Probability in the Engineering and Informational Sciences, **36**, 2022, 217–236.
doi:10.1017/S0269964820000583

ANALYTICAL RESULTS ON THE SERVICE PERFORMANCE OF STOCHASTIC CLEARING SYSTEMS

Bo Wei

IORA, National University of Singapore, Singapore, Singapore

Sila Çetinkaya

Engineering Management, Information, and Systems, Southern Methodist University, Dallas, TX, USA
E-mail: sila@lyle.smu.edu

Daren b. h. Cline

Department of Statistics, Texas A&M University College Station, College Station, TX, USA

Stochastic clearing theory has wide-spread applications in the context of supply chain and service operations management. Historical application domains include bulk service queues, inventory control, and transportation planning (e.g., vehicle dispatching and shipment consolidation). In this paper, motivated by a fundamental application in shipment consolidation, we revisit the notion of service performance for stochastic clearing system operation. More specifically, our goal is to evaluate and compare service performance of alternative operational policies for clearing decisions, as quantified by a measure of timely service referred to as *Average Order Delay* (*AOD*). All stochastic clearing systems are subject to service delay due to the inherent clearing practice, and AOD can be thought of as a benchmark for evaluating timely service. Although stochastic clearing theory has a long history, the existing literature on the analysis of AOD as a service measure has several limitations. Hence, we extend the previous analysis by proposing a more general method for a generic analytical derivation of AOD for any renewal-type clearing policy, including but not limited to alternative shipment consolidation policies in the previous literature. Our proposed method utilizes a new martingale point of view and lends itself for a generic analytical characterization of AOD, leading to a complete comparative analysis of alternative renewal-type clearing policies. Hence, we also close the gaps in the literature on shipment consolidation via a complete set of analytically provable results regarding AOD which were only illustrated through numerical tests previously.

Keywords: martingales, renewal-type clearing policies, shipment consolidation, stochastic clearing theory, truncated random variables

 0269-9648/19 $25.00

1. INTRODUCTION AND MOTIVATION

"Stochastic clearing systems are characterized by a stochastic input process and an output mechanism that intermittently clears the system, that is, instantaneously restores the net quantity in the system to zero [22]." In the seminal paper by Stidham [22], the system is cleared when the quantity in the system, exceeds a fixed threshold, and the explicit expression of the limiting distribution of the quantity in the system is derived. The concepts of stochastic clearing theory have found wide-spread applications in the context of supply chain and service operations management. Historical application domains include bulk service queues, inventory control, and transportation planning (e.g., vehicle dispatching and shipment consolidation).

Applications to bulk service queues and (s, S) inventory systems are introduced in Ref. [23] where the goal is to compute the optimal operational policy parameters to minimize the average cost including the fixed clearing and variable holding costs. Also, see Ref. [11] for more recent work, addressing a stochastic clearing problem arising in the context of a queueing theoretic transportation application, along with the related literature. In Ref. [14], the stock level process is assumed to be a superposition of a drifted Brownian motion and a compound Poisson process, reflected at zero, and some cost functionals for this system are introduced under several clearing policies.

In the context of vehicle dispatching, the problem is to determine the capacity and instants in time at which vehicles are dispatched leading to a dispatching (clearing) policy [18,19,24,26]. In another seminal paper by Ross [19], the goal is to compute an optimal dispatching policy under a Poisson input process $N(t)$ with rate $\lambda(t)$, where all items are dispatched at time T. An intermediate dispatch time needs to be selected to minimize the total waiting time of all items, and it is demonstrated that the optimal intermediate dispatch time should be the smallest t, such that $N(t) \geq \lambda(t)(T - t)$. Vehicle dispatching with non-stationary Poisson arrivals is studied in Ref. [18] where the optimal dispatching policy is characterized using the principles of stochastic control theory. Three alternative policies for vehicle dispatching are discussed in Ref. [24]: (1) a C-capacity policy (under which a vehicle is dispatched as soon as it is filled to capacity C); (2) a dispatching frequency policy (under which a vehicle is dispatched every T units of time assuming an infinite capacity); and (3) a (T, C) policy (under which a vehicle is dispatched every T units of time or whenever it is filled to capacity C, i.e., whichever happens first). The average cost models are derived under the three policies, and two firm models with cooperative and non-cooperative operation modes are discussed. In Ref. [26], a random vehicle dispatching problem with options to send rented vehicles is considered, and it determines the firm's optimal fleet size to minimize the average cost.

Shipment consolidation is the strategy of combining small size shipments or customer orders, that is, input process realizations, into a larger load with the goals of achieving scale economies and increasing resource, that is, truck, utilization [4] at the time of shipment-release. In the context of shipment consolidation, customer orders represent the stochastic input process. The consolidated loads are released at some specific times that correspond to clearing instances, and the consolidation practice is subject to fixed clearing (shipment-release) and variable holding costs. The problem at hand is to make decisions regarding the timing and quantity of consolidated shipments, leading to a clearing policy. Hence, a shipment consolidation system can be treated as a stochastic clearing system. The consolidation practice is well-justified because a fixed cost is incurred associated with each shipment-release, so it is not economical to dispatch the customer orders immediately, that is, one by one.

In this paper, motivated by a fundamental application in shipment consolidation, we revisit the notion of service performance for stochastic clearing system operation. All stochastic clearing systems are subject to service delay due to the inherent clearing practice. Our goal here is to evaluate and compare such delay under alternative operational policies for clearing decisions, as quantified by a basic measure of timely service (service performance), referred to as *Average Order Delay* (AOD). Although stochastic clearing theory has a long history, the existing literature on the analysis of service performance, in general, and AOD, in particular, has several limitations as we discuss next in Section 2. Hence, we extend the previous literature on stochastic clearing by proposing a more general method for a generic analytical derivation of AOD for any renewal-type clearing policy, including but not limited to alternative shipment consolidation policies investigated earlier.

The remainder of this paper is organized as follows. A detailed summary of previously established results are summarized in Section 2. In Section 3, using a martingale associated with the underlying stochastic input (Poisson order) process, we provide a unified formula to calculate AOD under different renewal-type clearing policies of interest. Section 4 includes the properties of truncated random variables of interest, which are the key to prove the comparative results among alternative policies in terms of AOD. The comparative results under a fixed shipment-release frequency are given in Section 5, and additional results under fixed policy parameters are discussed in Section 6. Finally, the paper concludes in Section 7.

2. RELEVANT LITERATURE AND CONTRIBUTIONS OF THE CURRENT PAPER

Within the existing literature on stochastic clearing, the paper that is most closely related to our work is [9] which is motivated by shipment consolidation applications. More specifically, in Ref. [9], the customer orders (stochastic input) follow a Poisson process $N(t)$ with rate λ and the problem at hand is to make decisions regarding the timing and quantity of consolidated shipments (clearing decisions). The paper places an emphasis on the importance of examining the timely delivery (i.e., service delay) implications of shipment consolidation under practically motivated policies with clearing characteristics. To the best of our knowledge, this is the only previous paper that pursues an *analytical characterization* of service performance in the context of a clearing system, in general, and a shipment consolidation application, in particular. *Simulation studies* addressing the notion of service delay (as it relates to inventory holding and customer waiting) in shipment consolidation can be found in Refs. [12] and [7]. For a detailed account of the analytical models for shipment consolidation, see Ref. [4]. More recent analytical work emphasizing computation of optimal shipment consolidation policies and their applications in vendor-managed inventory systems and last mile distribution can be found in Refs. [2,3,7,8,10,13,15,16,21,25] .

The analysis in Ref. [9] is based on the general theory of renewal/Poisson processes, and it offers some basic results for computing and comparing various service measures of practical interest. However, the approach has limitations in the sense that the analysis does not result in a full-scale analytical comparison of AOD among the policies. Instead, the comparative results are supported by a numerical investigation, and hence, the conclusions are dependent on the parametric settings considered.

Here, we revisit the exact same problem setting in Ref. [9] by treating it as a stochastic clearing system. We are able to extend the previous analysis by proposing a generic analytical approach amenable for a comparative analysis applicable under any renewal-type clearing policy, including but not limited to the alternative shipment consolidation policies identified and studied in the literature [2,3,9,17] . Unlike the previous work in the area, our proposed method utilizes a new martingale point of view and lends itself for a full analytical characterization of AOD, leading to a complete comparative analysis among renewal-type clearing policies. In particular, we are able to close the gaps in the previous literature via a complete set of analytically provable results (some of which were only demonstrated through numerical tests in Ref. [9]) along with a more general approach for examining service performance of stochastic clearing systems.

