



Production, Manufacturing and Logistics

# An inventory control model for modal split transport: A tailored base-surge approach



Chuanwen Dong, Sandra Transchel\*, Kai Hoberg

Kühne Logistics University, Großer Grasbrook 17, Hamburg 20457, Germany

## ARTICLE INFO

### Article history:

Received 10 March 2016

Accepted 8 May 2017

Available online 11 May 2017

### Keywords:

Inventory control

Modal split transport

Stochastic model applications

Tailored base-surge policy

## ABSTRACT

Firms are increasingly interested in transport policies that enable a shift in cargo volumes from road (truck) transport to less expensive, more sustainable, but slower and less flexible transport modes like railway or inland waterway transport. The lack of flexibility in terms of shipment quantity and delivery frequency may cause unnecessary inventories and lost sales, which may outweigh the savings in transportation costs. To guide the strategic volume allocation, we examine a modal split transport (MST) policy of two modes that integrates inventory controls.

We develop a single-product-single-corridor stochastic MST model with two transport modes considering a hybrid push–pull inventory control policy. The objective is to minimize the long-run expected total costs of transport, inventory holding, and backlogging. The MST model is a generalization of the classical tailored base-surge (TBS) policy known from the dual sourcing literature with non-identical delivery frequencies of the two transport modes. We analytically solve approximate problems and provide closed-form solutions of the modal split. The solution provides an easy-to-implement solution tool for practitioners. The results provide structural insights regarding the tradeoff between transport cost savings and holding cost spending and reveal a high utilization of the slow mode. A numerical performance study shows that our approximation is reasonably accurate, with an error of less than 3% compared to the optimal results. The results also indicate that as much as 85% of the expected volume should be split into the slow mode.

© 2017 Elsevier B.V. All rights reserved.

## 1. Introduction

In recent years, companies have identified the potential of modal split transport (MST) for optimizing the allocation of cargo into more than one transport mode. Rather than shipping all the cargo by truck, there is an increasing interest in moving transport volumes to trains or barges. There are numerous reasons for this prospective paradigm shift. First, road transport is generally more expensive per unit of cargo shipped, and its cost is still forced upward by factors such as congestions and empty running (American Transportation Research Institute, 2014; McKinnon & Ge, 2006). Second, the shortage of truck drivers is limiting the supply of truck capacity and causing structural fleet management issues (BCG, 2015; Sheffi, 2015). Third, firms' sustainability agendas and carbon reduction targets facilitate the shift to "greener" transport modes that favor trains or ships over trucks (Dekker, Bloemhof, & Mallidis, 2012; Dey, LaGuardia, & Srinivasan, 2011).

Despite the increasing emphasis that MST receives, shifting volume away from the road remains challenging. Statistics demonstrate that since 1995, there has been no significant change in modal split ratios among road, rail, and waterway in the EU-28 zone (EUROSTAT, 2015). In contrast, rail transport on certain routes had to be closed after several years of operation because the rigid schedule could not cope with the practical demand changes (Lammgård, 2012). Shippers hesitate to implement train transport due to the concern that there will not be sufficient volume to secure a cheap price (Pallme, Lambert, Miller, & Lipinski, 2014). The timetable of rail or barge is rigid, and it is therefore almost impossible to send an extra train when demand surges (Reis, Meier, Pace, & Palacin, 2013). Compared to other transport modes, truck transport is still the most flexible mode in terms of delivery time, routes, and quantity.

To obtain further insights into the challenges of MST, we partnered with a consumer goods company that further inspired our research. This company currently consigns almost all of the transport volume into trucks. On a daily basis, the distribution centers (DC) order inventory from the plant and expect instant deliveries within a short lead time. Such a "pull" inventory system al-

\* Corresponding author.

E-mail address: [Sandra.Transchel@the-klu.org](mailto:Sandra.Transchel@the-klu.org) (S. Transchel).

lows the DCs' to easily adapt their orders from day to day in line with demand; however, this creates high fluctuations in shipment volume. The company is interested in shifting more transport volume from trucks to trains or barges with the intent of saving cost and operating more sustainably. From interviews with managers, a practical challenge with the implementation of MST is to synchronize the more rigid slow transport modes with the more flexible fast transport mode without harming service levels or increasing inventories.

More specifically, train or barge operations are subject to restrained schedules and often have lower delivery frequencies than trucks. These schedules generally remain fixed over a long period (e.g., half of a year to 1 year), and firms are required to commit a fixed loading quantity over the period in advance to obtain a low transport cost. For example, a shipper needs to fix ten containers on the train from Antwerp (Belgium) to Hamburg (Germany) every Monday for the entire calendar year. Therefore moving from truck transport to MST also implies a change in the inventory control policy from a pure "pull" strategy to a hybrid "push–pull" strategy. Due to the long-term commitment, the slow mode shipments can be viewed as the inventory that is pushed to the DCs, while the more flexible fast mode shipments contain inventory that is pulled by the DCs. Against this background, we develop an MST policy that should consider the simultaneous usage of both modes, i.e., trucks and trains/barges in a single transport corridor, and incorporate the costs from inventory management. Although transport and inventory decisions require an integrated approach, practitioners often struggle to holistically implement the required policies. The fundamental objective of this research is to develop an insightful and easy-to-implement modal split policy to guide practitioners in real-world MST problems.

In this paper, we develop a single-product–single-corridor stochastic MST model with two transport modes considering a hybrid push–pull inventory policy. The model covers the following setting. A firm delivers a product from a plant to a DC and has to decide how to split the delivery quantity into two transport modes: a slow mode that is rather rigid in terms of time and delivery quantity, i.e., the firm has to commit to a fixed quantity to be shipped at specific time points, and a fast mode that operates every period and has full flexibility in terms of delivery quantity but also at higher transport cost than the slow mode. Whereas the "fast mode" in our research clearly indicates truck, the "slow mode" is not necessarily a certain mode but can also mean a mixed strategy using trucks and trains/barges. The firm aims to minimize the expected transport- and inventory-related costs by optimizing the fixed slow mode quantity that is committed in advance (push) and the delivery policy for the more flexible fast mode (pull).

Our MST model has a structural form comparable to the tailored base-surge (TBS) model studied in the dual-sourcing literature where firms split their orders into a fixed "base" quantity ordered from a cheap overseas supply source and a flexible "surge" quantity ordered from a more expensive but fast supply source (Allon & Van Mieghem, 2010).

The primary difference between our MST model and the classical TBS model is that the TBS model assumes that both slow mode and fast mode orders have identical delivery frequencies. Our MST model considers different delivery frequencies of the two modes based on the fact that trains/barges operate less frequently than trucks. Therefore, our MST model is a generalization of the classical TBS model. To the best of our knowledge, this is the first paper that makes this generalizing assumption.

However, previous studies have shown that the TBS model is not amenable to exact analysis, mainly due to the tractability of the expected overshoot analysis (e.g., Allon and Van Mieghem,

2010; Janakiraman, Seshadri, and Sheopuri, 2014; Janssen and de Kok, 1999, and Boute & Van Mieghem, 2014). The authors exclusively rely on a "heavy traffic" analysis in a GI/G/1 queue to derive a closed-form expression for the expected overshoot. Unfortunately, this "heavy traffic" phenomenon cannot be guaranteed in our MST model since slow mode deliveries are less frequent than fast mode deliveries. The different delivery frequencies result in all periods within a cycle (the time between two slow mode deliveries) being structurally different in a steady state. Therefore, the approximations of the classical TBS problem do not successfully work for the more general MST problem.

To obtain an analytical solution that is applicable in a practical environment, we use the deterministic benchmark (i.e., demand is perfectly known) as a starting point to identify key drivers and determinants of the volume allocation between the two transport modes. Based on these findings, we propose different tailored approximations of the cost function for different ranges of cost parameters (mainly with respect to transportation cost savings and inventory holding cost). These approximations allow us to derive closed-form expressions for the modal split policy, i.e., a fixed shipment quantity allocated to the slow mode and a base stock policy for the fast mode, which is an easy-to-implement tool for supply chain managers.

A numerical performance study with a wide range of parameters, suggested by the company, reveals that our approximation has sufficient accuracy compared to optimal solutions calculated using complete enumeration. On the test bed, the approximation error is less than 3%, and the computing time is only a fraction of the complete enumeration.

The analytic characterizations of our results capture the key trade-off of the MST problem: a *commitment effect* and a *cycle stock effect*. The commitment effect refers to the long-term commitment of the constant quantity in the slow mode enabling the reduction of transport cost compared to the fast mode. The cycle stock effect refers to the higher shipping quantity in the slow mode that potentially increases the inventory holding cost. Interestingly, the marginal effects can be simply determined by two parameters that frame the solution for the MST problems. We characterize these key drivers of volume allocation in the slow mode as follows: (i) the unit transportation cost savings of the slow mode compared to the fast mode and (ii) the volatility of the stochastic demand. This appears counterintuitive to many supply chain managers' beliefs: they often assume that the size of this fixed volume should not exceed the lower bound of the demand over the entire period when committing a constant volume in the slow mode in the long run. The presumption is that the volume that is delivered in the slow mode should always be consumed before the next slow mode delivery arrives. This is a major disadvantage of the practitioners who only treat MST as a pure transport problem.

Further insights from the numerical study reveal that for a typical "Runner" product of the industry with high expected demand and low demand variability (Relph & Milner, 2015), the optimal volume allocated to the slow mode could be as high as 85% of the expected demand. This surprisingly high ratio supports our findings and indicates that a holistic approach to jointly decide on inventory and transport mode is essential.

The remainder of this paper is organized as follows. In Section 2, we review the relevant literature. Next, we formulate the MST model in Section 3. In Section 4, we analyze the MST policy and derive approximate analytic solutions. In Section 5, we provide numerical results that highlight the error of our approximation and the potential volume split for both modes. We also present a model extension that considers volume-dependent transportation cost. In Section 6, we summarize our research and discuss further avenues of MST.

## 2. Literature

Our research is built on two streams of literature: (i) the freight mode choice literature, which studies problems in a context similar to ours, and (ii) the dual-sourcing literature, which studies a different sourcing model with multiple suppliers but has a similar mathematical structure as ours.

The freight mode choice literature analyzes the selection and usage of different transport modes in certain freight corridors or networks. It can be characterized in two categories: freight mode choice with multimodal transport, which focuses on the trade-off of the characteristics of different transport modes, e.g., cost, capacity, and lead time, and decides on how to simultaneously use them in a given freight network, and freight mode choice of alternative single transport modes, which studies the trade-off between inventory and transport costs and proposes alternative decisions in fast or slow mode.

Studies involving multimodal transport generally focus on transportation issues only and disregard stochastic inventory decisions. For example, [Verma and Verter \(2010\)](#) study how to choose between road and rail transport modes in a certain transport network by minimizing the total transport costs of the two modes and subject to a set of pre-specified lead time constraints. Interested readers are referred to [Bontekoning, Macharis, and Trip \(2004\)](#); [Crainic and Kim \(2007\)](#), and most recently, [StadieSeifi, Dellaert, Nuijten, Van Woensel, and Raoufi \(2014\)](#), for an extend literature review.

Studies that incorporate inventory decisions focus mainly on alternative single transport modes. [Baumol and Vinod \(1970\)](#) are one of the first to raise the idea that the transport mode choice needs to be made with inventory considerations. They develop a so-called “inventory-theoretic” model, and illustrate that a change in usage from a fast to a slow transport mode will theoretically increase a firm’s inventory costs. [Larson \(1988\)](#) studies the same model and determines the optimal delivery quantity by minimizing the joint cost of transport and inventory. [Speranza and Ukovich \(1994\)](#) consider how to minimize the sum of transport and inventory costs in a multi-product setting with the freight mode choice between full and partial truckload. They find that when multiple products are consolidated, firms might take advantage of economies of scale and change from partial truckload to full truckload. [Kutanoglu and Lohiya \(2008\)](#) study a model with freight mode choice, inventory decisions, and service-level constraints with the objective of minimizing total inventory and transport costs. They find that this problem is an NP-hard problem and develop feasible heuristics. They show that the firm can save total costs by making alternative decisions in fast, medium, and slow modes. [Lloret-Battle and Combes \(2013\)](#) empirically examine firms’ freight mode choice decisions based on a French shipper survey dataset including more than 10,000 shipments, and they find that by incorporating inventory costs, firms improve their decisions in freight mode choice.

Our research differs from this stream of literature by studying a freight transport problem with a simultaneous usage of two modes (multimodal) and incorporating inventory decisions with stochastic demand. We consider a model that minimizes the long-run average total costs, including transport cost, inventory holding, and backorder cost under stochastic demand, by jointly making decisions on the inventory policy and optimal split between the two transport modes. This setting is mathematically similar to the classical dual-sourcing inventory model, where a firm simultaneously orders from two sources (e.g., suppliers) in which one source provides a cheaper product with a longer lead time and the other source offers a higher price but with a shorter lead time. Previous studies in this field are numerous.

The first dual-sourcing model dates back to [Barankin \(1961\)](#), who analyzes this problem in a single-period case. [Whittemore and Saunders \(1977\)](#) find that the optimal policy is complex and that analytic solutions can only be found when the lead time difference is one. A detailed review of dual sourcing is summarized by [Minner \(2003\)](#). Recent dual-sourcing research mainly focuses on approximations or heuristics of practical policies, e.g., the dual index policy (DIP) and tailored base-surge (TBS) policy. In both policies, a special phenomenon of “overshoot” is observed, which leads to a complication of the problems. An overshoot occurs when the inventory position exceeds the base stock levels. DIP policies allow both sources to replenish following base stock policies and typically obtain overshoots via simulations. We refer to [Arts, van Vuuren, and Kiesmüller \(2011\)](#); [Sheopuri, Janakiraman, and Seshadri \(2010\)](#); [Veeraraghavan and Scheller-Wolf \(2008\)](#), and [Arts and Kiesmüller \(2013\)](#).

