FACULTY OF ECONOMICS AND BUSINESS

Characterizing order processes of (*s*,*S*) and (*r*,*nQ*) policies

Noblesse A, Boute R, Lambrecht M, Van Houdt B.

KBI_1309

Characterizing order processes of (s, S) and (r, nQ) policies

Ann M. Noblesse[a,*], Robert N. Boute[a,b], Marc R. Lambrecht[a], Benny Van Houdt[c]

[a]*Research Center for Operations Management, University KU Leuven, Belgium*
[b]*Technology & Operations Management Area, Vlerick Business School, Belgium*
[c]*Department of Mathematics and Computer Science, University of Antwerp, Belgium*

Abstract

In this paper we provide a novel approach to characterize the order process of an (s, S) and an (r, nQ) inventory policy. These inventory policies are commonly used when a fixed order cost is present. We provide two distinct perspectives. First, we analyze the time between subsequent replenishments, as well as the size of the replenishment orders, and characterize the distribution of both. Second, we look at the number of units ordered over an interval t, and compare it with the demand observed over the same period. Both perspectives are complementary and provide useful information for the upstream supplier observing this order process. We are in particular interested in the impact of the batching parameter (Q or $S - s$) on the order variance amplification ratio. It is commonly accepted in the supply chain literature that batching of orders creates a bullwhip effect. Our analysis shows that the answer is more subtle. For instance, the squared coefficient of variation (scv) of time between subsequent orders is smaller than or equal to the scv of time between subsequent demand arrivals for both policies. The order quantities can exhibit either variance amplification or dampening, compared to the demand quantities, depending on the distribution of the demand sizes and the value of the batching parameter. The bullwhip ratio, defined as the variance of the number of units ordered compared to the variance of the number of units demanded, strongly depends on the time window in which these numbers are observed. Our methodology is based on the properties of the batch-Markovian arrival process.

Keywords: Inventory policy, Batching, Bullwhip effect, Stochastic processes

1. Introduction

More than sixty years ago, Arrow et al. (1951) introduced the (s, S) inventory policy as a way to balance the cost of inventory with the cost of placing orders, hence exploiting the

*Corresponding author

Email addresses: ann.noblesse@kuleuven.be (Ann M. Noblesse), robert.boute@kuleuven.be (Robert N. Boute), marc.lambrecht@kuleuven.be (Marc R. Lambrecht), benny.vanhoudt@ua.ac.be (Benny Van Houdt)

economies of scale when ordering. In an (s, S) inventory policy, a lower stock point s and an upper stock point S are established: no order is placed until inventories fall to s or below, whereupon an order is placed to restore the inventory position to the level S. In other words, orders are placed with a batch size, always larger than or equal to $S - s$. The larger the discrepancy between S and s, the larger the average batch size. In an (s, S) policy, order sizes are stochastic: the more the inventory level falls below s (which happens, for instance, in case of a large demand size), the higher the order quantity will exceed $S - s$; we call this the *overshoot*.

In the next decades, several authors showed, in different settings, that an (s, S) policy is optimal when a fixed order cost is present (Scarf 1960, Iglehart 1963, Veinott 1966, Porteus 1971). Today, the (s, S) inventory policy is still of main importance to inventory theory and ordering policies and is incorporated in business software of many companies all over the world (Caplin and Leahy 2010).

When materials flow in fixed batch sizes, such as full truckloads or containers, an (r, nQ) inventory policy can be adopted, which is comparable, but slightly different to the (s, S) policy (Federgruen and Zheng 1992, Li and Sridharan 2008). In an (r, nQ) inventory policy, if the inventory position reaches the order point r, an order is placed equal to the smallest multiple of Q that raises the inventory position above r.

It is commonly accepted in the supply chain literature that batching of orders creates a bullwhip effect, indicating that the variability in the order process is larger than the variability in demand (Burbidge 1961, Lee et al. 1997, Potter and Disney 2006). This distorted information can lead to serious inefficiencies upstream in the supply chain (Lee et al. 1997). It is argued that larger batch sizes create more bullwhip (e.g., Burbidge 1961), and hence batching is to be avoided. However, Potter and Disney (2006) state: *"While it is often advocated that batch sizes should be reduced as much as possible, there has been limited investigation into the impact of batching on bullwhip. Placing an order of a single item is not always possible, due to technical, economical or social reasons."* Therefore, it is worthwhile to gain insight how batch sizes have an impact on the variability of the order process.

In this paper, we characterize the order process that is generated by an (s, S) or by an (r, nQ) inventory policy. This order process is the demand process observed by the upstream supplier. We provide two approaches to characterize this order process. In the first approach, we look at the time between orders and the size of the orders placed and we characterize the distribution of both. In some settings, this information is of major importance. For instance, a production facility producing on order, wants to know the mean and variance of the inter-arrival times between orders and of the size of the orders to be produced. These parameters determine the arrival rate of orders at the queue and thus the production lead times. This separate characterization of time between orders and order quantities is also found to be useful when, for instance, the upstream supplier decides on the parameters of his own inventory policy as they are both needed to characterize the intermittent demand process, that is created by the (s, S) or (r, nQ) policy of his customer.

In the second approach, we look at the total number of units ordered in a time interval t. This can be useful, for instance, when the upstream supplier reviews its orders on a periodic basis of interval t and as such aggregates the number of units that are ordered during this time interval. We characterize the variance of this order process and compare it with the variance of demand observed in the same period.

We briefly illustrate both approaches for an (s, S) policy. Figure 1 (solid line) illustrates a simulated compound Poisson demand pattern during 100 periods with on average 0.5 customer arrivals per period and Poisson demand sizes with an average of 8 units. Assume a continuous review (s, S) policy with $S - s = 15$ (because of a supposed set-up cost). Every time a demand occurs, the inventory position is evaluated: if the inventory position is smaller than or equal to s, an order quantity equal to the difference between S and the inventory position is placed; if it is larger than the order point s, we do not order. "Not ordering" at the moment of a demand arrival can be interpreted as "placing an order with an order quantity equal to zero", as is illustrated in Figure 1(a). However, this is not how the upstream supplier observes his orders. From his perspective, "not ordering" merely increases the time between orders; he does not receive an order with order quantity equal to zero. Therefore, Figure 1(b), rather than Figure 1(a) reflects how the upstream supplier looks at the order pattern, which exhibits a lot less fluctuations. As such, the conjecture of variance amplification (bullwhip), as is always predicted in the literature, does not seem that straightforward anymore and a closer analysis is needed. In this paper, we characterize both the distribution of order quantities and of the time between orders.

Insert Figure 1 about here.

In the second approach, we look at the total number of units that is demanded and ordered during a time interval t (see Figure 2 with the same demand pattern and (s, S) policy as in Figure 1). In case $t = 1$, this implies the number of units demanded (resp. ordered) every period (Figure 2(a)). Clearly, if $S - s$ grows, the variability of the number of units ordered is likely to increase as the number of zero-orders goes up and the discrepancy between an order of size zero and an order size of at least $S - s$ units has grown. Depending on the context, larger values for t might be appropriate. For instance, when the supplier reviews his inventory every time interval t and thus wants to know how many units were ordered in this time window. Figure 2(b) illustrates the number of units ordered every time interval $t = 10$. Clearly this has an impact on the observed order process: the longer the interval t, the lower the probability of observing an interval with zero-orders and the more aggregation on the number of units ordered. This leads to less fluctuations in the observed order process. In other words, the choice for t has a great impact on the bullwhip measure.

Insert Figure 2 about here.

The remainder of this paper is organized as follows. In section 2 we review the related literature. In section 3 we introduce our model and notations. Section 4 is devoted to the analysis of the order process of an (s, S) and an (r, nQ) inventory policy. In section 5 we provide some numerical illustrations for uniform, Poisson and bimodal distributed demand sizes. Section 6 concludes.

2. Literature review

This paper focuses on the order batching effect, which can be induced by an (s, S) or an (r, nQ) inventory policy. Lee et al. (1997) describe order batching as one of the key causes of the bullwhip effect. The bullwhip effect, which refers to the increased volatility of orders compared to the observed demand, has been observed in many supply chains and has a number of highly undesirable cost implications (Lee et al. 2004). Already in 1961, Burbidge (1961) recognized order batching as a main cause of the bullwhip effect. Blinder (1981) showed empirically the existence of the bullwhip effect between wholesalers and retailers and suggested that the use of (s, S) inventory models by retailers could be a possible explanation for the increased variability of the order process, compared to the demand process.

The majority of the literature studying the impact of order batching on the bullwhip effect considers a periodic review (s, S) or (r, nQ) policy. When a periodic review (r, nQ) policy is in use, Cachon (1999) finds that the coefficient of variation of order quantities in the review period can be reduced by increasing the review period and by decreasing the fixed batch size Q. Cachon (1999) assumes that Q is a multiple of the average demand. Also Li and Sridharan (2008) observe a bullwhip effect in a periodic review (r, nQ) policy: they show that the variance of order quantities in the review period is larger than or equal to the variance of demand sizes in the review period. They also show that the variance of order quantities is non-decreasing when the batch size is increased from Q_1 to mQ_1 (with m an integer > 1). Chen and Lee (2012) find that the bullwhip ratio depends on the amount of batch-ordering, the characteristics of the demand process and the presence of capacity constraints. When demand is i.i.d. and no capacity constraints exist, batch-ordering results in a bullwhip effect.

Potter and Disney (2006) analyze a variation of the periodic review (r, nQ) inventory policy, where the order is either rounded up *or down* to the multiple of the fixed batch size which is closest to the order point r (and thus must not necessarily raise the inventory position above r). The authors derive a closed form expression for bullwhip when demand is deterministic.

Schultz (1983) shows that in a periodic review (s, S) inventory policy, the variance of order quantities placed at the end of a review period is larger than the variance of demand sizes per review period, and this order variance is increasing in $S - s$. Based on the result of Schultz (1983), Schneider et al. (1995) derive an approximation for the variance of order quantities in a review period for large $S - s$ using renewal theory. Also Kelle and Milne (1999) use the approximation obtained by Schneider et al. (1995) in order to analyze the effect of an (s, S) ordering policy using gamma and normal distributed demand.

The literature concerning the order variability in a continuous review (s, S) policy is scarce. Only Caplin (1985) analyzes the variability of aggregate demand with continuous review (s, S) inventory policies. Assuming single unit demand, he finds that the variance of the number of units ordered in an interval t compared to the number of units demanded in that interval, linearly increases in the batching parameter Q, when the interval t is chosen such that either Q units are ordered or no order is placed in that time interval. For an arbitrary interval t, he finds that the variance of the number ordered exceeds the variance of the number demanded in the same time interval, indicating to a bullwhip effect. In our paper we confirm and extend Caplin's results for random batch sizes and general time intervals.