We consider the three general types of clearing policies common in shipment consolidation practice but applicable in the context of other stochastic clearing applications in queuing systems, inventory control, and transportation: (1) the first one is quantity-based policy (QP) which achieves economies of scale; (2) the second is called time-based policy (TP) which assures timely delivery; and (3) the last one is the hybrid policy (HP) which is aimed at balancing the tradeoff between economies of scale and timely delivery. For the sake of completeness, we recall the definitions of the specific policies examined here and in the previous literature as described exactly in Ref. [9] where the term (1) "order" refers to "input/input process," (2) "consolidated load/shipment" refers to "accumulated input," and (3) "release/shipment-release" refers to "clear/clearing."

- QP is aimed at consolidating a load of q units before releasing a shipment;
- Under TP1, a shipment is released every T units of time, and all orders that arrive between the two shipment-release epochs are consolidated;
- Under TP2, the arrival time of the first order after a shipment-release decision is recorded, and the next shipment is released T time units after the arrival time of the first order;
- Under HP1, the goal is to consolidate a load of size q. However, if the time since the last shipment-release epoch exceeds T, then a shipment-release decision is made anyways;
- Under HP2, the goal is also to consolidate a load of size q; but if the waiting time of the first order after the last shipment-release epoch exceeds T, then a consolidated load is released immediately.

Under TP1 and HP1, there may be empty shipments, which happens when $N(T) = 0$. In order to address this issue, we propose two revised policies, that is, revised TP1 and revised HP1, which do not allow empty shipments.

- Under revised TP1, a shipment is released every T units of time as long as the consolidated load is not 0. However, if there is no order arriving within T units of time since the last shipment, we do not release a shipment, but continue consolidating for another multiple of T units of time.
- Under revised HP1, the goal is to consolidate a load of size q. However, if the time since the last shipment epoch exceeds T and the consolidated load is positive, then the load is dispatched; on the other hand, if the time since the last shipment exceeds T and the consolidated load is zero, we do not release a shipment and the system restarts.

Observe that we have three sets of policies: HP1, QP, TP1, HP2, QP, TP2, and revised HP1, QP, revised TP1. Notice that within each set, HPs would degenerate to QP when the time parameter goes to infinity while degenerate to TPs when the quantity parameter goes to infinity.

As we have noted earlier, all stochastic clearing systems are subject to service delay due to the inherent clearing practice. More specifically for the case of shipment consolidation practices, all of the above policies are implemented at the expense of customer waiting since a prolonged order holding is needed to accumulate a large load [4]. Hence, a key indicator of service performance is indeed AOD, which is the average waiting time of orders before delivery [9]. Under all of the policies described above, the system is cleared once the consolidated load is released, and, hence, the consolidated load under each policy forms a regenerative process. In turn, the policies of interest are renewal-type clearing policies, and the AOD can be obtained by applying the Renewal Reward Theorem, that is,

$$\text{AOD} = \frac{\mathbb{E}[\text{Cumulative waiting per consolidation cycle}]}{\mathbb{E}[\text{Number of orders arriving in a consolidation cycle}]} = \frac{\mathbb{E}[W]}{\lambda\mathbb{E}[C]},$$

where W denotes the sum of the waiting times of the orders within a consolidation cycle and C denotes the consolidation cycle length. We index AOD, W, and C by policy type as needed. Our specific objective here is to provide a comparative analysis of AOD under the alternative policies HP1, QP, TP1, HP2, QP, TP2, and revised HP1, QP, revised TP1.

In Ref. [9], the expectation of cumulative waiting per consolidation cycle under each policy is calculated individually, and the expressions of AOD under HP1 and HP2 are rather complex/involved limiting the ability to offer a complete analytical comparison of the policies at hand. Here, by characterizing each renewal-type clearing policy as a stopping time, we are able to provide a unified method to calculate AOD from a martingale point of view with the aid of optional stopping theorem. Furthermore, we are able to derive the explicit expressions of AOD under HP1 and HP2, which include terms involving first and second moments of truncated Poisson random variables whose refined properties, to the best of our knowledge, have not been examined previously. More specifically, throughout this work, for a non-negative integer-valued random variable X, and a positive integer q, we denote

$$X_q \triangleq \min(X, q)$$

as the truncated random variable of interest. To provide the comparative results in terms of AOD under alternative policies, we develop refined properties of these truncated random variables. The resulting properties, in turn, are essential for the AOD comparisons among alternative policies. Specifically, we obtain the following interesting results of the preservation property, which have their own merits:

- Given a non-negative integer-valued random variable Y with $\mathbb{VAR}[Y] \leq \mathbb{E}[Y] < \infty$, for any positive integer N, we have $\mathbb{VAR}[Y_N] \leq \mathbb{E}[Y_N]$ (see Lemma 4.2). This, in turn, implies that the relationship of $\mathbb{VAR}[Y] \leq \mathbb{E}[Y]$ is preserved after Y is truncated.
- Let X, Y be two non-negative integer-valued random variables, and X is stochastically larger than Y. If $\mathbb{E}[X_q] \leq \mathbb{E}[Y_{q+1}]$, where q is a positive integer, then $\mathbb{E}[X_q^2] \leq \mathbb{E}[Y_{q+1}^2]$ (see Lemma 4.3). This property implies that the relationship between first moments of two random variables is preserved in second moment setting.
- Suppose $X \sim \text{Poisson}(\lambda)$ and N is a positive integer, then $\mathbb{E}[X_N^2]/\mathbb{E}[X_N]$ is increasing with respect to λ (see Lemma 4.4). We know $\mathbb{E}[X^2]/\mathbb{E}[X] = \lambda + 1$, which is increasing with respect to λ. That is, after X is truncated, the property is still preserved.

Based on the above-refined properties, we provide the following comparative, insightful results among alternative policies in terms of AOD: For a fixed expected consolidation cycle length:

- QP outperforms all renewal-type clearing policies in terms of AOD (see Theorem 5.1).

https://doi.org/10.1017/S0269964820000583 Published online by Cambridge University Press

- The general class of HPs performs better than the general class of counterpart TPs in terms of AOD (see Theorems 5.3 and A.2. Three sets: HP1, TP1, HP2, TP2, and revised HP1, revised TP1).
- The HP1 with larger quantity parameter would achieve larger AOD than the HP1 with smaller quantity parameter, and the same property holds for HP2 and revised HP1 (see Theorems 5.5 and A.2).

Furthermore, with fixed parameters, we show that

- The general class of HPs has less AOD than the general classes of counterpart QPs and TPs (see Theorems 6.1 and 6.2. Three sets: HP1, QP, TP1, HP2, QP, TP2, and revised HP1, QP, revised TP1).
- HP1 has the same AOD as revised HP1, and less AOD than HP2 (see Theorem 6.3 and Appendix).

Next, we proceed with the details of the new analytical approach for examining AOD which has provided the foundation of the insightful results summarized above.

3. AVERAGE ORDER DELAY

In this section, based on a martingale associated with the Poisson input process, we provide a unified method to calculate the AOD for any renewal-type clearing policy. The following lemma reveals this martingale.

LEMMA 3. 1: *Let $N(t)$ be a Poisson process with rate λ, and define $W(t) \triangleq \int_0^t N(u)\,du$. Then,*

$$\left\{ W(t) - \frac{1}{2\lambda} N^2(t) + \frac{1}{2\lambda} N(t) \right\}_{t \geq 0}$$

is a martingale with respect to the natural filtration $\{\mathcal{G}_t\}$, which is the σgenerated by the family of random variables $\{N(s), s \in [0,t]\}$.

PROOF: Since the Poisson process has stationary independent increment, for $s < t$, we have,

$$\begin{aligned} \mathbb{E}\left[\int_0^t N(u)\,du \,|\, \mathcal{G}_s\right] &= \int_0^s N(u)\,du + \mathbb{E}\left[\int_s^t N(u)\,du \mid \mathcal{G}_s\right] \\ &= \int_0^s N(u)\,du + (t-s)N(s) + \mathbb{E}\left[\int_0^{t-s} N(u)\,du\right] \\ &= \int_0^s N(u)\,du + (t-s)N(s) + \frac{1}{2}\lambda(t-s)^2, \\ \frac{1}{2\lambda}\mathbb{E}[N^2(t) \mid \mathcal{G}_s] &= \frac{1}{2\lambda}(N^2(s) + 2\lambda(t-s)N(s) + \lambda(t-s) + \lambda^2(t-s)^2), \end{aligned}$$

and

$$\frac{1}{2\lambda}\mathbb{E}[N(t) \mid \mathcal{G}_s] = \frac{1}{2\lambda}(N(s) + \lambda(t-s)).$$

Therefore,

$$\mathbb{E}\left[\int_0^t N(u)\,du - \frac{1}{2\lambda}N^2(t) + \frac{1}{2\lambda}N(t) \mid \mathcal{G}_s\right] = \int_0^s N(u)\,du - \frac{1}{2\lambda}N^2(s) + \frac{1}{2\lambda}N(s),$$

which shows that $W(t) - \frac{1}{2\lambda}N^2(t) + \frac{1}{2\lambda}N(t)$ is a martingale. ■