In the TBS model, a firm periodically orders a fixed quantity from the cheap but slow source, whereas the fast but expensive source orders follow a base stock policy. [Janssen and de Kok \(1999\)](#) study a similar policy by modeling the inventory system in a GI/G/1 queue. They solve the model using simulations. [Allon and Van Mieghem \(2010\)](#) optimize the TBS analytically and find that “the economic optimization naturally brings the system into a parameter regime called heavy traffic”, where the GI/G/1 queue is always busy with waiting customers. This key finding allows for the utilization of Kingman’s bound [Kingman \(1970\)](#) in heavy traffic analysis and further obtains an approximate analytic solution of the “base” volume. [Klosterhalfen, Kiesmüller, and Minner \(2011\)](#) use the term “constant order policy” rather than TBS to describe the same inventory control policy. They compare the policy with DIP numerically and find that either can outperform the other under some parameter settings. [Boute and Van Mieghem \(2014\)](#) make use of linear control theory for analytic solutions without relying on heavy traffic analysis. They find that TBS and other dual-sourcing policies smooth orders when capacity cost, inflexibility, or longer lead time difference is considered. [Janakiraman, Seshadri, and Sheopuri \(2014\)](#) analyze TBS in a discrete time period model and approximate the “base” volume in a closed form by using an upper bound for the cost function. They further numerically examine the accuracy of their approximation with different parameter settings.

Our study differs from the existing dual-sourcing literature by incorporating different delivery frequencies of the two sources/modes. Our model reflects the practical characteristics that in the transport world, trains or barges deliver less frequently than trucks. To the best of our knowledge, our model is the first dual-sourcing model that considers different delivery frequencies of the two modes/sources.

## 3. Model formulation

In this section, we formulate the model that is used for our analysis. The notation is summarized in the following table.

$\xi$	Demand per period (an i.i.d random variable)
$\mu$	Mean of $\xi$
$\sigma$	Standard deviation of $\xi$
$\Phi(\cdot)/\phi(\cdot)$	CDF/PDF of $\xi$
$c^s$	Shipment cost per unit of the slow mode
$c^f$	Shipment cost per unit of the fast mode
$\Delta$	Cost savings per unit of the modal shift, $\Delta = c^f - c^s$
$h$	Holding cost per unit per period
$b$	Penalty cost per unit per period
$s^B$	Optimal base stock level when only the fast mode is utilized
$t$	Index of time, $t = 1, 2, \dots, T$

$i$	Index of period in a steady state cycle, $i = 1, 2$
$B$	Symbol of the optimal base stock policy, when only the fast mode is used
$\mathcal{D}$	Symbol of the model split policy, when both the fast and slow modes are used
$C^B$	Expected total cost of two periods (a cycle) following policy $B$
$C^{\mathcal{D}}$	Expected total cost of two periods (a cycle) following policy $\mathcal{D}$
$\Pi$	Cost savings of policy $\mathcal{D}$ over $B$ , i.e., $\Pi = C^B - C^{\mathcal{D}}$
$O_t$	Overshoot in period $t$
$O_{\infty, i}$	Steady state overshoot in period $i$ of a cycle
$L$	Expected mismatch cost
$Q$	Fixed delivery quantity of slow mode in each cycle (decision variable)
$S_i$	Base stock level in period $i$ of a steady state cycle (decision variable)

We consider a distribution center (DC) of a firm that periodically orders from its manufacturing plant ( $M$ ) with unlimited capacity. The demand per period at the DC is denoted by the random variable  $\xi$  and is independently and identically distributed (i.i.d.) with mean  $\mu$  and standard deviation  $\sigma$ . The cumulative distribution function (CDF) and probability density function (PDF) of  $\xi$  are  $\Phi(\cdot)$  and  $\phi(\cdot)$  respectively.

We consider a baseline model in which every period the DC places an order with  $M$  and ships this order via the *fast mode* at a unit delivery cost  $c^f$ .  $M$  has sufficient capacity to fulfill every order with an overnight shipment, which implies a zero lead time with the fast mode transport. The inventory level at the end of each period is charged a holding cost of  $h$  per unit, and any unmet demand is backlogged at a unit cost of  $b$ . Furthermore, we assume that  $b > c^f$ , indicating that a pure accumulation of backorders without any deliveries is not the optimal solution. The DC needs to decide on the order/delivery quantities that minimize the expected total cost in transport and inventory mismatch. It is already proven by Karlin (1960) that a base stock policy with an order-up-to level of  $S^B = \Phi^{-1}\left(\frac{b}{b+h}\right)$  is optimal for such a system. We denote the optimal fast-mode-only policy as baseline policy  $B$ . The expected total cost for a cycle of two periods is then

$$C^B = 2(c^f\mu + L(S^B)) \quad (1)$$

where  $2c^f\mu$  represents the expected transport cost and  $2L(S^B)$  represents the expected inventory mismatch cost for the two periods of the cycle with  $L(S^B) = h \int_0^{S^B} (S^B - \xi)\phi(\xi)d\xi + b \int_{S^B}^{\infty} (\xi - S^B)\phi(\xi)d\xi$ . Note that  $C^B$  remains constant in the baseline model.

Given that an additional slow mode option exists, the firm wants to implement a *modal split transport (MST)* policy (denoted as  $\mathcal{D}$ ) in the following way: the firm has an offer from a *logistics service provider (LSP)* to regularly ship a constant quantity of  $Q$  items via the *slow mode* at a unit delivery cost  $c^s$  with a cost savings  $\Delta = c^f - c^s > 0$  per unit shipped. The parameter  $c^s$  includes transport cost and all other incremental costs associated with slow mode shipment, e.g., higher working capital cost for additional in-transit inventories or higher handling costs. Although the slow mode has a lower total landed unit cost, it is not available every period and only operates at a lower delivery frequency. Whereas the fast mode can be used every period, we assume that the slow mode is only available every other period. We define the time interval between two consecutive slow mode shipments as a *delivery cycle*. The slow mode shipment can be viewed as a form of a *commitment* for the firm in a way that every other period, the firm ships a constant quantity of  $Q$  units from  $M$  to DC via the slow mode. Because a constant  $Q$  is shipped and also received at a fixed delivery frequency, the lead time of the slow mode can be

neglected. In addition to the slow mode shipments, the DC can still place flexible orders via the fast mode in every period following a base stock policy. The firm's objective is to determine the optimal slow mode quantity and the optimal base stock policy to minimize the expected total costs, consisting of the expected transport cost (slow mode and fast mode) and inventory mismatch cost. The difference in terms of the fast-mode-only policy is that we have to consider the expected cost during a cycle because the two periods within a cycle are not identical.

The aforementioned problem is structurally similar to the TBS policy in the dual-sourcing literature (see, e.g., Allon & Van Mieghem, 2010; Janakiraman, Seshadri, & Sheopuri, 2014; Janssen & de Kok, 1999; Klosterhalfen, Kiesmüller, & Minner, 2011). The TBS policy is specified by two parameters: a constant order quantity  $Q$  that is placed with a cheap supplier with a long lead time (comparable to our slow mode) and a base stock level  $S$  that determines the flexible order policy from a more expensive supplier with a lower lead time (comparable to our fast mode). In all the TBS models discussed thus far, the order frequencies of the two source modes are assumed to be identical, i.e., the fast mode and the slow mode are available every period, which is not the case in our problem. In the MST problem, the slow mode deliveries are available less frequently compared to the fast mode. Therefore, we cannot simply minimize the expected cost per period but have to consider a cycle structure where for every period in that cycle, the base stock levels for using the fast mode could be different. In the following, the index  $i$  denotes the period of a cycle, i.e.,  $i = 1$  is the first period of the cycle where the slow mode delivery arrives and  $i = 2$  defines the second period of a cycle where only fast mode ordering is available. With an ongoing numbering, every "odd" period is a period with a slow mode delivery, while in every "even" period, only fast mode is available. Since fast mode orders are available every period following a base stock policy, we denote  $S_i$ ,  $i \in \{1, 2\}$  as the base stock levels in period  $i$  of a steady state cycle. Thus, our MST policy is addressed by three parameters:  $Q$ ,  $S_1$ , and  $S_2$ .

Similar to the classical TBS policy, a special phenomenon of our MST policy is the *overshoot*. The overshoot in an inventory system is defined as the amount by which the inventory position exceeds the base stock levels (Allon & Van Mieghem, 2010). We use  $O_t$ , ( $t = 1, 2, \dots, \infty$ ), to denote the overshoot in  $t$ , which describes a stochastic process. It is well known that in an inventory system following a base stock policy, orders are placed to bring the current inventory positions up to the predefined base stock levels. This is why a base stock policy is also called an "order-up-to" policy. In an MST system, however, a constant quantity  $Q$  is cyclically pushed into the system regardless of the levels of the inventory positions and might shoot the inventory position "over" the base stock levels. Fig. 1 shows an example of how overshoots behave depending on the pre-assumed  $Q$ ,  $S_1$ ,  $S_2$ , and the starting inventory  $I_1$  in period 1. In periods 1 and 2, the firm places orders in the fast mode,  $z_1 > 0$ ,  $z_2 > 0$ , to bring the inventory positions up to  $S_1$  and  $S_2$ , respectively, which is similar to the situation in the baseline models. However, in period 3, the arrival of  $Q$  shoots the inventory position over  $S_1$ , and an overshoot,  $O_3 > 0$ , is observed. Although  $Q$  only arrives in the odd periods, it might also indirectly drive the overshoot in the even periods, as shown in period 4 of the example. Still, the overshoots in odd and even periods are structurally different from each other because  $Q$  only arrives in the odd periods and its impact on the even periods is indirect.

If  $t$  is an odd period (i.e., period  $t = 1, 3, 5, \dots$ ), then the inventory position at the end of this period can be written as a function of  $O_t$ ,  $S_1 + O_t - \xi$ ; if  $t$  is an even period (i.e., period  $t = 2, 4, 6, \dots$ ), then the inventory position at the end of  $t$  is  $S_2 + O_t - \xi$ . Therefore, the overshoot in period  $t + 1$ ,  $O_{t+1}$ , is the excessive amount (if any) over  $S_2$  (when  $t$  is odd) or  $S_1$  (when  $t$  is even) and can be



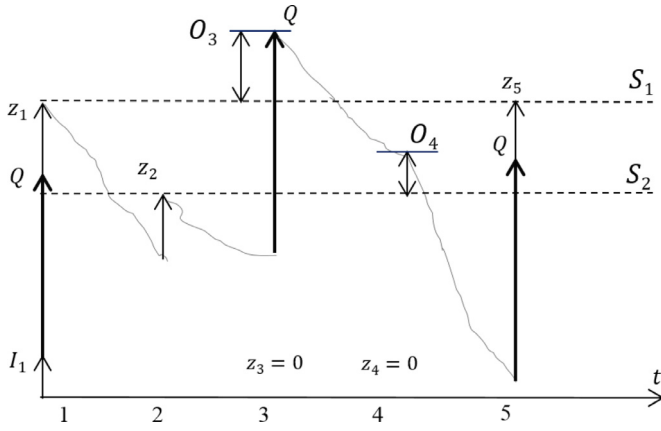


Fig. 1. An example of the overshoots.

recursively written as:

$$O_{t+1} = \begin{cases} (O_t + S_1 - \xi - S_2)^+ & \text{when } t \text{ is odd} \\ (O_t + S_2 + Q - \xi - S_1)^+ & \text{when } t \text{ is even} \end{cases} \quad (2)$$

where  $()^+$  denotes the positive part of a real number. As a comparison, the classical TBS policy from the dual-sourcing literature does not consider different delivery frequencies of the two modes, and it therefore only optimizes a single base stock level. Consequently, the recursion of the overshoot in a TBS policy has a simpler form as described in (1) of Janakiraman, Seshadri, and Sheopuri (2014), and its calculation should be reasonably easier compared to the overshoot in our MST problem.

We assume that  $Q < 2\mu$ , indicating that the supply from the slow mode should not exceed the expected demand<sup>1</sup>. Because the firm will adjust fast mode deliveries based on overshoots, it will not accumulate overshoots infinitely, and the magnitudes of the overshoots are therefore bounded. Denote  $O_{\infty,1}$  and  $O_{\infty,2}$  as the overshoots in periods one and two of a steady state cycle, and both  $E[O_{\infty,1}]$  and  $E[O_{\infty,2}]$  are hence non-negative and finite. Let  $i$  denote the index of the period within a cycle, i.e.,  $i = 1$  is the first period of the cycle where the slow mode delivery arrives and  $i = 2$  is the second period of the cycle with fast mode order only. In period  $i$ ,  $i \in \{1, 2\}$  of the steady state cycle, the expected inventory position before demand realization is  $S_i + E[O_{\infty,i}]$ . Note that because of the overshoots, the expected inventory positions are “over”  $S_i$ . Given the stochastic demand  $\xi$ , the expected mismatch cost  $L_i$  is:

$$L_i(S_i + E[O_{\infty,i}]) = hE[(S_i + E[O_{\infty,i}] - \xi)^+] + bE[(\xi - S_i - E[O_{\infty,i}])^+]. \quad (3)$$

The firm aims to find the optimal combination of  $(Q^*, S_1^*, S_2^*)$  to minimize the total expected cost of a steady state cycle, formalized as

$$C^D = c^s Q + c^f (2\mu - Q) + \sum_{i=1}^2 L_i(S_i + E[O_{\infty,i}]) \quad (4)$$

where  $c^s Q$  is the (expected) transport cost via the slow mode,  $c^f (2\mu - Q)$  is the expected transport cost via the fast mode, and  $\sum_{i=1}^2 L_i(S_i + E[O_{\infty,i}])$  is the expected total mismatch cost over the entire cycle. Comparing the cost savings of the MST policy to the baseline policy, i.e.,  $\Pi = C^B - C^D$ , optimizing the MST policy is

equivalent to maximizing the cost savings:

$$\Pi = C^B - C^D = \Delta Q + 2L(S^B) - \sum_{i=1}^2 L_i(S_i + E[O_{\infty,i}]). \quad (5)$$

Eq. (5) captures the trade-off of the MST problem:  $\Delta Q$  is the transport cost savings from the slow mode, and  $2L(S^B) - \sum_{i=1}^2 L_i(S_i + E[O_{\infty,i}])$  is the expected net effect from the inventory cost. The holistic supply chain is optimized by trading off transport and inventory decisions.

