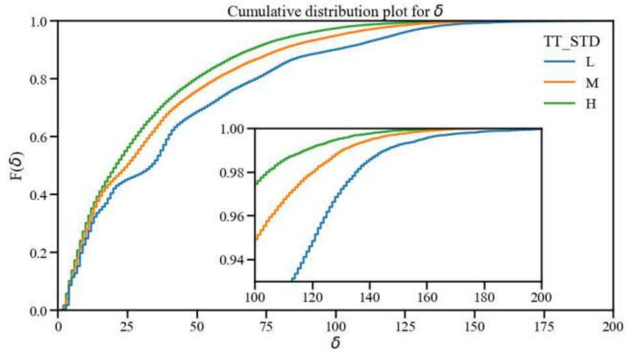


**Figure 7.** (Color online) $\delta^*$ for Varying Levels of Request Duration Variability (TT_STD)



exist scenarios with limited flexibility of utilizing fewer assets for customer requests. Finally, it is interesting to note that when the spot-resource capacity constraints are tight, the solutions have a larger number of regular assets in order to make the problem feasible. Consequently, this increases the total cost and diminishes the positive effects of service-duration variability.

## 8. Extensions

Our core formulation can be extended to include additional costs and different operational characteristics. In this section, we give examples of how fixed asset costs and operational costs can be incorporated in our model. Further, we show how our model can be extended to incorporate asset routing.

### 8.1. Fixed Costs and Outsourcing Operations

Model (1)–(8) can accommodate fixed costs of deployed assets by adding the term $\sum_{n\in N}\sum_{t\in T}\sum_{s\in S} f_n^a p_{ts} y_{nts}$ in the Objective Function (1), where $f_n^a$ is the cost of deploying asset $n$ in a time period. In this case, the optimal solution of the new model remains identical to that of (1)–(8) because the latter minimizes the total number of assets deployed in each period. However, because assets have different fixed deployment costs $f_n^a$, some relabeling may be required to ensure that the least expensive $\sum_n y_{nts}$ assets are deployed under scenario $s$ in period $t$. This can be achieved by presorting the assets in increasing order of their deployment costs. Note also that fixed costs of service requests are inconsequential for the formulation because each customer has a fixed number of requests per period and scenario. As such, having a fixed cost of service requests adds a constant term in the Objective Function (1).

More interesting is the case where spot assets represent outsourced operations. To see the difference, consider an LSP who requests a fleet of 10 trucks to operate for a single week, at a fixed price per truck. This is the situation of spot-asset *acquisition*, which implies that the operational costs of using the trucks are identical, regardless of whether the trucks are owned by the LSP or rented for one week, and is captured by Model (1)–(8). If, however, the same fleet of trucks is operated by an external third party, which charges a fixed price per truck and a price per trip (which can be different among spot assets), the LSP needs to consider the additional operational costs. For this, let $f_{cn}^o$ be the marginal operational costs (relative to operating own assets) of a trip to customer $c$ using spot asset $n$, $z_{nts} \in \{0,1\}$ a variable that captures whether asset $n$ is outsourced in period $t$ under scenario $s$, and $o_{cnts}$ the number of requests/trips of customer $c$ accommodated by spot asset $n$ in period $t$ under scenario $s$. The corresponding model is an extension of (1)–(8) and can be formulated as follows.

$$\min \quad \Lambda\omega + \Gamma\sum_{t\in T}\sum_{s\in S} p_{ts}\zeta_{ts} + \sum_{t\in T}\sum_{s\in S}\sum_{n\in N}\sum_{c\in C} p_{ts} f_{cn}^o o_{cnts}, \tag{18}$$

$$\text{s.t.} \quad \omega + \zeta_{ts} \geq \sum_{n\in N} y_{nts}, \qquad \forall t\in T,\ \forall s\in S, \tag{19}$$

$$\sum_{n\in N} a_{cnts} = \theta_{cts}, \qquad \forall c\in C,\ \forall t\in T,\ \forall s\in S, \tag{20}$$

$$\sum_{c\in C} \tau_{cts} a_{cnts} \leq \rho y_{nts}, \quad \forall n\in N,\ \forall t\in T,\ \forall s\in S, \tag{21}$$

$$z_{nts} \leq y_{nts}, \qquad \forall n\in N,\ \forall t\in T,\ \forall s\in S, \tag{22}$$

$$\sum_{n\in N} z_{nts} = \zeta_{ts}, \qquad \forall t\in T,\ \forall s\in S, \tag{23}$$