The shipment-release time is always a stopping time with respect to the stochastic customer order process. For a consolidation policy with shipment-release time τ, the cumulative waiting time within one consolidation cycle is

$$W(\tau) = \int_0^{\tau} N(u)\, du.$$

Throughout this work, for two real numbers a and b denote $a \wedge b \triangleq \min(a, b)$. From Lemma 3.1 and the martingale stopping theorem [20] Thm. 6.2.2, we have that for any stopping time τ and any fixed $t > 0$,

$$\mathbb{E}[W(\tau \wedge t)] = \frac{1}{2\lambda}\mathbb{E}[N^2(\tau \wedge t) - N(\tau \wedge t)]. \tag{1}$$

Also, applying again the martingale stopping theorem on another martingale $\{N(t) - \lambda t\}_{t\geq 0}$, we have that for any stopping time τ and any fixed $t > 0$,

$$\mathbb{E}[N(\tau \wedge t)] = \lambda \mathbb{E}[\tau \wedge t]. \tag{2}$$

We consider a renewal-type clearing policy with finite cycle mean, that is, τ is of finite mean. From the monotone convergence theorem [1] Thm. 2.3.4,

$$\lim_{t\to\infty} \mathbb{E}[W(\tau \wedge t)] = \mathbb{E}[W(\tau)]$$

and

$$\lim_{t\to\infty} \mathbb{E}[N(\tau \wedge t)(N(\tau \wedge t) - 1)] = \mathbb{E}[N(\tau)(N(\tau) - 1)].$$

Noticing Eq. (1), we obtain that the expectation of cumulative waiting time within one consolidation cycle is

$$\mathbb{E}[W(\tau)] = \frac{1}{2\lambda}\mathbb{E}[N^2(\tau) - N(\tau)]. \tag{3}$$

Similarly, from Eq. (2), we have the expectation of consolidation cycle is

$$\mathbb{E}[\tau] = \frac{1}{\lambda}\mathbb{E}[N(\tau)]. \tag{4}$$

From the above discussion, we can deduce the AOD for any renewal-type shipment consolidation policy. Now, we calculate the AOD under different shipment consolidation policies of interest here as follows. For two random variables X and Y, we use $X \sim Y$ to denote that they follow the same distribution.

1. QP with integer parameter $q \geq 1$: $\tau = \tau_q$, the time until the q-th order, q is a positive integer; $N(\tau_q) = q$. So

$$\begin{aligned}
\mathbb{E}[W_{\mathrm{QP}}] &= \mathbb{E}[W(\tau_q)] = \frac{1}{2\lambda}q(q-1),\\
\mathbb{E}[C_{\mathrm{QP}}] &= \mathbb{E}[\tau_q] = \frac{q}{\lambda}.
\end{aligned}$$

2. TP1 with parameter T: $\tau = T$, a constant; $N(T) \sim \mathrm{Poisson}(\lambda T)$. So

$$\begin{aligned}
\mathbb{E}[W_{\mathrm{TP1}}] &= \mathbb{E}[W(T)] = \frac{1}{2\lambda}\mathbb{E}[N^2(T) - N(T)] = \frac{1}{2}\lambda T^2,\\
\mathbb{E}[C_{\mathrm{TP1}}] &= T.
\end{aligned}$$

Table 1. Summary of the expressions of AOD, where $Y \sim \text{Poisson}(\lambda T)$

AOD_τ	$=$	$\dfrac{\mathbb{E}[W(\tau)]}{\lambda \mathbb{E}[C_\tau]} = \dfrac{\mathbb{E}[N^2(\tau) - N(\tau)]/(2\lambda)}{\mathbb{E}[N(\tau)]}$
AOD_{QP}	$=$	$\dfrac{\mathbb{E}[W_{\text{QP}}]}{\lambda \mathbb{E}[C_{\text{QP}}]} = \dfrac{(q-1)q/2\lambda}{q} = \dfrac{q-1}{2\lambda}$
AOD_{TP1}	$=$	$\dfrac{\mathbb{E}[W_{\text{TP1}}]}{\lambda \mathbb{E}[C_{\text{TP1}}]} = \dfrac{\lambda T^2/2}{\lambda T} = \dfrac{1}{2}T$
AOD_{TP2}	$=$	$\dfrac{\mathbb{E}[W_{\text{TP2}}]}{\lambda \mathbb{E}[C_{\text{TP2}}]} = \dfrac{T + \lambda T^2/2}{1 + \lambda T}$
AOD_{HP1}	$=$	$\dfrac{\mathbb{E}[W_{\text{HP1}}]}{\lambda \mathbb{E}[C_{\text{HP1}}]} = \dfrac{\mathbb{E}[Y_q(Y_q - 1)]/(2\lambda)}{\mathbb{E}[Y_q]}$
AOD_{HP2}	$=$	$\dfrac{\mathbb{E}[W_{\text{HP2}}]}{\lambda \mathbb{E}[C_{\text{HP2}}]} = \dfrac{\mathbb{E}[Y_{q-1}(Y_{q-1} + 1)]/(2\lambda)}{\mathbb{E}[Y_{q-1} + 1]}$

3. TP2 with parameter T: $\tau = \tau_1 + T$; $N(\tau_1 + T) \sim 1 + N(T)$. So

$$\mathbb{E}[W_{\text{TP2}}] = \mathbb{E}[W(\tau_1 + T)] = \frac{1}{2\lambda}\mathbb{E}[N^2(T) + N(T)] = \frac{1}{2}\lambda T^2 + T,$$
$$\mathbb{E}[C_{\text{TP2}}] = \frac{1}{\lambda} + T.$$

4. HP1 with parameters q and T: $\tau = \tau_q \wedge T$; Let $Y \sim \text{Poisson}(\lambda T)$, $N(\tau_q \wedge T) \sim N(T) \wedge q \sim Y_q$, which is a truncated Poisson random variable. So

$$\mathbb{E}[W_{\text{HP1}}] = \mathbb{E}\left[W\left(\tau_q \wedge T\right)\right] = \frac{1}{2\lambda}\mathbb{E}[Y_q(Y_q - 1)],$$
$$\mathbb{E}[C_{\text{HP1}}] = \frac{1}{\lambda}\mathbb{E}[Y_q].$$

5. HP2 with parameters q and T: $\tau = \tau_q \wedge (\tau_1 + T)$; Let $Y \sim \text{Poisson}(\lambda T)$, $N(\tau_q \wedge (\tau_1 + T)) \sim (1 + N(T)) \wedge q \sim Y_{q-1} + 1$. So

$$\mathbb{E}[W_{\text{HP2}}] = \mathbb{E}\left[W\left(\tau_q \wedge (\tau_1 + T)\right)\right] = \frac{1}{2\lambda}\mathbb{E}[Y_{q-1}(Y_{q-1} + 1)],$$
$$\mathbb{E}[C_{\text{HP2}}] = \frac{1}{\lambda}\mathbb{E}[Y_{q-1} + 1].$$

In Table 1, we summarize the AOD for different shipment consolidation policies. We would notice that the expressions of AOD_{HP1} and AOD_{HP2} involve the truncated Poisson random variables, which are much simplified than the expressions in Ref. [9]. Note that under TP1 and HP1, the consolidation cycle clock starts over, even if no order arrives within the previous cycle. We consider the correspondingly revised policies in the Appendix, which do not allow empty dispatches.

4. PROPERTIES OF TRUNCATED RANDOM VARIABLES

In this section, we investigate the properties of truncated random variables, which are connected to the comparison of different shipment consolidation policies of interest here in terms of AOD.

LEMMA 4. 1: *Given an integer-valued random variable Y, and a positive integer M, we have*

$$\mathbb{VAR}[Y] - \mathbb{VAR}[Y_M] = \mathbb{VAR}[Y - Y_M] + 2(M - \mathbb{E}[Y_M])(\mathbb{E}[Y] - \mathbb{E}[Y_M]).$$

In particular, $\mathbb{VAR}[Y_M] \leq \mathbb{VAR}[Y]$.

PROOF: Since Y is integer-valued and M is an integer, we have

$$Y - Y_M = (Y - M)1_{Y \geq M+1}$$

and

$$\mathbb{E}[Y_M(Y - Y_M)] = M\mathbb{E}[(Y - M)1_{Y \geq M+1}] = M(\mathbb{E}[Y] - \mathbb{E}[Y_M]).$$

Therefore,

$$\begin{aligned}\mathbb{COV}(Y_M, Y - Y_M) &= \mathbb{E}[Y_M(Y - Y_M)] - \mathbb{E}[Y_M]\mathbb{E}[Y - Y_M] \\ &= (M - \mathbb{E}[Y_M])(\mathbb{E}[Y] - \mathbb{E}[Y_M]).\end{aligned}$$

We have

$$\begin{aligned}\mathbb{VAR}[Y] &= \mathbb{VAR}[Y_M + (Y - Y_M)] \\ &= \mathbb{VAR}[Y_M] + \mathbb{VAR}[Y - Y_M] + 2\mathbb{COV}(Y_M, Y - Y_M) \\ &= \mathbb{VAR}[Y_M] + \mathbb{VAR}[Y - Y_M] + 2(M - \mathbb{E}[Y_M])(\mathbb{E}[Y] - \mathbb{E}[Y_M]).\end{aligned}$$

■

Based on Lemma 4.1, we establish the following result, which will be useful in the comparison between the general class of HPs and the general class of counterpart TPs in terms of AOD.

