Coordinating lead-time and safety stock decisions in a two-echelon supply chain with autocorrelated consumer demand

ROBERT N. BOUTE [* 1,2], STEPHEN M. DISNEY [3],
MARC R. LAMBRECHT [2] and BENNY VAN HOUDT [4]

[1] *Operations & Technology Management Center, Vlerick Leuven Gent Management School. Vlamingenstraat 83, 3000 Leuven, Belgium.*

[2] *Research Center for Operations Management, Katholieke Universiteit Leuven. Naamsestraat 69, 3000 Leuven, Belgium.*

[3] *Logistics Systems Dynamics Group, Cardiff Business School, Cardiff University, Aberconway Building, Colum Drive, Cardiff, CF10 3EU, UK. E-mail: disneysm@cardiff.ac.uk.*

[4] *Department of Mathematics and Computer Science, University of Antwerp, Middelheimlaan 1, 2020 Antwerpen, Belgium. Email: benny.vanhoudt@ua.ac.be.*

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

Coordinating lead-time and safety stock decisions in a two-echelon supply chain with autocorrelated consumer demand

In this paper we study a two-echelon supply chain with a retailer serving a consumer who is sensitive to marketing and pricing promotions. This results in either positively or negatively autocorrelated demand. Based on the observed consumer demand, the retailer replenishes with an adaptive order-up-to inventory policy satisfying a pre-specified fill rate. We assume the manufacturer produces the retailer's orders on a make-to-order basis and he decides on the lead time based on the retailer's order pattern. We analyze the interaction between the consumer demand process, the retailer's replenishment decision (and corresponding safety stock decision), and the manufacturer's production lead time. We encounter a lead time/ safety stock dependency problem – the retailer's replenishment decision depends on the expected manufacturer's lead time, whereas the actual manufacturing lead time depends on the replenishment decision (order size) – and develop an exact iterative procedure to solve this interaction effect. Surprisingly, given equal variability, a negatively autocorrelated, period-to-period oscillatory consumer demand provides shorter lead times and lower safety stocks as opposed to a positively autocorrelated, meandering consumer demand.

Keywords: production/inventory system, supply chain coordination, lead-time/safety stock interaction, operations/marketing interface

1. Introduction

Supply chain coordination has been a central research theme in numerous papers over the last five to ten years. Bernstein et al. (2006) analyze supply chain coordination through pricing schemes and Cachon (2003) describes coordination with contracts. Supply chains can also be coordinated by controlling the upstream variability propagation (known as the bullwhip effect), see e.g. Balakrishnan et al. (2004). In this paper we also study a supply chain coordination problem, more specifically the issue of coordination the retailer's safety stock requirements and the manufacturer's lead time decisions. It is commonly known that the manufacturer's lead time has a direct impact on the retailer's safety stocks: longer and more variable lead times require larger safety stocks. But there is also an impact in the opposite direction: the lead time will vary according to the order

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

stream of the retailer and its variability. Hence we have an interaction effect between safety stocks (retailer level) and lead times (manufacturing level) requiring a joint – coordinated – decision process. The resulting production/inventory system with endogenous lead time poses some challenging methodological issues. This is the main research issue in this paper.

We consider a basic retailer-manufacturer supply chain with a consumer demand sensitive to marketing and pricing promotions, resulting in either positively autocorrelated (meandering) sales or negatively autocorrelated (period-to-period oscillatory) behavior. We analyzed a large number of consumer demand patterns (weekly POS data) for consumer packaged non food products. We examined both branded products and private label products. For the regular 'turn' business, positively autocorrelated demand patterns seem to dominate, which is confirmed by Disney et al. (2006) who analyzed P&G's home care and family care product categories. In the presence of weekly promotions, however, we may obtain negative autocorrelation, due to consumers stockpiling during the promotion period and deceleration before and after the promotion. In the marketing literature, this is referred to as pre- and postpromotion dips (Macé and Neslin, 2004). Stockpiling is the propensity of consumers to increase their inventories above normal levels either by purchasing the category earlier, or by purchasing greater-than-normal quantities (Neslin et al., 1985). Deceleration is the willingness of consumers to deplete their inventories below normal levels by 'holding out' for an anticipated promotion (Mela et al., 1998). These behaviors create negatively autocorrelated demand behavior.