$$o_{cnts} \leq M z_{nts}, \qquad \forall c\in C,\ \forall n\in N,\ \forall t\in T,\ \forall s\in S, \tag{24}$$

$$o_{cnts} \geq a_{cnts} + (z_{nts} - 1)M, \qquad \forall c\in C,\ \forall n\in N,\ \forall t\in T,\ \forall s\in S, \tag{25}$$

$$0 \leq \zeta_{ts} \leq K_t; \quad \zeta_{ts}\ \text{integer} \qquad \forall t\in T,\ \forall s\in S, \tag{26}$$

$$y_{nts}, z_{nts} \in \{0,1\}, \qquad \forall n\in N,\ \forall t\in T,\ \forall s\in S, \tag{27}$$

$$a_{cnts}, o_{cnts} \geq 0; \quad a_{cnts}, o_{cnts}\ \text{integer} \qquad \forall c\in C,\ \forall n\in N,\ \forall t\in T,\ \forall s\in S, \tag{28}$$

$$\omega \geq 0; \quad \omega\ \text{integer}. \tag{29}$$

The Objective Function (18) minimizes the regular- and spot-asset costs and the expected operational costs. Equations (22) indicate that an asset $n$ can only be outsourced if it is used, and (23) posits that the total number of outsourced assets should equal the number of spot assets. Then, (24) and (25) enforce that when spot asset $n$ is outsourced, the number of trips it performs, $a_{cnts}$, is recorded in variable $o_{cnts}$ and essentially linearizes the relationship $o_{cnts} = a_{cnts}z_{nts}$ using a big $M$ parameter. In summary, (18)–(29) uses two new sets of variables: $z_{nts} \in \{0,1\}$, which captures whether an asset is outsourced, and $o_{cnts}$, which accounts for the additional operational costs that are incurred when an asset is outsourced. The following proposition gives insight on the structure of optimal solutions of (18)–(29).

**Proposition 4.** *An optimal solution of* (18)–(29) *has* $y_{nts}$ *values that satisfy* (13), *and, therefore, minimizes the total number of assets in each period-scenario pair.*

**Proof.** Fix $t$ and $s$. Suppose there is an optimal solution of (18)–(29) with some $y_{nts}$ such that $\sum_n y_{nts} > \sum_n y^*_{nts}$, where $\sum_n y^*_{nts}$ is an optimal solution of (13). Note that $y^*_{nts}$ is feasible for (20) and (21) for the given $t$ and $s$, and, therefore, substituting $y_{nts}$ with $y^*_{nts}$ results in a feasible solution for (18)–(29). We will then show that this solution cannot have a higher objective than the assumed optimal one. There are two cases. First, if $\zeta_{ts} = 0$, then both solutions have an equal objective value because no assets are outsourced in pair $(t, s)$. Second, if $\zeta_{ts} > 0$, given that $\sum_n y_{nts} > \sum_n y^*_{nts}$, there exists some asset $\overline{n}$ that is not used in $y^*_{nts}$ and is used in $y_{nts}$, which, due to (22) and (23), implies that we can eliminate asset $\overline{n}$ from the spot trains and obtain $\zeta^*_{ts} < \zeta_{ts}$, and, therefore, improve the objective value by $\Gamma p_{ts}$. Finally, if a spot asset $\overline{n}$ can be eliminated, the customer requests that it serves will be redistributed to either other spot assets, to regular assets, or to a combination of both (otherwise, $y^*_{nts}$ would had been infeasible). In any case, the objective function value can only improve. □

Proposition 4 shows that an important decomposability property of our original Model (1)–(8) is preserved in this setting—namely, that the $y_{nts}$ values can be found by minimizing the total number of assets in pair $(t, s)$. However, because operational costs are not homogeneous, we need to take care to use spot assets that have the lowest operational costs in priority. The key idea is then to bring this model into a form similar to the Reduced Newsvendor Model (9)–(12), but where changes in the values of $\zeta_{ts}$ capture increasing operational costs. The objective function of the new model will take the form $\sum_{t\in T}\sum_{s\in S} p_{ts} G_{ts}(\zeta_{ts})$ and is calculated for a fixed $(t, s)$ pair by the following procedure:

- Obtain $y^*_{nts}$ and $a^*_{cnts}$ from (13).
- For each asset $n \in N$, calculate its operational cost, assuming it is outsourced: $g^o_n = \sum_c f^o_{cn} a^*_{cnts}$.
- Sort assets from smaller to largest costs $g^o_n$. Let $n_1, \dots n_\tau$ and $g^o_{n_1}, \dots, g^o_{n_\tau}$ be the corresponding asset and cost rankings, respectively.
- Define the function $G : \mathbb{Z} \to \mathbb{R}$ with $G(\zeta) = G(\zeta - 1) + \Gamma + g^o_{n_\zeta}$ for $\zeta > 0$ and $G(\zeta) = 0$, for $\zeta = 0$. By construction, $G$ has increasing differences.