LEMMA 4. 2: *Given a non-negative integer-valued random variable Y with $\mathbb{VAR}[Y] \leq \mathbb{E}[Y] < \infty$, then, for any positive integer N, we have $\mathbb{VAR}[Y_N] \leq \mathbb{E}[Y_N]$. In particular, if Y is a Poisson random variable, $\mathbb{VAR}[Y_N] < \mathbb{E}[Y_N]$.*

PROOF: Define $f(N) \triangleq \mathbb{VAR}[Y_N] - \mathbb{E}[Y_N]$, for $N \geq 1$. Noticing

$$Y_N = Y_{N+1} \wedge N, \quad Y_{N+1} - Y_N = 1_{Y \geq N+1},$$

and applying Lemma 4.1, we have

$$\begin{aligned}f(N+1) - f(N) &= (\mathbb{VAR}[Y_{N+1}] - \mathbb{VAR}[Y_N]) - (\mathbb{E}[Y_{N+1}] - \mathbb{E}[Y_N]) \\ &= \mathbb{VAR}[Y_{N+1} - Y_N] + 2(N - \mathbb{E}[Y_N])(\mathbb{E}[Y_{N+1}] - \mathbb{E}[Y_N]) - (\mathbb{E}[Y_{N+1}] - \mathbb{E}[Y_N]) \\ &= \mathbb{VAR}[Y_{N+1} - Y_N] + (2N - 2\mathbb{E}[Y_N] - 1)(\mathbb{E}[Y_{N+1}] - \mathbb{E}[Y_N]) \\ &= \mathbb{P}(Y \geq N+1)\mathbb{P}(Y \leq N) + (2N - 2\mathbb{E}[Y_N] - 1)\mathbb{P}(Y \geq N+1) \\ &= (2\mathbb{E}[\max(N - Y, 0)] - \mathbb{P}(Y \geq N+1))\mathbb{P}(Y \geq N+1).\end{aligned}$$

Obviously, $2\mathbb{E}[\max(N - Y, 0)] - \mathbb{P}(Y \geq N+1)$ is increasing in N and $\mathbb{P}(Y \geq N+1) \geq 0$. This, in turn, implies that $f(N+1) - f(N)$ changes sign at most once with respect to N: either from negative to positive or always positive. In the first case, $f(N)$ is first non-increasing and then non-decreasing; in the second case, $f(N)$ is non-decreasing. Noting that

$$\lim_{N\to\infty} f(N) = \lim_{N\to\infty} (\mathbb{VAR}[Y_N] - \mathbb{E}[Y_N]) = \mathbb{VAR}[Y] - \mathbb{E}[Y] \leq 0$$

and

$$f(1) = \mathbb{VAR}[1_{Y\geq 1}] - \mathbb{E}[1_{Y\geq 1}] = \mathbb{P}(Y \geq 1)(1 - \mathbb{P}(Y \geq 1)) - \mathbb{P}(Y \geq 1) \leq 0,$$

we conclude $f(N) \leq 0$ for all N, that is, $\mathbb{VAR}[Y_N] \leq \mathbb{E}[Y_N]$.

If Y is a Poisson random variable, by the same reasoning as above, $f(N)$ is either first decreasing and then increasing or always increasing. Moreover, $f(1) < 0$ and $\lim_{N\to\infty} f(N) = 0$. Thus, $f(N) = \mathbb{VAR}[Y_N] - \mathbb{E}[Y_N] < 0$ for any positive integer N. ■

Next, we provide a result which would be essential in comparing the same type HP with different parameters in terms of AOD, with a fixed expected consolidation cycle length $\mathbb{E}[C]$.

Lemma 4. 3: *Assume X, Y are two integer-valued random variables, and X is stochastically larger than Y. If $\mathbb{E}[X_q] \leq \mathbb{E}[Y_{q+1}]$, where q is a positive integer, then $\mathbb{E}[X_q^2] \leq \mathbb{E}[Y_{q+1}^2]$.*

Proof: From

$$Y_{q+1}^2 - Y_q^2 = (2q+1)(Y_{q+1} - Y_q),$$

we have

$$\mathbb{E}[Y_{q+1}^2] - \mathbb{E}[Y_q^2] = (2q+1)(\mathbb{E}[Y_{q+1}] - \mathbb{E}[Y_q]) \geq (2q+1)(\mathbb{E}[X_q] - \mathbb{E}[Y_q]).$$

Therefore,

$$\mathbb{E}[Y_{q+1}^2] - \mathbb{E}[X_q^2] \geq \mathbb{E}[Y_q^2] - \mathbb{E}[X_q^2] + (2q+1)(\mathbb{E}[X_q] - \mathbb{E}[Y_q]). \tag{5}$$

Since X is stochastically larger than Y, it follows immediately that X_q is also stochastically larger than Y_q. From Ref. [20] Prop. 9.2.2, we always can find two random variables X' and Y', such that X' has the same probability distribution as X_q, Y' has the same probability distribution as Y_q, and $X' \geq Y'$ almost surely. From Eq. (5), we have

$$\begin{aligned}\mathbb{E}[Y_{q+1}^2] - \mathbb{E}[X_q^2] &\geq \mathbb{E}[Y'^2] - \mathbb{E}[X'^2] + (2q+1)(\mathbb{E}[X'] - \mathbb{E}[Y'])\\ &= \mathbb{E}[(X' - Y')(2q + 1 - X' - Y')] \geq 0,\end{aligned}$$

where the last inequality follows from $X' \leq q$, $Y' \leq q$, and $X' \geq Y'$ almost surely. ■

The following lemma characterizes how the ratio between the second moment and the first moment of a truncated Poisson random variable changes with respect to the Poisson rate parameter, which will be used when we compare HP1 and HP2 under fixed policy parameters, in terms of AOD.

Lemma 4. 4: *Suppose $X \sim Poisson(\lambda)$ and N is a positive integer, then $\mathbb{E}[X_N^2]/\mathbb{E}[X_N]$ is increasing with respect to λ.*

Proof: Let $Y \sim \text{Poisson}(\lambda_1)$, $Z \sim \text{Poisson}(\lambda_2)$, where $\lambda_1 < \lambda_2$. When $k < m < N$,

$$\mathbb{P}(Z_N = m)\mathbb{P}(Y_N = k) - \mathbb{P}(Y_N = m)\mathbb{P}(Z_N = k) = \frac{e^{-\lambda_1-\lambda_2}}{m!k!}(\lambda_2^m \lambda_1^k - \lambda_1^m \lambda_2^k) > 0, \quad \textbf{(6)}$$

and when $k < N$,

$$\begin{aligned}
&\mathbb{P}(Z_N = N)\mathbb{P}(Y_N = k) - \mathbb{P}(Y_N = N)\mathbb{P}(Z_N = k) \\
&\quad = \sum_{j \geq N} (\mathbb{P}(Z = j)\mathbb{P}(Y = k) - \mathbb{P}(Y = j)\mathbb{P}(Z = k)) \\
&\quad = \sum_{j \geq N} \frac{e^{-\lambda_1-\lambda_2}}{j!k!}(\lambda_2^j \lambda_1^k - \lambda_1^j \lambda_2^k) > 0. \qquad \textbf{(7)}
\end{aligned}$$

Note that for any non-negative integer-valued random variable W, we have

$$\mathbb{E}[W^2] = \sum_{m=1}^{\infty} m^2 \mathbb{P}(W = m) = \sum_{m=1}^{\infty} \sum_{j=1}^{m} m\mathbb{P}(W = m) = \sum_{j=1}^{\infty} \sum_{m=j}^{\infty} m\mathbb{P}(W = m).$$

Therefore, we obtain

$$\begin{aligned}
&\mathbb{E}[Z_N^2]\mathbb{E}[Y_N] - \mathbb{E}[Y_N^2]\mathbb{E}[Z_N] \\
&\quad = \sum_{j=1}^{N} \sum_{m=j}^{N} m\mathbb{P}(Z_N = m) \sum_{k=1}^{N} k\mathbb{P}(Y_N = k) - \sum_{j=1}^{N} \sum_{m=j}^{N} m\mathbb{P}(Y_N = m) \sum_{k=1}^{N} k\mathbb{P}(Z_N = k) \\
&\quad = \sum_{j=1}^{N} \sum_{m=j}^{N} \sum_{k=1}^{j-1} mk[\mathbb{P}(Z_N = m)\mathbb{P}(Y_N = k) - \mathbb{P}(Y_N = m)\mathbb{P}(Z_N = k)] > 0,
\end{aligned}$$

where the second equality is derived from

$$\sum_{m=j}^{N} \sum_{k=j}^{N} mk[\mathbb{P}(Z_N = m)\mathbb{P}(Y_N = k) - \mathbb{P}(Y_N = m)\mathbb{P}(Z_N = k)] = 0,$$

and the last inequality holds by Eqs. (6) and (7).