#### 4. Model analysis

The difficulty in the model analysis lies in finding analytic expressions for the stationary expected overshoot  $E[O_{\infty,i}]$  for  $i = 1, 2$ . Analyzing the overshoot is already challenging in the classical TBS policy where slow mode and fast mode ordering have the same delivery frequency. Janssen and de Kok (1999) use an approximation based on the equivalence of the overshoot to the waiting time in a GI/G/1 queueing model. Allon and Van Mieghem (2010) evaluate the system using a GI/G/1 queue and find that the optimization of the TBS policy naturally brings the queue into a “heavy traffic” state, where the slow mode is heavily utilized ( $Q$  close to its upper bound  $\mu$ ). The authors further show that analytic closed-form expressions are not tractable. To solve the problem, the authors perform an asymptotic analysis and use one of Kingman’s bounds (Kingman, 1970) to bound the expected steady state overshoot. Klosterhalfen, Kiesmüller, and Minner (2011) pursue an exact approach by modeling the overshoot as a Markov chain with an infinite state space. Boute and Van Mieghem (2014) state that the evaluation of a GI/G/1 queue in TBS is required; otherwise, a simulation analysis needs to be used. Janakiraman, Seshadri, and Sheopuri (2014) also use Kingman’s bound to bound the expected steady state overshoot and thus the average cost per period to analyze the effectiveness of the best TBS policy relative to the optimal policy over all feasible policies. To the best of our knowledge, the heavy traffic analysis is currently the only method mentioned in the literature for approximating and characterizing an exact TBS analysis.

As a generalized TBS problem, the MST system unfortunately cannot guarantee the heavy traffic phenomenon. First, because the slow mode delivery  $Q$  arrives only once every two periods due to the different delivery frequencies, it may induce a “cycle inventory”, i.e., inventory that is stocked to exploit lower transport costs at the beginning of the first period. It is difficult to analytically distinguish between the cycle stock and overshoot. Second, the steady state overshoot in the even periods is structurally different from the overshoot in odd periods and thus requires different treatments. Third, if the transport cost savings  $\Delta$  are much smaller than the additional holding cost  $h$ , the firm tends to commit a smaller  $Q$  for the slow mode to reduce the overshoot risk in the second period in (2) and the consequential inventory holding cost in (5); therefore, a slow mode cannot be heavily utilized. From a practical perspective, a small  $\Delta$  is not a rare event. Many sites do not have a direct connection with trains or barges and thus require a multimodal transport mode, e.g., truck-train-truck, which requires additional costs for extra handling, waiting, and detour, among others, and offsets or sometimes even outweigh the cost savings from the utilization of a slow mode. Macharis, Van Hoeck, Pekin, and Van Lier (2010) support this observation based on academic research. Therefore, a new methodology is required to characterize the generalized TBS problem.

##### 4.1. Deterministic demand

We first restrict our attention to MST under deterministic demand, i.e.,  $\sigma = 0$ , and capture the trade-off effects. The optimal

<sup>1</sup> If  $Q > 2\mu$ , the system diverges with infinite inventory because there is more supply than demand. If  $Q = 2\mu$ , the system is in the state of “unstable equilibrium”, where any deviation from the uncertainty of random demand will bring the system into a state of chaos.

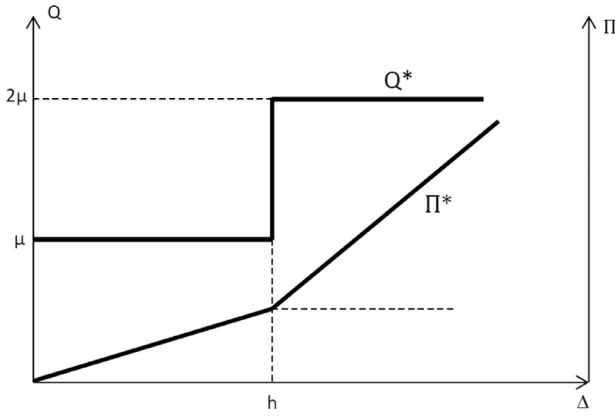


Fig. 2.  $Q^*$  and  $\Pi^*$  for deterministic demand.

solution clearly has a bang-bang structure. If the savings by using slow mode transport cost are greater than the additional holding cost during a cycle, i.e.,  $\Delta > h$ , then the firm ships the demand of both periods via the slow mode, i.e.,  $Q^* = 2\mu$ . The consequence is an additional holding cost for an inventory level of  $\mu$  at the end of the first period (cycle inventory). If the transport cost savings are less than the additional holding cost, i.e.,  $\Delta < h$ , then the firm only ships the demand of the first period via the slow mode, i.e.,  $Q^* = \mu$ , and the fast mode shipment at the beginning of the second period equals the demand of the second period  $\mu$ . If  $\Delta = h$ , the firm is indifferent in terms of the two options. Fig. 2 illustrates the optimal slow mode order  $Q^*$  and the profit  $\Pi^*$ . The question arises as to how  $Q^*$  changes when the demand is stochastic and additional inventory is required to buffer uncertainty.

An additional insight from the deterministic case is that the profit function is not continuously differentiable in the decision variable  $Q$ , which needs to be considered in the stochastic case when  $\sigma \rightarrow 0$ .

#### 4.2. Stochastic demand

When demand is stochastic, i.e.,  $\sigma > 0$ , there is an additional uncertainty of not meeting the demand in a period. Hence, inventory at the end of a period can have two functions: (i) it may serve as safety stock to buffer against demand uncertainty, and (ii) it serves as cycle stock to exploit lower transport cost from the slow mode, which runs every other period. We will utilize the results of the deterministic case to solve the stochastic version of the problem. For the solution, we distinguish three scenarios A, B and C that differ based on the relationship between  $h$  and  $\Delta$ .

##### 4.2.1. Scenario A: $\Delta > h$

If  $\Delta > h$ , i.e., the unit savings in the transport cost from the slow mode are higher than the unit holding cost per period. Intuitively, the firm is inclined to commit to a larger  $Q$  shipped via the slow mode. “Large” here means that after the delivery of  $Q$ , the inventory on hand is expected to satisfy not only the demand in the first period but also part<sup>2</sup> of the demand in the second period of the cycle despite extra inventory holding costs at the end of the first period. Hence, the probability of using the fast mode transport and/or running out of stock at the end of the first period

of a cycle tends to be zero. Consequently, we eliminate the decision variable  $S_1$  and assume no fast mode ordering in period one. We replace the expected mismatch cost at the end of the first period by  $h(S_2 + E[O_{\infty,2}] + Q - 2\mu)$  (no backorders occur at the end of the first period). The expected mismatch cost at the end of the first period is approximated as  $L_1(S_1 + E[O_{\infty,1}]) = h(I_1 + Q - \mu)$ , where  $I_1$  is the expected starting inventory in period one and  $I_1 = S_2 + E[O_{\infty,2}] - \mu$ . Thus, the cost savings of  $C^D$  over  $C^{B*}$ , derived from (5), can be approximated as

$$\hat{\Pi} = \Delta Q + 2L(S^B) - h(S_2 + E[O_{\infty,2}] + Q - 2\mu) - L_2(S_2 + E[O_{\infty,2}]) \quad (6)$$

and the long-run policy can be reduced to two decision variables,  $Q$  and  $S_2$ .

Because we do not consider fast mode ordering in the first period and thus no  $S_1$  and  $E[O_{\infty,1}]$ ,  $S_2$  only depends on the slow mode quantity  $Q$ , such that  $S_2(Q)$  is the solution to the newsvendor formula of  $\frac{dL_2(S_2 + E[O_{\infty,2}])}{dS_2} = 0$ , which yields:

$$S_2 = \Phi^{-1}\left(\frac{b}{b+h}\right) - E[O_{\infty,2}] \quad (7)$$

where  $E[O_{\infty,2}]$  is a function of  $Q$ . Modeling the cyclic inventory system as a GI/D/1 queue, we can show that the expected overshoot  $E[O_{\infty,2}]$  does not depend on  $S_2$  and can be approximated with the following proposition (the proofs of all propositions can be found in the appendix).

**Proposition 1.** The expected overshoot in the second period of a steady state cycle is independent of the base stock level  $S_2$  and can be bounded by:

$$E[O_{\infty,2}] \leq \frac{\sigma^2}{2\mu - Q}. \quad (8)$$

To determine the optimal slow mode level  $Q$ , we use the upper bound approximation by Janakiraman, Seshadri, and Sheopuri (2014): For any  $(x, y)$  for which  $y \geq 0$ , it follows that  $(x+y)^+ \leq x^+ + y$  and  $(x+y)^- \leq x^-$ . Therefore, we obtain an upper bound for the expected mismatch cost of the second period as  $L_2(S_2 + E[O_{\infty,2}]) = hE[(S_2 + E[O_{\infty,2}] - \xi)^+] + bE[(S_2 + E[O_{\infty,2}] - \xi)^-] \leq hE[(S_2 - \xi)^+] + hE[O_{\infty,2}] + bE[(S_2 - \xi)^-] = L_2(S_2) + hE[O_{\infty,2}]$ , which implies a lower bound for the cost savings as follows

$$\hat{\Pi} \geq \Delta Q + 2L(S^B) - h(S_2 + Q - 2\mu) - 2hE[O_{\infty,2}] - hE(S_2 - \xi)^+ - bE(S_2 - \xi)^-. \quad (9)$$

Taking the first-order derivative of this lower bound with respect to (w.r.t.)  $Q$ , we obtain  $\Delta - h - 2h \frac{\partial E[O_{\infty,2}]}{\partial Q} = 0$  with  $\frac{\partial E[O_{\infty,2}]}{\partial Q} = \frac{\sigma^2}{(2\mu - Q)^2}$ . Thus, the approximate optimal slow mode shipment  $\hat{Q}$  is

$$\hat{Q} = 2\mu - \sigma \sqrt{\frac{2h}{\Delta - h}}, \quad (10)$$

and in combination with (7), we obtain the approximate optimal base stock level

$$\hat{S}_2 = \Phi^{-1}\left(\frac{b}{b+h}\right) - \sigma \sqrt{\frac{\Delta - h}{2h}}. \quad (11)$$

The slow mode is advantageous in terms of cost savings. When the cost difference  $\Delta$  increases, it is intuitively beneficial to ship more volume in the slow mode. The disadvantage of the slow mode is a lack of flexibility due to operational constraints. When the demand is more volatile, the firm needs to decrease the share in the slow mode to tackle the flexibility of the fast mode. This explains why  $\hat{Q}$  decreases in  $\sigma$ , as shown by (10).

<sup>2</sup> The constant delivery quantity  $Q$  via the slow mode is not expected to satisfy the demand in both periods, i.e.,  $Q < 2\mu$ . If  $Q > 2\mu$ , the system diverges because there is more supply than demand; if  $Q = 2\mu$ , the system is in the state of “unstable equilibrium”, where any deviation from the uncertainty of random demand will bring the system into a state of chaos.

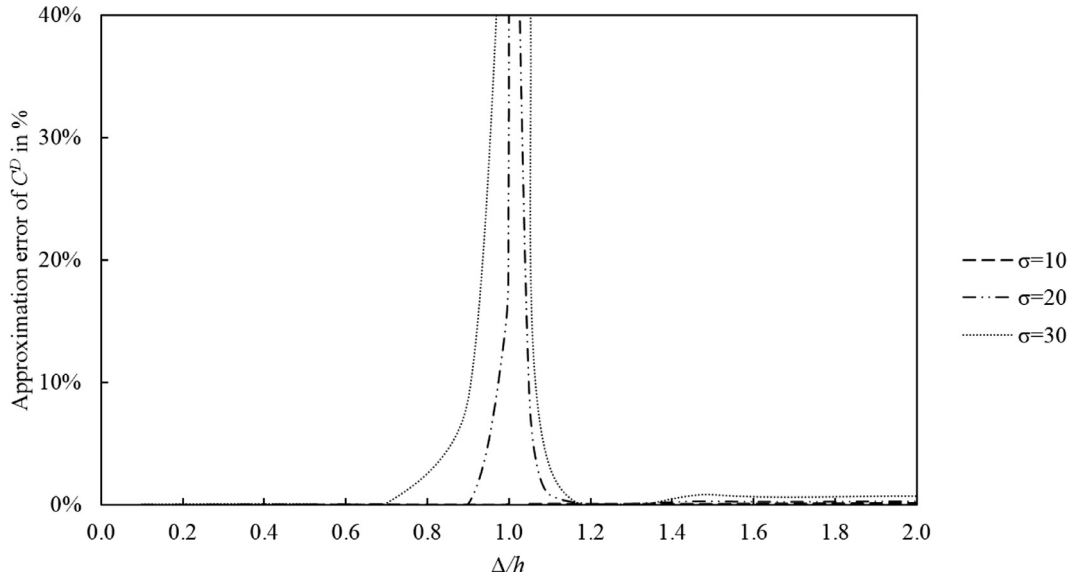


Fig. 3. The approximation error of  $C^D$  in %.

Practitioners tend to believe that if they must commit a constant slow mode volume over a long period, this size of the volume should not exceed the lower bound of the demand during the cycle. Otherwise, the slow mode will not be fully loaded, and consequently, penalties might occur. Because demand is stochastic, the lower bound over a period could be rather small and approach zero. Our finding in (10) reveals that the size of the commitment in the slow mode is actually independent of the lower bound of demand, and it could surprisingly take a larger value close to  $2\mu$ .

