Ann Oper Res (2010) 181: 1–19
DOI 10.1007/s10479-009-0662-9

IPA derivatives for a discrete model of make-to-stock production-inventory systems with backorders

Benjamin Melamed · Yihong Fan · Yao Zhao · Yorai Wardi

Published online: 20 November 2009

Abstract We consider a class of single-stage, single-product *Make-to-Stock production-inventory system* (*MTS* system) with backorders. The system employs a continuous-review base-stock policy which strives to maintain a prescribed base-stock level of inventory. In a previous paper of Zhao and Melamed (*Methodology and Computing in Applied Probability* 8:191–222, 2006), the *Infinitesimal Perturbation Analysis* (*IPA*) derivatives of inventory and backorders time averages with respect to the base-stock level and a parameter of the production-rate process were computed in *Stochastic Fluid Model* (*SFM*) setting, where the demand stream at the inventory facility and its replenishment stream from the production facility are modeled by stochastic rate processes. The advantage of the SFM abstraction is that the aforementioned IPA derivatives can be shown to be unbiased. However, its disadvantages are twofold: (1) on the modeling side, the highly abstracted SFM formulation does not maintain the identity of transactions (individual demands, orders and replenishments) and has no notion of lead times, and (2) on the applications side, the aforementioned IPA derivatives are brittle in that they contain instantaneous rates at certain hitting times which

B. Melamed (✉)
Rutgers Business School—Newark and New Brunswick, Department of Supply Chain Management and Marketing Sciences, Rutgers University, 94 Rockafeller Rd., Piscataway, NJ 08854, USA
e-mail: melamed@rbs.rutgers.edu

Y. Fan
Rutgers Business School—Newark and New Brunswick, Department of Management Science and Information Systems, Rutgers University, 1 Washington Park, Newark, NJ 07102, USA
e-mail: fanyihon@andromeda.rutgers.edu

Y. Zhao
Rutgers Business School—Newark and New Brunswick, Department of Supply Chain Management and Marketing Sciences, Rutgers University, 1 Washington Park, Newark, NJ 07102, USA
e-mail: yaozhao@andromeda.rutgers.edu

Y. Wardi
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
e-mail: wardi@ee.gatech.edu

are rarely known, and consequently, need to be estimated. In this paper, we remedy both disadvantages by using a discrete setting, where transaction identity is maintained, and order fulfillment from inventory following demand arrivals and inventory restocking following replenishment arrivals are modeled as discrete jumps in the inventory level. We then compute the aforementioned IPA derivatives with respect to the base-stock level and a parameter of the lead-time process in the discrete setting under any initial system state. The formulas derived are shown to be unbiased and directly computable from sample path observables, and their computation is both simple and computationally robust.

Keywords Infinitesimal perturbation analysis (IPA) · IPA derivatives · Make-To-Stock production-inventory system (MTS system) · Discrete model

1 Introduction

We consider a class of single-stage, single-product *Make-to-Stock* (*MTS*) production-inventory systems with backorders. An MTS system consists of a production facility coupled to an inventory facility: the inventory facility is visited by a stream of demands and the production facility replenishes the inventory facility. The system is driven by random demand and possibly random production processes. We assume that the production facility has an unlimited supply of raw material, so it never starves. The system employs a continuous-review *base-stock* policy which strives to maintain a prescribed base-stock level, S, of inventory as follows: while the inventory level is below S, replenishment is turned on, and otherwise, it is turned off.

The general description above of an MTS system can be modeled using two related paradigms, the *Stochastic Fluid Model* (*SFM*) paradigm and the discrete paradigm. These paradigms will be briefly described next, and their relative merits and shortcomings will be discussed later on.

In the discrete MTS model with backorders, the demand process consists of an inter-arrival time process and a random demand size process. Demands arrive at the inventory facility and are satisfied from inventory on hand (if available). When an inventory shortfall is encountered, the arriving demand receives the entire inventory on hand (if any), and the shortfall is backordered from the production facility; the demand then waits in a first-come-first-serve (FCFS) buffer at the inventory facility until sufficient replenishment arrives. A continuous-review base-stock policy governs the orders placed with the production facility, based on the prevailing *inventory position* (inventory on hand plus replenishments en route minus backorders). Initially, the system's inventory position is allowed to exceed S, a situation referred to as *overage*. Similarly, when the inventory position lies strictly below S, we refer to this situation as *underage*. When a demand arrival results in an underage, the inventory facility immediately sends an order that raises the inventory position back to S; thus, overage can only occur initially, but once the system encounters the first underage, overage can never recur. Replenishments corresponding to previously placed orders arrive at the inventory facility after a random lead time, which may include manufacturing time, transportation time, etc. We make no assumptions on the processing order of replenishment orders in the production facility. Both demand satisfaction (following demand arrival) and inventory restocking (following replenishment arrival) are instantaneous. Consequently, the inventory level and inventory position processes are piecewise constant with jumps occurring when a demand arrives (downward jump) or replenishment arrives (upward jump). Such jumps will be referred to as *sample-path events*. A schematic of the discrete MTS system is depicted in Fig. 1.

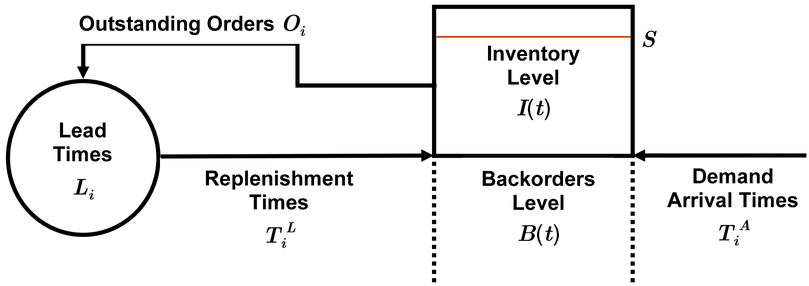

Fig. 1 A schematic of the discrete MTS system with backorders under the base stock policy

In contrast, an SFM counterpart is a highly abstracted model, which does not maintain the identity of transactions (individual demands, orders and replenishments) and has no notion of lead times. Rather, the flows of inventory into the inventory facility and out of it are modeled as fluid-flow. More specifically, the stream of outgoing inventory is modeled by a demand arrival-rate process that draws down on the inventory level, while the stream of replenishment is modeled by a replenishment-rate process that restocks the inventory. The rate processes are often assumed to be piecewise-constant, and consequently, the inventory level and inventory position processes are piecewise linear. Backordering is captured by allowing the inventory position to be negative, while the base-stock policy is implemented by requiring that the replenishment rate does not lead to an overage following an underage.

The subject matter of this paper is *Infinitesimal Perturbation Analysis* (*IPA*) of MTS systems with backorders under the base-stock policy. IPA is a technique for obtaining sample path derivatives of a random variable $L(\theta)$ with respect to some parameters of interest, θ; see Fu and Hu (1997) and Cassandras and Lafortune (1999) for comprehensive discussions of IPA derivatives and their applications. For IPA-based application to be statistically accurate, it is necessary that the IPA derivative should be unbiased. Specifically, letting $L(\theta)$ be a random variable, an IPA derivative is said to be *unbiased* if the expectation and differentiation operators commute, i.e., $E[\frac{d}{d\theta}L(\theta)] = \frac{d}{d\theta}E[L(\theta)]$; otherwise, it is said to be *biased*. Sufficient conditions for unbiased IPA derivatives are given in the following result.

Fact 1 (See Rubinstein and Shapiro 1993, Lemma A2, p. 70) *An IPA derivative* $\frac{d}{d\theta}L(\theta)$ *is unbiased*, *if*

(a) *For each* θ, *the IPA derivatives* $\frac{d}{d\theta}L(\theta)$ *exist w.p.*1 (*with probability* 1),
(b) *W.p.*1, $L(\theta)$ *is Lipschitz continuous in* Θ, *and the* (*random*) *Lipschitz constants have finite first moments*.