Therefore,

$$\frac{\mathbb{E}[Z_N^2]}{\mathbb{E}[Z_N]} - \frac{\mathbb{E}[Y_N^2]}{\mathbb{E}[Y_N]} = \frac{\mathbb{E}[Z_N^2]\mathbb{E}[Y_N] - \mathbb{E}[Y_N^2]\mathbb{E}[Z_N]}{\mathbb{E}[Y_N]\mathbb{E}[Z_N]} > 0,$$

which implies that $\mathbb{E}[X_N^2]/\mathbb{E}[X_N]$ is increasing with respect to λ. ■

5. COMPARISON OF AOD UNDER A FIXED EXPECTED CYCLE LENGTH, $\mathbb{E}[C]$

In **(O10)** of Ref. [9], through a numerical study, there is an observation that for a given $\mathbb{E}[C]$, the QP performs the best and TPs perform the worst in terms of AOD. In this section, we analytically show that for a given $\mathbb{E}[C]$, QP provides superior service compared with any other shipment consolidation policy in terms of AOD, not limited to HPs and TPs. Furthermore, we rigorously compare HPs and TPs in terms of AOD, for a given $\mathbb{E}[C]$. In addition, for a given $\mathbb{E}[C]$, we provide the comparative result between the same type HP

https://doi.org/10.1017/S0269964820000583 Published online by Cambridge University Press

policies with different parameters, in terms of AOD. The readers who are interested in the managerial motivation of this comparison of AOD are referred to Ref. [9].

THEOREM 5. 1: *For a fixed expected consolidation cycle length* $\mathbb{E}[C] = q/\lambda$ *with* q *as a positive integer, QP with parameter* q *outperforms all the other renewal-type clearing policies in terms of AOD.*

PROOF: From Table 1, we know AOD of a shipment consolidation policy with shipment-release time τ is

$$\mathrm{AOD}_\tau = \frac{\mathbb{E}[N^2(\tau) - N(\tau)]/(2\lambda)}{\mathbb{E}[N(\tau)]}.$$

From $\mathbb{E}[\tau] = \mathbb{E}[C] = q/\lambda$ and Eq. (4), we have that $\mathbb{E}[N(\tau)] = q$ is fixed. Furthermore, $\mathbb{E}[N^2(\tau)] \geq \mathbb{E}^2[N(\tau)] = q^2$, where the equality holds if and only if the random variable $N(\tau)$ is the constant q. Thus,

$$\mathrm{AOD}_\tau = \frac{1}{2\lambda}\left(\frac{\mathbb{E}[N^2(\tau)]}{\mathbb{E}[N(\tau)]} - 1\right) \geq \frac{1}{2\lambda}(\mathbb{E}[N(\tau)] - 1) = \frac{q-1}{2\lambda},$$

where the inequality is an equality if and only if $N(\tau)$ is the constant q. This, in turn, implies QP with parameter q achieves the least AOD with a fixed expected consolidation cycle length. ■

REMARK 5. 2: *If there is a consolidation policy with shipment-release time* τ*, which has the same expected cycle length as a quantity-based policy with positive integer parameter* q*, that is* $\mathbb{E}[\tau] = q/\lambda$*, the average cost associated with this policy is*

$$\frac{A_D + C_D\mathbb{E}[N(\tau)] + \omega\mathbb{E}[W(\tau)]}{\mathbb{E}[\tau]},$$

where A_D *is the fixed cost for each shipment-release,* C_D *is the unit transportation cost, and* ω *is the waiting cost per unit per unit time. With fixed* $\mathbb{E}[\tau]$, $\mathbb{E}[N(\tau)]$ *is also fixed. From Theorem 5.1, we conclude that the corresponding quantity-based policy achieves less average cost than this policy with shipment-release time* τ.

One disadvantage of QP is that it has no upper bound on the cycle length, in contrast, HP is of more practical importance since by definition it has an upper bound on the cycle length. In the next result, we show that the general class of HPs outperforms the general class of counterpart TPs in terms of AOD, under a fixed expected consolidation cycle length.

THEOREM 5. 3: *For a fixed expected consolidation cycle length* $\mathbb{E}[C]$*, HP1 performs better than TP1, and HP2 performs better than TP2 in terms of AOD.*

PROOF: We consider a fixed $\mathbb{E}[C]$ and use the following notation for the corresponding policy parameters under this $\mathbb{E}[C]$ value: TP1 with parameter T_1, TP2 with parameter T_2, HP1 with parameters q_{H1} and T_{H1}, and HP2 with parameters q_{H2} and T_{H2}. Recalling the

$\mathbb{E}[C]$ expressions in Table 1, we note that, by assumption,

$$\frac{1}{\lambda}\mathbb{E}[X_{q_{\mathrm{H}1}}] = T_1 \tag{8}$$

and

$$\frac{1}{\lambda}\mathbb{E}[1 + Z_{q_{\mathrm{H}2}-1}] = \frac{1}{\lambda} + T_2, \tag{9}$$

where $X \sim \text{Poisson}(\lambda T_{\mathrm{H}1})$, $Z \sim \text{Poisson}(\lambda T_{\mathrm{H}2})$. Next, recalling the results in Table 1 and the assumption of fixed $\mathbb{E}[C]$ values for all the policies of interest, we need to show that

$$\mathbb{E}[X_{q_{\mathrm{H}1}}(X_{q_{\mathrm{H}1}} - 1)] < \lambda^2 T_1^2 \tag{10}$$

and

$$\mathbb{E}[Z_{q_{\mathrm{H}2}-1}(Z_{q_{\mathrm{H}2}-1} + 1)] < 2\lambda T_2 + \lambda^2 T_2^2. \tag{11}$$

In fact, by recalling Eqs. (8) and (9), we have

$$\mathbb{E}[X_{q_{\mathrm{H}1}}(X_{q_{\mathrm{H}1}} - 1)] = \mathbb{VAR}[X_{q_{\mathrm{H}1}}] + \mathbb{E}^2[X_{q_{\mathrm{H}1}}] - \mathbb{E}[X_{q_{\mathrm{H}1}}] < \mathbb{E}^2[X_{q_{\mathrm{H}1}}] = \lambda^2 T_1^2$$

and

$$\begin{aligned}\mathbb{E}[Z_{q_{\mathrm{H}2}-1}(Z_{q_{\mathrm{H}2}-1} + 1)] &= \mathbb{VAR}[Z_{q_{\mathrm{H}2}-1}] + \mathbb{E}^2[Z_{q_{\mathrm{H}2}-1}] + \mathbb{E}[Z_{q_{\mathrm{H}2}-1}] \\ &< 2\mathbb{E}[Z_{q_{\mathrm{H}2}-1}] + \mathbb{E}^2[Z_{q_{\mathrm{H}2}-1}] \\ &= 2\lambda T_2 + \lambda^2 T_2^2,\end{aligned}$$

where the inequalities are derived from Lemma 4.2. ■

REMARK 5. 4: *Following the same argument as in Remark 5.2, in terms of average cost criterion, the general class of HPs also outperforms the general class of counterpart TPs.*

From Lemma 4.3, we deduce a stronger result as follows, which allows us to compare two HP policies of the same type under a fixed $\mathbb{E}[C]$.

THEOREM 5. 5: *For a fixed expected consolidation cycle length $\mathbb{E}[C]$, the HP1 with a larger quantity parameter achieves larger AOD than the HP1 with a smaller quantity parameter, and the similar result holds for HP2.*

PROOF: We consider a fixed $\mathbb{E}[C]$ and use the following notation for the corresponding policy parameters under this $\mathbb{E}[C]$ value: the first HP1 with parameters q_{H} and T_{H}, the second HP1 with parameters $q_{\mathrm{H}} + 1$ and T'_{H}. Recalling the $\mathbb{E}[C]$ expressions in Table 1, we note that, by assumption,

$$\mathbb{E}[X_{q_{\mathrm{H}}}] = \mathbb{E}[Y_{q_{\mathrm{H}}+1}], \tag{12}$$

where $X \sim \text{Poisson}\,(\lambda T_{\mathrm{H}})$, $Y \sim \text{Poisson}\,(\lambda T'_{\mathrm{H}})$. Clearly, $T_{\mathrm{H}} > T'_{\mathrm{H}}$.