We find that the base stock level in even periods  $\hat{S}_2$  is less than that from the baseline model  $S^B$ . When managers already know that in the next period a constant quantity of  $Q$  will be committed to arrive, they would order less units in the current period to maintain a relatively lower inventory level compared to the baseline model.  $\hat{S}_2$  decreases in  $\Delta$ , indicating an increasing effect when the slow mode is economically more attractive. The disadvantage of this approximation is that when  $\Delta$  approaches  $h$  from the right side,  $\hat{Q}$  drops to  $-\infty$  in (10). Later in the paper, we provide a lower bound of  $\hat{Q}$  for this scenario to prevent this divergence and denote this as Scenario C.

#### 4.2.2. Scenario B: $\Delta < h$

If  $\Delta < h$ , i.e., the marginal savings in transport cost from using the slow mode are larger than the unit holding cost per period, the firm is inclined to commit to a lower  $Q$  shipped via the slow mode. By this logic, it is more likely that the fast mode order will be utilized in the second period of the cycle, diminishing the occurrence of an overshoot. We approximate this scenario by eliminating the overshoot at the beginning of the second period of the cycle. Therefore, the expected cyclic cost function can be approximated as  $C^D = c^S Q + c^F(2\mu - Q) + L_1(S_1 + E[O_{\infty,1}]) + L_2(S_2)$ , and the cost savings function of (5) is estimated as:

$$\hat{\Pi} = \Delta Q + 2L(S^B) - L_1(S_1 + E[O_{\infty,1}]) - L_2(S_2). \quad (12)$$

By assuming no overshoot in the second period, the expected overshoot in the first period can be approximated as:

$$\begin{aligned} E[\hat{O}_{\infty,1}] &= E[S_2 + Q - S_1 - \xi]^+ \\ &= \int_0^{S_2+Q-S_1} (S_2 + Q - S_1 - \xi) \phi(\xi) d\xi, \end{aligned} \quad (13)$$

i.e., the non-negative expected difference of the inventory position of the first period ( $I_1 = S_2 - \xi + Q$ ) and the base stock level

$S_1$ . Plugging (13) into (12) and taking the first-order condition of  $\hat{\Pi}$  with respect to  $Q$ ,  $S_1$  and  $S_2$  separately, we derive the approximated expressions of  $(\hat{Q}, \hat{S}_1, \hat{S}_2)$ . The result of Scenario B is concluded using the following proposition:

**Proposition 2.** When  $\Delta < h$ , the optimal decision variables of the MST policy can be approximated as:

$$\begin{cases} \hat{S}_1 = \Phi^{-1}\left(\frac{b+\Delta}{b+h}\right) \\ \hat{S}_2 = \Phi^{-1}\left(\frac{b-\Delta}{b+h}\right) \\ \hat{Q} = \Phi^{-1}\left(\frac{b+\Delta}{b+h}\right) - \Phi^{-1}\left(\frac{b-\Delta}{b+h}\right) + \Phi^{-1}\left(\frac{\Delta}{h}\right) \end{cases} \quad (14)$$

Similar to Scenario A,  $\hat{Q}$  increases in  $\Delta$ , indicating a positive relationship between cost savings and the utilization of the slow mode. Moreover, a decrease of  $\hat{S}_2$  in  $\Delta$  is also consistent with the results from Scenario A. Similar to the situation of Scenario A, the approximation deteriorates when  $\Delta$  is close to  $h$ . Specifically, when  $\Delta$  approaches  $h$  from the left side,  $\hat{Q}$  spikes to  $\infty$  in (14). Therefore, we next provide an upper bound of  $\hat{Q}$  for Scenario B based on Scenario C.

#### 4.2.3. Scenario C: $\Delta = h$ and revised Scenarios A' and B'

The previous methods in Scenarios A and B rely on the trade-off of commitment and cycle stock effects. By comparing their relative sizes, two different approximations, based on eliminating expected overshoot terms, are implemented, and the analytic expressions are obtained. When  $\Delta = h$ , both effects break even, and none of the analytic expressions in Scenarios A or B provide an answer at this point. A separate method is required for Scenario C. Furthermore, the results of the previous two scenarios already deteriorate when  $\Delta$  approaches  $h$  from either the left or right side. Accordingly, we revise the range over which the expressions from A and B are applied and refer to Scenarios A' and B'. Fig. 3 shows an example of this divergence, whereas  $\Delta$  approaches  $h$ , i.e., the ratio  $\Delta$  over  $h$  approaches one, the approximation error tends toward infinity.

Additionally, when  $\sigma$  increases, the firm needs to cope with the increased volatile demand with more flexibility in transport decisions. This requires a higher utilization of the fast mode. However,

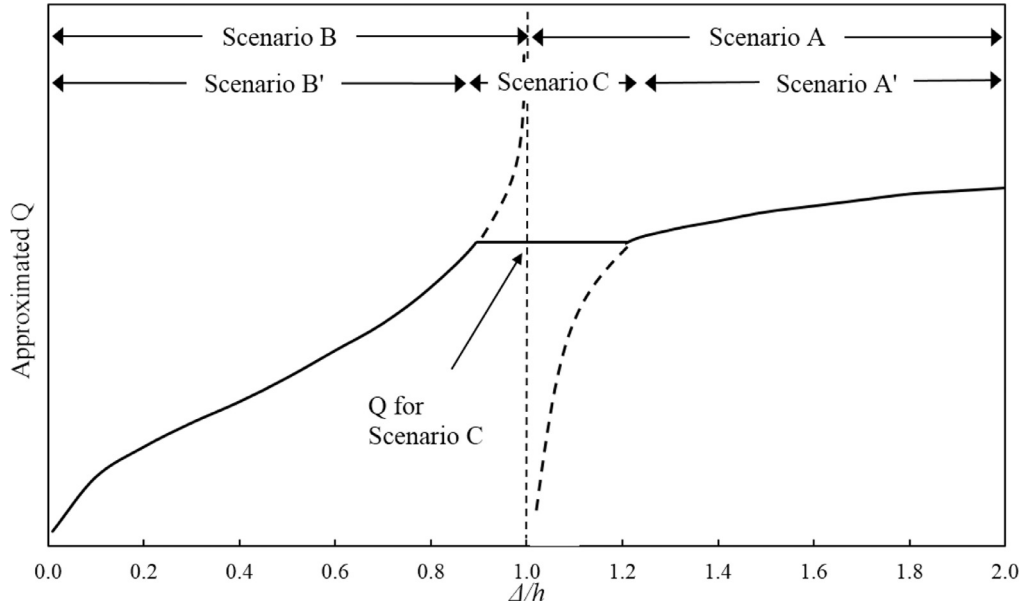


Fig. 4. Approximation of  $\hat{Q}^C$  and combined Scenarios A', B' and C.

the approximation in Scenario A forbids the utilization of the fast mode in the first period, and this leads to a larger approximation error in  $\sigma$ . A more volatile demand  $\xi$  increases the chance of overshoot, as indicated in the overshoot recursion function (2). Ignoring overshoot in period two therefore gives rise to more error in the approximation, as described in Scenario B. This explains the increase in approximation error for this scenario when  $\sigma$  increases. The results obtained from Scenario C are then used as a correction for Scenarios A and B to improve the approximation accuracy in this region.

Managerially,  $\Delta = h$  is a break-even point at which managers are indifferent in terms of choosing the solutions from either Scenario A or B. Comparing (10) and (14), only the approximation of  $S_2$  is tractable. We take  $\hat{S}_2 = \Phi^{-1}\left(\frac{b-\Delta}{b+h}\right)$  from (14) of Scenario B. (Note that the solution from Scenario A, i.e.,  $\hat{S}_2 = \Phi^{-1}\left(\frac{b}{b+h}\right)$  from (10), can also be applied, and it offers a similar accuracy in our numerical tests.) The base stock level obtained in Scenario B is slightly smaller than that in Scenario A because in Scenario B, the overshoot term  $O_{\infty,2}$  is assumed to be zero. If overshoot is not eliminated, the order-up-to level in period two should be  $S_2 + O_{\infty,2}$ . The advantage of using the result from Scenario B is that in this Scenario,  $O_{\infty,2}$  can be assumed to be zero.

An indifferent choice of  $S_1$  is also expected. In other words, whether or not we order via the fast mode in the first period does not affect the expected cost of the modal split policy. We use  $\hat{S}_1 = \hat{S}_2$  for the following reason: a free choice of  $\hat{S}_1$  also indicates the independence of the overshoot  $O_{\infty,1}$  with respect to  $\hat{S}_1$  or  $\hat{S}_2$ . The set of  $\hat{S}_1 = \hat{S}_2$  eliminates this dependency in the definition of  $O_{\infty,1}$  in (2), where the overshoot is shown to be a function of the difference between two base stock levels. Given  $\hat{S}_1$  and  $\hat{S}_2$ , we further obtain  $\hat{Q}$  as follows.

In the deterministic case, regardless of the value that  $Q^*$  takes in the range of  $[\mu, 2\mu]$ , the value of the objective function remains the same because both commitment and cycle stock effect break even.  $Q^*$  cannot be any value outside of this range: if  $Q$  is smaller than  $\mu$ , an extra backorder cost exists at the end of period one; if  $Q$  is larger than  $2\mu$ , an extra spending on holding cost at the end of period two is expected. When the demand is stochastic, these two costs are brought into the system with certain probabilities. We denote the sum of these two costs  $\bar{C}$ , and  $Q$  is set to minimize these costs. Denote  $\hat{I}_1$  as the expected starting inventory position

at the beginning of a steady state cycle, where  $\hat{I}_1 = \hat{S}_2 - \mu$ .

$$\bar{C}(Q) = b \int_{\hat{I}_1+Q}^{\infty} (\xi - \hat{I}_1 - Q) \phi_{\xi}(\xi) d\xi + h \int_0^{\hat{I}_1+Q} (\hat{I}_1 + Q - \eta) \phi_{\eta}(\eta) d\eta \quad (15)$$

where  $\eta = 2\xi$  denotes the accumulated demand in the steady state cycle with two periods,  $\phi_{\eta}$  is the pdf of  $\eta$ , and  $\phi_{\xi}$  is the pdf of  $\xi$ . By taking the first-order condition of  $\bar{C}$  with respect to  $Q$ , the optimal volume in the slow mode  $\hat{Q}$  can be approximated by solving the following transcendental equation:

$$h\Phi_{\eta}(\hat{S}_2 - \mu + \hat{Q}) + b\Phi_{\xi}(\hat{S}_2 - \mu + \hat{Q}) = b. \quad (16)$$

In summary, the result of Scenario C is:

$$\begin{cases} \hat{S}_1 = \Phi^{-1}\left(\frac{b-h}{b+h}\right) \\ \hat{S}_2 = \Phi^{-1}\left(\frac{b-h}{b+h}\right) \\ h\Phi_{\eta}(\hat{S}_2 - \mu + \hat{Q}) + b\Phi_{\xi}(\hat{S}_2 - \mu + \hat{Q}) = b \end{cases} \quad (17)$$

The policy in (17) can further be used as an alternative result for Scenarios A and B when the analytic approximations deteriorate. Because  $\hat{Q}$  increases in  $\Delta$  as shown in (10) and (14),  $\hat{Q}$  obtained from Scenario C can be set as the lower bound of Scenario A and upper bound of Scenario B (Fig. 4), and therefore, the ill-measured  $\hat{Q}$  in the two Scenarios when  $\Delta$  is close to  $h$  (the dotted line) is then discarded. This leads to revised ranges for Scenarios A and B, which we refer to as A' and B', as indicated in Fig. 4. More specifically, we denote  $\hat{Q}^A$ ,  $\hat{Q}^B$ , and  $\hat{Q}^C$  as the approximated  $Q$  from Scenarios A', B' and C, respectively. If  $\hat{Q}^A < \hat{Q}^C$ , we then use (17) to replace the result from (10); if  $\hat{Q}^B > \hat{Q}^C$ , we use (17) to replace the result from (14). We set the range for Scenario C according to how we apply the bounded  $\hat{Q}^C$ .

## 5. Numerical analysis

In this section, we perform a numerical study to obtain further insights into the MST policy. Specifically, we analyze the approximation accuracy relative to the optimal solution. We therefore determine the optimal solution using complete enumeration. In ad-



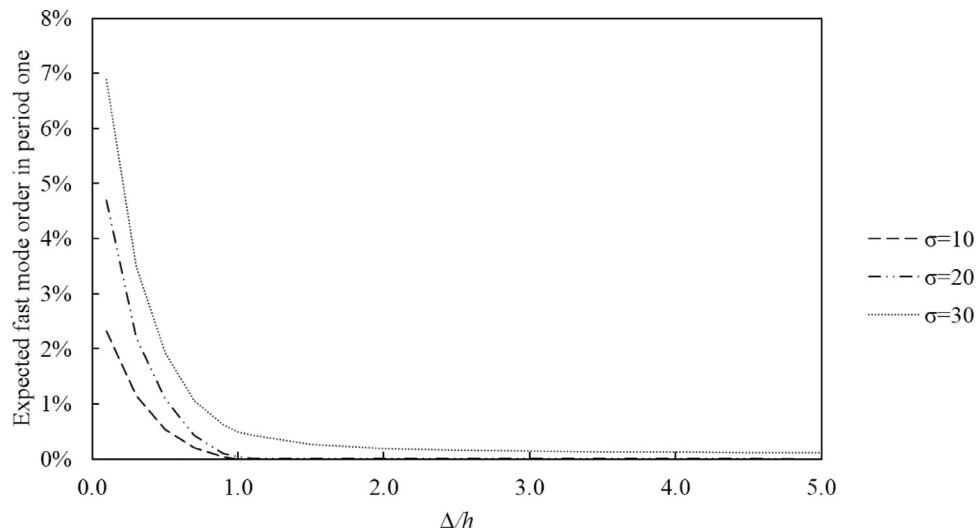


Fig. 5. Expected fast mode order in period one as a % of the total demand of a cycle.

dition, we use the parameters suggested by the company to understand the expected modal split and the cost savings induced by the MST policy. We also show an analysis when the unit delivery cost of the fast mode  $c^f$  depends on its delivery quantity.