In this paper we use a batch-Markovian arrival process to characterize the order process of continuous review (s, S) and (r, nQ) policies, assuming a compound Poisson demand process. We provide two distinct, yet complementary approaches. First, we describe the orders placed based on the distribution of time between orders, and the distribution of order quantities separately. Note that this approach is commonly used for slow moving items with intermittent demand (e.g., Teunter and Duncan 2008). Secondly, we define the total number of units ordered in a time interval t and characterize the variance of this number in function of the interval length t.

3. Model description and notations

We assume a single product model with customer demand characterized by a compound (batch) Poisson process. This means that the time between subsequent demand arrivals is exponentially distributed with parameter λ and demand sizes are discrete identically distributed random variables. Our analysis holds for any arbitrary finite demand size distribution with maximum demand size m (in section 5 we provide numerical illustrations for a discrete uniform, Poisson and bimodal demand size distribution). We use D to denote the random demand size variable, d_i the probability of a demand of size i, μ_{dk} the average demand size and σ^2_{dk} its variance.

Inventory is controlled by an (s, S) or by an (r, nQ) inventory policy. It is assumed that inventory positions are continuously monitored, and consequently orders can be placed at any time. In case of a stockout, unmet demand is backlogged. Let O denote the random order size variable. Then, orders in an (s, S) inventory policy are of size $O \in \{S - s, S - s + 1, \dots S - s + m - 1\}$, depending on the observed customer demand prior to the order was placed (and its *overshoot*). We denote $\pi(i)$ as the probability to order size i in an (s, S) policy, μ_{ok} the average order size and σ^2_{ok} its variance.

In an (r, nQ) policy, the orders have random size $O \in \{Q, 2Q, \dots \lceil \frac{m}{Q} \rceil Q\}$. Denote $\widehat{\pi}(i)$ as the probability to order size i in an (r, nQ) policy, $\widehat{\mu}_{ok}$ the average order size and $\widehat{\sigma}^2_{ok}$ its variance. Note that in case the batching parameter Q exceeds the maximum demand size m, $O = Q$. When $S - s = 1$ or $Q = 1$, the (s, S) and the (r, nQ) inventory policy reduce to an order-up-to or base-stock policy. This policy, which is optimal in absence of a fixed order cost, is also referred to as an $(S - 1, S)$ policy (Nahmias 1997, Zipkin 2000).

For both an (s, S) and (r, nQ) policy, the time between orders placed is stochastic. We use μ_{ot} and σ^2_{ot} to denote the average time and variance of the time between orders placed in an (s, S) policy. Similarly, $\widehat{\mu}_{ot}$ and $\widehat{\sigma}^2_{ot}$ denote the average time and variance of the time between orders placed in an (r, nQ) policy.

Finally, define $M(t)$ as the number of demand arrivals received during an interval t, and $M_B(t)$ the number of units that are demanded in interval t. Similarly, define $N(t)$ as the number of orders placed during an interval t, and $N_B(t)$ the number of units that are ordered in interval t. p_k denotes the probability that an order consists of k demands.

We use $c^2 = \frac{\mu^2}{\sigma^2}$ to denote the squared coefficient of variation in any of the above defined variables.

4. Order process of a continuous review (s, S) and (r, nQ) policy

We make use of a batch Markovian arrival process (BMAP) to characterize the order process of (s, S) and (r, nQ) inventory policies. We first describe how the order process can be characterized by a BMAP process in section 4.1. This approach allows us to look at the order process from two distinct perspectives. In section 4.2 we look at the replenishment orders when they are placed, and we analyze the distribution of the time between orders placed and the distribution of the order sizes. In section 4.3 we look at the number of units that are placed during a certain time window and compare it with the observed demand during the same period.

4.1. The order process as a batch Markovian arrival process

The batch Markovian arrival process (BMAP) is a stochastic point process which allows for arrivals of batch order quantities, correlated batch sizes, dependent inter-arrival times and non-exponential inter-arrival time distributions (Lucantoni 1993, Cordeiro and Kharoufeh 2011). The order process of both an (s, S) and an (r, nQ) inventory policy can be modeled as a BMAP as follows.

An (s, S) inventory policy can be characterized by a continuous time Markov chain with the states equal to the inventory positions $\{S, S-1, \dots, s+1\}$. Define C as its $(S-s)\times(S-s)$ transition probability matrix. A transition from state j to state $j' < j$ occurs when demand depletes inventory, its rate $(C)_{j,j'}$ is defined by the demand rate λ and the probability $d_{j-j'}$ of observing a demand of size $j - j'$. As soon as the inventory position reaches the order point s (or less), an order is placed to raise the inventory position to S. (As replenishments occur instantly, the time spent in inventory positions equal to or smaller than s is negligible.) A transition from state j to state S occurs at rate $(C)_{j,S}$. Define $(C_i)_{j,S}$ as the transition rate from inventory position j to S while an order of size i is placed, for $i \in \{S-s, S-s+1, \dots, S-s+m-1\}$. Then, C_i contains the rates of inventory position changes that are accompanied by an order of size i. If an order is placed, the inventory position immediately raises to the order-up-to level S. Hence, the transition matrix C_{S-s+k} for $k \in \{0, \dots, m-1\}$, is given by

$$C_{S-s+k} = \lambda \begin{bmatrix} d_{S-s+k} & 0 & \dots & 0 \\ d_{S-s+k-1} & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ d_{1+k} & 0 & \dots & 0 \end{bmatrix}. \tag{1}$$

For $i \in \{1, \dots, S-s-1\}$, $C_i = \mathbf{0}$, since the order size is always greater than or equal to $S - s$. Let C_0 contain the rates of inventory position changes that are not accompanied by an order,

$$C_0 = \lambda \begin{bmatrix} d_0 - 1 & d_1 & \dots & d_{S-s-1} \\ 0 & \ddots & \ddots & \vdots \\ \vdots & & \ddots & d_1 \\ 0 & \dots & & d_0 - 1 \end{bmatrix}. \tag{2}$$

The order process of an (s, S) policy is then characterized by its $(S-s) \times (S-s)$ infinitesimal generator matrix C:

$$C = \sum_{i=0}^{S-s+m-1} C_i. \tag{3}$$

Remark that the greatest common divisor of the set $\{S-s\} \cup \{i | d_i \neq 0\}$ must be non-zero, otherwise C is not irreducible. Let $\theta = [\theta_S, \theta_{S-1}, \ldots, \theta_{s+1}]$ be the steady state vector of C, defining the probability to be in a specific inventory position at a random point in time: θ is the unique positive solution of $\theta C = \mathbf{0}$ and $\theta e = 1$, with $\mathbf{0}$ a $1 \times (S-s)$ zero vector and e a $(S-s) \times 1$ vector of ones.

In a similar way, an (r, nQ) inventory policy can be characterized by a continuous time Markov chain with the states equal to the inventory positions $\{r+Q, r+Q-1, \ldots, r+1\}$. In an (r, nQ) inventory policy, an order is placed as soon as state r or a lower state is reached; a quantity equal to the smallest multiple of Q is ordered such that the inventory position raises above the order point r. We denote its $Q \times Q$ transition matrix as $\widehat{C}$. A transition from state j to state $j^{'} < j$ occurs when a demand depletes inventory; its rate $(\widehat{C})_{j,j'}$ is equivalent to $(C)_{j,j'}$ in an (s, S) policy. However, when state r or a lower state is reached, a replenishment is made up to inventory position $j^{'}$ with $j^{'} \in \{r+Q, \ldots, r+1\}$, and not up to S like in an (s, S) policy. Define the $Q \times Q$ matrices $\widehat{C}_i$ as the matrices containing the rates at which orders of size i are placed with $i \in \{Q, 2Q \ldots, \lceil \frac{m}{Q} \rceil Q\}$. When inventory position $r-k$ is reached with $k \in \{(n-1)Q, \ldots, nQ-1\}$ and $n \in \{1, 2, \ldots\}$, one places an order equal to nQ, such that the inventory position raises to $r-k+nQ$. Therefore,

$$\widehat{C}_{nQ} = \lambda \begin{bmatrix} d_{nQ} & \cdots & d_{(n+1)Q-1} \\ d_{nQ-1} & \cdots & d_{(n+1)Q-2} \\ \vdots & \ddots & \vdots \\ d_{(n-1)Q+1} & \cdots & d_{nQ} \end{bmatrix}. \tag{4}$$

For $i \in \{(n-1)Q+1, \ldots, nQ-1\}$ with $n \in \{1, 2, \ldots\}$, $\widehat{C}_i = \mathbf{0}$, since the order size is always a multiple of Q. $\widehat{C}_0$ provides the transition rates at which the inventory position changes without placing an order, and is similar to (2), where $S - s = Q$.

The order process of the (r, nQ) inventory policy is characterized by its $Q \times Q$ infinitesimal generator matrix $\widehat{C}$:

$$\widehat{C} = \sum_{i=0}^{\lceil \frac{m}{Q} \rceil Q} \widehat{C}_i. \tag{5}$$

Note, $\widehat{C}$ is a circulant matrix with its first row identical to $\sum_{n \geq 0}(d_{nQ}, d_{nQ+1}, \ldots, d_{nQ+Q-1}) - (1, 0, \ldots, 0)$, which sums to zero. Remark that the greatest common divisor of the set $\{Q\} \cup \{i | d_i \neq 0\}$ must be non-zero, otherwise $\widehat{C}$ is not irreducible.

Let $\widehat{\theta}$ be the steady state probabilities of the BMAP at a random point in time. Vector $\widehat{\theta}$ is the unique positive solution of $\widehat{\theta}\widehat{C} = \mathbf{0}$ and $\widehat{\theta} e = 1$, with $\mathbf{0}$ a $1 \times Q$ zero vector and e a $Q \times 1$ vector of ones. As $\widehat{C}$ is circulant with its row and column sums equal to zero, one finds that $\widehat{\theta} = \left[\widehat{\theta}_{r+Q}, \widehat{\theta}_{r+Q-1}, \ldots, \widehat{\theta}_{r+1}\right] = \frac{1}{Q}[1, \ldots, 1]$, which was also found by Li and Sridharan (2008).

Based on the characterization as a BMAP, we are now in a position to derive properties of this order process.

4.2. First approach: Distribution of time between orders and order quantities

In our first approach we characterize the order process of (s, S) and (r, nQ) policies based on the distribution of time between orders placed and the size of the order quantities placed. This separate characterization is often used for slow moving items. Also in a production setting, this detailed information might prove to be useful.