The impact of (price) promotions on consumer demand behavior is extensively described in the marketing literature. It is important to study these dynamics, since such behaviors influence profitability because they decrease the incremental sales generated by promotions (Blattberg and Neslin, 1993; Hendel and Nevo, 2006). Macé and Neslin (2004) empirically studied the relationships between pre- and postpromotion dips in weekly store data, and find that these dips are stronger for high-priced, frequently promoted, mature, high-market-share products.

In this paper we leave the exact reason for positively or negatively autocorrelated demand processes to the Marketing researchers, but we show that it has an important impact on the performance of the supply chain in terms of safety stocks and lead times. This reinforces once more the importance of coordinating marketing and operations decisions along the chain. Contrary to what might be intuitively expected, we find that negatively autocorrelated, and hence period-to-period oscillatory, sales results in shorter lead times and lower safety stocks as opposed to a positively autocorrelated, meandering, consumer demand. Note that in this paper we study the impact of autocorrelation, rather than the overall variability in demand, which can be caused by (price) promotions. We refer to Raju (1992) who relates the promotional activity in a product category to its variability in sales and Boute et al. (2007) who analyze the operational impact of this demand variability on lead times and safety stocks.

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

The issue of coordinating the retailer's safety stocks with the manufacturer's lead times is valid in a setting where the manufacturer produces his retailer's orders on a make-to-order basis. Lotus Bakeries, an industrial bakery, inspired us to conduct this research. They produce authentic specialities in the biscuit and cake world: caramelized biscuits, waffles, frangipane, and cake specialities among others. For certain products, a make-to-order policy is employed for a major retailer due to specific packaging requirements with the retailer's label on the product, sometimes combined with a specific, temporary, promotion. As the products have a limited shelf live and the retailer's orders fluctuate every period, a make-to-stock policy is excluded for these products. Clearly, short lead times and low inventories are important to guarantee to freshness of the products.

In this type of setting, the variability of the order pattern (interarrival times and order quantities), combined with the variability of the production process and the utilization level of the manufacturing system all have an impact on the lead time (Hopp and Spearman, 2001). This lead time is in turn a prime determinant in setting the safety stocks at the retailer. Unfortunately, much of the operations management literature separates the issues of production and inventory control decisions. However, inventory replenishment decisions influence production by initiating orders, and production decisions influence inventory by completing and delivering orders to inventory. We therefore model a two-echelon (retailer-manufacturer) supply chain as a production/inventory system and as such we treat lead times as endogenous variables. This means that we do not merely assume the replenishment lead time to be a random exogenous variable, but we include the impact of the replenishment decision on the production lead times and use these lead times in our inventory model. We propose an iterative procedure to solve this interaction effect.

The paper is organized as follows. The next section presents a brief overview of the relevant literature. Section 3 describes our research model and derives expressions for the orders generated by the retailer. Section 4 presents an iterative procedure to determine manufacturing lead times and section 5 is devoted to the analysis of the combined production/inventory system. Section 6 provides a numerical experiment and Section 7 concludes.

2. Literature review

There are three streams of research related to our work: (1) replenishment rules and forecasting under autocorrelated demand; (2) production/inventory systems with endogenous lead times; and (3) phase type (PH) distributions and queueing models using matrix-analytic techniques.

Several papers discuss supply chains under an autocorrelated demand. Kahn (1987) and Lee et al. (1997) were among the first to demonstrate the existence of variance amplification upstream in the chain (bullwhip) when the retailer follows a base-stock policy and demand is positively correlated. Dejonckheere et al. (2003) extend this result by showing that an exponential smoothing (ES)

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

or moving average (MA) forecast produces bullwhip for all demand processes (including AR and ARMA). Zhang (2004b) studies the role of forecasting for AR(1) demands and concludes that the minimum Mean Squared Error (MSE) forecasting method minimizes the variance of the forecasting error among all linear forecasting methods, and therefore leads to the lowest inventory costs. Alwan et al. (2003) employ this optimal MSE forecasting scheme and determine the underlying time-series model of the resulting ordering process. They show that when consumer demand is negatively correlated (AR demand), the variability in order quantities is dampened with respect to the observed demand, as opposed to the ES and MA forecasting methods, which always create bullwhip independent of the demand process. This result is of great importance for our paper.