The following assumptions are used in this section. The  $DC$  serves a large region; thus, the accumulated demand from all customers is pooled and can be assumed to follow a gamma distribution. The mean of the demand is standardized to 100 units. A unit could be an item, a carton, or a pallet in practice. Three different standard deviation values, 10, 20, and 30, are separately tested. This type of demand with high mean and low variability is of particular interest for the company because it focuses on high volume products, which maximize the impact for the company's supply chain innovation projects. The gamma distribution with mean 100 and standard deviation 30 represents a typical Runner SKU of the industry. Recall that the objective of the numerical analysis is to understand the accuracy of the approximation, and the ratio of modal split, the exact magnitude of the demand does not affect the percentage findings.

The unit holding cost  $h$  is generally measured as a percentage of the product value and is endogenously fixed by the firm. It is normalized to 1 in the numerical analysis. The unit backorder cost  $b$  is used to indirectly secure the firm's non-stockout probability ( $\alpha$ -service level) at 95%, which gives by using the newsvendor ratio  $\frac{b}{b+h}$  a backorder cost  $b = 19$ . The unit transport cost saving of the slow mode over the fast mode,  $\Delta$ , is an exogenous variable because it depends not only on the LSP's transport cost offers but also on the packaging and loading method of the firm. According to the firm's experience and discarding the negative values it could take,  $\Delta$  could be as high as 500% of  $h$ . A range of  $\Delta/h$  from 0 to 5 is tested.

### 5.1. Approximation accuracy

To understand the quality of the approximations, we conduct three different analyses. Our solution in Scenario A' is built on the assumption that no fast mode is utilized in the first period of the steady state cycle. We validate this approximation by calculating the share of fast mode orders in period one as an expected percentage of the total demand during a cycle for the unconstrained model. Similarly, we examine the practicability of the approximation in Scenario B' by analyzing the expected percentage of overshoot over demand. Finally, we validate the overall accuracy of our MST solution against the optimal solution.

To understand the approximation in Scenario A (and A'), we test the expected ordering quantity via the fast mode in period one as a percentage of the expected cyclic demand  $2\mu$ . Fig. 5 shows that the expected fast mode orders in period one decrease in  $\Delta/h$ . The values in the region of  $\Delta > h$  are all significantly smaller than those in the region of  $\Delta < h$ . More specifically, for  $\sigma = 10, 20$ , fast mode ordering in the first period drops to zero for  $\Delta < h$ . Recall that the parameter regime  $\Delta > h$  is our Scenario A, where we approximate the problem using a TBS model that assumes zero utilization of the fast mode. Accordingly, the approximation appears to be reasonable.

To understand the approximation in Scenario B (and B'), we numerically test the average overshoot in period two as a percentage of the expected cyclic demand  $2\mu$ . The results are shown in Fig. 6. When  $\Delta < h$ , the curves are flat and close to zero. This behavior provides numerical support for our Scenario B, where we approximate the MST problem with a TBS policy that has zero overshoot in the second period. When  $\Delta > h$ , the curves increase substantially, and the overshoot values cannot be eliminated. This effect becomes stronger as  $\sigma$  increases.

To validate the overall accuracy of the tool outlined in Section 4.2, we calculate the percentage error in the average total cost per steady state cycle, i.e.,  $\%C^D = \frac{C^{D*} - \hat{C}^D}{C^{D*}}$ . The optimal result  $C^{D*}$  is calculated using a complete enumeration. The approximation errors are shown in Fig. 7.

In general, our method provides a robust estimation. When  $\Delta < h$ , our approximated analytical expressions (14) in Scenario B provide very good results, and the approximation error is almost zero. When  $\Delta > h$ , the approximate result in (10) is less accurate than that in Scenario B but still provides robust results with an approximation error less than 3%.

We have to distinguish between two different types of errors in our overall solution: the approximation error when the TBS approach is implemented to solve MST problems and the endogenous error of the TBS policy itself. The first type of error is prominent around  $\Delta = h$  under the assumption that the fast mode is not utilized in Scenario A and the overshoot is neglected in Scenario B. By introducing a separate solution in Scenario C, the error shown in Fig. 3 is smoothed, as shown in Fig. 7. The second type of error occurs when  $\Delta > h$  and a TBS approach with heavy traffic analysis is applied. Allon and Van Mieghem (2010) and Janakiraman, Seshadri, and Sheopuri (2014) observe this error in their TBS studies. They find that the error of the TBS policy increases when the cost difference between the two modes increases. Our solutions are

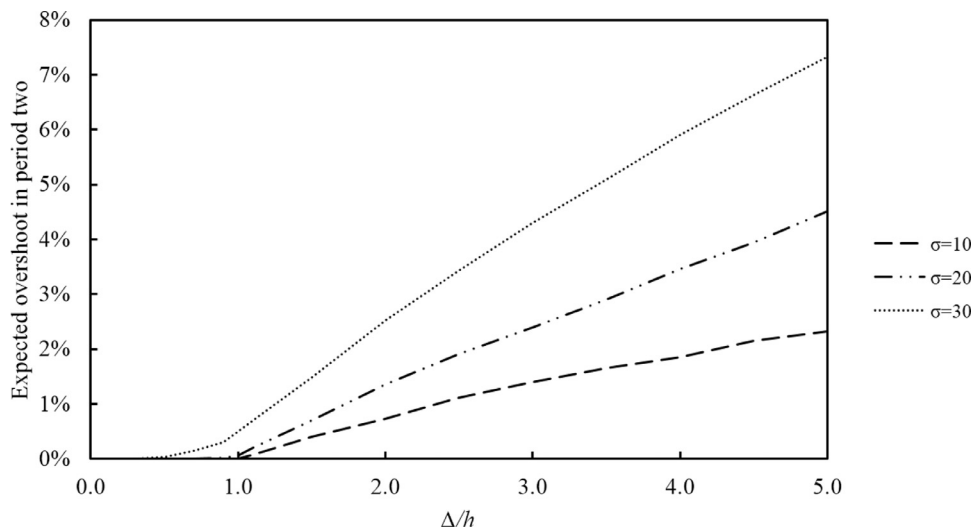


Fig. 6. Expected overshoot in period two as % of the total demand of a cycle.

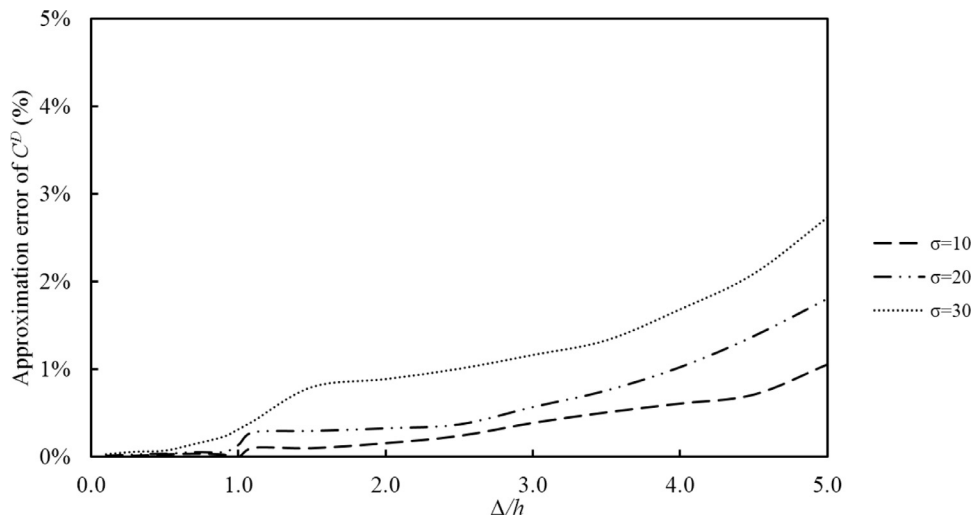


Fig. 7. The approximation error of  $C^D$  as a % compared to the optimal results.

aligned with their findings, showing an increase in error in  $\Delta$  in Fig. 7.

## 5.2. MST policy results

Next, we aim to obtain detailed insights into the results of the MST policy. We are particularly interested in the expected volume that is shipped in the slow mode at the optimum cost. This is particularly important as it provides an indication of the practical feasibility of the MST policy: the slow mode can only be organized for sufficiently large volumes. Next, we study the overall cost savings induced by the MST policy. Finally, we would like to understand the expected volume shipped in the slow mode in case the firm aims to implement the MST with the objective of zero cost savings. This case is important if a firm aims to apply the slow mode in line with her sustainability agenda rather than for cost minimization.

Fig. 8 shows the optimal slow mode volume expressed as a percentage of the mean demand per cycle. The solid line characterizes (as shown in Section 4.1) the deterministic benchmark scenario where either 50% of the cyclic demand (if  $\Delta < h$ ) or 100% of the cyclic demand (if  $\Delta > h$ ) is shipped via the slow mode.

When  $\Delta$  increases, i.e., the slow mode shipment becomes cheaper, the expected optimal volume in the slow mode increases.

Interestingly, the expected ratio in the slow mode is surprisingly high: for a typical SKU with  $(\mu, \sigma) = (100, 30)$ , even when the ratio of transport cost savings to holding cost is rather low, i.e.,  $\Delta/h = 0.01$ , approximately 22% of the expected transport volume per cycle can be shifted to the slow mode; when  $\Delta/h = 5$ , this ratio increases to approximately 85%. These ratios do not depend on the lower bound of the stochastic demand but are a function of its mean and the cost savings of the slow mode. Supply chain managers could work more aggressively on modal split projects with larger consignments on slow modes.

A further observation is that the slow mode split has different behaviors in  $\sigma$  at different  $\Delta/h$ -ratios. When the demand is more volatile ( $\sigma$  increases), the policy prefers more flexibility in MST and favors a larger volume in the fast mode. This is the intuitive reason why  $Q$  decreases in  $\sigma$ . Interestingly, we find that there is a certain area, when  $\Delta$  is close to  $h$  from the left side, where  $Q$  increases in  $\sigma$ . This is mainly due to the existence of a third effect: the *safety stock effect*. When the demand is more volatile, more safety stock is required to buffer uncertainty. Because of the cost advantage, the firm would prefer to use the slow mode to replenish the safety stock. When  $\Delta$  is close to  $h$ , both commitment and cycle stock effect almost cancel each other out and the safety

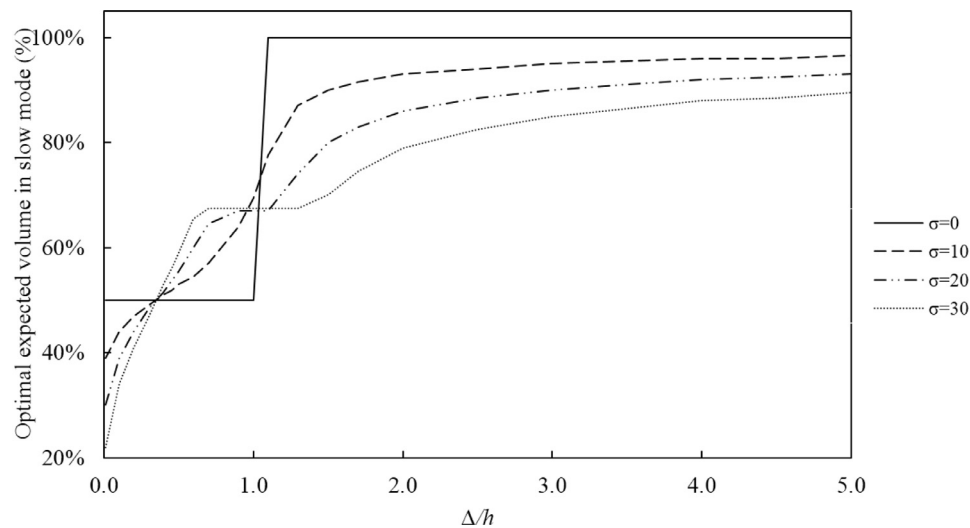


Fig. 8. Expected slow mode volume in % of the expected demand per cycle under profit maximization.

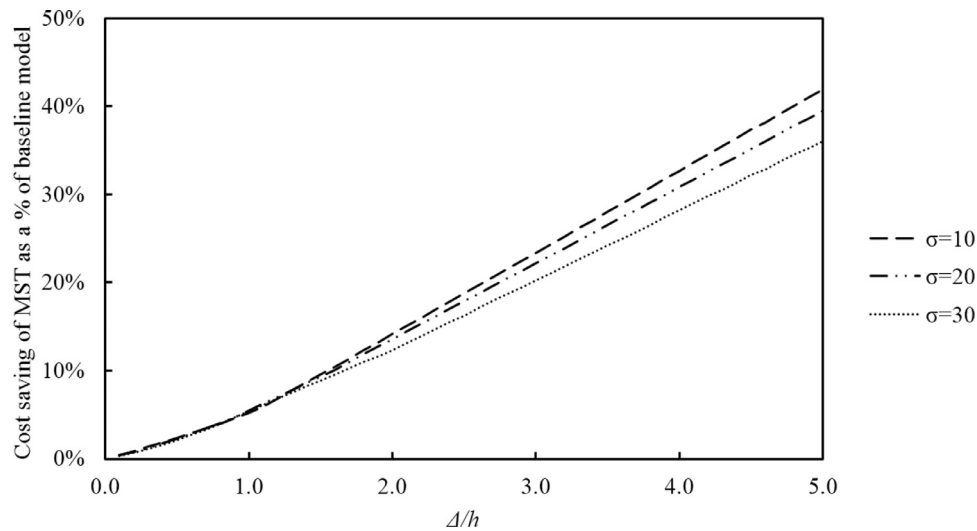


Fig. 9. Expected total cost savings using the MST policy.

stock effect dominates, increasing the size of  $Q$  when  $\sigma$  increases. This phenomenon occurs in an area where  $\Delta$  approaches  $h$  from the left side because of the following: 1). the commitment effect and the overstocking effect almost cancel out (close to the break-even point  $\Delta = h$ ), so that the safety stock effect is left alone, and 2). the commitment effect is still small ( $\Delta < h$ ), so that an increase in the slow mode volume does not lead to too much inflexibility in the system.