Next, recalling the results in Table 1 and the assumption of fixed $\mathbb{E}[C]$ values for all the policies of interest, we need to show that

$$\mathbb{E}[X_{q_{\mathrm{H}}}(X_{q_{\mathrm{H}}} - 1)] < \mathbb{E}[Y_{q_{\mathrm{H}}+1}(Y_{q_{\mathrm{H}}+1} - 1)]. \tag{13}$$

From Lemma 4.3 and recalling Eq. (12), we have

$$\mathbb{E}[X_{q_{\mathrm{H}}}^2] \leq \mathbb{E}[Y_{q_{\mathrm{H}}+1}^2],$$

so that inequality (13) is verified.

The same procedure can be applied to prove the similar result between two HP2 policies. ■

https://doi.org/10.1017/S0269964820000583 Published online by Cambridge University Press

6. COMPARISON OF AOD UNDER FIXED PARAMETERS Q AND/OR T

In Ref. [9], the authors analytically show that under fixed parameters, the general class of HPs outperforms the general classes of counterpart QP and TPs in terms of AOD. In this section, we provide a simpler proof of the above statement based on the rewritten expressions in Table 1. Furthermore, we show under fixed parameters, HP1 outperforms HP2 in terms of AOD. The readers who are interested in the managerial implications of this comparison of AOD are referred to Ref. [9]. Note that with fixed parameter $q = 1$, the AOD of QP, HP1 and HP2 are all 0, and thus, we only consider the cases where $q \geq 2$.

THEOREM 6. 1: *With fixed parameters $q \geq 2$ and T, HP1 performs better than QP and TP1 in terms of AOD.*

PROOF: On the one hand, we need to show HP1 performs better than QP in terms of AOD with the same parameters q, T, from Table 1, that is,

In fact, $(q-1)\mathbb{E}[Y_q] - \mathbb{E}[Y_q(Y_q-1)] = q\mathbb{E}[Y_q] - \mathbb{E}[Y_q^2] = \mathbb{E}[(q-Y_q)Y_q] > 0$.

On the other hand, we need to show HP1 performs better than TP1 in terms of AOD with the same parameters q, T, from Table 1, that is,

$$\frac{\mathbb{E}[Y_q(Y_q-1)]}{\mathbb{E}[Y_q]} < \lambda T.$$

In fact, from Lemma 4.2, we have $\mathbb{VAR}[Y_q] < \mathbb{E}[Y_q]$, which can written as

$$\mathbb{E}[Y_q(Y_q-1)] < \mathbb{E}^2[Y_q].$$

Furthermore, we have $\mathbb{E}[Y_q] < \lambda T$ since $Y \sim \text{Poisson}(\lambda T)$. Thus, we arrive at the desired inequality. ■

THEOREM 6. 2: *With fixed parameters $q \geq 2$ and T, HP2 performs better than QP and TP2 in terms of AOD.*

PROOF: On the one hand, we need to show HP2 performs better than QP in terms of AOD with the same parameters q, T, from Table 1, that is,

In fact, $(q-1)\mathbb{E}[1+Y_{q-1}] - \mathbb{E}[Y_{q-1}(Y_{q-1}+1)] = \mathbb{E}[(q-1-Y_{q-1})(Y_{q-1}+1)] > 0$.

On the other hand, we need to show HP2 performs better than TP2 in terms of AOD with the same parameters q, T, from Table 1, that is,

$$\frac{\mathbb{E}[Y_{q-1}(Y_{q-1}+1)]}{\mathbb{E}[1+Y_{q-1}]} < \frac{2\lambda T + \lambda^2 T^2}{1+\lambda T}.$$

In fact, from Lemma 4.2, we have $\mathbb{VAR}[Y_{q-1}] < \mathbb{E}[Y_{q-1}]$, which can be written as

$$\mathbb{E}[Y_{q-1}(Y_{q-1}+1)] < \mathbb{E}^2[Y_{q-1}+1] - 1.$$

Furthermore, due to $\mathbb{E}[Y_{q-1}] < \lambda T$, we have

$$\mathbb{E}[Y_{q-1}+1] - \frac{1}{\mathbb{E}[Y_{q-1}+1]} < \lambda T + 1 - \frac{1}{\lambda T + 1}.$$

Thus, we arrive at the desired inequality. ■

The following result allows us to compare HP1 and HP2 with fixed parameters q and T, which relies on Lemma 4.4.

THEOREM 6. 3: *With fixed parameters $q \geq 2$ and T, HP1 performs better than HP2 in terms of AOD.*

PROOF: From Table 1, we need to show

$$\frac{\mathbb{E}[Y_q(Y_q-1)]}{\mathbb{E}[Y_q]} < \frac{\mathbb{E}[Y_{q-1}(Y_{q-1}+1)]}{\mathbb{E}[1+Y_{q-1}]}.$$

After simplification, it suffices to show

$$\frac{\mathbb{E}[Y_q^2]}{\mathbb{E}[Y_q]} < \frac{\mathbb{E}[(Y_{q-1}+1)^2]}{\mathbb{E}[Y_{q-1}+1]}. \tag{14}$$

Note for $X \sim \text{Poisson}(\mu)$, we have

$$\frac{d}{d\mu}\mathbb{E}[g(X)] = \mathbb{E}[g(X+1)] - \mathbb{E}[g(X)],$$

for any appropriate function $g(x)$. Let $\mu = \lambda T$, $g_1(x) = (x \wedge q)^2$, and $g_2(x) = x \wedge q$, for $x \geq 0$, we have

$$\begin{aligned}\frac{d}{d\mu}\mathbb{E}[Y_q^2] = \frac{d}{d\mu}\mathbb{E}[g_1(Y)] &= \mathbb{E}[g_1(Y+1)] - \mathbb{E}[g_1(Y)] \\ &= \mathbb{E}[(Y_{q-1}+1)^2] - \mathbb{E}[Y_q^2],\end{aligned}$$

and

$$\begin{aligned}\frac{d}{d\mu}\mathbb{E}[Y_q] = \frac{d}{d\mu}\mathbb{E}[g_2(Y)] &= \mathbb{E}[g_2(Y+1)] - \mathbb{E}[g_2(Y)] \\ &= \mathbb{E}[Y_{q-1}+1] - \mathbb{E}[Y_q].\end{aligned}$$

Hence,

$$\begin{aligned}\frac{d}{d\mu}\frac{\mathbb{E}[Y_q^2]}{\mathbb{E}[Y_q]} &= \frac{(\mathbb{E}[(Y_{q-1}+1)^2] - \mathbb{E}[Y_q^2])\mathbb{E}[Y_q] - \mathbb{E}[Y_q^2](\mathbb{E}[Y_{q-1}+1] - \mathbb{E}[Y_q])}{\mathbb{E}^2[Y_q]} \\ &= \frac{\mathbb{E}[(Y_{q-1}+1)^2]\mathbb{E}[Y_q] - \mathbb{E}[Y_q^2]\mathbb{E}[Y_{q-1}+1]}{\mathbb{E}^2[Y_q]}.\end{aligned}$$

From Lemma 4.4, we know $(d/d\mu)(\mathbb{E}[Y_q^2]/\mathbb{E}[Y_q]) > 0$, thus

$$\mathbb{E}[(Y_{q-1}+1)^2]\mathbb{E}[Y_q] - \mathbb{E}[Y_q^2]\mathbb{E}[Y_{q-1}+1] > 0,$$

which implies inequality (14) is satisfied. ∎

https://doi.org/10.1017/S0269964820000583 Published online by Cambridge University Press

7. CONCLUSION

Motivated by applications in shipment consolidation, we first provide a new unified method to compute AOD for any renewal-type clearing policy based on a martingale associated with the Poisson process and the martingale stopping theorem. Our goal is to provide a complete analytical comparison of alternative clearing policies (of type QP, TP, and HP) in terms of AOD. Our proposed method lends itself for a generic analytical characterization of AOD, leading to a complete comparative analysis of the policies of interest. In particular, we demonstrate that, under a fixed expected consolidation cycle length, QP outperforms any other renewal-type clearing policy in terms of AOD, not limited to HPs and TPs (see Theorem 5.1). Also, we complete the proof for the comparison of AOD between HPs and TPs under a fixed expected consolidation cycle length (see Theorems 5.3 and A.2), and we provide a simplified proof for the AOD comparison among HPs, TPs, and QP under fixed parameters (see Theorems 6.1 and 6.2), which are related to a property of truncated Poisson random variables: for a truncated Poisson random variable $Y_N = \min(Y, N)$, $\mathbb{VAR}[Y_N] < \mathbb{E}[Y_N]$ (see Lemma 4.2).