4.2.1. Distribution of the time between orders

Proposition 1. *The time between orders placed in an (s, S) policy is phase-type distributed, characterized by the triple $(S-s, C_0, \alpha)$, with C_0 an $(S-s)\times(S-s)$ substochastic matrix, defined by (2); and $\alpha = [1, 0, \ldots, 0]$.*

The average and the variance of time between subsequent orders equal:

$$\mu_{ot} = \frac{1}{\lambda}\sum_{j=0}^{S-s-1} b_{S-j}, \tag{6}$$

$$\sigma^2_{ot} = \frac{1}{\lambda^2}\left[\left(2\sum_{j=0}^{S-s-1} b_{S-j}\sum_{i=0}^{S-s-1-j} b_{S-i}\right) - \left(\sum_{j=0}^{S-s-1} b_{S-j}\right)^2\right], \tag{7}$$

$$\text{with } b_{S-j} = \begin{cases} b_S\sum_{i=1}^{j} d_i \cdot b_{S-j+i} & \text{for } j \geq 1,\\ \frac{1}{1-d_0} & \text{for } j = 0.\end{cases} \tag{8}$$

PROOF OF PROPOSITION 1. In an (s, S) policy the time between two orders placed is equivalent to the time it takes for the Markov chain to move from the order-up-to level S until absorption in the order point s or below. Since the time between successive transitions in a continuous time Markov chain is exponentially distributed, the distribution of the time until absorption to s is the sum of a random number of exponentially distributed time intervals, and thus follows a phase-type (PH) distribution. The number of phases is equivalent to the number of inventory positions before placing an order, $S-s$. As we always order up to S, the initial vector $\alpha = [1, 0, \ldots, 0]$. C_0 contains the transition rates between the transient states ranging from S to $s+1$ (see (2)).

The expected time spent in each inventory position before placing an order is given by $\alpha(-C_0)^{-1}$ (Latouche and Ramaswami 1999), where the matrix $(-C_0)$ is an $(S-s)\times(S-s)$ upper triangular Toeplitz matrix. The inverse of $(-C_0)$ is again an upper triangular Toeplitz matrix, defined as

$$(-C_0)^{-1} = \frac{1}{\lambda}\begin{bmatrix} b_S & b_{S-1} & b_{S-2} & \cdots & b_{s+1} \\ 0 & b_S & b_{S-1} & \cdots & b_{s+2} \\ 0 & 0 & b_S & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & b_{S-1} \\ 0 & 0 & 0 & \ddots & b_S \end{bmatrix},$$

with $\frac{1}{\lambda}b_j$ the expected time spent in inventory position j for $j \in \{s+1, \ldots, S\}$. Using backward substitution, the entries b_j of $(-C_0)^{-1}$ are defined by (8).

Using the definition of the mean and the variance of a phase-type distribution (i.e., $\mu = \alpha\left[(-C_0)^{-1} e\right]$ and $\sigma^2 = 2\alpha C_0^{-2} e - \left(\alpha\left[C_0^{-1} e\right]\right)^2$, with e an $(S-s) \times 1$ unit vector (Latouche and Ramaswami 1999), we find (6) and (7). □

For some simple demand size distributions, we can obtain a closed-form expression for b_{S-j}. In appendix we provide results when demand sizes are Poisson or uniform distributed. In case of single unit demands, the time between orders follows an Erlang distribution $\sim$ *Erlang*(Q, λ) with $\mu_{ot} = Q/\lambda$ and $\sigma^2_{ot} = Q/\lambda^2$.

From Proposition 1, we see that the average time between two subsequent orders is equal to the sum of the average time spent in all inventory positions j prior to the order point s, which are given by $\frac{1}{\lambda}b_j$, with $S \geq j \geq s+1$. As all b_j are positive and $b_S \geq 1$ (as $b_S = \frac{1}{1-d_0}$), we find that $\mu_{ot} \geq \frac{1}{\lambda}$. Or, in other words, the average time between two orders is larger than or equal to the average time between two demand arrivals, i.e. $\mu_{ot} \geq \mu_{dt}$.

As we are particularly interested in the order variance amplification, created by an (s, S) policy, we derive the following corollaries from Proposition 1:

Corollary 1. *In an (s, S) policy, the variance of the time between subsequent orders is larger than or equal to the variance of time between subsequent demand arrivals, i.e., $\sigma^2_{ot} \geq \sigma^2_{dt}$.*

PROOF OF COROLLARY 1. As $\sigma^2_{dt} = 1/\lambda^2$, it suffices to show that $\sigma^2_{ot} \geq 1/\lambda^2$. In fact we will prove that $\sigma^2_{ot} \geq \mu_{ot}/\lambda$, which is a somewhat stronger result as $\mu_{ot} \geq 1/\lambda$. Let p_k be the probability that an order consists of k demands (where some demands may be of size zero if $d_0 > 0$). Provided that an order consists of k demands, the inter-arrival time follows an Erlang-k distribution. As the second moment of an Erlang-k distribution equals $(k+k^2)/\lambda^2$, we can write σ^2_{ot} as

$$\begin{aligned} \sigma^2_{ot} &= \sum_{k\geq 1} p_k \frac{k+k^2}{\lambda^2} - \left(\sum_{k\geq 1} p_k \frac{k}{\lambda}\right)^2 \\ &= \frac{\mu_{ot}}{\lambda} + \sum_{k\geq 1} p_k \frac{k^2}{\lambda^2} - \sum_{k\geq 1} p_k \sum_{i\geq 1} p_i \frac{ki}{\lambda^2}, \end{aligned}$$

since $\mu_{ot} = \sum_{k\geq 1} p_k \frac{k}{\lambda}$. Further, $ki \leq (k^2+i^2)/2$ as $(k-i)^2 \geq 0$, which yields

$$\begin{aligned} \sigma^2_{ot} &\geq \frac{\mu_{ot}}{\lambda} + \sum_{k\geq 1} p_k \frac{k^2}{\lambda^2} - \sum_{k\geq 1} p_k \sum_{i\geq 1} p_i \frac{k^2+i^2}{2\lambda^2} \\ &= \frac{\mu_{ot}}{\lambda}, \end{aligned}$$

as $\sum_{i\geq 1} p_i = 1$. □

Corollary 2. *In an (s,S) policy, the squared coefficient of variation (scv) of time between subsequent orders is smaller than or equal to the scv of time between subsequent demand arrivals, i.e., $c^2_{ot} \leq c^2_{dt}$.*

Proof of Corollary 2. As demand arrives according to a compound Poisson process, it follows that $c^2_{dt} = 1$. Hence, to prove corollary 2, we need to show that $\sigma^2_{ot} \leq (\mu_{ot})^2$. Substituting Eqs. (6-7), we need to show that

$$\frac{1}{\lambda^2}\left[\left(2\sum_{j=0}^{S-s-1} b_{S-j} \sum_{i=0}^{S-s-1-j} b_{S-i}\right) - \left(\sum_{j=0}^{S-s-1} b_{S-j}\right)^2\right] \leq \left[\frac{1}{\lambda}\sum_{j=0}^{S-s-1} b_{S-j}\right]^2,$$

or by rewriting,

$$\sum_{j=0}^{S-s-1} b_{S-j}\left(\sum_{i=0}^{S-s-1-j} b_{S-i} - \sum_{i=0}^{S-s-1} b_{S-i}\right) \leq 0.$$

As all b_i are positive, we know that $\sum_{i=0}^{S-s-1-j} b_{S-i} \leq \sum_{j=0}^{S-s-1} b_{S-j}$, which proves Corollary 2. □

In a similar way, we can characterize the time between orders for an (r,nQ) policy.

Proposition 2. *The time between orders placed in an (r,nQ) policy is phase-type distributed, characterized by the triple $(Q, \widehat{C}_0, \widehat{\alpha})$, with $\widehat{C}_0$ a $Q\times Q$ substochastic matrix, defined by (2); and*

$$\widehat{\alpha} = \frac{1}{X}\left[1-d_0, 1-d_0-d_1, \ldots, 1-\sum_{k=0}^{Q-1} d_k\right], \tag{9}$$

with $X = Q - \sum_{i=0}^{Q-1}(Q-i)\,d_i$. The average and the variance of time between subsequent *orders equal:*

$$\widehat{\mu}_{ot} = \frac{1}{\lambda X} \sum_{i=0}^{Q-1} \left[\left(1 - \sum_{j=0}^{i} d_j \right) \sum_{k=0}^{Q-1-i} b_{S-k} \right] \tag{10}$$

$$\widehat{\sigma}_{ot}^2 = \frac{1}{\lambda^2 X} \left[2 \sum_{i=0}^{Q-1} \left(1 - \sum_{j=0}^{i} d_j \right) \sum_{k=0}^{Q-1-i} b_{S-k} \sum_{l=0}^{Q-1-i-k} b_{S-l} - \frac{1}{X} \left[\sum_{i=0}^{Q-1} \left(1 - \sum_{j=0}^{i} d_j \right) \sum_{k=0}^{Q-1-i} b_{S-k} \right]^2 \right], \tag{11}$$

with b_{S-j} as defined in (8).

Proof of Proposition 2. The proof is analogous to Proposition 1, with $s = r$ and $Q = S - r$. In an (r, nQ) policy the time between two orders placed is equivalent to the time it takes for the Markov chain until absorption in the order point r or below. $\widehat{C}_0$ describes how the inventory position decreases prior to reaching one of the absorbing states. However, when state r (or lower) is reached, a replenishment is made up to inventory position j' with $j' \in \{r+Q, \ldots, r+1\}$, and not up to S like in an (s, S) policy. Therefore, the steady state probabilities $\widehat{\alpha}$ of the inventory positions immediately after a replenishment are

$$\begin{aligned} \widehat{\alpha} &= \widehat{\theta} \sum_{i \geq 1} \widehat{C}_i / (\widehat{\theta} \sum_{i \geq 1} \widehat{C}_i e) \\ &= \widehat{\theta}(\widehat{C} - \widehat{C}_0) / \widehat{\theta}(\widehat{C} - \widehat{C}_0) e \\ &= \widehat{\theta}\widehat{C}_0 / \widehat{\theta}\widehat{C}_0 e \\ &= e\widehat{C}_0 / e\widehat{C}_0 e, \end{aligned}$$

as $\widehat{\theta}\widehat{C} = \mathbf{0}$; This leads to (9). Using the definition of the mean and the variance of a phase-type distribution, we find (10) and (11). □

As can be intuitively expected, also in an (r, nQ) policy, it can be shown that the average time between two subsequent orders is larger than or equal to the average time between two demand arrivals, i.e. $\widehat{\mu}_{ot} \geq \mu_{dt}$. Indeed, using (10), the expression

$$\frac{1}{\lambda X} \sum_{i=0}^{Q-1} \left[\left(1 - \sum_{j=0}^{i} d_j \right) \sum_{k=0}^{Q-1-i} b_{S-k} \right] \geq \frac{1}{\lambda} \tag{12}$$

holds if

$$\sum_{i=0}^{Q-1} \left[\left(1 - \sum_{j=0}^{i} d_j \right) \sum_{k=0}^{Q-1-i} b_{S-k} \right] \geq Q - \sum_{i=0}^{Q-1} (Q-i)\, d_i. \tag{13}$$

Eq. (13) can be proven by induction on Q: For $Q = 1$, Eq. (13) is true as $b_S \geq 1$. (13) also holds for $Q = x + 1$ if (13) is true for $Q = x$, as

$$\sum_{i=0}^{x-1} \left[\left(1 - \sum_{j=0}^{i} d_j\right) b_{S-(x-i)} \right] + \left(1 - \sum_{j=0}^{x} d_j\right) b_S \geq 1 - \sum_{j=0}^{x} d_j.$$

We also find that the variance of the time between subsequent orders is larger than or equal to the variance of time between subsequent arrivals of demand, i.e. $\widehat{\sigma}_{ot}^2 \geq \sigma_{dt}^2$. The proof is identical to the proof of Corollary 1.