The expected total cost savings of MST compared to the baseline fast-mode-only model are shown in Fig. 9. In the region  $\Delta > h$ , the curves are sensitive to  $\sigma$ , indicating a cost advantage of MST when the demand is less volatile and the commitment effect dominates. In the region  $\Delta < h$ , however, the cost savings of MST are rather insensitive to demand volatility. This result is mainly because of the existence of the safety stock effect, as also observed in Fig. 8. When demand is more volatile, the increasing safety stock drives to a larger share in the low-cost slow mode and further leads to a cost reduction. This safety stock effect offsets the commitment effect, and therefore, the total cost savings of MST are not sensitive to  $\sigma$ . Practitioners generally believe that the slow mode transport is always favorable for products with more stable demand patterns. Our numerical results reveal that this is only part of the story—if the commitment effect dominates, we find this to

be the case; if the cycle stock effect dominates, the volatility of the demand is indeed insensitive to the total cost savings of MST. Again, the measurement of the sizes of the two effects can be simply achieved by a comparison of the two parameters  $\Delta$  and  $h$ .

Thus far, we have assumed that firms intend to minimize the total expected costs using the MST policy. However, firms could have an alternative objective: rather than minimizing the cost, they could intend to improve their sustainability or carbon agenda by minimizing the use of fast transport modes. In this case, the MST policy could be applied to cost-neutrally shift volumes to the slow mode transport. Fig. 10 shows the resulting cost savings from shifting volume to the slow mode transport for  $\sigma = 20$  at different transport cost savings  $\Delta$ . The slow mode maximizing volume is reached for the zero-cost case with  $\Pi = 0$ . The numerical results highlight that a large fraction of the total volume can be shipped using the slow mode if no cost decrease compared to the baseline setting is required: even for low transport cost savings compared to holding costs  $\Delta = 0.1h$ , almost 60% of the volume is shipped in the slow mode. For higher  $\Delta$ , the overall cost savings decrease with the slow mode volume but only become negative as the slow mode volume approaches 100%. Here, the full reliance on the slow mode leads to a complete loss of flexibility.

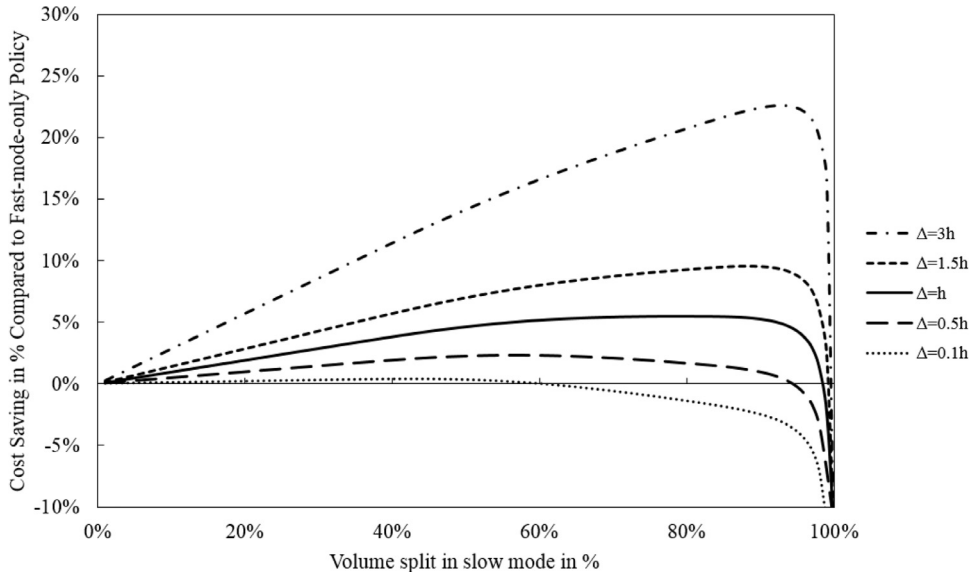


Fig. 10. Cost savings based on the expected slow mode volume in % of mean demand.

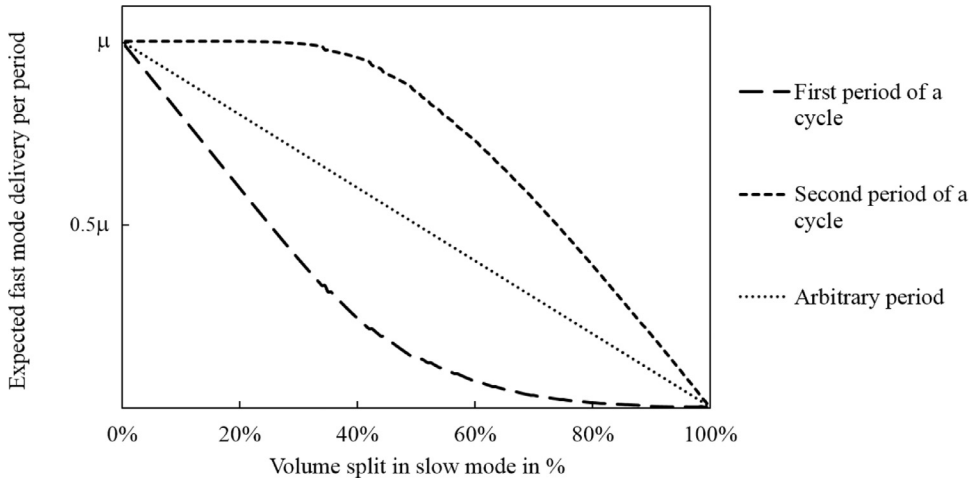


Fig. 11. Expected fast mode delivery in the first and second periods of a cycle and for an arbitrary period.

### 5.3. MST policy implications

Similar to previous literature on dual sourcing, e.g., Allon and Van Mieghem (2010); Veeraraghavan and Scheller-Wolf (2008), and Janakiraman, Seshadri, and Sheopuri (2014), our model thus far assumes that the unit transport cost of the fast mode,  $c^f$ , is constant and does not depend on other external factors such as volume or variability changes. However, many LSPs offer freight rates that depend on the volume shipped (Coyle, Novack, B., & J., 2015): The more volume a LSP receives from a shipper, the cheaper is the unit transport cost. Likewise, an LSP could also raise the unit transport cost as the variability of the freight volume increases. We next analyze the volume and variability implications of the MST policy.

**Volume implication.** Fig. 11 shows the mean usage of the fast mode, separately illustrated for the first and second periods of a steady state cycle, as well as for a general period without distinguishing between the cycle periods. It can be observed that when  $Q$  is rather low and is increased, the average fast mode delivery in the first period of a cycle is immediately affected due to the slow mode order arriving in that period. In contrast, the expected fast mode quantity of the second period of the cycle remains almost the same if  $Q$  is rather low. It only increases for higher  $Q$  as slow

mode quantities are carried over to the second period of the cycle. For an arbitrary period (generally, the average implication of the first and second periods) we find that the average fast mode delivery almost linearly decreases in  $Q$ . As  $Q$  reaches the mean demand of the full cycle, the expected fast mode delivery in both periods tends to zero.

Suppose the fast mode LSP installs a volume-dependent freight rate tariff with an  $m+1$ -tier structure depending on the expected fast mode usage per cycle. More specifically, the LSP charges a unit transport cost  $c_j^f$  if the expected fast mode shipment per cycle, denoted as  $Q^f$ , exceeds or is equal to a threshold  $\bar{Q}_{m-j}^f$ , for  $j = 0, \dots, m$ . The fast mode transport cost per unit is then:

$$c^f = \begin{cases} c_0^f & \text{if } \bar{Q}_m^f \leq Q^f \leq 2\mu \\ c_1^f & \text{if } \bar{Q}_{m-1}^f \leq Q^f < \bar{Q}_m^f \\ \vdots & \vdots \\ c_m^f & \text{if } 0 \leq Q^f < \bar{Q}_1^f \end{cases}, \quad (18)$$

where  $c_0^f < c_1^f < \dots < c_m^f$ , indicating that the less expected volume  $Q^f$  the fast mode LSP receives from the shipper, the higher is the



unit fast mode transport cost  $c^f$ . The above volume-dependent tariff is richly studied in the literature (see, e.g., Swenseth and Godfrey, 2002 and Ventura, Valdebenito, & Golany, 2013).

Since the expected fast mode shipment per cycle is a linear function of the slow mode shipment, i.e.,  $Q^f = 2\mu - Q$ , the fast mode cost per unit is  $c_j^f$  if the slow mode shipment  $Q \leq 2\mu - \bar{Q}_{m-j}^f$ . Therefore, the expected cost function of an MST policy with a volume-dependent tariff is multi-fold with  $C^D = c^s Q + c_j^f (2\mu - Q) + \sum_{i=1}^2 L_i (S_i + E[O_{\infty,i}])$ , for  $Q \leq 2\mu - \bar{Q}_{m-j}^f$  and  $j$  from 0, ...,  $m$ . We assume in the baseline situation where all volume are shipped in the fast mode, a lowest unit cost  $c_0^f$  is charged by the fast mode LSP, such that the expected cost is  $C^B = c_0^f 2\mu + 2L(S^B)$ . The functional form of the expected cost savings of the MST policy  $D$  over the baseline policy  $B$  depends on the fast mode tariff and changes to

$$\begin{aligned} \Pi_j(Q, S_1, S_2) &= (c_j^f - c^s)Q + (2L(S^B) \\ &\quad - \sum_{i=1}^2 L_i (S_i + E[O_{\infty,i}])) - (c_j^f - c_0^f)2\mu. \end{aligned} \quad (19)$$

Compared to the case without the volume-dependent tariff, there is an additional term  $-(c_j^f - c_0^f)2\mu$  that reduces the total cost savings, which characterizes the additional fast mode cost that the firm faces due to the higher price caused by the lower expected fast mode volume.

This problem can be solved by using a modified version of the classical algorithm for the standard economic order quantity (EOQ) problem with quantity discount, illustrated in Silver, Pyke, and Peterson (1998, p. 162). For all cost levels  $c_j^f$ , we calculate the transport cost savings  $\Delta_j = c_j^f - c^s$  and the resulting  $\hat{Q}_j, \hat{S}_{j1}, \hat{S}_{j2}$ . The problem is that this solution may be infeasible if  $\hat{Q}_j > 2\mu - \bar{Q}_{m-j}^f$  since this cost savings function is only valid for  $\hat{Q}_j \leq 2\mu - \bar{Q}_{m-j}^f$ . In such a case, the slow mode volume for this cost level is set to  $\hat{Q}_j = 2\mu - \bar{Q}_{m-j}^f$  and  $\hat{S}_{j1}, \hat{S}_{j2}$  are calculated accordingly based on the previously discussed scenarios A, B, and C. The cost savings,  $\Pi_j$ , for cost level  $c_j^f$  are then determined by (19), and the optimal approximate solution is

$$(\hat{Q}, \hat{S}_1, \hat{S}_2) = \arg \max \{ \Pi_j(\hat{Q}_j, \hat{S}_{j1}, \hat{S}_{j2}) | j = 0, \dots, m \}. \quad (20)$$

Installing a volume-dependent tariff by the fast mode LSP may lead to both a higher or lower volume shift to the slow mode depending on the specific cost tariff. The explanation is the following: Given the unit transport cost in the baseline model is  $c_0^f$ , the total expected fast mode volume is  $2\mu$ . Once the shipper implements MST and ships  $Q$  units via the slow mode, the expected fast mode shipments during a two period cycle would decrease to  $Q^f = 2\mu - Q$ , which may have negative consequences to the fast mode LSP. By installing the volume-dependent tariff, the fast mode LSP charges higher prices, if the expected fast mode shipments fall below certain thresholds. Depending on the magnitude of these price increases, the shipper may either shift even more volume to the slow mode, if the relative cost savings between the two transport modes is even more beneficial; or the shipper may shift volume back from the slow mode to the fast mode in order to utilize the better fast mode price due to a larger volume.

The following numerical example will illustrate these two effects. We use the same data as in Section 5, i.e. the mean demand per period follows a Gamma distribution with  $\mu = 100$  and  $\sigma = 30$ . Suppose the unit fast mode cost in the baseline model is  $c_0^f = 10$ . As the shipper implements MST, the optimal volume shifted to the slow mode LSP is  $A = 179$  as illustrated in Fig. 12. We now assume that the fast mode LSP installs a two-fold tariff as such  $c_0^f = 10$

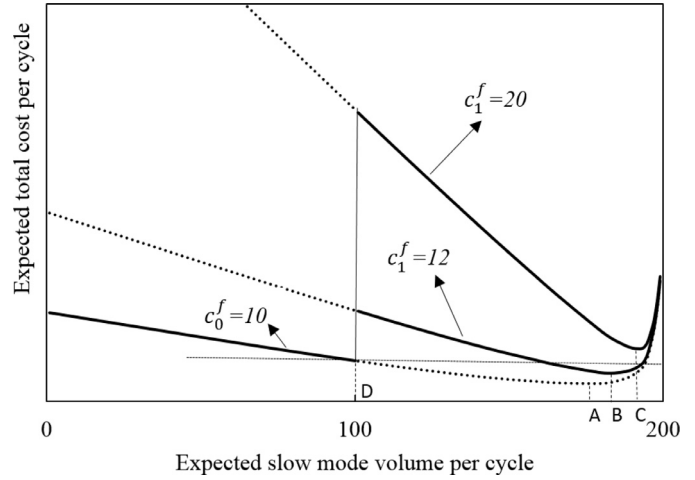


Fig. 12. Expected total cost per period with a volume-dependent tariff of the fast mode.

only if  $Q^f \geq 100$  (equivalently, if the slow mode volume  $Q \leq 100$ ). If  $Q^f < 100$  (equivalently,  $Q > 100$ ), we distinguish two alternative unit cost scenarios (i)  $c_1^f = 12$  and (ii)  $c_1^f = 20$ . The volume-dependent tariff can be formalized as follows:

$$c^f = \begin{cases} c_0^f = 10 & \text{if } 100 \leq Q^f \leq 2\mu \\ c_1^f \in \{12, 20\} & \text{if } 0 \leq Q^f < 100 \end{cases} \quad (21)$$

Fig. 12 illustrates the expected total cost per cycle as a function of slow mode volume  $Q$  for the different unit fast mode costs. The solid line represents the region in which this cost function is feasible, hence the dotted line represents the region in which this cost function is infeasible. For  $Q \leq 100$ , the expected fast mode volume  $Q^f$  exceeds the threshold of 100, therefore, the lowest unit cost  $c_0^f = 10$  is feasible and the lowest total cost is reached at  $Q = 100$  (point D). For  $Q > 100$ , the expected fast mode volume  $Q^f$  would be less than the threshold of 100 such that the higher price  $c_1^f \in \{12, 20\}$  is charged. If  $c_1^f = 12$ , one can see that the cost-minimal slow mode quantity is at B, i.e.,  $Q = 183$ , whose cost is lower than the expected cost at D. Therefore, if  $c_1^f = 12$ , the shipper would shift even more volume to the slow mode LSP (from A to B) with the consequence that the expected volume for the fast mode LSP decreases. However, if  $c_1^f = 20$ , the cost-minimal expected cost is at  $Q = 192$  (point C), which is larger than the expected cost at the threshold  $Q = 100$  where the lower cost  $c_0^f = 10$  is still feasible (point D). In this case, the shipper would move volume back from the slow mode to the fast mode (from A to D) to benefit from the lower price.