IPA derivatives are nonparametric in the sense that they are computed from sample paths without any knowledge of the underlying probability law. Consequently, they can be computed from simulation runs or in real-life systems deployed in the field, and the values can potentially be used in stochastic optimization. This property holds out the promise of utilizing IPA derivative formulas to provide sensitivity information on system metrics with respect to control parameters of interest, and can serve as the theoretical underpinning for offline design algorithms and online control algorithms. However, when contemplating an appropriate modeling paradigm, (discrete or SFM), the modeler is faced with a basic tradeoff. A discrete model is a priori preferable, because it requires far less abstraction than its SFM counterpart.

Moreover, experience shows that IPA derivatives in SFM setting can contain computationally brittle terms in the form of instantaneous rates at certain hitting times, which are rarely known, and consequently, need to be estimated (e.g., Theorems 1 and 2 in each of Zhao and Melamed 2006, 2007). This is not the case in the discrete setting, since all quantities in the associated IPA derivatives are directly observable from system sample paths. On the other hand, IPA derivatives in discrete setting tend to be biased (Heidelberger et al. 1988), while their SFM counterparts tend to be unbiased. Thus, provided the associated IPA derivatives are unbiased, the modeler would generally prefer the discrete paradigm, and otherwise the modeler would resort to the SFM paradigm (e.g., Wardi et al. 2002, Cassandras et al. 2002, 2003).

IPA in SFM setting has been previously applied to MTS systems. Paschalidis et al. (2004) treats a tandem supply chain with MTS inventories at each stage and demand at the last stage, subject to backordering. The paper seeks to optimize the overall inventory costs, and the solution combines IPA with large deviations. Additionally, the relation between these theories is elucidated. Panayiotou and Cassandras (2006) devises online algorithms to optimize inventory capacities with respect to an objective function that balances inventory carrying and stockout costs. This paper shows that the requisite IPA derivatives are unbiased and simpler than their discrete counterparts (which are generally biased). The class of MTS systems under the base-stock policy is treated in Zhao and Melamed (2006, 2007) in SFM setting, where IPA derivatives are computed for the time averaged inventory level and time-averaged backorder level with respect to the base-stock level, as well as a parameter of the production rate process (the first paper addresses MTS systems with backorders, and the second addresses MTS systems with lost sales). In both papers, the IPA derivatives are shown to be unbiased, and their formulas turn out to be fairly simple. However, as mentioned above, for some initial conditions, the corresponding IPA derivatives include terms that contain instantaneous flow rates at certain hitting times (cf. Theorems 1 and 2 in each of Zhao and Melamed 2006, 2007). Unfortunately, estimating unknown instantaneous rates at specific time points in SFM setting is not straightforward.

The contribution of this paper to the subject is to exhibit a *discrete model formulation* for MTS systems with backorders under the base-stock policy, such that the corresponding IPA derivatives of the time averaged inventory level and backorder level with respect to the base-stock level and a parameter of the lead-time process are derived and shown to be *unbiased*. Thus, the current paper improves on Zhao and Melamed (2006) in two ways: (1) by postulating a higher-fidelity model of the actual MTS system under study, and (2) by developing *unbiased* and *computationally robus*t IPA formulas which are *directly* computable from sample path observables, thereby obviating the need for rate estimation.

The main tool for computing the IPA derivatives is the generalized Leibniz integral rule given below.

Fact 2 (See Lemma A.1 and Corollary A.1 in Fan et al. 2009) *Let the function* $f(t,\theta)$ *be defined in the rectangle* $[A,B]\times[\varphi,\psi]$*, and let* $a_i(\theta)$, $1\le i\le n$, *be differentiable functions on* $[\varphi,\psi]$*, satisfying* $A\le a_1(\theta)\le a_2(\theta)\le\cdots\le a_n(\theta)\le B$ *for every* $\theta\in[\varphi,\psi]$*. For* $1\le i\le n-1$*, let* $D_{1,i}=\{(t,\theta):\theta\in[\varphi,\psi],t\in(a_i(\theta),a_{i+1}(\theta))\}$ *and* $D_{2,i}=\{(t,\theta):\theta\in[\varphi,\psi],t\in[a_i(\theta),a_{i+1}(\theta)]\}$*. Suppose that for each* $1\le i\le n-1$, $f(t,\theta)$ *is continuous on* $D_{1,i}$ *with a continuously differentiable partial derivative* $\frac{\partial}{\partial\theta}f(t,\theta)$ *on* $D_{1,i}$*. Suppose further that there exist functions* $g_i(t,\theta)$, $1\le i\le n-1$, *such that*

1. *For each* $1\le i\le n-1$, $g_i(t,\theta)$ *is continuous on* $D_{2,i}$ *with a continuously differentiable partial derivative* $\frac{\partial}{\partial\theta}g_i(t,\theta)$ *on* $D_{2,i}$*.*
2. *For every* $\theta\in[\varphi,\psi]$, $f(t,\theta)=g_i(t,\theta)$ *on each interval* $(a_i(\theta),a_{i+1}(\theta))$, $1\le i\le n-1$.

Then the function

$$F(\theta) = \int_{a_1(\theta)}^{a_n(\theta)} f(t,\theta)dt$$

is differentiable on $[\varphi, \psi]$, *and its derivative is given by*

$$\begin{aligned}\frac{d}{d\theta}F(\theta) &= \frac{d}{d\theta}\int_{a_1(\theta)}^{a_n(\theta)} f(t,\theta)dt \\ &= \int_{a_1(\theta)}^{a_n(\theta)} \frac{\partial}{\partial\theta} f(t,\theta)dt \\ &\quad + \sum_{i=1}^{n-1}\left[f(a_{i+1}(\theta)-,\theta)\frac{d}{d\theta}a_{i+1}(\theta) - f(a_i(\theta)+,\theta)\frac{d}{d\theta}a_i(\theta)\right]. \qquad (1.1)\end{aligned}$$

Throughout the paper, we use the following notational conventions and terminology. $N(x)$ denotes a neighborhood of x, where x may be vector valued. A function $f(x)$ is said to be *locally differentiable* at x if it is differentiable in a neighborhood of x; it is said to be *locally independent* of x if it is constant in a neighborhood of x. The indicator function of set A is denoted by 1_A, $x^+ = \max\{x, 0\}$ and $x^- = -\min\{x, 0\}$.

The rest of the paper is organized as follows. Section 2 presents the discrete MTS model. Section 3 describes the performance metrics and parameters of interest. Section 4 obtains IPA derivative formulas and shows them to be unbiased. Section 5 concludes the paper.

2 The discrete MTS model

Consider a discrete MTS system with backorders, under the base-stock policy, over a finite time horizon $[0, T]$. We assume that the system satisfies the following regularity assumption:

Assumption 1 Demand-arrival and product-replenishment sample-path events do not occur simultaneously, almost surely. Consequently, no simultaneous jumps occur in the inventory and backorder processes, w.p.1.

We distinguish between stochastic processes that specify the model (*defining processes*) and those which are functions thereof (*derived processes*).

2.1 Defining processes

The defining processes of the discrete MTS system under study are listed below.

- $\{T_i^A : i \geq 1\}$ is the *demand arrival-time process*, where T_i^A is the arrival time of the i-th demand. Note that this is also the time of placing the corresponding replenishment order.
- $\{D_i : i \geq 1\}$ is the *demand-size process*, where D_i is the quantity of the i-th demand.
- $\{O_i : i \geq 1\}$ is the *order-size process*, where O_i is the quantity of the i-th order placed by the inventory facility. We assume that at time 0, there are N_0 pending orders (N_0 is generally random). These pending orders, called the *initial pending orders*, are assumed to have been placed prior to time 0, but have not yet arrived. The order-size process is enumerated by order-placement time.