4.2.2. Distribution of order quantities

In this section, we derive the distribution of order quantities in an (s, S) inventory policy and an (r, nQ) policy. We first focus on the (s, S) policy.

Proposition 3. *In an (s, S) policy, the probability to order quantity i is defined as*

$$\pi(i) = \begin{cases} 0 & \text{for } i \in \{1, \ldots, S-s-1\}, \\ \dfrac{\sum_{j=0}^{S-s-1} d_{i-j} b_{S-j}}{\sum_{i=S-s}^{S-s+m-1} \sum_{j=0}^{S-s-1} d_{i-j} b_{S-j}} & \text{for } i \in \{S-s, \ldots, S-s+m-1\}, \end{cases} \tag{14}$$

with b_{S-j} as defined in (8).

Proof of Proposition 3. The probability to place an order with size $S-s$ is equal to the probability that the inventory position exactly reaches the order point s. This probability is given by the sum for all j (ranging from 0 to $S - s - 1$) of the probabilities of being in inventory position $S - j$ and a demand of size $S - s - j$ realizes, in which case the inventory position drops to s and a replenishment of size $S - s$ is placed. Therefore, the probability to order a quantity $S - s$ equals $\theta\left(C_{S-s}e\right)/(\theta \sum_{i \geq S-s} C_i e)$, with θ the steady state probabilities of inventory positions S to $s + 1$ at a random point in time.

In general, the probability to order a quantity i, for all $i \in \{S - s, \ldots, S - s + m - 1\}$, is equal to

$$\begin{aligned} \pi(i) &= \theta C_i e/(\theta \sum_{i \geq S-s} C_i e) \\ &= \theta C_i e/(\theta(-C_0)e). \end{aligned}$$

Due to the probabilistic interpretation (or by means of $\theta = \alpha(-C_0)^{-1}/(\alpha(-C_0)^{-1}e)$), we have $\theta = (b_S, b_{S-1}, \ldots, b_{s+1})/(\sum_{i=s+1}^{S} b_i)$. Substituting (2) and (1) leads to Proposition 3. □

The expected order quantity is larger than or equal to the expected demand size, i.e. $\mu_{ok} \geq \mu_{dk}$. Indeed, as the average order rate must equal the average demand rate in order to have a stable system, $\mu_{ok}\frac{1}{\mu_{ot}} = \mu_{dk}\frac{1}{\mu_{dt}}$, we know that $\mu_{ok} \geq \mu_{dk}$, since in the previous section we showed that $\mu_{ot} \geq \mu_{dt}$.

If we want to analyze the variance in order quantities, we can do so using its definition:

$$\sigma_{ok}^2 = \left[\sum_{i \geq S-s}^{S-s+m-1} (i)^2 \cdot \pi(i)\right] - \left[\sum_{i \geq S-s}^{S-s+m-1} (i) \cdot \pi(i)\right]^2 \tag{15}$$

The above confirms that in case of only single-unit demands, there is no overshoot, and Eqs. (14) and (15) reduce to $\pi(S-s) = 1$ and $\sigma_{ok}^2 = 0$. We always order a fixed order quantity equal to $S - s$.

In a similar way, the distribution of order quantities of an (r, nQ) inventory policy can be derived.

Proposition 4. *In an (r, nQ) policy, the probability to order quantity i is defined as:*

$$\begin{aligned} \widehat{\pi}(i) &= \begin{cases} 0 & \text{for } i \notin \{Q, 2Q, \ldots\} \\ \frac{1}{X} \sum\limits_{j=-(Q-1)}^{Q-1} (Q - |j|) d_{i+j} & \text{for } i \in \{Q, 2Q, \ldots\} \end{cases} \\ \text{with } X &= Q - \sum_{i=0}^{Q-1} (Q-i) d_i. \end{aligned} \tag{16}$$

Proof of Proposition 4. As soon as the inventory position reaches the order point r an order is placed. The order quantity is the smallest multiple of Q such that the inventory position is at least $r+1$. Therefore, we know that $\widehat{\pi}(Q) = \widehat{\theta}\widehat{C}_Q e / \widehat{\theta} \sum_{i \geq Q} \widehat{C}_i e$.

In general, for $k \geq Q$,

$$\begin{aligned} \widehat{\pi}(k) &= \widehat{\theta}\widehat{C}_k e / \widehat{\theta} \sum_{i \geq Q} \widehat{C}_i e \\ &= e\widehat{C}_k e / e \sum_{i \geq Q} \widehat{C}_i e \\ &= e\widehat{C}_k e / e(-\widehat{C}_0) e, \end{aligned}$$

as $\widehat{\theta} = \frac{1}{Q}[1, \ldots, 1]$. Substituting (2) and (4) in the above leads to Proposition 4. □

Observe that, if $m \in \{1, \ldots, Q\}$, then $\widehat{\pi}(Q) = 1$ and $\widehat{\sigma}_{ok}^2 = 0$, as the inventory position is always larger than or equal to $r - m + 1$ in a continuous review setting.

Also here, we have $\widehat{\mu}_{ok} \geq \mu_{dk}$, analogous to the (s, S) policy. Looking at the variance amplification in the order sizes, generated by the (s, S) and (r, nQ) policy, we numerically find that we can have both order variance dampening, as well as amplification with respect to the demand sizes. We find that this strongly depends on the demand size distribution and the value of $S - s$ or Q. We refer to Section 5 for some numerical illustrations.

The same holds for the squared coefficient of variation in orders: our numerical illustrations show that we can have variance amplification in the order sizes for small batch sizes, but we primarily find variance dampening in the order sizes compared to the demand sizes. Again, this is dependent on the demand size distribution and the batch size.

4.3. Second approach: total number of units ordered in a time interval t

Next to characterizing the order process from the perspective of the time between orders and the size of the orders placed, we can also look at the total number of units ordered in an arbitrary interval of length t. This can be useful when the upstream supplier reviews its orders on a periodic basis of interval t and as such aggregates the number of units that are ordered during this time interval. We compare the variance of the total number of units ordered in a time interval t with the variance of the number of units demanded in the same interval. This ratio can be used as a measure for the bullwhip effect. Note that this ratio will differ depending on the length of the interval t. Indeed, the longer the interval t, the more these variances will concur.

Let $M(t)$ and $N(t)$ be the number of demand arrivals and the number of orders placed in an arbitrary interval of length t, and let $M_B(t)$ and $N_B(t)$ be the number of units demanded and the number of units ordered in an arbitrary interval of length t, respectively. Then, we can define the order variance amplification in function of the time interval t as

$$Bullwhip(t) = \frac{Var[N_B(t)]}{Var[M_B(t)]} = Var\left[\sum_{i=1}^{N(t)} O_i\right] \Bigg/ Var\left[\sum_{i=1}^{M(t)} D_i\right], \tag{17}$$

with D_i the size of the i-th demand in such an interval and O_i the size of the i-th order. As the demand sizes are independent of the time between demand arrivals, we have

$$Var[M_B(t)] = E[M(t)]\sigma_{dk}^2 + \mu_{dk}^2 Var[M(t)] = \lambda t[\sigma_{dk}^2 + \mu_{dk}^2], \tag{18}$$

as $E[M(t)] = Var[M(t)] = \lambda t$ for a compound Poisson demand process. The order sizes O_i are, however, not independent of the time between orders placed, meaning $N(t)$ is not independent of O_i. To compute $Var[N_B(t)]$ we will therefore rely on the fact that the order process is a BMAP characterized by a set of matrices C_i, for $i \geq 0$. As shown in Neuts and Li (1996) the variance of the counting process $N_B(t)$ of a stationary[1] BMAP can be

[1] A stationary BMAP is a BMAP with the additional assumption that the state at time zero is distributed

expressed as

$$Var[N_B(t)] = [\lambda_2 - 2(\lambda^*)^2 + 2cC_1^* e]t - 2c[I - e^{Ct}]d, \tag{19}$$

with $\lambda_2 = \theta C_2^* e$, $\lambda^* = \theta C_1^* e$, $c = \theta C_1^*(e\theta - C)^{-1}$, $d = (e\theta - C)^{-1} C_1^* e$ and $C_i^* = \sum_{n\geq 1} n^i C_n$, for $i = 1, 2$. Note, this expression is valid for the (s, S) policy, but also for the (r, nQ) policy provided that we replace C_n by $\widehat{C}_n$ for all n, C by $\widehat{C}$ and θ by $\widehat{\theta}$.

For the (r, nQ)-policy we can also derive a different expression for $Var[N_B(t)]$.

Proposition 5. *In an (r, nQ) policy, we have*

$$Var[N_B(t)] = Var[M_B(t)] + E[N_{B,Q}(t)(Q - N_{B,Q}(t))],$$

with $Var[M_B(t)] = \lambda t[\sigma_{dk}^2 + \mu_{dk}^2]$, as defined in (18), and $N_{B,Q}(t) = (N_B(t) \mod Q)$, i.e., $N_{B,Q}(t)$ is the number of units demanded modulo Q in an arbitrary interval of length t.

Proof of Proposition 5. This property is proven in a manner similar to Proposition 4.3 in Caplin (1985). As $E[N_B(t)] = E[M_B(t)] = \lambda t \mu_{dk}$, we have

$$\begin{aligned} Var[N_B(t)] \quad - \quad & Var[M_B(t)] = \\ & \sum_{a=0}^{\infty} \sum_{b=0}^{Q-1} \left(E[N_B^2(t) | M_B(t) = aQ + b] - (aQ + b)^2 \right) \Pr[M_B(t) = aQ + b]. \end{aligned}$$

Denote $I(t)$ the inventory position at the start of the interval of length t, i.e., $\Pr[I(t) = p] = \widehat{\theta}_i = 1/Q$. Then, for $M_B(t) = aQ + b$ and $I(t) = p$, we have $N_B(t) = (a + 1)Q$ if $p = s + 1, \ldots, s + b$ and $N_B(t) = aQ$ for $p > s + b$ under the (r, nQ)-policy. As a result,

$$\begin{aligned} E[N_B^2(t) | M_B(t) = aQ + b] &= \sum_{p=s+1}^{S} E[N_B^2(t) | M_B(t) = aQ + b, I(t) = p] \Pr[I(t) = p], \\ &= \frac{b}{Q}((a + 1)Q)^2 + \frac{Q - b}{Q}(aQ)^2 = (aQ + b)^2 - b^2 + Qb. \end{aligned}$$

This implies

$$Var[N_B(t)] - Var[M_B(t)] = \sum_{a=0}^{\infty} \sum_{b=0}^{Q-1} b(Q - b) \Pr[M_B(t) = aQ + b],$$

which completes the proof. □

Although it may be more effective to compute $Var[N_B(t)]$ via (19), the above property is particularly useful as it allows us to prove the following corollary:

---

as θ with $\theta C = 0$. Since, we are interested in the number of orders placed in an arbitrary interval of length t, the BMAP of interest is in fact the stationary BMAP.