By this procedure, (18)–(29) can be written as

$$\min \ \Lambda\omega + \sum_{t\in T}\sum_{s\in S} p_{ts} G_{ts}(\zeta_{ts}), \tag{30}$$

$$\text{s.t. } \omega + \zeta_{ts} \geq \delta_{ts}, \qquad \forall t \in T, \quad \forall s \in S, \tag{31}$$

$$\omega \geq 0, \qquad \omega \text{ integer}, \tag{32}$$

$$\zeta_{ts} \in \{0, 1, \dots, K_t\}, \qquad \forall t \in T, \quad \forall s \in S. \tag{33}$$

Because of the convexity of the objective function, it holds that $\zeta_{ts} = \max\{\delta_{ts} - \omega, 0\}$, and, therefore, the objective function can be written as

$$f(\omega) = \Lambda\omega + \sum_{t\in T}\sum_{s\in S} p_{ts} G_{ts}(\max\{\delta_{ts} - \omega, 0\}),$$

which is convex and can be solved with single-variable search methods. In conclusion, incorporating the case of outsourcing can be accommodated by our model with some additional care to take into account the operational cost of spot assets.

## 8.2. Routing

Finally, the modular structure of our formulation allows us to modify the period/scenario subproblems and model richer transportation settings. One common such setting is when assets can be used to serve several customer requests combined, via routing. In Online Appendix H, we show how the asset-minimizing Subproblem (13) can be extended to incorporate routing, giving rise to a periodic vehicle routing problem (Cordeau, Gendreau, and Laporte 1997; Baldacci et al. 2011; Archetti, Fernández, and Huerta-Muñoz 2017), where each period-scenario pair is partitioned to microperiods, in which the routing occurs. For this model, we report on the results of two computational studies that lead to some key insights. First, the impact of average request duration can be split in two regimes: Below a certain threshold (0.5 in our experiment), there is no impact on cost, whereas above this threshold, cost increases with average request duration in a convex fashion. The absence of impact for small average values can be explained from the relative robustness of the objective, which minimizes the total number of assets used within a period. Second, the impact of customer-request variability follows a more involved pattern, in that higher variability does not necessarily imply lower costs. For small average request durations, their variability does not influence cost. However, for averages within a certain range (0.5 and 0.65 in our experiment), lower variability leads to lower costs, whereas beyond this range, this relationship is reversed, and lower variability is associated with higher costs. This reversal can be explained by the probability that a route is time-feasible, which we calculate analytically. In particular, we show that lower variability almost guarantees the existence of a time-feasible route for average durations up to a threshold (0.60), but beyond this threshold, the probability that a time-feasible route exists is reduced very fast: It drops to 65% for 0.65, and then to 20%. For higher variability levels, the probability of time-feasible routing is generally lower, but it drops at a slower pace as time variability increases (to 57% and 38%, respectively). Intuitively, the higher the average request duration, the less routing opportunities exist and the more the problem resembles the no-routing setting, where higher variability is associated with lower cost.

## 9. Discussion and Future Research

We have studied the strategic acquisition of transportation assets, a problem faced by LSPs, emphasizing that demand for services is expressed in the form of service requests and time during which the servicing asset is occupied. The motivation for developing this formulation originates in the problem described by Wagenaar, Fragkos, and Zuidwijk (2021), where optimizing the integrated tactical and operational planning of multimodal networks has to abide by the constraints imposed by wagon acquisitions made during strategic planning. This situation—namely, resource constraints imposed by decisions taken at the strategic level—is often found in an array of other transportation problems, and, as such, the trade-off captured by our formulation is pertinent.