- $\{T_i^L : i \geq 1\}$ is the *replenishment arrival-time process*, where T_i^L is the arrival time of the i-th replenishment.
- $\{V_i : i \geq 1\}$ is the *replenishment-size process*, where V_i is the quantity of the i-th replenishment. The replenishment-size process is enumerated by replenishment arrival time.
- $\{L_i : i \geq 1\}$ is the *replenishment lead-time process*, where L_i is the lead time of the i-th replenishment. The lead time starts when an order is placed and ends when the corresponding replenishment arrives. For initial pending orders, the lead times are the residual lead times starting from time zero.

The processes $\{D_i\}, \{O_i\}$ and $\{V_i\}$ are related. The relation between $\{O_i\}$ and $\{V_i\}$ is given by

$$O_i = V_{C_i}, \tag{2.1}$$

where $\{C_i : i \geq 1\}$ is the random process that maps the index of the order (when it is placed) to the index of the corresponding replenishment (when it arrives). Note that (2.1) admits overtaking in replenishment relative to order placement. The relation between $\{D_i\}$ and $\{O_i\}$ will be exhibited in (2.9).

2.2 Derived processes

The following derived processes pertain to the discrete MTS system.

- $\{W(t) : t \geq 0\}$ is the *extended inventory-level process* (see Zhao and Melamed 2006, 2007), where $W(t)$ is defined by

$$W(t) = W(0) - \sum_{i=1}^{\infty} 1_{\{T_i^A \leq t\}} D_i + \sum_{j=1}^{\infty} 1_{\{T_j^L \leq t\}} V_j, \tag{2.2}$$

  and the initial extended inventory level, $W(0)$, is a given random variable.
- $\{I(t) : t \geq 0\}$ is the *inventory-level process*, where $I(t)$ is the volume of inventory on-hand at time t, given by

$$I(t) = W(t)^+ = W(t) 1_{\{W(t) \geq 0\}}. \tag{2.3}$$

- $\{B(t) : t \geq 0\}$ is the *backorder-level process*, where $B(t)$ is the volume of backorders at time t, given by

$$B(t) = W(t)^- = -W(t) 1_{\{W(t) \leq 0\}}. \tag{2.4}$$

- $\{P(t) : t \geq 0\}$ is the *inventory-position process*, where $P(t)$ is the amount of inventory on-hand plus all pending replenishment sizes minus all outstanding backorder sizes at time t.
- $\{P_-(t) : t \geq 0\}$ is the *inventory short-position process*, where $P_-(t)$ is the inventory position at time t minus the order size placed at time t (if any). In particular, the initial inventory short position is given by

$$P_-(0) = W(0) + \sum_{i=1}^{N_0} O_i. \tag{2.5}$$

Note that $W(t)$ determines both $I(t)$ and $B(t)$ (and vice versa), since (2.3) and (2.4) imply the relation

$$W(t) = I(t) - B(t) = \begin{cases} I(t), & \text{if } B(t) = 0, \\ -B(t), & \text{if } I(t) = 0. \end{cases} \tag{2.6}$$

2.3 Model construction

We make the following assumptions without loss of practical generality.

Assumption 2

(a) W.p.1, $O_i(\theta), V_i(\theta), D_i(\theta) \le O^*$, $i \ge 1$, where O^* is a positive deterministic constant independent of $\theta \in \Theta$ and i.
(b) W.p.1, the number of sample-path events is bounded by a deterministic constant $K^*(T)$ depending on T, but independent of $\theta \in \Theta$.

These assumptions state that order sizes (and consequently, replenishment sizes) cannot be arbitrarily large, and as such capture common operational rules in actual supply chains. We also assume that all demand sizes satisfy these bounds.

We next proceed to exhibit some relations and define auxiliary random variables. By Assumptions 1 and 2, the order-size process $\{O_i\}$ satisfies the following relationships:

- The first N_0 orders are the initial pending orders, $O_1, \ldots, O_{N_0}$.
- For $i = N_0 + 1$ (the first order placed at or after time 0),

$$O_{N_0+1} = \begin{cases} S - P_-(0), & \text{on } \{S - P_-(0) > 0\}, \\ S - P_-(T_H^A), & \text{otherwise}, \end{cases} \tag{2.7}$$

  where T_H^A is the arrival time for the H-th demand, and the random index H is given by

$$H = \begin{cases} \min\{j \ge 1 : S - P_-(T_j^A) > 0\}, & \text{if the minimum exists}, \\ \infty, & \text{otherwise}. \end{cases} \tag{2.8}$$

  Note that H may be strictly greater than 1, because the initial inventory position is allowed to engender overage. However, if the initial position corresponds to an underage, then an initial order is immediately placed at time 0.
- For $i > N_0 + 1$, the relationship between $\{O_i\}$ and $\{D_i\}$ after time T_H^A, is given by

$$O_{N_0+1+j} = D_{H+j}, \quad j \ge 1. \tag{2.9}$$

Let $[0, T]$ be a finite time interval for some prescribed T. Suppressing the dependence on ω and T, define the auxiliary random intervals $(Q_j(\theta), R_j(\theta))$, $j = 0, 1, \ldots, J(\theta)$, where $J(\theta)$ is the number of such intervals that have a non-empty intersection with $[0, T]$. For $j = 1, \ldots, J(\theta)$, let $(Q_j(\theta), R_j(\theta))$ be the ordered extremal subintervals of $[0, T)$, such that $P(t, \theta) = S > W(t, \theta)$ for all $t \in (Q_j(\theta), R_j(\theta))$, namely, the endpoints, $Q_j(\theta)$ and $R_j(\theta)$, are obtained via the inf and sup functions, respectively. For $j = 0$, we define $(Q_0(\theta), R_0(\theta)) = (0, Q_1(\theta))$. Note that if $W(0) < S$, then $Q_0(\theta) = R_0(\theta) = Q_1(\theta) = 0$. By conventi‘on, if any of these endpoints does not exist, then it is set to ∞. For any finite time interval $[0, T)$, $J(\theta) < \infty$ w.p.1 by part (b) of Assumption 2, and

$$Q_1(\theta) < R_1(\theta) < Q_2(\theta) < R_2(\theta) < \cdots < Q_{J(\theta)}(\theta) < R_{J(\theta)}(\theta) \quad \text{w.p.1}. \tag{2.10}$$

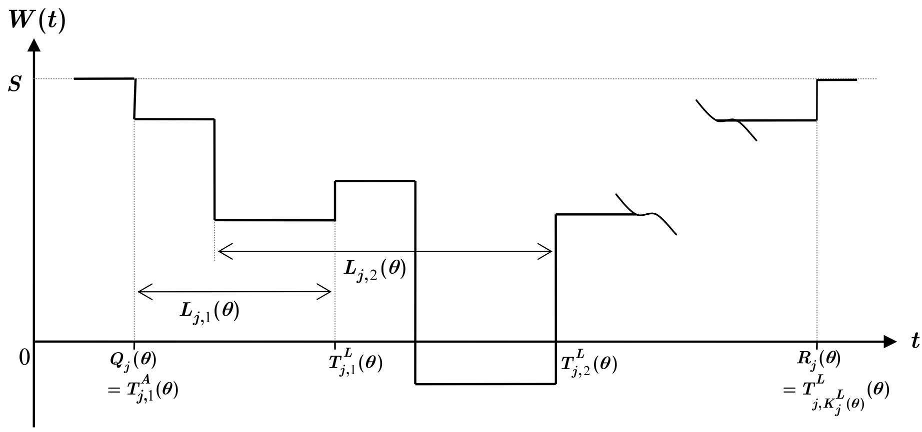

Fig. 2 Sample path of a discrete MTS system

For $j = 0, 1, \ldots, J(\theta)$, let $K_j^L(\theta)$ be the number of replenishments in the interval $(Q_j(\theta), R_j(\theta))$, and for each replenishment $k = 1, \ldots, K_j^L(\theta)$ in the interval $(Q_j(\theta), R_j(\theta))$, let $L_{j,k}(\theta)$ be the k-th replenishment's lead time, $T_{j,k}^L(\theta)$ its arrival time, and $V_{j,k}(\theta)$ its size. The enumerations of $L_{j,k}$ and L_i correspond to different indexing schemes of lead times, such that if $L_i = L_{j,k}$, then