Corollary 3. *In an (r, nQ) policy, we have $Bullwhip(t) \geq 1$, and $\lim_{t\to\infty} Bullwhip(t) = 1$. Further,*

$$Bullwhip(t) = \lambda(\sigma_{dk}^2 + \mu_{dk}^2)t - 2c[I - e^{\widehat{C}t}]d,$$

and

$$\lim_{t\to\infty} (Var[N_B(t)] - Var[M_B(t)]) = 2cd - 2(\lambda^*)^2.$$

Proof of Corollary 3. Proposition 5 and (18) imply

$$Bullwhip(t) = 1 + \frac{E[N_{B,Q}(t)(Q - N_{B,Q}(t))]}{\lambda t(\sigma_{dk}^2 + \mu_{dk}^2)}.$$

As $0 \leq N_{B,Q}(t) \leq Q - 1$, we find $Bullwhip(t) \geq 1$ and

$$Bullwhip(t) \leq 1 + \frac{Q^2}{\lambda t(\sigma_{dk}^2 + \mu_{dk}^2)} \to 1,$$

as $t \to \infty$. Note, $Bullwhip(t) = 1$ if and only if all the demand sizes are multiples of Q. By noting that $e^{\widehat{C}t}$ converges to $e\widehat{\theta}$ and that $ce = \lambda^* = \widehat{\theta}d$, we get

$$\lim_{t\to\infty} 2ce^{\widehat{C}t}d = 2(\lambda^*)^2.$$

When combined with (19) and the fact that $\lim_{t\to\infty} Bullwhip(t) = 1$, this proves that

$$\lambda_2 - 2(\lambda^*)^2 + 2c\widehat{C}_1^*e = \lambda[\sigma_{dk}^2 + \mu_{dk}^2],$$

and this suffices to conclude that

$$\lim_{t\to\infty} Var[N_B(t)] - Var[M_B(t)] = \lim_{t\to\infty} 2c(I - e^{\widehat{C}t})d.$$

□

Remark that, in case of single unit demand sizes, i.e., $d_1 = 1$, the (r, nQ) and (s, S) policy behave identical and the above results therefore also apply for the (s, S) policy if $d_1 = 1$. Unfortunately, the above proof does not allow us to demonstrate the same for the (s, S) policy with random demand sizes.

5. Numerical illustration

In this section we numerically illustrate the impact of the batching effect, induced by an (s, S) or an (r, nQ) inventory policy, on its order process. We use a setting where demand is characterized by a compound Poisson process, with arrival rate $\lambda = 1$ and demand sizes according to resp. a discrete uniform ($D \sim U\{1,\ldots,19\}$), Poisson ($D \sim Pois\{10\}$) and bimodal distribution, each with $\mu_{dk} = 10$. For the bimodal demand size distribution, we assume 50% of demand arrivals has a discrete uniform demand size distribution $D \sim U\{1,\ldots,5\}$, and 50% has a discrete uniform demand size distribution $D \sim U\{15,\ldots,19\}$. We also benchmark the results for a Poisson distributed demand process, parameterized by a demand arrival rate $\lambda = 1$ and single unit demand sizes (in this setting, no difference exists between the order process of an (s,S) or an (r,nQ) policy, i.e., we always order Q units).

We first illustrate the order process, making a distinction between the order quantities and the time between orders (our first approach). Recall that when demand is characterized by a (single unit) Poisson process, the average order quantity equals $S - s$ or Q and the variance of order quantities is zero. The time between orders follows an Erlang distribution with average Q/λ and variance Q/λ^2, indicating that they both increase linearly in Q. If demand is characterized by a compound Poisson process, the distribution of time between orders and order quantities is less straightforward.

Figure 3 plots the ratio of the average order size compared to the average demand size in function of the batching parameter $S - s$ or Q, for both the (s,S) and (r,nQ) policy. Note that this ratio coincides with the ratio of the average time between orders to the average time between demand arrivals, as in order to have a stable system, the average rate of items ordered must equal the average rate of items demanded, or $\mu_{ok}/\mu_{ot} = \mu_{dk}/\mu_{dt}$, which leads to $\mu_{ot}/\mu_{dt} = \mu_{ok}/\mu_{dk}$. From Figure 3 we see that these ratios increase with the batching parameter, but not always in a linear way. For an (r,nQ) policy, we do have a linear increase in Q for $Q \geq m$, as for in that case one always orders Q units. Observe as well that the average order size μ_{ok} and its corresponding time between orders μ_{ot} are larger for an (s,S) policy compared to $\widehat{\mu}_{ok}$ and $\widehat{\mu}_{ot}$ for an (r,nQ) policy. This is intuitive as one always orders up to S in an (s,S) policy, whereas in an (r,nQ) policy one orders at most up to $r + Q$.

Insert Figure 3 about here.

Figure 4 illustrates the variance amplification ratio of the time between orders placed compared to the time between demand arrivals. Here, we observe that we always have variance amplification, which increases in Q, albeit not always linearly (recall that for a single unit demand process, the increase is linear in Q). We also see that we have more variance amplification in the time between orders placed created by an (s,S) policy, compared to the (r,nQ) policy. Figure 5 provides the scv of the time between orders placed to the time between demand arrivals (which equals 1 for a compound Poisson demand). Interestingly, it shows that the relative variance of the time between orders is actually less that the relative variance of the time between demands, indicating that we have variance dampening (smoothing) in the time between orders.

Insert Figures 4 and 5 about here.

Figure 6 shows the ratio of the variance of order sizes compared to the variance of demand sizes in function of $S-s$ or Q. Here we do not always see variance amplification. Instead we observe a wavy pattern: for some values of Q, we may observe variance amplification, but we find as well order variance smoothing. Observe that the variance of the order sizes converges to a limiting value in an (s,S) policy, which can either indicate bullwhip or smoothing. For an (r,nQ) policy, the variance of the order sizes is zero when $Q \geq m$.

If we look at the ratio of the scv of order sizes compared to the scv of demand sizes (Figure 7), we can make the same observation, but we see that we primarily have smoothing in the order sizes. This indicates that the relative variance of the order sizes is smaller that than the relative variance of the demand sizes. Although this ratio generally goes down when Q increases, this decrease is not strict in Q; we observe some wavy behavior in function of Q (albeit less outspoken).

Insert Figures 6 and 7 about here.

We finally illustrate the order variance amplification ratio, looking at the number of units ordered in a time interval t, compared to the number of units demanded in that same time window (our second approach). Figure 8 assumes time interval $t = 1$, whereas in Figure 9, the time interval $t = 60$. We find variance amplification in orders, which is generally increasing in Q, albeit not always linearly and not always strict increasing in Q (for some values of Q, we even observe a slight decrease of the bullwhip effect in Q).

Figure 10 illustrates how the time interval t impacts the bullwhip ratio for single unit demand sizes. For small intervals t, we see an increase in function of Q, but this is not necessarily the case for large time intervals t and small values of Q. Recall that Caplin (1985), assuming single unit demands, finds a linear increase when the time interval is chosen such that either Q or 0 units are ordered; indeed, this assumption is more likely to hold if the batch size Q is larger than the time interval t. This linear increase is also observed in Figure 10 when Q exceeds t. However, if Q is smaller than t, the bullwhip ratio does not increase linearly in Q.

Insert Figures 8, 9 and 10 about here.

6. Conclusion

In this paper, we presented a novel approach to characterize the order process of continuous review (s,S) and (r,nQ) inventory policies, using the properties of the batch Markovian arrival process. We characterized the order process from two distinct, yet complementary perspectives, which can be useful for the upstream supplier to characterize its demand. In our first approach, we derived the distribution of order quantities and time between orders separately. Splitting a stochastic process into information concerning timing and quantities is common practice for slow moving items with intermittent demand. This separate characterization can also be of particular interest if the orders are sent to production (e.g., if the upstream supplier produces on order), in which case the production facility can have a better insight in the arrival rate of orders at the queue, which is defined by the time between arrivals of orders and the order sizes. These are prime determinants of the production lead time. We find that the *relative* variance of the time between orders, as measured by its squared coefficient of variation, is less that the relative variance of the time between demands. Also the order quantities do not always display a larger variance compared to the demand sizes; we observe a wavy pattern in function of Q, and find variance dampening in many instances.

We also characterized the order process by observing the number of units ordered over a time interval t, and compared it with the number of units demanded over the same period. This can be useful if the supplier reviews its orders on a periodic basis and as such aggregates the number of units that are ordered during this time interval. From this perspective, we always have order variance amplification, which is generally increasing in Q, albeit not always linearly and not always strict increasing in Q. We also find that the degree of order variance amplification is significantly less as the time interval t is longer (and it converges to 1 if $t \to \infty$). Our numerical analysis revealed that these measures strongly depend on the distribution of demand sizes.

Appendix

The calculation of the order process characteristics are to a large extent based on the values for b_{S-j}, where $\frac{1}{\lambda} b_{S-j}$ refers to the expected time spent in inventory position $S-j$, for $S-j \in \{s+1, \ldots, S\}$. The values for b_{S-j} can be found using the recursive equations in (8). When the demand size follows a simple distribution, we can find closed-form expressions for these values. In this appendix we illustrate this for a Poisson distributed ($D \sim Pois\{\mu_{dk}\}$) or uniform distributed ($D \sim U\{1, \ldots, m\}$) demand size.