An interesting avenue of research is to investigate in more detail the impact of uncertainty in requests and request duration on the planning cost. The current work shows computationally that variability in requests and in duration has a negative and a positive effect, respectively, but adding hard capacity constraints on the availability of spot resources diminishes the positive effect of duration variability, whereas when asset routing is feasible, this relationship can be reversed. It would be interesting to see in which settings this insight persists and in which ones other factors exert larger effects on the cost. Our results should be refined further, when additional factors are at play. For example, in our computational study, we consider time-stationary distributions and homogeneous customers, but it would be more realistic to utilize a forecasting model with spot seasonality and trend. Other important extensions are when operational constraints span multiple periods and when acquisition of regular assets can occur in every period for any number of remaining periods. Such models are significantly harder because the LSP needs to have a decision policy that maps every possible future information state to an action vector of resource acquisitions, and they call for different modeling approaches and solution methods.

In addition, the effect of average and variance of demand can be generalized to higher-order moments, especially in cases of heavy-tailed distributions, such as power laws. From a methodology perspective, characterizing analytically the distribution of the minimum number of assets $\delta_t^*$ using the distributions of requests, $\theta_{ct}$, and duration, $\tau_c$, can lead to further insights regarding how these uncertainties influence the magnitude and structure of optimal solutions. Padmanabhan et al. (2021) compute the cumulative distribution of the optimal objective value and provide complexity results for some classes of problems, when information about the marginal distributions of the objective coefficients is available. Our setting is more involved because uncertainty lies in the technology-matrix coefficients. Finally, another potential extension of our model is to consider cases where assets such as wagons can be merged together to serve customer requests. This extension require the solution of one-dimensional cutting-stock problems with multiple stock lengths (Poldi and Arenales 2009), which is significantly harder than the archetypal, single-length version. We hope our research inspires further advancements in such directions.

## Acknowledgments

The authors thank SURFsara (www.surfsara.nl) for the support in using the Lisa Compute Cluster.

## References

Archetti C, Fernández E, Huerta-Muñoz DL (2017) The flexible periodic vehicle routing problem. *Comput. Oper. Res.* 85:58–70.

Baldacci R, Bartolini E, Mingozzi A, Valletta A (2011) An exact algorithm for the period routing problem. *Oper. Res.* 59(1):228–241.

Baldacci R, Battarra M, Vigo D (2008) Routing a heterogeneous fleet of vehicles. Golden B, Rahgavan S, Wasil E, eds. *The Vehicle Routing Problem: Latest Advances and New Challenges* (Springer, Boston), 3–27.

Baykasoğlu A, Subulan K, Taşan AS, Dudaklı N (2019) A review of fleet planning problems in single and multimodal transportation systems. *Transportmetrica A Transportation Sci.* 15(2):631–697.

Berge C, Johnson EL (1977) Coloring the edges of a hypergraph and linear programming techniques. *Ann. Discrete Math.* 1:65–78.

Bertsimas D, Delarue A, Jaillet P, Martin S (2019) Travel time estimation in the age of big data. *Oper. Res.* 67(2):498–515.

Boin R, Gavin R, Rau P, Stoffels J (2020) Getting the price right in logistics. Accessed March 2, 2023, https://www.mckinsey.com/industries/travel-logistics-and-infrastructure/our-insights/getting-the-price-right-in-logistics.

Chen RR, Cheng T, Choi TM, Wang Y (2016) Novel advances in applications of the newsvendor model. *Decision Sci.* 47(1):8–10.

Chen Z, Fan WD (2020) Analyzing travel time distribution based on different travel time reliability patterns using probe vehicle data. *Internat. J. Transportation Sci. Tech.* 9(1):64–75.

Cordeau JF, Gendreau M, Laporte G (1997) A tabu search heuristic for periodic and multi-depot vehicle routing problems. *Networks* 30(2):105–119.

Crainic TG, Fomeni FD, Rei W (2021) Multi-period bin packing model and effective constructive heuristics for corridor-based logistics capacity planning. *Comput. Oper. Res.* 132:105308.

Crainic TG, Gobbato L, Perboli G, Rei W (2016) Logistics capacity planning: A stochastic bin packing formulation and a progressive hedging meta-heuristic. *Eur. J. Oper. Res.* 253(2):404–417.

Crainic TG, Gobbato L, Perboli G, Rei W, Watson JP, Woodruff DL (2014) Bin packing problems with uncertainty on item characteristics: An application to capacity planning in logistics. *Procedia Soc. Behav. Sci.* 111:654–662.