$$i = i(j,k) = \sum_{m=0}^{j-1} K_m^L(\theta) + k. \tag{2.11}$$

The indexing relationship of (2.11) also applies to the corresponding replenishment arrival times (T_i^L and $T_{j,k}^L$) and replenishment sizes (V_i and $V_{j,k}$). Let $K^L(\theta) = \sum_{j=0}^{J(\theta)} K_j^L(\theta)$ be the total number of replenishments in $[0, T]$, the last one being $V_{K^L(\theta)}(\theta)$. In a similar vein, let $T_{j,k}^A(\theta)$, $j = 0, \ldots, J(\theta)$, $k = 1, \ldots, K_j^A(\theta)$, denote the arrival time of the k-th demand in the interval $[Q_j(\theta),\ R_j(\theta))$. Let $K^A(\theta) = \sum_{j=0}^{J(\theta)} K_j^A(\theta)$ be the total number of orders in $[0, T]$, the last one being $D_{K^A(\theta)}(\theta)$. A typical sample path of the discrete MTS is illustrated in Fig. 2.

3 Performance metrics and parameters

Let $[0, T]$ be a finite time interval for some prescribed time T, and let $\theta \in \Theta$ denote a generic parameter of interest from a closed and bounded domain Θ. We shall be interested in the following performance metrics:

- *Inventory time average.* The time average of the volume of inventory on-hand over the interval $[0, T]$, given by

$$M_I(T, \theta) = \frac{1}{T} \int_0^T I(t, \theta)\, dt. \tag{3.1}$$

  We point out that this metric can be used to compute inventory holding costs over the interval $[0, T]$, often computed as $h \int_0^T I(t, \theta)\, dt$, where h is the inventory holding cost coefficient per inventory item per unit time.

- *Backorder time average.* The time average of the volume of backorders over the interval $[0, T]$, given by

$$M_B(T,\theta) = \frac{1}{T}\int_0^T B(t,\theta)dt. \tag{3.2}$$

We point out that this metric can be used to compute backorder penalties over the interval $[0, T]$, similarly to the previous metric.

The IPA parameters of interest are:

- *Base stock level parameter.* The base stock level of inventory, $S(\theta) = \theta$.
- *Lead time parameter.* A parameter θ of the lead-time process, $\{L_i(\theta)\}$.

In this paper, the functional form of the lead time derivatives $\frac{d}{d\theta}L_i(\theta) = L_i'(\theta)$ is provided by the modeler. Three typical choices are presented in (3.3), (3.4) and (3.6) in the next three subsections: concurrent, sequential and inverse quadratic.

3.1 Concurrent lead time derivatives

A *concurrent lead time derivative* has the form

$$L_i'(\theta) = 1, \quad i = 1, \ldots, K^L(\theta). \tag{3.3}$$

In this case, lead times are modeled as service times in an infinite server group, so that multiple replenishments progress concurrently (in parallel). Consequently, all lead times are varied linearly as function of θ. Note that in this case, replenishment completions do not necessarily occur in the order that their corresponding orders were placed.

3.2 Sequential lead time derivatives

A *sequential lead time derivative* has the form

$$L_{j,k}'(\theta) = 1_{\{1\}}(j)1_{\{K_0^L(\theta)<N_0\}}K_0^L(\theta) + k, \quad j = 0, \ldots, J(\theta),\ k = 1, \ldots, K_j^L(\theta). \tag{3.4}$$

In this case, lead times are modeled as sojourn times in a single-server FCFS queue as follows:

- The lead time of an order consists of a waiting time in the queue (if any), W_i, plus a manufacturing time, S_i, when the order reaches the head of the queue (there is no transportation time).
- $\frac{d}{d\theta}S_i(\theta) = 1$, for all $i \ge 1$, that is, the manufacturing times, $S_i(\theta)$, are varied linearly as function of θ.

To see that these assumptions imply (3.4), observe the following:

1. Over the intervals $[0, T^L_{0,K_0^L(\theta)}(\theta))$ and $[Q_j(\theta), R_j(\theta)]$, $j = 1, \ldots, J(\theta)$, the production queue is always busy.
2. Over the interval $[T^L_{0,K_0^L(\theta)}(\theta), Q_1(\theta)]$, the production queue is busy on the event $\{K_0^L(\theta) < N_0\}$, and empty on the event $\{K_0^L(\theta) = N_0\}$.
3. Over the intervals $[R_j(\theta), Q_j(\theta)]$, $j = 1, \ldots, J(\theta)$, the production queue is empty.

Next, let A_i be the interarrival time between orders i and $i+1$ at the production queue (it is also the interarrival time between demands at the inventory facility). Then by Lindley's Integral Equation (Kleinrock 1975, p. 276), $W_1(\theta)=0$ is locally independent of θ, and while the queue is busy,

$$W_{i+1}(\theta)=W_i(\theta)+S_i(\theta)-A_{i+1}, \quad i>1. \tag{3.5}$$

Finally, differentiating the above equation and noting that $\frac{d}{d\theta}S_i(\theta)=1$ by assumption, yields (3.4) via a straightforward induction.

3.3 Inverse quadratic lead time derivatives

An *inverse quadratic lead time derivative* has the form

$$L'_i(\theta)=-\frac{V_i}{\theta^2}, \quad i=1,\dots,K^L(\theta). \tag{3.6}$$

As will be shown, this definition provides a bridge between the lead time and a corresponding replenishment rate (it is often more convenient to design an MTS system in terms of replenishment rates than lead times, as they often stand for production capacity). To this end, note that if $r(t,\theta)$ is the instantaneous replenishment rate in an SFM, then the following relation must hold

$$\int_{T_i^L(\theta)-L_i(\theta)}^{T_i^L(\theta)} r(t,\theta)dt=V_i. \tag{3.7}$$

Next, define the *imputed replenishment rate* during the i-th lead time by

$$r_i(\theta)=\frac{V_i}{L_i(\theta)}, \quad i=1,\dots,K^L(\theta). \tag{3.8}$$

To justify this definition, note that on substituting the $r_i(\theta)$ for $r(t,\theta)$ in (3.7), that equation holds trivially. Furthermore, the replenishment rate $r(t,\theta)$ is rarely known in practice, while the imputed replenishment rates $r_i(\theta)$ are always computable from observed data. Thus, the $r_i(\theta)$ can be used as proxies for the $r(t,\theta)$ over the i-th lead time. Finally, if we wish to vary $r_i(\theta)$ linearly in θ, that is, $r_i(\theta)=\theta$, then (3.6) readily follows from (3.8) by differentiation.

4 IPA derivatives

To obtain the requisite derivatives, we shall analyze pairs of systems: the original discrete MTS system (as function of θ) and a perturbed discrete MTS system (as function of $\theta\pm\Delta\theta$), both starting from the same initial state. Our objective in this section is first to derive formulas for the IPA derivatives and then to prove them unbiased.

4.1 IPA derivatives with respect to the base-stock level parameter

To obtain the IPA derivatives for $M_I(T,\theta)$ and $M_B(T,\theta)$ with respect to the base stock level parameter, θ, we make the following assumptions.

Assumption 3

(a) $S(\theta)=\theta$, where $\theta\in\Theta$.
(b) The demand arrival-time process, $\{T_i^A\}$, and the demand-size process, $\{D_i\}$, are independent of θ.

(c) The replenishment lead-time process $\{L_i\}$ is independent of θ.
(d) The initial random variables $W(0)$, $P_-(0)$, and $O_1, \ldots, O_{N_0}$ are independent of θ.
(e) As the order size tends to zero, so does the associated lead time.