Proposition 6. *If $D \sim Pois\{\mu_{dk}\}$, then*

$$b_{S-j} = \frac{\mu_{dk}^j}{j!} \left(\sum_{k=0}^{\infty} k^j e^{-k\mu_{dk}} \right).$$

PROOF OF PROPOSITION 6. The proof of Proposition 6 is based on induction. For $j = 0$, Proposition 6 holds as

$$b_S = \frac{1}{1-d_0} = \frac{1}{1-e^{-\mu_{dk}}} = \sum_{k=0}^{\infty} e^{-k\mu_{dk}}.$$

Given that Proposition 6 holds for $b_{S-(j-i)}$, we prove that it also holds for b_{S-j}. From (8) we find that

$$\begin{aligned}
b_{S-j} &= \frac{1}{1-e^{\mu_{dk}}} \sum_{i=1}^{j} \left(\frac{\mu_{dk}^i}{i!} e^{-\mu_{dk}} \right) b_{S-(j-i)} \\
&= \frac{1}{1-e^{\mu_{dk}}} \sum_{i=1}^{j} \left(\frac{\mu_{dk}^i}{i!} e^{-\mu_{dk}} \right) \frac{\mu_{dk}^{j-i}}{(j-i)!} \left(\sum_{k=0}^{\infty} k^{j-i} e^{-k\mu_{dk}} \right) \\
&= \frac{\mu_{dk}^j}{j!} \frac{1}{1-e^{-\mu_{dk}}} \sum_{i=1}^{j} \binom{j}{i} \sum_{k=0}^{\infty} k^{j-i} e^{-(k+1)\mu_{dk}} \\
&= \frac{\mu_{dk}^j}{j!} \frac{1}{1-e^{-\mu_{dk}}} \sum_{k=0}^{\infty} e^{-(k+1)\mu_{dk}} \left((k+1)^j - k^j \right) \\
&= \frac{\mu_{dk}^j}{j!} \left(\sum_{k=0}^{\infty} k^j e^{-k\mu_{dk}} \right)
\end{aligned}$$

□

For uniform distributed demand sizes, we provide the closed-form formulae for b_{S-j} when $j \in \{0, \ldots, 2m+1\}$). Note that one can derive the formulae for larger values of j as well, but they soon become increasingly complex.

Proposition 7. *If $D \sim U\{1, \ldots, m\}$, then*

$$b_{S-j} = \begin{cases} 1 & \text{if} \quad j = 0, \\ \frac{1}{m}\left(1+\frac{1}{m}\right)^{j-1} & \text{if} \quad j \in \{1, \ldots, m\}, \\ \frac{1}{m}\left[\left(1+\frac{1}{m}\right)^{j-1} - \left(1+\frac{1}{m}\right)^{j-m-1} - \frac{j-m-1}{m}\left(1+\frac{1}{m}\right)^{j-m-2}\right] & \text{if} \quad j \in \{m+1, \ldots, 2m+1\}. \end{cases} \tag{20}$$

Proof of proposition 7. As $d_0 = 0$, we immediately find $b_S = 1$ from (8).

For $j \in \{1, \ldots, m\}$, we find b_{S-j} through induction. From (8), we know that $b_{S-1} = \frac{1}{m} b_S = \frac{1}{m}\left(1+\frac{1}{m}\right)^0$. Assume that b_{S-j} with $j \in \{1, \ldots, m\}$ holds for $j = x$, then from (8) it follows that

$$\begin{aligned}
b_{S-x-1} &= \frac{1}{m}\left(b_{S-x} + \sum_{i=1}^{x} b_{S-x+i}\right) \\
&= \frac{1+m}{m} b_{S-x} \\
&= \left(\frac{1}{m}+1\right)\left(\frac{1}{m}\left(1+\frac{1}{m}\right)^{x-1}\right),
\end{aligned}$$

which proves (20) for $j = x+1$.

We find b_{S-j} for $j \in \{m+1, \ldots, 2m+1\}$ also based on induction. If $j = m+1$, we have $d_j = 0$. Then, Eq. (8) gives

$$b_{S-m-1} = \sum_{i=1}^{m} \frac{1}{m} b_{S-(m+1)+i} = \frac{1}{m}\left(\sum_{i=1}^{m+1} b_{S-(m+1)+i} - b_S\right) = \frac{1}{m}\left[\left(1+\frac{1}{m}\right)^m - 1\right],$$

as

$$\sum_{i=1}^{m+1} b_{S-(m+1)+i} = \left(b_{S-m} + \sum_{i=1}^{m} b_{S-m+i}\right) = (1+m) b_{S-m},$$

with b_{S-m} derived above. Next, we show that b_{S-j} with $j \in \{m+1, \ldots, 2m+1\}$ holds for $j = x+1$, if it holds for $j = x$:

$$\begin{aligned}
b_{S-x-1} &= \frac{1}{m} \sum_{i=1}^{m} b_{S-x-1+i} \\
&= \frac{1}{m}\left(b_{S-x} - b_{S-x+m} + \sum_{i=1}^{m} b_{S-x+i}\right) \\
&= \frac{1}{m}\left(b_{S-x} - b_{S-x+m} + m b_{S-x}\right) \\
&= \left(1+\frac{1}{m}\right) b_{S-x} - \frac{1}{m^2}\left(1+\frac{1}{m}\right)^{x-m-1} \\
&= \frac{1}{m}\left[\left(1+\frac{1}{m}\right)^{x} - \left(1+\frac{1}{m}\right)^{x-m} - \frac{x-m}{m}\left(1+\frac{1}{m}\right)^{x-m-1}\right]
\end{aligned}$$

□

References

Arrow, K.J., T. Harris, J. Marschak. 1951. Optimal inventory policy. *Econometrica* **19**(3) 250–272.

Blinder, A.S. 1981. Retail inventory behavior and business fluctuations. *Brookings Papers on Economic Activity* (2) 443–520.

Burbidge, J.L. 1961. The new approach to production. *Production Engineer* **40**(12) 769–784.

Cachon, G.P. 1999. Managing supply chain demand variability with scheduled ordering policies. *Management science* 843–856.

Caplin, A.S. 1985. The variability of aggregate demand with (s,S) inventory policies. *Econometrica* **53**(6) 1395–1409.

Caplin, A.S., J. Leahy. 2010. Economic theory and the world of practice: a celebration of the (s,S) model. *The Journal of Economic Perspectives* **24**(1) 183–202.

Chen, L., H.L. Lee. 2012. Bullwhip effect measurement and its implications. *Operations Research* **60** 771–784.

Cordeiro, J.D., J.P. Kharoufeh. 2011. *Wiley encyclopedia of operations research and management science*. John Wiley & Sons, New York.

Federgruen, A., Y.S. Zheng. 1992. An efficient algorithm for computing an optimal (r,Q) policy in continuous review stochastic inventory systems. *Operations research* **40**(4) 808–813.

Iglehart, D.L. 1963. Optimality of (s,S) policies in the infinite horizon dynamic inventory problem. *Management Science* **9**(2) 259–267.

Kelle, P., A. Milne. 1999. The effect of (s,S) ordering policy on the supply chain. *International Journal of Production Economics* **59**(1-3) 113 – 122.

Latouche, G., V. Ramaswami. 1999. *Introduction to matrix analytic methods in stochastic modeling*. ASA-SIAM series on statistics and applied probability, Society for Industrial and Applied Mathematics.

Lee, H. L., V. Padmanabhan, S. Whang. 2004. Comments on "information distortion in a supply chain: the bullwhip effect". *Management Science* **50**(12) 1887–1893.

Lee, H.L., V. Padmanabhan, S. Whang. 1997. Information distortion in a supply chain: the bullwhip effect. *Management science* 546–558.

Li, X., V. Sridharan. 2008. Characterizing order processes of using (r,nQ) inventory policies in supply chains. *Omega* **36**(6) 1096–1104.

Lucantoni, D. 1993. The BMAP/G/1 queue: a tutorial. *Performance Evaluation of Computer and Communication Systems* 330–358.

Nahmias, S. 1997. *Production operation analysis, third edtion*. McGraw Hill International.

Neuts, M.F., J.M. Li. 1996. An algorithm for the P(n,t) matrices of a continuous BMAP. S.R. Chakravarthy, A.S. Alfa, eds., *Matrix-analytic methods in stochastic models*. Marcel Dekker, 7–19.

Porteus, E.L. 1971. Optimality of generalized (s,S) policies. *Management Science Series A-theory* **17**(7) 411–426.

Potter, A., S.M. Disney. 2006. Bullwhip and batching: An exploration. *International Journal of Production Economics* **104**(2) 408–418.

Scarf, H. 1960. The optimality of (s,S) policies in the dynamic inventory problem. *Mathematical Methods in the Social Sciences* 196–202.

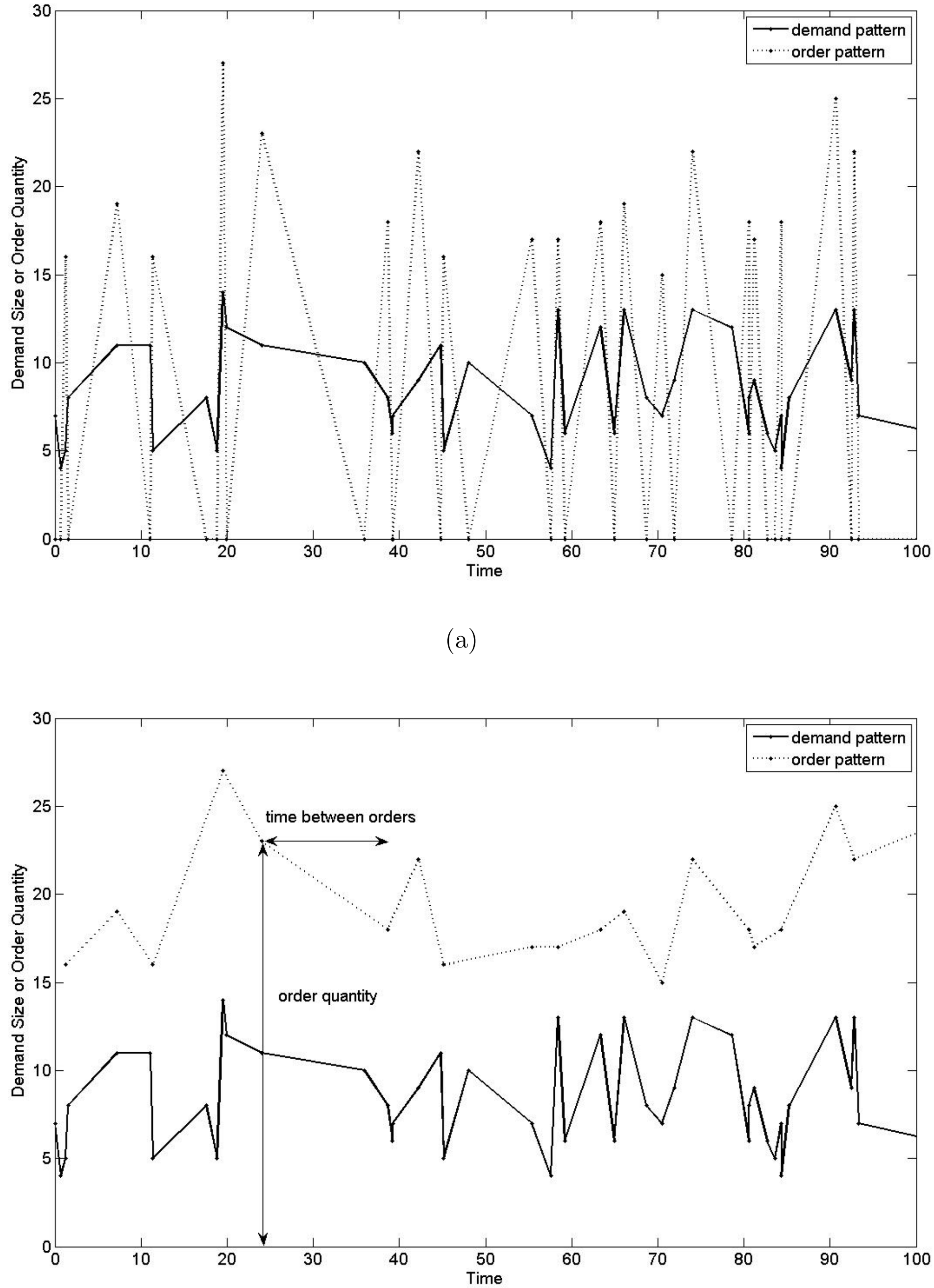
Schneider, H., D.A.N.B. Rinks, P. Kelle. 1995. Power approximations for a two-echelon inventory system using service levels. *Production and Operations Management* **4**(4) 381–400.