In short, the volume-dependent tariff can lead to both more and less usage of the slow mode. The impact depends on the specific design of the tariff.

**Variability implication.** Besides the volume implication, implementing an MST policy also affects the variability of the fast mode shipments. When fast mode shipments become more volatile, the fast mode LSP must put more effort in capacity management, e.g., for empty truck reposition and truck driver retention (Jonathon, Taylor, Usher, English, & Roberts, 2002), and might consequently raise the freight rate.

Suppose the fast mode LSP needs to plan its fleet capacity based on more volatile orders from the shipper. The expected fast mode shipment during a two period cycle is  $Q^f$  with  $Q^f = 2\mu - Q$ , and the standard deviation of the shipment is  $\sigma^f$ . Assuming that the fast mode LSP plans its capacity, denoted as  $C^f$ , via a simple newsvendor-style model:  $C^f = Q^f + z\sigma^f$ . The parameter  $z$  is linked to the safety capacity factor to be met by the fast mode

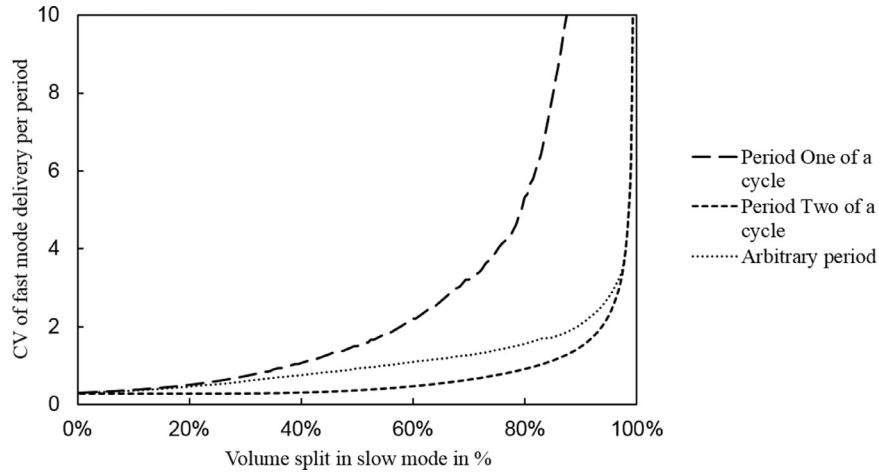


Fig. 13. Coefficient of variation of fast mode delivery in the first and second periods of a cycle and for an arbitrary period.

LSP. If  $\bar{c}$  is the cost to install one unit of fleet capacity, then the fast mode LSP's average capacity cost for one unit shipped is:  $\bar{c}^f = \bar{c} \cdot C^f / Q^f$ . We can simplify the equation by substituting the capacity  $C^f = Q^f + z\sigma^f$  and the coefficient of variation of the fast model shipment  $CV^f = \sigma^f / Q^f$ . Therefore, the average capacity cost for one unit shipped can be written as:  $\bar{c}^f = \bar{c}(1 + z \cdot CV^f)$ , which shows the importance of the coefficient of variation of the fast mode shipment.

Fig. 13 shows  $CV^f$ , separately illustrated for the first and second period of a steady state cycle, as well as for a general period without distinguishing between the cycle periods. It can be observed that all three curves increase monotonously, indicating that the fast mode shipment is more volatile when more volume is shifted to the slow mode. In addition, the  $CV^f$  of the first period of a cycle is larger than that of the second period, because the mean fast mode shipment in the first period is smaller (shown in Fig. 13), provided that a substantial part of demand is already satisfied by the slow mode in this period.

To consider the variability implication in the MST policy, an approach similar to the one implemented for the volume implication analysis can be applied. To be more specific, the fast mode LSP could install the following  $CV^f$ -dependent tariff with the unit transport cost  $c_j^f$ :

$$c_j^f = \begin{cases} c_0^f & \text{if } 0 \leq CV^f < \bar{CV}_1^f \\ c_1^f & \text{if } \bar{CV}_1^f \leq CV^f < \bar{CV}_2^f \\ \vdots & \vdots \\ c_m^f & \text{if } \bar{CV}_m^f \leq CV^f \end{cases} \quad (22)$$

where  $c_0^f < c_1^f < \dots < c_m^f$ , indicating that the more volatile the fast mode shipment, the higher the unit fast mode transport cost  $c^f$ .

The implication of the variability-dependent tariff is similar to that of the volume-dependent tariff. Suppose the fast mode LSP observes a higher variability in its shipment due to the shipper's MST implementation, and wants to raise the freight rate according to tariff (22). Similar to the volume-dependent implication, the shipper may either increase or decrease the volume shifted to the slow mode LSP, depending on the freight rate increase (see Fig. 12). If the unit transport cost increase (due to the variability-dependent tariff) is not too large, the shipper will shift even more volume to the slow mode LSP since it minimizes its total cost. If the unit transport cost increase (due to the variability-dependent tariff) is large, then the shipper would shift the volume back to the last threshold where he still obtains a lower fast mode freight rate. Similar to the case with the volume-dependent tariff, the fi-

nal decision of the shipper depends on the specific parameters of the variability-dependent tariff (22). Note, a LSP could combine the volume-dependent and variability-dependent tariffs to account for both slow mode implications. Since both implications lead to either more or less usage of the slow mode, a combination of both tariffs will hence be expected to have a similar impact.

In summary, due to the slow mode shipment, the LSP observes lower freight volumes and higher variability in the fast mode. Consequently he might raise the freight rate under a specified tariff. We find that the specific design of the freight tariff can lead to either an increase or a decrease of the shipper's slow mode usage.

## 6. Summary

The main contribution of this paper is the development of a modal split transport (MST) policy that enables volume allocation into two transport modes and that integrates inventory controls. An attractive feature of the MST policy is the cost-efficient slow mode shipment in the first period of each two-period cycle. This constant slow mode volume is complemented by flexible volumes that can be shipped via the more costly fast mode every period. Accordingly, the MST model needs to consider different delivery frequencies of the two modes and therefore has an extended mathematical structure compared to the classical tailored base-surge problem, which assumes that the cheap and the expensive supplier deliver in every period. However, a "heavy traffic" phenomenon, shown in the TBS problem, is not guaranteed in the MST model and requires new solution procedures. We approximate the steady state cost function of the MST problem and derive closed-form expressions for the modal split policy using three distinct scenarios. This solution provides a simple and easy-to-implement tool for practitioners. We find that the trade-off between the commitment effect and the cycle stock effect drives the optimal level for the slow mode quantity. When a constant quantity is committed in the slow mode, the firm on one side gains a transport cost savings every cycle, but on the other side endures extra cycle stock cost in the first period of the cycle. Using a numerical study with practically based data from a company, we find that the solution approach provides an approximation error of less than 3%. We also analyze the cost reductions and find that they can be significant based on the relationship between the transport cost savings in the slow mode and inventory holding costs.

The proposed MST policy has important managerial implications as it bridges a key dilemma that MST has been facing in practice over the years: despite receiving increasing attention from the industry, government and academia, real-world implementa-

tions have ground to a halt. Based on our discussions with the company, we found that practitioners were particularly challenged by integrating multi-modal transport decisions with inventory controls. The proposed model can provide practitioners with the confidence that these decisions can be well integrated and handled. The numerical results are encouraging with respect to the practical feasibility of the slow mode. The supply chain managers that we interviewed were concerned about the volume split between the fast and the slow transport modes in the cost-optimal solution. Managers tend to believe that if they need to commit a constant volume in the slow mode cyclically in the long run, the size of this fixed volume should not exceed the lower bound of the demand over the entire period. In contrast to their expectations, we found that the volumes on the slow mode are sufficiently large to justify the implementation of new slow mode transport routes. For Runner products with high demand and relatively low demand variation, the volume for the slow mode can be as high as 85% of the total volume. Our insights are also relevant for firms that apply the MST policy to reduce carbon emissions via the slow transport mode rather than costs. If the objective is to implement MST in a cost-neutral way, the numerical results indicate that large volumes are shipped in the slow mode even if the transport cost savings are relatively small. This is encouraging because it provides the potential to trigger an even larger demand for the slow mode that is required to facilitate the MST approach.

Multiple extensions of the MST study could be considered in future research. Our model is limited because it assumes a slow mode delivery frequency that is half of that of the fast mode and a cycle has two periods, i.e., the steady state cost function of the MST policy has at most three decision variables, the slow mode volume  $Q$  and the base stock levels  $S_1$  and  $S_2$  of the fast mode policy. Our approximated analytic solution builds on the fact that under certain circumstances (relationship between transportation cost savings and holding cost) specific elements of the cost function (in particular expected overshoot terms) can be omitted, which allows us to derive closed-form expressions for the relevant decision variables. When  $n > 2$ , the tradeoff between transport cost savings and inventory costs still exists and a similar approximation method could be applied. However, the number of decision variables increases linearly, i.e., the number of decision variables is  $n + 1$ , which makes the analytical tractability more difficult. To keep the model analytically tractable and to derive closed-form expressions, more terms of the cost functions have to be eliminated, which is theoretically possible but would increase the approximation error. Therefore, this type of approximation would not be appropriate for larger cycle lengths. Numerical solutions are recommended for more accurate results. Theoretically, our analysis can be extended to any arbitrary  $n$ ; however, when  $n$  increases, in the absence of any other relevant adjustments, the stochasticity of the cyclic demand will increase, and this might deteriorate the performance of the approximations.

From a managerial perspective, we expect our results to extend to cases of  $n > 2$ , i.e., for similar transport and inventory costs the volume shipped in the slow mode is rather high compared to the volume shipped in the fast mode. For example, consider our current model with  $n = 2$  assuming that a week is separated into two periods: a slow mode shipment on a Monday (period 1) and fast mode shipments on Monday and Thursday (periods 1 and 2). Now, we extend the model to  $n = 6$ , i.e., keeping the slow mode on Monday and allowing fast mode shipments each day from Monday to Saturday. There is no indication that the slow mode volume would considerably decrease as inventory holding cost rates are adjusted to individual days rather than periods of 3 days. However, more detailed research is required in the future, particularly as expected demand is likely to differ across periods.

Thus far,  $n$  is only studied as an exogenous parameter. Occasionally, larger shippers with substantial volume could have strong bargaining power in the transport market and can negotiate the delivery frequencies of the slow mode with LSPs. The delivery frequency will then be an endogenous decision variable to be optimized, rather than an exogenous parameter. This entails further changes of the MST model, e.g., cost structure of the slow mode, and optimality decisions of the fast mode. The approximation used in this article might not work, and further solution methods are required.

The scope of MST could be further broadened by incorporating multiple products. The solution of the single-product model can be applied as a first approximation for the multi-product problem by optimizing the modal split for each individual product separately. However, this procedure does not fully capture the advantages of multi-product management, e.g., when the demand of one product drops, that of the others might still be adequate to fill in the slow mode. The aggregated demand might have a lower coefficient of variation and drives a pooling effect. The firm could then reserve a pooled capacity in the slow mode for all products, rather than booking capacities for each product separately. The pooling effect might foster a larger modal split into the slow mode. Still, the downside of the multi-product management is that the firm needs to make extra decisions in the allocation of products into the pool. Extra handling costs will occur in the allocation operations, e.g., to pack, transit, declare, and load the products with different sizes, weights and safety instructions. Additional data and parameters are needed to further analyze the tradeoff and profitability of the multi-product problem.

## Appendix A

### A.1. Proof of Proposition 1

We take one cycle of two periods as the inventory review time unit. Because there are no fast mode orders in the first period of a cycle, both the fast and slow modes are utilized only once in a cycle, and  $S_2$  is the only order-up-to level in this model. The demand of one cycle follows an i.i.d distribution with mean  $2\mu$  and standard deviation  $\sqrt{2}\sigma$ . The dynamics of this inventory model are shown in Fig. A.14.