Assumption (e) above is a technical condition designed to remove a practical difficulty in the IPA derivative with respect to S. Specifically, suppose that the initial position is S or an arriving demand lowers the inventory position from overage to exactly S. Then when the base-stock level parameter is perturbed from a reference level S (in a reference MTS system) to $S + \Delta S$ (in a perturbed MTS system), a new (infinitesimal) order of size ΔS and associated lead time will be created in the perturbed system, which do not exist in the reference system. Since we observe only the reference MTS system, while its perturbed counterparts are mathematical constructs without physical reality, there is no way to observe the lead time associated with such a new order. Unfortunately, unless the new (unknown) lead time is also infinitesimal, it would appear in the IPA derivative. Again, infinitesimal orders are mathematical constructs that do not occur in reality. Assumption (e) above effectively serves to remove such vanishingly small orders (along with their limiting lead times) from consideration in a practically reasonable manner. Consequently, these infinitesimal random variables can be excluded from consideration in the corresponding processes associated with the reference MTS system. For additional discussion, refer to Sect. 5.

In this section we make use of the hitting time, $T_S(\theta)$, defined by

$$T_S(\theta) = \begin{cases} \min\{t \in [0, Q_1(\theta)) : P(0) \geq S, P(t,\theta) = S(\theta)\}, & \text{if the minimum exists,} \\ \infty, & \text{otherwise.} \end{cases} \tag{4.1}$$

Thus, $T_S(\theta)$, when it exists, is the first time prior to $Q_1(\theta)$ that the inventory position process reaches the base stock level from above.

Proposition 1 *W.p.*1,

(a) K^A, $\{T_i^L\}$ *and* K^L *are independent of* θ.
(b) $\{O_i\}$ *is independent of* θ *except for* $O_{N_0+1}(\theta)$ (*placed either at time* 0 *or at time* T_H^A). *Furthermore*,

$$\frac{d}{d\theta^-} O_{N_0+1}(\theta) = 1, \tag{4.2}$$

$$\frac{d}{d\theta^+} O_{N_0+1}(\theta) = \begin{cases} 0, & \text{on } \{T_S(\theta) < \infty\}, \\ 1, & \text{on } \{T_S(\theta) = \infty\}. \end{cases} \tag{4.3}$$

(c) $\{V_i\}$ *is independent of* θ *except for* $V_{C_{N_0+1}}(\theta)$. *Furthermore*,

$$\frac{d}{d\theta^-} V_{C_{N_0+1}}(\theta) = 1, \tag{4.4}$$

$$\frac{d}{d\theta^+} V_{C_{N_0+1}}(\theta) = \begin{cases} 0, & \text{on } \{T_S(\theta) < \infty\}, \\ 1, & \text{on } \{T_S(\theta) = \infty\}. \end{cases} \tag{4.5}$$

Proof Part (a) follows immediately from parts (b) and (c) of Assumption 3 and the relation $T_i^L = T_i^A + L_i$.

To prove part (b), note that $\{O_i\}$ is independent of θ except for $O_{N_0+1}(\theta)$, because the first N_0 orders, $O_1, \ldots, O_{N_0}$, are independent of θ by part (d) of Assumption 3, while the orders O_i, $i > N_0 + 1$, are independent of θ by (2.9) and part (b) of Assumption 3. We next consider the order $O_{N_0+1}(\theta)$. In the perturbed MTS system corresponding to $\theta + \Delta\theta$, we have $O_{N_0+1}(\theta + \Delta\theta) = D_H$ on the event $\{T_S(\theta) < \infty\}$, since the inventory position will immediately increase from θ to $\theta + \Delta\theta$ at $T_S(\theta)$ by part (e) of Assumption 3. This implies the first case of (4.3) since $\{D_i\}$ is independent of θ by part (b) of Assumption 3. Next, observe that $P_-(0)$ is independent of θ by part (d) of Assumption 3, while $P_-(T_H^A) = P_-(0) - \sum_{i=1}^{H} D_i + \sum_{j=1}^{K^L} 1_{\{T_j^L \le T_H^A\}} V_j$ is independent of θ because the right-hand side is independent of θ by part (b) of Assumption 3 and part (a) above, and noting that $\{V_i\}$ is independent of θ on the events $\{T_j^L \le T_H^A\}$ by part (d) of Assumption 3. Equation (4.2) and the second case of (4.3) now follow from (2.7) and (2.8).

Finally, part (c) is implied by part (b) above and (2.1). □

Lemma 1 *Consider a discrete MTS system with backorders under the base-stock policy, subject to Assumptions 1, 2 and 3. Then for any* $\theta \in \Theta$, $0 < T < \infty$, $t \in [0, T]$,

(a) *On the event* $A(\theta) = \{T_S(\theta) = \infty\} \cap \{0 < t < T^L_{C_{N_0+1}}(\theta)\}$,

$$\frac{\partial}{\partial\theta} I(t,\theta) = 0, \tag{4.6}$$

$$\frac{\partial}{\partial\theta} B(t,\theta) = 0. \tag{4.7}$$

(b) *On the event* $B(\theta) = \{T_S(\theta) = \infty\} \cap \{t > T^L_{C_{N_0+1}}(\theta)\} \cap \{I(t,\theta) > 0\}$,

$$\frac{\partial}{\partial\theta} I(t,\theta) = 1, \tag{4.8}$$

$$\frac{\partial}{\partial\theta} B(t,\theta) = 0. \tag{4.9}$$

(c) *On the event* $C(\theta) = \{T_S(\theta) = \infty\} \cap \{t > T^L_{C_{N_0+1}}(\theta)\} \cap \{B(t,\theta) > 0\}$,

$$\frac{\partial}{\partial\theta} I(t,\theta) = 0, \tag{4.10}$$

$$\frac{\partial}{\partial\theta} B(t,\theta) = -1. \tag{4.11}$$

(d) *On the event* $D(\theta) = \{T_S(\theta) = \infty\} \cap \{t > T^L_{C_{N_0+1}}(\theta)\} \cap \{I(t,\theta) = B(t,\theta) = 0\}$,

$$\frac{\partial}{\partial\theta^+} I(t,\theta) = 1, \tag{4.12}$$

$$\frac{\partial}{\partial\theta^+} B(t,\theta) = 0, \tag{4.13}$$

and

$$\frac{\partial}{\partial\theta^-} I(t,\theta) = 0, \tag{4.14}$$

$$\frac{\partial}{\partial\theta^-} B(t,\theta) = -1. \tag{4.15}$$

(e) *On the event* $E(\theta) = \{T_S(\theta) < \infty\} \cap \{0 < t < T_S(\theta)\}$,

$$\frac{\partial}{\partial\theta} I(t,\theta) = 0, \tag{4.16}$$

$$\frac{\partial}{\partial\theta} B(t,\theta) = 0. \tag{4.17}$$

(f) *On the event* $F(\theta) = \{T_S(\theta) < \infty\} \cap \{t > T_S(\theta)\} \cap \{B(t,\theta) = 0\}$,

$$\frac{\partial}{\partial\theta^+} I(t,\theta) = 1, \tag{4.18}$$

$$\frac{\partial}{\partial\theta^+} B(t,\theta) = 0. \tag{4.19}$$

(g) *On the event* $G(\theta) = \{T_S(\theta) < \infty\} \cap \{t > T_S(\theta)\} \cap \{B(t,\theta) > 0\}$,

$$\frac{\partial}{\partial\theta^+} I(t,\theta) = 0, \tag{4.20}$$

$$\frac{\partial}{\partial\theta^+} B(t,\theta) = -1. \tag{4.21}$$

(h) *On the event* $H(\theta) = \{T_S(\theta) < \infty\} \cap \{T_S(\theta) < t < T^L_{C_{N_0+1}}(\theta)\}$,

$$\frac{\partial}{\partial\theta^-} I(t,\theta) = 0, \tag{4.22}$$

$$\frac{\partial}{\partial\theta^-} B(t,\theta) = 0. \tag{4.23}$$

(i) *On the event* $I(\theta) = \{T_S(\theta) < \infty\} \cap \{t > T^L_{C_{N_0+1}}(\theta)\} \cap \{I(t,\theta) > 0\}$,

$$\frac{\partial}{\partial\theta^-} I(t,\theta) = 1, \tag{4.24}$$

$$\frac{\partial}{\partial\theta^-} B(t,\theta) = 0. \tag{4.25}$$

(j) *On the event* $J(\theta) = \{T_S(\theta) < \infty\} \cap \{t > T^L_{C_{N_0+1}}(\theta)\} \cap \{I(t,\theta) = 0\}$,

$$\frac{\partial}{\partial\theta^-} I(t,\theta) = 0, \tag{4.26}$$

$$\frac{\partial}{\partial\theta^-} B(t,\theta) = -1. \tag{4.27}$$

Proof To prove part (a), note that on $\{0 < t < T^L_{C_{N_0+1}}(\theta)\}$,

$$W(t,\theta) = W(0) - \sum_{i=1}^{K^A} 1_{\{T_i^A \le t\}} D_i + \sum_{j=1}^{K^L} 1_{\{T_j^L \le t\}} V_j. \tag{4.28}$$