Schultz, C.R. 1983. Computing demand properties at the wholesale warehouse level. *Naval Research Logistics Quarterly* **30**(1) 37–48.

Teunter, R.H., L. Duncan. 2008. Forecasting intermittent demand: a comparative study. *Journal of the Operational Research Society* **60**(3) 321–329.

Veinott, A.F. 1966. On the optimality of (s,S) inventory policies: new conditions and a new proof. *SIAM Journal on Applied Mathematics* **14**(5) pp. 1067–1083.

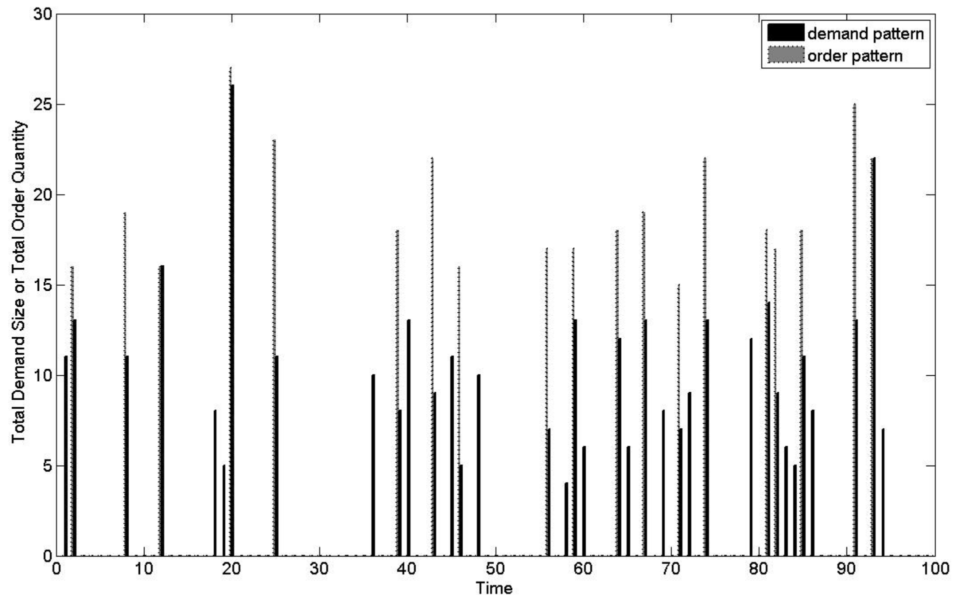
Zipkin, P.H. 2000. *Foundations of inventory management*. Irwin.



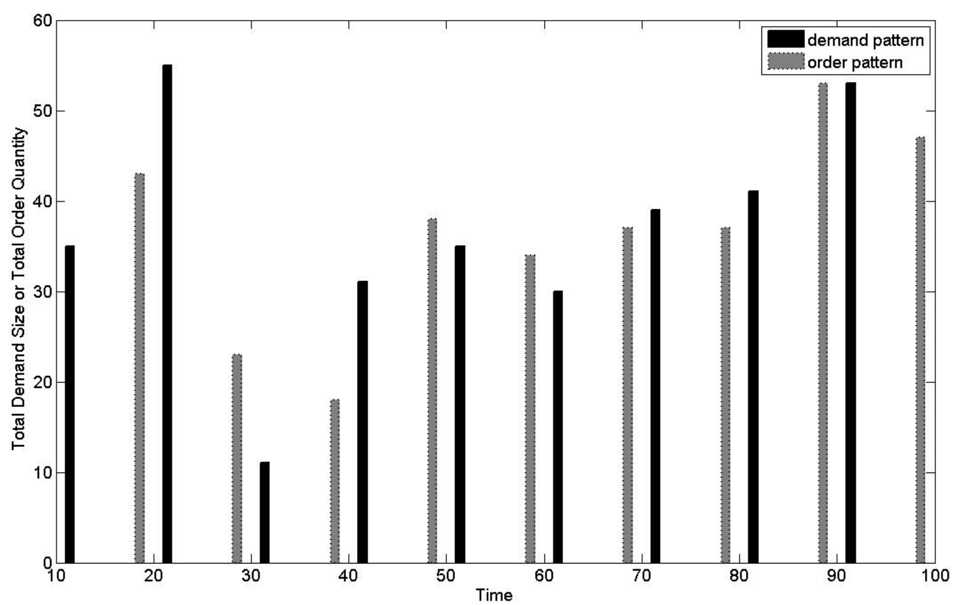
(a)

(b)

Figure 1: Simulation of a compound Poisson demand and its corresponding orders, placed by an (s, S) policy with $S - s = 15$, looking at demand arrivals and orders placed (first approach).



(a) $t = 1$



(b) $t = 10$

Figure 2: Simulation of a compound Poisson demand and its corresponding orders, placed by an (s, S) policy with $S - s = 15$, looking at the number of units demanded and ordered during interval t (second approach).

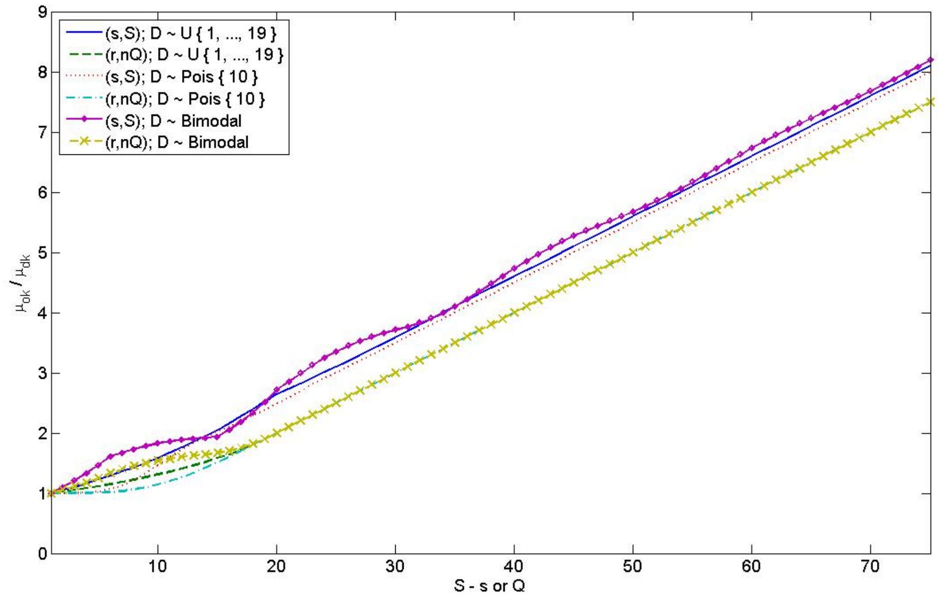


Figure 3: Ratio of the average order size over the average demand size (equivalent to time between orders over time between demand arrivals) in function of the batching parameter $Q = S - s$

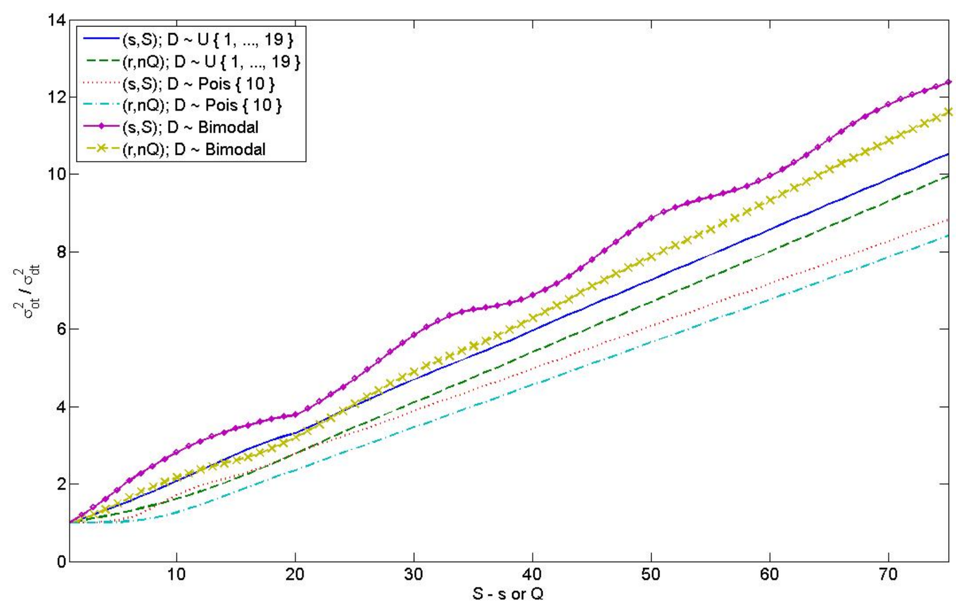


Figure 4: Variance amplification ratio of the time between orders placed versus the time between demand arrivals