The overshoot of the cycle is explained as follows: denote as  $O_k$ , ( $O_k \geq 0$ ) the overshoot in cycle  $k$ , and the inventory position after the deliveries from both modes and before the realization of demand is  $S_2 + O_k$ . During the cycle, a demand of  $2\xi$  is realized, and the inventory is  $S_2 + O_k - 2\xi$  at the end of the cycle. At the beginning of the next cycle,  $Q$  arrives and the overshoot in the next cycle is calculated in (23). Note that the overshoot recursion does not depend on any inventory base stock levels.

$$O_{k+1} = (O_k + Q - 2\xi)^+ \quad (23)$$

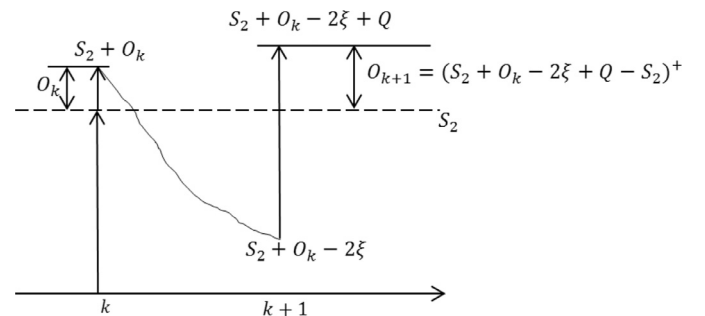


Fig. A.14. The overshoot in a cyclic inventory model.

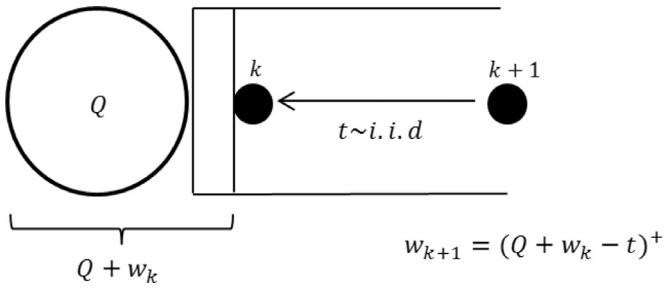


Fig. A.15. The waiting time in a GI/D/1 queue.

As a comparison of the inventory system, we consider a GI/D/1 queue in Fig. A.15, where the customer inter-arrival time  $t$  follows an i.i.d distribution with mean  $\mu_t$  and standard deviation  $\sigma_t$ , and the service time  $Q$  is deterministic.

Assume that at time 0, customer  $k$  arrives and that her waiting time is  $w_k$  with  $w_k \geq 0$ . The queue will be busy in the next time interval  $Q + w_k$ . In period  $t$ , customer  $k+1$  arrives. If  $t \geq Q + w_k$ , the next customer has zero waiting time; if  $t \leq Q + w_k$ , he has to wait for a time interval of  $Q + w_k - t$ . Consequently, the waiting time of this customer is:

$$w_{k+1} = (w_k + Q - t)^+ \quad (24)$$

Comparing (23) with (24), the stochastic variables  $O_k$  and  $w_k$  have analogous recursion functions. Consequently, we can model the cyclic inventory model as a GI/D/1 queue, where the demand  $2\xi$  represents the inter-arrival time of customers, the constant delivery  $Q$  represents the service time, and most importantly, the overshoot is the waiting time of the queue.

Kingman (1970) provides a bound for the expected waiting time of a general GI/G/1 queue:  $E[w] \leq \frac{\sigma_t^2}{2(2\mu_t - Q)}$ . The GI/D/1 queue and the cyclic inventory model are connected by  $\sigma_t^2 = 2\sigma^2$  and  $\mu_t = 2\mu$ , and the expected overshoot at the end of a cycle is then  $E[O_{\infty,2}] \leq \frac{\sigma^2}{2\mu - Q}$ .

Substituting  $\tilde{S}_2 = S_2 + E[O_{\infty,2}]$  within (6) and calculating the first-order derivative with respect to  $\tilde{S}_2$  gives  $h + h\Phi(\tilde{S}_2) - b(1 - \Phi(\tilde{S}_2))$ , which implies  $\tilde{S}_2$  as follows:  $\tilde{S}_2 = \Phi^{-1}\left(\frac{b-h}{b+h}\right)$ . Therefore, the base stock level is  $S_2 = \Phi^{-1}\left(\frac{b-h}{b+h}\right) - E[O_{\infty,2}]$ .

## A.2. Proof of Proposition 2

Implement the following rule: for any  $(x, y)$  such that  $y \geq 0$ ,  $(x+y)^+ \leq x^+ + y$  and  $(x+y)^- \leq x^-$ ,  $L_1(S_1 + E[O_{\infty,1}]) = hE[(S_1 + E[O_{\infty,1}] - \xi)^+] + bE[(S_1 + E[O_{\infty,1}] - \xi)^-] \leq hE[(S_1 - \xi)^+] + hE[O_{\infty,1}] + bE[(S_1 - \xi)^-] = L_1(S_1) + hE[O_{\infty,1}]$ .  $\hat{\Pi}$  in (12) can be approximated as:

$$\hat{\Pi} = \Delta Q + 2L(S^B) - L_1(S_1) - hE[O_{\infty,1}] - L_2(S_2) \quad (25)$$

$E[O_{\infty,1}]$  is approximated in (13):  $E[\hat{O}_{\infty,1}] = \int_0^{S_2+Q-S_1} (S_2 + Q - S_1 - \xi)\phi(\xi)d\xi$ , with  $\frac{dE[\hat{O}_{\infty,1}]}{dQ} = \frac{dE[\hat{O}_{\infty,1}]}{dS_2} = -\frac{dE[\hat{O}_{\infty,1}]}{dS_1} = \Phi(S_2 - S_1 + Q)$ .

The first-order condition of (25) w.r.t.  $Q$  is:  $\frac{d\hat{\Pi}}{dQ} = \Delta - h\Phi(S_2 - S_1 + Q) = 0$ .

The first-order condition of (25) w.r.t.  $S_2$  is:  $\frac{d\hat{\Pi}}{dS_2} = -h\Phi(S_2 - S_1 + Q) - (b+h)\Phi(S_2) + b = 0$ .

The first-order condition of (25) w.r.t.  $S_1$  is:  $\frac{d\hat{\Pi}}{dS_1} = h\Phi(S_2 - S_1 + Q) - (b+h)\Phi(S_1) + b = 0$ .

Combine the conditions  $\frac{d\hat{\Pi}}{dQ}$  and  $\frac{d\hat{\Pi}}{dS_2}$ :  $\hat{S}_2 = \Phi^{-1}\left(\frac{b-h}{b+h}\right)$ .

Combine the conditions  $\frac{d\hat{\Pi}}{dQ}$  and  $\frac{d\hat{\Pi}}{dS_1}$ :  $\hat{S}_1 = \Phi^{-1}\left(\frac{b+h}{b+h}\right)$ .

Plug in  $\hat{S}_1$  and  $\hat{S}_2$  into the equation  $\frac{d\hat{\Pi}}{dQ} = 0$ :  $\hat{Q} = \Phi^{-1}\left(\frac{b+h}{b+h}\right) - \Phi^{-1}\left(\frac{b-h}{b+h}\right) + \Phi^{-1}\left(\frac{b}{h}\right)$ .

## References

- Allon, G., & Van Mieghem, J. A. (2010). Global dual sourcing: Tailored base-stock allocation to near-and offshore production. *Management Science*, 56(1), 110–124.
- American Transportation Research Institute (2014). An analysis of the operational costs of trucking: A 2014 update (Accessed in September, 2014) <http://www.atrri-online.org/wp-content/uploads/2014/09/ATRI-Operational-Costs-of-Trucking-2014-FINAL.pdf>.
- Arts, J., & Kiesmüller, G. P. (2013). Analysis of a two-echelon inventory system with two supply modes. *European Journal of Operational Research*, 225(2), 263–272.
- Arts, J., van Vuuren, M., & Kiesmüller, G. (2011). Efficient optimization of the dual-index policy using Markov chains. *IIE Transactions*, 43(8), 604–620.
- Barankin, E. W. (1961). A delivery-lag inventory model with an emergency provision (the single-period case). *Naval Research Logistics Quarterly*, 8(3), 285–311.
- Baumol, W. J., & Vinod, H. D. (1970). An inventory theoretic model of freight transport demand. *Management Science*, 16(7), 413–421.
- BCG (2015). A hard road: Why CPG companies need a strategic approach to transportation (Accessed July, 2015) <https://www.bcgperspectives.com/content/articles/consumer-products-supply-chain-management-a-hard-road/>.
- Bontekoning, Y. M., Macharis, C., & Trip, J. J. (2004). Is a new applied transportation research field emerging?—A review of intermodal rail-truck freight transport literature. *Transportation Research Part A: Policy and Practice*, 38(1), 1–34.
- Boute, R., & Van Mieghem, J. A. (2014). Global dual sourcing and order smoothing: The impact of capacity and lead times. *Management Science*, 61(9), 2080–2099.
- Coyle, J. J., Novack, R. A., B., G., & J., B. E. (2015). *Transportation: A global supply chain perspective* (8th). Mason, OH: South-Western College Pub.
- Crainic, T. G., & Kim, K. H. (2007). Intermodal transportation. In C. Barnhart, & G. Laporte (Eds.), *Handbooks in operations research and management science: vol. 14* (pp. 467–537). Elsevier.
- Dekker, R., Bloemhof, J., & Mallidis, I. (2012). Operations research for green logistics—An overview of aspects, issues, contributions and challenges. *European Journal of Operational Research*, 219(3), 671–679.
- Dey, A., LaGuardia, P., & Srinivasan, M. (2011). Building sustainability in logistics operations: A research agenda. *Management Research Review*, 34(11), 1237–1259.
- EUROSTAT (2015). EU transport in figures—Statistical pocketbook 2015. <http://ec.europa.eu/transport/facts-fundings/statistics/doc/2015/pocketbook2015.pdf>.
- Janakiraman, G., Seshadri, S., & Sheopuri, A. (2014). Analysis of tailored base-stock policies in dual sourcing inventory systems. *Management Science*, 61(7), 1547–1561.
- Janssen, F., & de Kok, T. (1999). A two-supplier inventory model. *International Journal of Production Economics*, 59(1), 395–403.
- Jonathon, T. F., Taylor, G. D., Usher, J. S., English, J. R., & Roberts, J. N. (2002). Forecasting freight demand using economic indices. *International Journal of Physical Distribution & Logistics Management*, 32(4), 299–308.
- Karlin, S. (1960). Dynamic inventory policy with varying stochastic demands. *Management Science*, 6(3), 231–258.
- Kingman, J. (1970). Inequalities in the theory of queues. *Journal of the Royal Statistical Society. Series B (Methodological)*, 32(1), 102–110.
- Klosterhalfen, S., Kiesmüller, G., & Minner, S. (2011). A comparison of the Constant-Order and Dual-Index Policy for dual sourcing. *International Journal of Production Economics*, 133(1), 302–311.
- Kutanoglu, E., & Lohiya, D. (2008). Integrated inventory and transportation mode selection: A service parts logistics system. *Transportation Research Part E: Logistics and Transportation Review*, 44(5), 665–683.
- Lammgård, C. (2012). Intermodal train services: A business challenge and a measure for decarbonisation for logistics service providers. *Research in Transportation Business & Management*, 5, 48–56.
- Larson, P. D. (1988). The economic transportation quantity. *Transportation Journal*, 28(2), 43–48.
- Lloret-Battle, R., & Combes, F. (2013). Estimation of an inventory-theoretic model of mode choice in freight transport. *Journal of the Transportation Research Board*, 2378, 13–21.
- Macharis, C., Van Hoeck, E., Pekin, E., & Van Lier, T. (2010). A decision analysis framework for intermodal transport: Comparing fuel price increases and the internalisation of external costs. *Transportation Research Part A: Policy and Practice*, 44(7), 550–561.
- McKinnon, A., & Ge, Y. (2006). The potential for reducing empty running by truck: A retrospective analysis. *International Journal of Physical Distribution & Logistics Management*, 36(5), 391–410.
- Minner, S. (2003). Multiple-supplier inventory models in supply chain management: A review. *International Journal of Production Economics*, 81–82, 265–279.
- Pallme, D., Lambert, B., Miller, C., & Lipinski, M. (2014). A review of public and private intermodal railroad development in the Memphis region. *Research in Transportation Business & Management*, 14, 44–55.
- Reis, V., Meier, J. F., Pace, G., & Palacin, R. (2013). Rail and multi-modal transport. *Research in Transportation Economics*, 41(1), 17–30.
- Rolph, G., & Milner, C. (2015). *Inventory management: Advanced methods for managing inventory within business systems*. London, UK: Kogan Page Limited.
- Sheffi, Y. (2015). What price truck driver shortage (Accessed in August, 2015) <http://supplychainmit.com/2015/08/20/what-price-truck-driver-shortages/>.



- Sheopuri, A., Janakiraman, G., & Seshadri, S. (2010). New policies for the stochastic inventory control problem with two supply sources. *Operations Research*, 58(3), 734–745.
- Silver, E., Pyke, D., & Peterson, R. (1998). *Inventory management and production planning and scheduling* (3rd). Hoboken, NJ: John Wiley & Sons.
- Speranza, M. G., & Ukovich, W. (1994). Minimizing transportation and inventory costs for several products on a single link. *Operations Research*, 42(5), 879–894.
- StadieSeifi, M., Dellaert, N. P., Nuijten, W., Van Woensel, T., & Raoufi, R. (2014). Multimodal freight transportation planning: A literature review. *European Journal of Operational Research*, 233(1), 1–15.
- Swenseth, S., & Godfrey, M. (2002). Incorporating transportation costs into inventory replenishment decisions. *International Journal of Production Economics*, 77(2), 113–130.
- Veeraraghavan, S., & Scheller-Wolf, A. (2008). Now or later: Dual Index Policies for capacitated dual sourcing systems. *Operations Research*, 56(4), 850–864.
- Ventura, J., Valdebenito, V., & Golany, B. (2013). A dynamic inventory model with supplier selection in a serial supply chain structure. *European Journal of Operational Research*, 230(2), 258–271.
- Verma, M., & Verter, V. (2010). A lead-time based approach for planning rail-truck intermodal transportation of dangerous goods. *European Journal of Operational Research*, 202(3), 696–706.
- Whittemore, A. S., & Saunders, S. C. (1977). Optimal inventory under stochastic demand with two supply options. *SIAM Journal on Applied Mathematics*, 32(2), 293–305.