It follows that $W(t,\theta)$ is independent of θ on $A(\theta)$, because $\{T_i^A\}$, $\{D_i\}$ are independent of θ by part (b) of Assumption 3 and $W(0)$ is independent of θ by part (d) of Assumption 3, while $\{T_i^L\}$ is independent of θ by part (a) of Proposition 1, and $\{V_i\}$ is independent of θ on $\{0 < t < T^L_{C_{N_0+1}}(\theta)\}$ by part (c) of Proposition 1. By (2.6), both $I(t,\theta)$ and $B(t,\theta)$ are also independent of θ on $A(\theta)$, implying (4.6) and (4.7).

To prove parts (b), (c) and (d), note that on $\{t > T^L_{C_{N_0+1}}(\theta)\}$,

$$W(t,\theta) = W(T^L_{C_{N_0+1}}(\theta),\theta) - \sum_{i=1}^{K^A} 1_{\{T^L_{C_{N_0+1}}(\theta) < T_i^A \le t\}} D_i + \sum_{j=1}^{K^L} 1_{\{T^L_{C_{N_0+1}}(\theta) < T_j^L \le t\}} V_j, \tag{4.29}$$

where $W(T^L_{C_{N_0+1}}(\theta),\theta) = W(T^L_{C_{N_0+1}}(\theta)-,\theta) + V_{C_{N_0+1}}(\theta)$. Differentiating the previous equation on $B(\theta), C(\theta)$ and $D(\theta)$ yields

$$\frac{d}{d\theta} W(T^L_{C_{N_0+1}}(\theta),\theta) = 1, \tag{4.30}$$

because $W(T^L_{C_{N_0+1}}(\theta)-,\theta)$ is independent of θ by the proof of part (a), and $\frac{d}{d\theta} V_{C_{N_0+1}}(\theta) = 1$ by (4.4) and (4.5). Differentiating (4.29) with respect to θ with the aid of (4.30) we conclude that $\frac{\partial}{\partial\theta} W(t,\theta) = 1$ on $B(\theta), C(\theta)$ and $D(\theta)$, because $\{T_i^A\}$ and $\{D_i\}$, are independent of θ by part (b) of Assumption 3, while $\{T_i^L\}$ is independent of θ by part (a) of Proposition 1, and $\{V_i\}$ is independent of θ on $\{t > T^L_{C_{N_0+1}}(\theta)\}$ by part (c) of Proposition 1. Parts (b), (c) and (d) then follow from (2.6). Note that the right and left derivatives differ on the event $D(\theta)$, since an increase in θ changes the inventory level from 0 to a positive value, while the backorder level remains 0, and a decrease in θ changes the backorder level from 0 to a positive value, while the inventory level remains 0.

To prove part (e), note that (4.28) holds on $E(\theta)$, and the rest of the proof of this part is similar to that of part (a).

To prove parts (f), and (g), note that on $\{t > T_S(\theta)\}$

$$W(t,\theta) = W(T_S(\theta),\theta) - \sum_{i=1}^{K^A} 1_{\{T_S(\theta) < T_i^A \le t\}} D_i + \sum_{j=1}^{K^L} 1_{\{T_S(\theta) < T_j^L \le t\}} V_j, \tag{4.31}$$

where on $\{T_S(\theta) < \infty\}$, we have

$$W(T_S(\theta),\theta) = P(T_S(\theta),\theta) - \sum_{i=1}^{N_0} O_i + \sum_{j=1}^{K^L} 1_{\{T_j^L < T_S(\theta)\}} V_j. \tag{4.32}$$

Differentiating (4.32) on $F(\theta)$ and $G(\theta)$ yields

$$\frac{d}{d\theta} W(T_S(\theta),\theta) = 1, \tag{4.33}$$

because $P(T_S(\theta),\theta) = S(\theta) = \theta$ by definition, $O_1,\ldots,O_{N_0}$ are independent of θ by part (d) of Assumption 3, while $\{T_i^L\}$ is independent of θ by part (a) of Proposition 1 and $\{V_i\}$ is independent of θ on $\{t < T_S(\theta)\}$ by part (c) of Proposition 1. Differentiating (4.31) with respect to θ with the aid of (4.33), we conclude that $\frac{\partial}{\partial\theta} W(t,\theta) = 1$ on $F(\theta)$ and $G(\theta)$. The rest of the proof of this part is similar to that of parts (b), (c) and (d).

To prove part (h), note that (4.28) holds on $H(\theta)$. The rest of the proof of this part is similar to that of part (a).

To prove parts (i) and (j), note that (4.29) holds on $I(\theta)$ and $J(\theta)$. The rest of the proof of this part is similar to that of parts (b), (c) and (d). □

Theorem 1 *Consider a discrete MTS system with backorders under the base-stock policy, subject to Assumptions 1, 2 and 3. Then for any $\theta \in \Theta$, $0 < T < \infty$, the IPA derivatives of the inventory time average and the backorder time average with respect to the base stock level parameter are given by*

$$\frac{\partial}{\partial \theta^-} M_I(T,\theta) = 1_{\{T^L_{C_{N_0+1}}(\theta) < T\}} \frac{1}{T} \int_{T^L_{C_{N_0+1}}(\theta)}^{T} 1_{\{I(t,\theta)>0\}} dt, \tag{4.34}$$

$$\begin{aligned} \frac{\partial}{\partial \theta^+} M_I(T,\theta) = {}& 1_{\{T_S(\theta) < T\}} \frac{1}{T} \int_{T_S}^{\min\{T^L_{C_{N_0+1}}(\theta), T\}} 1_{\{B(t,\theta)=0\}} dt \\ & + 1_{\{T^L_{C_{N_0+1}}(\theta) < T\}} \frac{1}{T} \int_{T^L_{C_{N_0+1}}(\theta)}^{T} 1_{\{B(t,\theta)=0\}} dt, \end{aligned} \tag{4.35}$$

and

$$\frac{\partial}{\partial \theta^-} M_B(T,\theta) = -1_{\{T^L_{C_{N_0+1}}(\theta) < T\}} \frac{1}{T} \int_{T^L_{C_{N_0+1}}(\theta)}^{T} 1_{\{I(t,\theta)=0\}} dt, \tag{4.36}$$

$$\begin{aligned} \frac{\partial}{\partial \theta^+} M_B(T,\theta) = {}& -1_{\{T_S(\theta) < T\}} \frac{1}{T} \int_{T_S}^{\min\{T^L_{C_{N_0+1}}(\theta), T\}} 1_{\{B(t,\theta)>0\}} dt \\ & - 1_{\{T^L_{C_{N_0+1}}(\theta) < T\}} \frac{1}{T} \int_{T^L_{C_{N_0+1}}(\theta)}^{T} 1_{\{B(t,\theta)>0\}} dt. \end{aligned} \tag{4.37}$$

Furthermore, the IPA derivatives are unbiased for each finite $T > 0$ and every $\theta \in \Theta$.