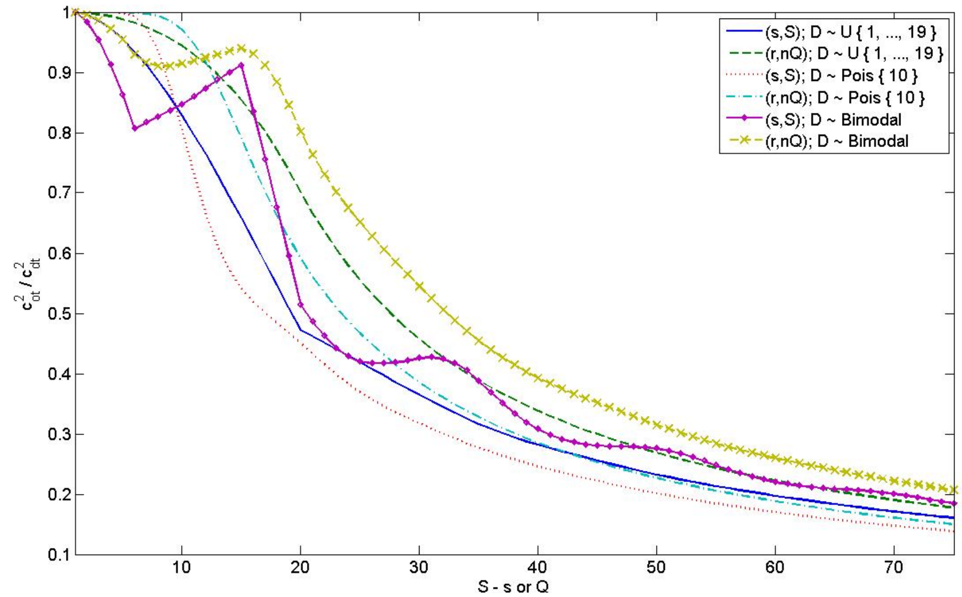


Figure 5: Squared coefficient of variation of the time between orders placed

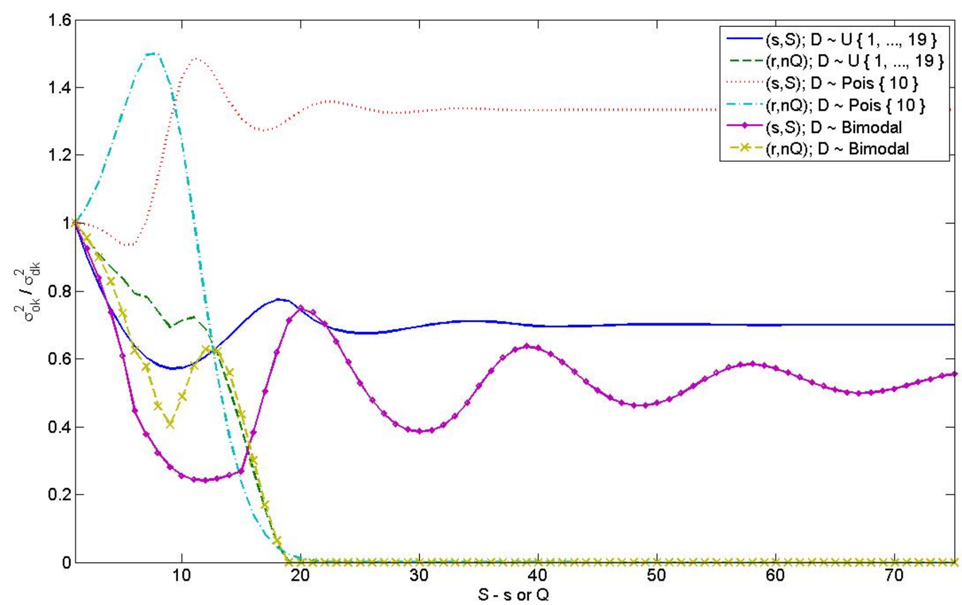


Figure 6: Variance amplification ratio of the order sizes versus the demand sizes

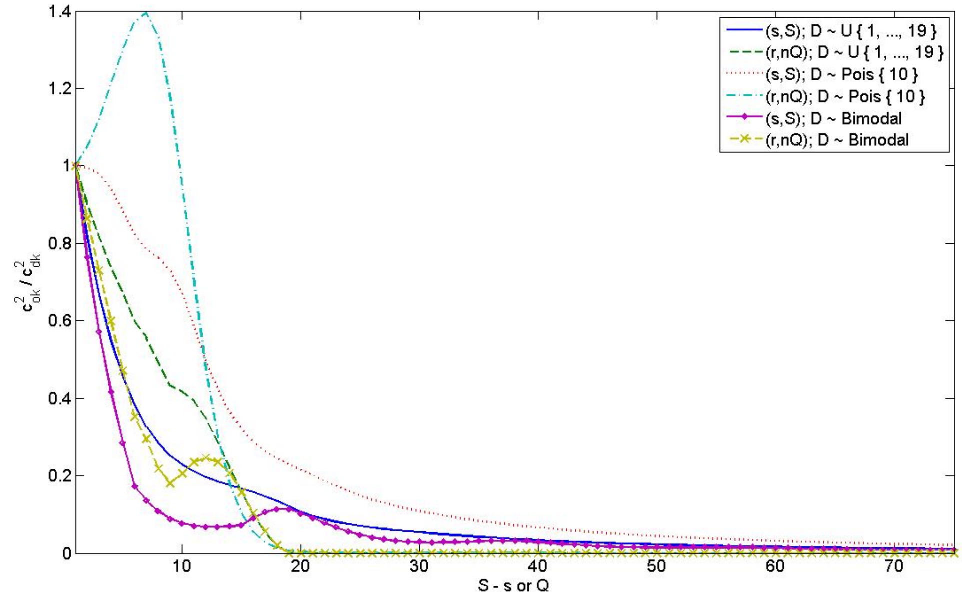


Figure 7: Ratio of the squared coefficient of variation of order sizes versus the scv of demand sizes

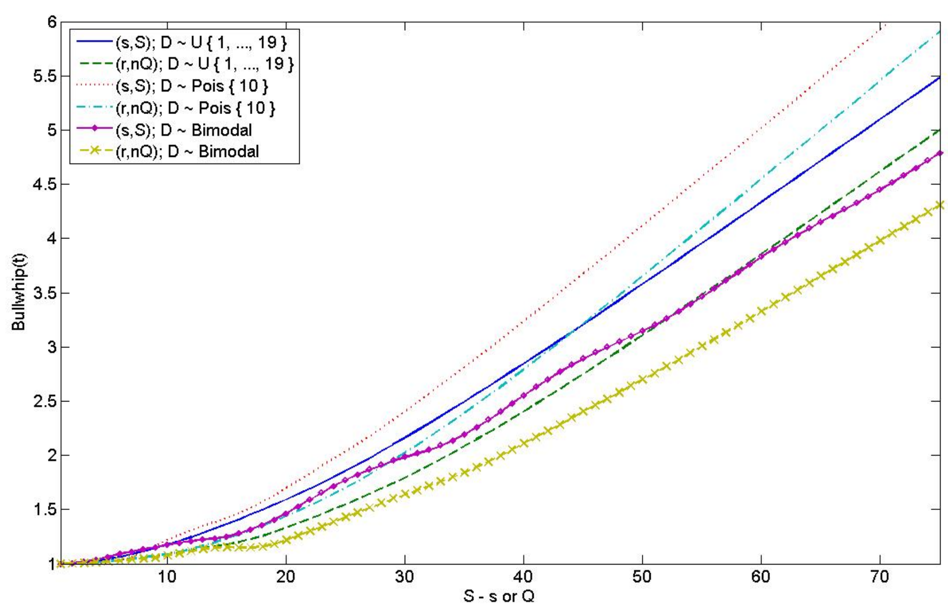


Figure 8: Variance amplification ratio of the number of items ordered versus the number of items demanded in interval $t = 1$

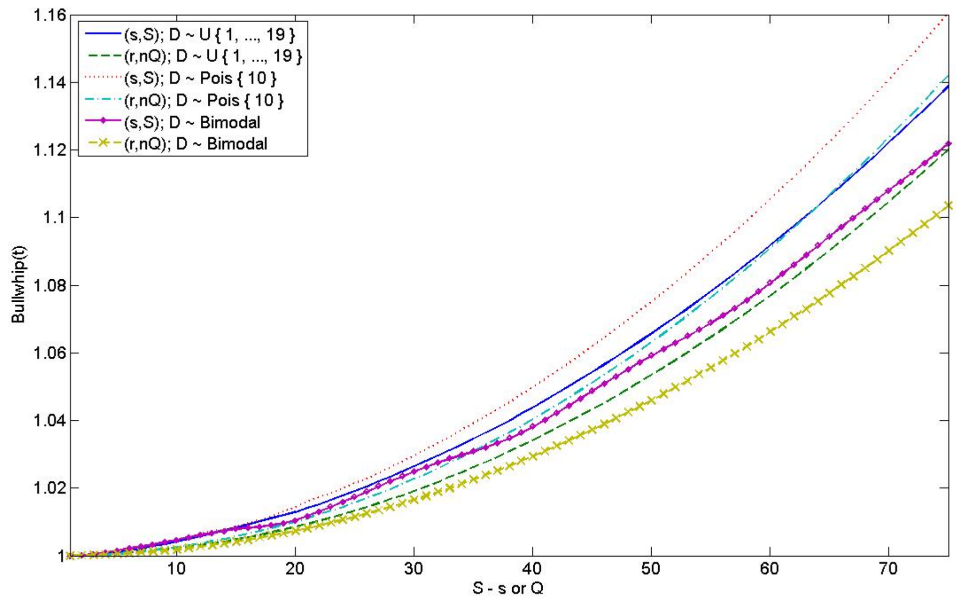


Figure 9: Variance amplification ratio of the number of items ordered versus the number of items demanded in interval $t = 60$

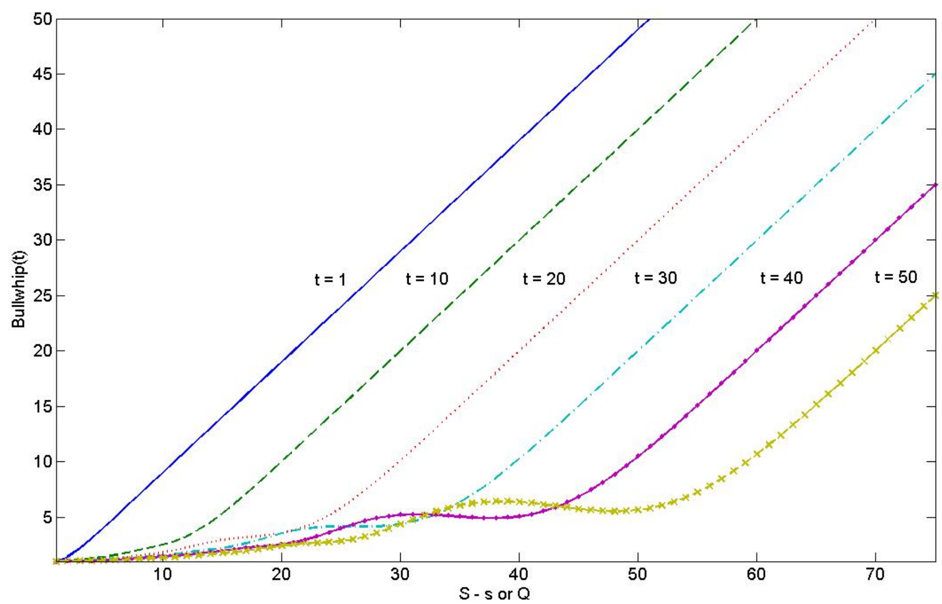


Figure 10: Variance amplification ratio of the number of items ordered versus the number of items demanded for single unit demand

FACULTY OF ECONOMICS AND BUSINESS
Naamsestraat 69 bus 3500
3000 LEUVEN, BELGIË
tel. + 32 16 32 66 12
fax + 32 16 32 67 91
info@econ.kuleuven.be
www.econ.kuleuven.be