Faro M (2018) Optimizing strategic decision making in a multimodal transportation setting. Master's thesis, Tilburg University, Tilburg, Netherlands. https://www.dropbox.com/sh/o579ummzbjy5z9a/AAD3y7MyBFqmeI-IETkjTN1ga?dl=0.

Fragkos I, De Reyck B (2016) Improving the maritime transshipment operations of the noble group. *Interfaces* 46(3):203–217.

Guessous Y, Aron M, Bhouri N, Cohen S (2014) Estimating travel time distribution under different traffic conditions. *Transportation Res. Procedia* 3:339–348.

Hollander Y, Liu R (2008) Estimation of the distribution of travel times by repeated simulation. *Transportation Res. Part C Emerging Tech.* 16(2):212–231.

Hua G, Wang S, Cheng T (2012) Optimal pricing and order quantity for the newsvendor problem with free shipping. *Internat. J. Production Econom.* 135(1):162–169.

Jabali O, Gendreau M, Laporte G (2012) A continuous approximation model for the fleet composition problem. *Transportation Res. Part B Methodological* 46(10):1591–1606.

Jiang Y, Shang J, Liu Y (2013) Optimizing shipping-fee schedules to maximize e-tailer profits. *Internat. J. Production Econom.* 146(2): 634–645.

Kantorovich LV (1960) Mathematical methods of organizing and planning production. *Management Sci.* 6(4):366–422.

Klosterhalfen S, Kallrath J, Fischer G (2014) Rail car fleet design: Optimization of structure and size. *Internat. J. Production Econom.* 157:112–119.

Loxton R, Lin Q, Teo KL (2012) A stochastic fleet composition problem. *Comput. Oper. Res.* 39(12):3177–3184.

Marcotte O (1985) The cutting stock problem and integer rounding. *Math. Program.* 33(1):82–92.

Martello S, Toth P (1990) Lower bounds and reduction procedures for the bin packing problem. *Discrete Appl. Math.* 28(1):59–70.

Meng Q, Wang S, Andersson H, Thun K (2014) Containership routing and scheduling in liner shipping: Overview and future research directions. *Transportation Sci.* 48(2):265–280.

Meng Q, Wang T (2011) A scenario-based dynamic programming model for multi-period liner ship fleet planning. *Transportation Res. Part E Logist. Transportation Rev.* 47(4):401–413.

Nitsche C, Scheithauer G, Terno J (1998) New cases of the cutting stock problem having MIRUP. *Math. Methods Oper. Res.* 48(1): 105–115.

Noori-daryan M, Taleizadeh AA, Govindan K (2018) Joint replenishment and pricing decisions with different freight modes considerations for a supply chain under a composite incentive contract. *J. Oper. Res. Soc.* 69(6):876–894.

Nourinejad M, Roorda MJ (2017) A continuous approximation model for the fleet composition problem on the rectangular grid. *OR Spectrum* 39(2):373–401.

Padmanabhan D, Damla Ahipasaoglu S, Ramachandra A, Natarajan K (2021) Extremal probability bounds in combinatorial optimization. Preprint, submitted September 3, https://arxiv.org/abs/2109.01591.

Pantuso G, Fagerholt K, Hvattum LM (2014) A survey on maritime fleet size and mix problems. *Eur. J. Oper. Res.* 235(2):341–349.

Poldi KC, Arenales MN (2009) Heuristics for the one-dimensional cutting stock problem with limited multiple stock lengths. *Comput. Oper. Res.* 36(6):2074–2081.

Scheithauer G, Terno J (1995) The modified integer round-up property of the one-dimensional cutting stock problem. *Eur. J. Oper. Res.* 84(3):562–571.

Simchi-Levi D (1994) New worst-case results for the bin-packing problem. *Naval Res. Logist.* 41(4):579–585.

Strathman JG, Hopper JR (1993) Empirical analysis of bus transit on-time performance. *Transportation Res. Part A Policy Practice* 27(2):93–100.

Wagenaar J, Fragkos I, Zuidwijk R (2021) Integrated planning for multimodal networks with disruptions and customer service requirements. *Transportation Sci.* 55(1):196–221.

Wolsey LA, Nemhauser GL (1999) *Integer and Combinatorial Optimization, Wiley-Interscience Series in Discrete Mathematics and Optimization*, vol. 55 (John Wiley & Sons, New York).

Yamada T, Russ BF, Castro J, Taniguchi E (2009) Designing multimodal freight transport networks: A heuristic approach and applications. *Transportation Sci.* 43(2):129–143.