Proof Applying the Leibniz integral rule to (3.1) and (3.2) yields

$$\frac{\partial}{\partial \theta} M_I(T,\theta) = \frac{1}{T}\frac{d}{d\theta}\int_0^T I(t,\theta)dt = \frac{1}{T}\int_0^T \frac{\partial}{\partial \theta} I(t,\theta)dt, \tag{4.38}$$

$$\frac{\partial}{\partial \theta} M_B(T,\theta) = \frac{1}{T}\frac{d}{d\theta}\int_0^T B(t,\theta)dt = \frac{1}{T}\int_0^T \frac{\partial}{\partial \theta} B(t,\theta)dt. \tag{4.39}$$

To see that, note that the end-points of the interval $[0, T]$ do not depend on θ and the hitting times, $\{T_i^A\}$ and $\{T_i^L\}$, are independent of θ by part (b) of Assumption 3 and part (a) of Proposition 1, respectively, so (4.38) and (4.39) follow readily from Fact 2. Equations (4.34), (4.35), (4.36) and (4.37) now follow by substituting the derivatives computed in Lemma 1 into (4.38) and (4.39).

We next prove that the IPA derivatives are unbiased using Fact 1. The proofs of (4.34), (4.35), (4.36) and (4.37) establish that Condition (a) of Fact 1 is satisfied for the sided derivatives of $M_I(T,\theta)$ and $M_B(T,\theta)$ for each finite $T > 0$. Next, note that (4.34), (4.35), (4.36) and (4.37) are each bounded uniformly by 1 in T and θ w.p.1. Since any differentiable function with such a bounded derivative is Lipschitz continuous and its Lipschitz constant trivially has a finite first moment, Condition (b) of Fact 1 holds, thereby completing the proof. □

4.2 IPA derivatives with respect to a lead time parameter

In this section we derive the IPA derivatives for $M_I(T,\theta)$ and $M_B(T,\theta)$ with respect to a lead time parameter, θ. We make the following assumptions.

Assumption 4

(a) The derivatives $\frac{d}{d\theta}L_i(\theta) = L_i'(\theta)$, $i = 1, \ldots, K^L(\theta)$, $\theta \in \Theta$, are given.
(b) W.p.1, $|L_i'(\theta)| \le L'$, where L' is a positive deterministic constant, independent of $\theta \in \Theta$ and $i \ge 1$.
(c) The base stock level, S, the demand arrival-time process $\{T_i^A\}$, and the demand-size process $\{D_i\}$ are all independent of θ.
(d) The initial random variables $W(0)$, $P_-(0)$, and $O_1, \ldots, O_{N_0}$ are independent of θ.

Observation 1

(a) By parts (c) and (d) of Assumption 4, the order-size process $\{O_i\}$ and the replenishment-size process $\{V_i\}$ are independent of θ.
(b) W.p.1, $K^L(\theta)$ is locally independent of θ.
(c) By part (c) of Assumption 4 and the relation $T_i^L(\theta) = T_i^A + L_i(\theta)$, we have

$$\frac{d}{d\theta}T_i^L(\theta) = \frac{d}{d\theta}L_i(\theta) = L_i'(\theta). \tag{4.40}$$

Lemma 2 *Consider a discrete MTS system with backorders under the base-stock policy, subject to Assumptions 1, 2 and 4. Then for any $\theta \in \Theta, 0 < T < \infty$, and almost every $t \in [0,T]$, one has w.p.1,*

$$\frac{\partial}{\partial\theta}I(t,\theta) = \frac{\partial}{\partial\theta}B(t,\theta) = 0. \tag{4.41}$$

Proof Let $\{\Gamma_n(\theta) : 0 \le n \le K^A + K^L(\theta) + 2\}$ be an increasing sequence of hitting times, where $\Gamma_0 = 0$, $\Gamma_{K^A+K^L(\theta)+2} = T$, and for $0 < n < K^A + K^L(\theta) + 2$, $\Gamma_n(\theta)$ is of the form T_i^A or $T_j^L(\theta)$. Next, write

$$W(t,\theta) = W(0) - \sum_{i=1}^{K^A} 1_{\{0<T_i^A<t\}} D_i + \sum_{j=1}^{K^L(\theta)} 1_{\{0<T_j^L(\theta)<t\}} V_j, \quad t \in [0,T]. \tag{4.42}$$

Equation (4.42) implies that $\{W(t,\theta)\}$ is locally independent of θ over all open time intervals of the form $(\Gamma_n(\theta), \Gamma_{n+1}(\theta))$. Furthermore, the same holds for $\{I(t,\theta)\}$ and $\{B(t,\theta)\}$, since (2.6) implies that $I(t,\theta) = W^+(t,\theta)$ and $B(t,\theta) = W^-(t,\theta)$. This completes the proof of the lemma. □

Theorem 2 *Consider a discrete MTS system with backorders under the base-stock policy, subject to Assumptions 1, 2 and 4. Then for any $\theta \in \Theta$, $0 < T < \infty$, the IPA derivatives of the inventory time average and the backorder time average with respect to a lead time parameter are given by*

$$\frac{\partial}{\partial\theta}M_I(T,\theta) = -\frac{1}{T}\sum_{i=1}^{K^L(\theta)} \min\{V_i, I(T_i^L(\theta),\theta)\}L_i'(\theta), \tag{4.43}$$

$$\frac{\partial}{\partial\theta}M_B(T,\theta)=\frac{1}{T}\sum_{i=1}^{K^L(\theta)}\max\{0, V_i-I(T_i^L(\theta),\theta)\}L_i'(\theta). \tag{4.44}$$

Furthermore, the IPA derivatives are unbiased for each finite $T>0$ and every $\theta\in\Theta$.

Proof To prove (4.43) and (4.44), we use Fact 2 in conjunction with (4.40)–(4.42) to deduce

$$\begin{aligned}\frac{\partial}{\partial\theta}\int_0^T W(t,\theta)dt &= \sum_{i=1}^{K^L(\theta)}\left[W(T_i^L(\theta)-,\theta)\frac{d}{d\theta}T_i^L(\theta)-W(T_i^L(\theta)+,\theta)\frac{d}{d\theta}T_i^L(\theta)\right]\\ &= \sum_{i=1}^{K^L(\theta)}[-V_iL_i'(\theta)].\end{aligned} \tag{4.45}$$

Next by (2.6), we have

$$\frac{d}{d\theta}\int_0^T I(t,\theta)dt=-\sum_{i=1}^{K^L(\theta)}\min\{V_i, I(T_i^L(\theta),\theta)\}L_i'(\theta), \tag{4.46}$$

$$\frac{d}{d\theta}\int_0^T B(t,\theta)dt=\sum_{i=1}^{K^L(\theta)}\max\{0, V_i-I(T_i^L(\theta),\theta)\}L_i'(\theta). \tag{4.47}$$

Equations (4.43) and (4.44) now follow from (3.1) and (3.2).

We next prove that the IPA derivatives are unbiased using Fact 1. The proofs of (4.43) and (4.44) establish that Condition (a) of Fact 1 is satisfied for both $M_I(T,\theta)$ and $M_B(T,\theta)$ for each finite $T>0$ and every $\theta\in\Theta$. It remains to show that these equations are uniformly bounded w.p.1. By part (b) of Assumption 2, $K^L(\theta)\le K^*(T)$ holds w.p.1 for each finite $T>0$, and every $\theta\in\Theta$. Applying this bound in conjunction with the bounds $V_i\le O^*$ from part (a) of Assumption 2 and $|L_i'(\theta)|\le L'$ from part (b) of Assumption 4 to (4.43) and (4.44), we obtain w.p.1,

$$\left|\frac{\partial}{\partial\theta}M_I(T,\theta)\right|,\left|\frac{\partial}{\partial\theta}M_I(T,\theta)\right|\le\frac{O^*L'K^*(T)}{T}<\infty,$$

where the right-hand side is independent of θ w.p.1, for each finite $T>0$. Since any differentiable function with such a bounded derivative is Lipschitz continuous and its Lipschitz constant trivially has a finite first moment, it follows that Condition (b) of Fact 1 holds for both $M_I(T,\theta)$ and $M_B(T,\theta)$, thereby completing the proof. □

Recall that the derivatives $L_i'(\theta)$ are provided as input parameters at the modeler's discretion. We next consider three special cases of particular interest (see Sects. 3.1–3.3).