Furthermore, we provide explicit and stronger comparative results between two HPs of the same type under a fixed expected consolidation cycle length (see Theorem 5.5), which rely on a property of truncated random variables: given two integer-valued random variables X, Y, X is stochastically larger than Y, if $\mathbb{E}[X_q] = \mathbb{E}[Y_{q+1}]$, where q is a positive integer, then $\mathbb{E}[X_q^2] \leq \mathbb{E}[Y_{q+1}^2]$ (see Lemma 4.3).

Last but not least, we analytically show HP1 performs better than HP2 in terms of AOD under fixed parameters (see Theorem 6.3), which is equivalent to another property of truncated Poisson random variables: $X \sim \text{Poisson}(\mu)$, then $\mathbb{E}[X_N^2]/\mathbb{E}[X_N]$ is increasing with respect to μ (see Lemma 4.4).

Our results offer insightful and analytically justifiable guidance for logistics managers in selecting an appropriate shipment consolidation policy with an eye on the resulting service performance. Several challenging extensions of the problem at hand remain open for future research including the case where shipment consolidation efforts are subject to multiple and/or more general input processes, for example, the case of a multi-class stochastic clearing system subject to Markov-modulated, renewal, or Brownian motion input processes without or with customer/order differentiation, for example, prioritization.

References

1. Athreya, K.B. & Lahiri, S.N. (2006). *Measure theory and probability theory*. New York, NY, USA: Springer Science & Business Media.
2. Bookbinder, J.H., Cai, Q., & He, Q.-M. (2011). Shipment consolidation by private carrier: the discrete time and discrete quantity case. *Stochastic Models* 27(4): 664–686.
3. Cai, Q., He, Q.M., & Bookbinder, J.H. (2014). A tree-structured Markovian model of the shipment consolidation process. *Stochastic Models* 30(4): 521–553.
4. Çetinkaya, S. (2005). Coordination of inventory and shipment consolidation decisions: a review of premises, models, and justification. In J. Geunes, E. Akçali, P.M. Pardalos, H.E. Romejin, & Z.J. Shen (eds), *Applications of supply chain management and e-commerce research*. New York: Springer, pp. 3–51.
5. Çetinkaya, S. & Bookbinder, J.H. (2003). Stochastic models for the dispatch of consolidated shipments. *Transportation Research Part B: Methodological* 37(8): 747–768.
6. Çetinkaya, S. & Lee, C.-Y. (2000). Stock replenishment and shipment scheduling for vendor-managed inventory systems. *Management Science* 46(2): 217–232.
7. Çetinkaya, S., Mutlu, F., & Lee, C.-Y. (2006). A comparison of outbound dispatch policies for integrated inventory and transportation decisions. *European Journal of Operational Research* 171(3): 1094–1112.
8. Çetinkaya, S., Tekin, E., & Lee, C.-Y. (2008). A stochastic model for joint inventory and outbound shipment decisions. *IIE Transactions* 40(3): 324–340.

9. Çetinkaya, S., Mutlu, F., & Wei, B. (2014). On the service performance of alternative shipment consolidation policies. *Operations Research Letters* 42(1): 41–47.
10. Cook, R.A. & Lodree, E.J. (2017). Dispatching policies for last-mile distribution with stochastic supply and demand. *Transportation Research Part E: Logistics and Transportation Review* 106: 353–371.
11. Economou, A. & Manou, A. (2013). Equilibrium balking strategies for a clearing queueing system in alternating environment. *Annals of Operations Research* 208(1): 489–514.
12. Higginson, J. & Bookbinder, J.H. (1994). Policy recommendations for a shipment-consolidation program. *Journal of Business Logistics* 15(1): 87–112.
13. Kaya, O., Kubalı, D., & Örmeci, L. (2013). Stochastic models for the coordinated production and shipment problem in a supply chain. *Computers & Industrial Engineering* 64(3): 838–849.
14. Kella, O., Perry, D., & Stadje, W. (2003). A stochastic clearing model with a Brownian and a compound Poisson component. *Probability in the Engineering and Information Sciences* 17: 1–22.
15. Lai, M., Weili, X., & Zhao, L. (2016). Cost allocation for cooperative inventory consolidation problems. *Operations Research Letters* 44(6): 761–765.
16. Mutlu, F. & Çetinkaya, S. (2010). An integrated model for stock replenishment and shipment scheduling under common carrier dispatch costs. *Transportation Research Part E: Logistics and Transportation Review* 46(6): 844–854.
17. Mutlu, F., Çetinkaya, S., & Bookbinder, J.H. (2010). An analytical model for computing the optimal time-and-quantity-based policy for consolidated shipments. *IIE Transactions* 42: 367–377.
18. Robin, M. & Tapiero, C.S. (1982). A simple vehicle dispatching policy with non-stationary stochastic arrival rates. *Transportation Research Part B* 16(6): 449–457.
19. Ross, S.M. (1969). Optimal dispatching of a Poisson process. *Journal of Applied Probability* 6: 692–699.
20. Ross, S.M. (1996). *Stochastic processes*, 2nd ed. New York: John Wiley & Sons.
21. Satır, B., Erenay, F.S., & Bookbinder, J.H. (2018). Shipment consolidation with two demand classes: rationing the dispatch capacity. *European Journal of Operational Research* 270(1): 171–184.
22. Stidham, S. (1974). Stochastic clearing systems. *Stochastic Processes and Their Applications* 2: 85–113.
23. Stidham, S. (1977). Cost models for stochastic clearing systems. *Operations Research* 25(1): 100–127.
24. Tapiero, C.S. & Zuckerman, D. (1979). Vehicle dispatching with competition. *Transportation Research Part B* 13: 207–216.
25. Ülkü, M.A. & Bookbinder, J.H. (2012). Optimal quoting of delivery time by a third party logistics provider: the impact of shipment consolidation and temporal pricing schemes. *European Journal of Operational Research* 221(1): 110–117.
26. Zuckerman, D. & Tapiero, C.S. (1980). Random vehicle dispatching with options and optimal fleet size. *Transportation Research Part B* 14: 361–368.

APPENDIX: THE CASE WITH NO EMPTY SHIPMENTS

Under TP1 and HP1, there may be empty shipments, which happens when $N(T) = 0$. In this Appendix, we consider the policies referred to as revised TP1 and revised HP as defined in Section 1. These policies do not allow empty shipments.

Under revised HP1 with parameters q, T, the following recursion equation about the expected consolidation cycle length $\mathbb{E}[C_{\mathrm{RHP1}}]$ is satisfied,

$$\mathbb{E}[C_{\mathrm{RHP1}}] = \mathbb{P}(N(T) \geq 1)\mathbb{E}[\tau_q \wedge T \mid N(T) \geq 1] + \mathbb{P}(N(T) = 0)(T + \mathbb{E}[C_{\mathrm{RHP1}}]). \qquad \textbf{(A.1)}$$

The equation means that if no order arrives within T units time, which happens with probability $\mathbb{P}(N(T) = 0)$, the consolidation cycle restarts; if there are orders arriving within T units time, which happens with probability $\mathbb{P}(N(T) \geq 1)$, the load is dispatched at stopping time $\tau_q \wedge T$. Noting that

$$\begin{aligned}\mathbb{E}[\tau_q \wedge T] &= \mathbb{P}(N(T) \geq 1)\mathbb{E}[\tau_q \wedge T \mid N(T) \geq 1] + \mathbb{P}(N(T) = 0)\mathbb{E}[\tau_q \wedge T \mid N(T) = 0] \\ &= \mathbb{P}(N(T) \geq 1)\mathbb{E}[\tau_q \wedge T \mid N(T) \geq 1] + \mathbb{P}(N(T) = 0)T,\end{aligned}$$

we have

$$\mathbb{P}(N(T) \geq 1)\mathbb{E}[\tau_q \wedge T \mid N(T) \geq 1] = \mathbb{E}[\tau_q \wedge T] - \mathbb{P}(N(T) = 0)T. \qquad \textbf{(A.2)}$$

https://doi.org/10.1017/S0269964820000583 Published online by Cambridge University Press

Now, using Eq. (A.2) in Eq. (A.1), and recalling $\mathbb{E}[C_{\mathrm{HP1}}]$ in Table 1, we have

$$\mathbb{E}[C_{\mathrm{RHP1}}] = \frac{\mathbb{E}[\tau_q \wedge T]}{1-\mathbb{P}(N(T)=0)} = \frac{\mathbb{E}[Y_q]}{\lambda(1-\mathbb{P}(Y=0))} = \frac{\mathbb{E}[C_{\mathrm{HP1}}]}{1-\mathbb{P}(Y=0)}, \tag{A.3}$$

where $Y \sim \mathrm{Poisson}(\lambda T)$.