Case 1: For concurrent lead times (see Sect. 3.1), the IPA derivatives in (4.43) and (4.44) become, in view of (3.3),

$$\frac{\partial}{\partial\theta}M_I(T,\theta)=-\frac{1}{T}\sum_{i=1}^{K^L(\theta)}\min\{V_i, I(T_i^L(\theta),\theta)\} \tag{4.48}$$

and

$$\frac{\partial}{\partial\theta}M_B(T,\theta)=\frac{1}{T}\sum_{i=1}^{K^L(\theta)}\max\{0,\,V_i-I(T_i^L(\theta),\theta)\}. \tag{4.49}$$

Case 2: For sequential lead times (see Sect. 3.2), the IPA derivatives in (4.43) and (4.44) become, in view of (3.4),

$$\frac{\partial}{\partial\theta}M_I(T,\theta)=-\frac{1}{T}\sum_{j=0}^{J(\theta)}\sum_{k=1}^{K_j^L(\theta)}\min\{V_{j,k},\,I(T_{j,k}^L(\theta),\theta)\}[1_{\{1\}}(j)1_{\{K_0^L(\theta)<N_0\}}K_0^L(\theta)+k] \tag{4.50}$$

and

$$\frac{\partial}{\partial\theta}M_B(T,\theta)=\frac{1}{T}\sum_{j=0}^{J(\theta)}\sum_{k=1}^{K_j^L(\theta)}\max\{0,\,V_{j,k}-I(T_{j,k}^L(\theta),\theta)\}[1_{\{1\}}(j)1_{\{K_0^L(\theta)<N_0\}}K_0^L(\theta)+k]. \tag{4.51}$$

Case 3: For inverse quadratic lead times (see Sect. 3.3), the IPA derivatives in (4.43) and (4.44) become, in view of (3.6),

$$\frac{\partial}{\partial\theta}M_I(T,\theta)=\frac{1}{T}\sum_{k=1}^{K^L(\theta)}\min\{V_i,\,I(T_i^L(\theta),\theta)\}\frac{V_i}{\theta^2} \tag{4.52}$$

and

$$\frac{\partial}{\partial\theta}M_B(T,\theta)=-\frac{1}{T}\sum_{i=1}^{K^L(\theta)}\max\{0,\,V_i-I(T_i^L(\theta),\theta)\}\frac{V_i}{\theta^2}. \tag{4.53}$$

5 Conclusion

This paper treats a discrete model of an MTS system under the continuous-review base-stock policy. It derives the IPA derivatives formulas of time average of inventory on hand and time average of backorders with respect to the base-stock level and a lead-time process parameter under any initial inventory state. Furthermore, the IPA derivatives are shown to be unbiased and easily computable.

In addition to deriving the aforementioned IPA derivatives, the paper discusses the advantages and shortcomings that are traded off when selecting a discrete model or an SFM counterpart. The choice of a discrete model or SFM for an MTS system is motivated by a fundamental tradeoff, stemming from the tendency of IPA derivatives to have singularities (and bias) in discrete models, while their SFM counterparts often require the estimation of unknown instantaneous rates at specific times. Rate estimation renders the corresponding IPA formulas in SFM versions computationally brittle, whereas their discrete-model counterparts are readily computable from sample path observables. The general recommendation is to use a discrete model to the extent possible (that is, provided the model has IPA or sided IPA derivatives that are unbiased), since those have readily computable IPA derivatives and do not require the estimation of specific instantaneous rates.

Finally, part (e) in Assumption 3 is an essential technical condition that cannot be replaced by more realistic assumptions without introducing singularities into some IPA derivatives. As a case in point, consider replacing part (e) above by imposing instead a minimal order size, o^*, in which case an order $O_i(\theta)$ is placed only when $O_i(\theta) \geq o^*$. However, the minimal-order assumption introduces a singularity into the resultant IPA derivatives with respect to the base-stock level, S, on an event of possibly positive probability (those with respect to a lead time parameter do not change). To see that, note that on the events $\{S - P_-(0) = o^*\}$ or $\{S - P_-(T_H^A) = o^*\}$ (to be referred to as *singularity events*), perturbing S to $S - \Delta S$, would eliminate an order from the order stream at time 0 or T_H^A with the next order picking up the slack. Since this is the case for all sufficiently small $\Delta S > 0$, the finite differences of the time-average metrics under study diverge as $\Delta S \downarrow 0$. This phenomenon is due to the fact that the minimum-order policy is inherently "discontinuous" in S, and the resulting singularities are irremovable. For the details of the attendant IPA derivative formulas (which on non-singularity events are similar to the ones developed in this paper), see Fan (2008).

References

Cassandras, C. G., & Lafortune, S. (1999). *Introduction to discrete event systems*. Boston: Kluwer Academic.

Cassandras, C. G., Wardi, Y., Melamed, B., Sun, G., & Panayiotou, C. G. (2002). Perturbation analysis for on-line control and optimization of stochastic fluid models. *IEEE Transactions on Automatic Control*, *47*, 1234–1248.

Cassandras, C. G., Sun, G., Panayiotou, C., & Wardi, Y. (2003). Perturbation analysis and control of two-class stochastic fluid models for communication networks. *IEEE Transactions on Automatic Control*, *48*(5), 770–782.

Fan, Y. (2008). Stochastic models for make-to-stock production-inventory systems. PhD Dissertation, Rutgers Business School—Newark and New Brunswick, Newark, New Jersey.

Fan, Y., Melamed, B., Zhao, Y., & Wardi, Y. (2009). IPA derivatives for make-to-stock production-inventory systems with backorders under the (R, r) policy. *Methodology and Computing in Applied Probability*, *11*(2), 159–179.

Fu, M. C., & Hu, J. Q. (1997). *Conditional Monte Carlo: gradient estimation and optimization applications*. Boston: Kluwer Academic.

Heidelberger, H., Cao, X. R., Zazanis, M., & Suri, R. (1988). Convergence properties of infinitesimal analysis estimates. *Management Science*, *34*, 1281–1302.

Kleinrock, L. (1975). *Queueing systems, Vol. 1: Theory*. New York: Wiley.

Panayiotou, C., & Cassandras, C. G. (2006). Infinitesimal perturbation analysis and optimization for make-to-stock manufacturing systems based on stochastic fluid models. *Discrete Event Dynamic Systems*, *16*(1), 109–142.

Paschalidis, I. C., Liu, Y., Cassandras, C. G., & Panayiotou, C. G. (2004). Inventory control for supply chains with service level constraints: a synergy between large deviations and perturbation analysis. *Annals of Operations Research*, *126*, 231–258 (Special Volume on Stochastic Models of Production-Inventory Systems).

Rubinstein, R. Y., & Shapiro, A. (1993). *Discrete event systems: sensitivity analysis and stochastic optimization by the score function method*. New York: Wiley.

Wardi, Y., Melamed, B., Cassandras, C. G., & Panayiotou, C. G. (2002). On-line IPA gradient estimators in stochastic continuous fluid models. *Journal of Optimization Theory and Applications*, *115*(2), 369–405.

Zhao, Y., & Melamed, B. (2006). IPA gradients for make-to-stock production-inventory systems with backorders. *Methodology and Computing in Applied Probability*, *8*, 191–222.

Zhao, Y., & Melamed, B. (2007). IPA gradients for make-to-stock production-inventory systems with lost sales. *IEEE Transactions on Automatic Control*, *52*(8), 1491–1495.