Next, we calculate the expected cumulative delay within one consolidation cycle under revised HP1, which is denoted as $\mathbb{E}[W_{\mathrm{RHP1}}]$. The following recursion equation is satisfied,

$$\mathbb{E}[W_{\mathrm{RHP1}}] = \mathbb{P}(N(T)\geq 1)\mathbb{E}\left[\int_0^{\tau_q\wedge T} N(t)\,dt \mid N(T)\geq 1\right] + \mathbb{P}(N(T)=0)\mathbb{E}[W_{\mathrm{RHP1}}]. \tag{A.4}$$

This equation means if no order arrives within T units time, which happens with probability $\mathbb{P}(N(T)=0)$, the consolidation system restarts; if there are orders arriving within T units time, which happens with probability $\mathbb{P}(N(T)\geq 1)$, the cumulative delay within one consolidation cycle is $\int_0^{\tau_q\wedge T} N(t)\,dt$.

By noticing

$$\begin{aligned}\mathbb{E}\left[\int_0^{\tau_q\wedge T} N(t)\,dt\right] &= \mathbb{P}(N(T)\geq 1)\mathbb{E}\left[\int_0^{\tau_q\wedge T} N(t)\,dt \,\middle|\, N(T)\geq 1\right] \\ &\quad + \mathbb{P}(N(T)=0)\mathbb{E}\left[\int_0^{\tau_q\wedge T} N(t)\,dt \,\middle|\, N(T)=0\right] \\ &= \mathbb{P}(N(T)\geq 1)\mathbb{E}\left[\int_0^{\tau_q\wedge T} N(t)\,dt \,\middle|\, N(T)\geq 1\right],\end{aligned} \tag{A.5}$$

and replacing Eq. (A.5) into Eq. (A.4), together with recalling $\mathbb{E}[W_{\mathrm{HP1}}]$ in Table 1, we have

$$\mathbb{E}[W_{\mathrm{RHP1}}] = \frac{\mathbb{E}[\int_0^{\tau_q\wedge T} N(t)\,dt]}{1-\mathbb{P}(N(T)=0)} = \frac{\mathbb{E}[W_{\mathrm{HP1}}]}{1-\mathbb{P}(Y=0)} = \frac{\mathbb{E}[Y_q(Y_q-1)]}{2\lambda(1-\mathbb{P}(Y=0))}. \tag{A.6}$$

Define a new random variable $\tilde{Y}$, which has the same distribution of $Y \mid Y>0$. We can rewrite

$$\mathbb{E}[C_{\mathrm{RHP1}}] = \frac{1}{\lambda}\mathbb{E}[\tilde{Y}_q], \tag{A.7}$$

$$\mathbb{E}[W_{\mathrm{RHP1}}] = \frac{1}{2\lambda}\mathbb{E}[\tilde{Y}_q(\tilde{Y}_q-1)]. \tag{A.8}$$

Similarly, we can obtain the expected cycle length under revised TP1 with parameters T is

$$\mathbb{E}[C_{\mathrm{RTP1}}] = \frac{\mathbb{E}[C_{\mathrm{TP1}}]}{1-\mathbb{P}(N(T)=0)} = \frac{T}{1-e^{-\lambda T}}, \tag{A.9}$$

and the cumulative delay with one consolidation cycle under revised TP1 with parameters T is

$$\mathbb{E}[W_{\mathrm{RTP1}}] = \frac{\mathbb{E}[W_{\mathrm{TP1}}]}{1-\mathbb{P}(N(T)=0)} = \frac{\lambda T^2}{2(1-e^{-\lambda T})}. \tag{A.10}$$

From Eqs. (A.3), (A.6), (A.9), and (A.10) and the definition of AOD, we have that AOD of revised HP1 is the same as HP1, AOD of revised TP1 is the same as TP1, if the parameters q, T are fixed. From Theorem 6.1, with fixed parameters q, T, revised HP1 also performs better than QP and revised TP1 in terms of AOD.

From Theorem 5.1, we can conclude that for a given expected consolidation cycle length, QP performs better than revised HP1 and revised TP1, in terms of AOD. In the following, we provide the comparison between revised HP1 and revised TP1 with a given expected consolidation cycle length.

Suppose $Y \sim \text{Poisson}(\lambda_1)$, $Z \sim \text{Poisson}(\lambda_2)$, and $\lambda_1 > \lambda_2$, we know Y is stochastically larger than Z. Define $\tilde{Y} \sim Y \mid Y > 0$ and $\tilde{Z} \sim Z \mid Z > 0$, we show $\tilde{Y}$ is also stochastically larger than $\tilde{Z}$ in the following lemma.

LEMMA A.1: *Let $Y \sim$ Poisson(λ), $\tilde{Y}$ is distributed as $Y \mid Y > 0$, then $\mathbb{P}(\tilde{Y} > n)$ is increasing in λ, for any integer $n \geq 1$.*

PROOF: Notice $(d/d\lambda)\mathbb{P}(Y > n) = \mathbb{P}(Y = n)$. Then, for $n \geq 1$,

$$\begin{aligned}
\frac{d}{d\lambda}\mathbb{P}(\tilde{Y} > n) &= \frac{d}{d\lambda}\frac{\mathbb{P}(Y > n)}{\mathbb{P}(Y > 0)} \\
&= \frac{\mathbb{P}(Y = n)\mathbb{P}(Y > 0) - \mathbb{P}(Y > n)\mathbb{P}(Y = 0)}{(\mathbb{P}(Y > 0))^2} \\
&= \frac{\mathbb{P}(Y = n) - e^{-\lambda}\mathbb{P}(Y \geq n)}{(\mathbb{P}(Y > 0))^2}.
\end{aligned}$$

In addition, by using $\mathbb{P}(Y = k) = (\lambda/k)\mathbb{P}(Y = k - 1)$, we have

$$\begin{aligned}
\mathbb{P}(Y = n) - e^{-\lambda}\mathbb{P}(Y \geq n) &= \frac{\lambda}{n}\mathbb{P}(Y = n - 1) - e^{-\lambda}\sum_{k=n}^{\infty}\frac{\lambda}{k}\mathbb{P}(Y = k - 1) \\
&> \frac{\lambda}{n}\left(\mathbb{P}(Y = n - 1) - e^{-\lambda}\mathbb{P}(Y \geq n - 1)\right).
\end{aligned}$$

Since $\mathbb{P}(Y = 0) - e^{-\lambda}\mathbb{P}(Y \geq 0) = 0$, it follows by induction that

$$\mathbb{P}(Y = n) - e^{-\lambda}\mathbb{P}(Y \geq n) > 0.$$

Therefore, $(d/d\lambda)\mathbb{P}(\tilde{Y} > n) > 0$. ■

THEOREM A.2: *For a given expected consolidation cycle length $\mathbb{E}[C]$, the revised HP1 with a larger quantity parameter achieves larger AOD than the revised HP1 with a smaller quantity parameter, in terms of AOD. In particular, revised HP1 achieves less AOD than revised TP1, under a given expected consolidation cycle length $\mathbb{E}[C]$.*

PROOF: We consider a fixed $\mathbb{E}[C]$ and use the following notation for the corresponding policy parameters under this $\mathbb{E}[C]$ value: a revised HP1 with parameters q_{H} and T_{H}, the other revised HP1 with parameters $q_{\mathrm{H}} + 1$ and T'_{H}. Recalling Eq. (A.7) and by assumption that the two revised HP1 have the same expected cycle length, we have,

$$\mathbb{E}[\tilde{U}_{q_{\mathrm{H}}}] = \mathbb{E}[\tilde{V}_{q_{\mathrm{H}}+1}], \tag{A.11}$$

where $\tilde{U}$ is distributed as $U \mid U > 0$, $U \sim \text{Poisson}(\lambda T_{\mathrm{H}})$, and $\tilde{V}$ is distributed as $V \mid V > 0$, $V \sim \text{Poisson}(\lambda T'_{\mathrm{H}})$. Clearly, $T_{\mathrm{H}} > T'_{\mathrm{H}}$. From Lemma A.1, $\tilde{U}$ is stochastically larger than $\tilde{V}$.

Next, recalling Eq. (A.8) and reiterating the assumption of fixed $\mathbb{E}[C]$, we proceed to show that

$$\mathbb{E}[\tilde{U}_{q_{\mathrm{H}}}(\tilde{U}_{q_{\mathrm{H}}} - 1)] \leq \mathbb{E}[\tilde{V}_{q_{\mathrm{H}}+1}(\tilde{V}_{q_{\mathrm{H}}+1} - 1)]. \tag{A.12}$$

From Lemma 4.3, and recalling Eq. (A.11), we have

$$\mathbb{E}[\tilde{U}_{q_{\mathrm{H}}}^2] \leq \mathbb{E}[\tilde{V}_{q_{\mathrm{H}}+1}^2],$$

so that inequality (A.12) is verified.

Revised TP1 can be considered as revised HP1 with quantity parameter ∞. Therefore, under the same expected consolidation cycle $\mathbb{E}[C]$, revised HP1 achieves less AOD than revised TP1. ■