The interaction between safety stocks and lead times is generally studied in production/inventory systems with endogenous lead times. Base-stock controlled production/inventory systems have been studied, among many others, by Song and Zipkin (1996), Sox et al. (1997), Jemaï and Karaesmen (2005) in continuous time with exponential (single unit) demand processes. Boute et al. (2007) propose a solution method for discrete time production/inventory systems with a random IID integer consumer demand. The interaction between order release models and lead times is somewhat related to this problem, see De Kok and Fransoo (2003), Pahl et al. (2005) and Selcuk et al. (2009). Graves (1988) provides an excellent review and critique of the research literature on safety stocks for manufacturing systems, and proposes a model to include consideration of the flexibility of the production stage in planning safety stocks.

The methodology in this paper is based on Phase Type (PH) distributions (see e.g. Horváth and Telek (2002)), Markov chains of the GI/M/1 type (Neuts, 1981) and matrix analytic methods (Latouche and Ramaswami, 1999). The domain of matrix analytic techniques was advocated by Neuts (1981, 1989). These methods are popular as modeling tools because they can be used to construct and analyze a wide class of stochastic models. They are applied in several areas, of which the performance analysis of telecommunication systems is one of the most notable. We refer to Bini et al. (2005) for its recent algorithmic developments. Software tools both in Fortran and MATLAB were made available by Bini et al. (2006).

The paper contributes to the existing literature in three ways. First, we compute the lead time distribution when production orders are generated by a periodic review base-stock policy and MSE forecasting for AR(1) demand processes. Second, we solve the lead time dependency problem that arises in this context: orders are dependent on the lead time distribution and vice versa. Third, we find an exact solution for the inventory distribution and the safety stock requirements of the corresponding production/inventory system, taking the correlation between demand and lead times into account. This paper illustrates the important interplay between the consumer demand process, the retailer's replenishment decision (and corresponding safety stocks), and the manufacturer's production lead time.

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

3. Model description

We study a basic two-echelon supply chain with one retailer and one manufacturer. Consumer demand D_t is observed at the beginning of a time period t, but it need not be fulfilled until the end of the period (unfilled demand is backordered). Retailer's inventory levels are reviewed after demand is satisfied, and an order O_t is placed and sent to the manufacturer's production. The manufacturer does not hold a finished goods inventory, but produces on a make-to-order basis. Once an entire order is produced, it replenishes the retailer's inventory (no transfer batch). The time from the period an order is placed to the period that it replenishes the retailer's inventory, is the lead time T_p.

Supply lead times are endogenously generated by the manufacturer's finite capacity production system. The production system is capacitated in the sense that there is a single processor that sequentially processes items one at the time on a first-come-first-served basis. When the server is busy, the order joins the queue of unprocessed orders. The queueing process at the manufacturer implies that the retailer's replenishment lead times are stochastic and correlated with the order quantity.

In the following, we describe in more detail the consumer demand process, the retailer's inventory policy and its forecasting model, the order process generated by the retailer and the production process at the manufacturer.

3.1 Consumer demand process

There are a number of potential stochastic processes that can be assumed for the consumer demand process, ranging from a simple IID process to a non-stationary process. One industrially relevant, flexible, correlative demand process that has often been studied in the supply chain literature is the first-order autoregressive or AR(1) model. Traditionally, an AR(1) demand is modeled as

$$D_t = \mu + \phi D_{t-1} + \varepsilon_t, \qquad |\phi| < 1, \tag{1}$$

where D_t is the demand observed in period t, ϕ is the first-order autocorrelation coefficient, μ is a constant that determines the mean of the demand, i.e., $E(D) = \mu/(1-\phi)$, and ε_t is an IID random error with mean 0 and variance σ^2. The assumption of $|\phi| < 1$ assures that the demand process is covariance stationary.

Sometimes, Eq. (1) is re-written as a mean-centered demand pattern, $D_t = E(D) + \phi(D_{t-1} - E(D)) + \varepsilon_t$, which omits the parameter μ.

For the purpose of this paper, we use a slightly different notation. We assume consumer demand

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

follows the following correlated process:

$$D_t = \phi\ D_{t-1} + (1-\phi)\, G_t, \tag{2}$$

where $G_t = (\mu + \varepsilon_t)/(1-\phi)$ is a random IID term with mean $E(G) = \mu/(1-\phi) > 0$ and variance $Var(G) = \sigma^2/(1-\phi)^2$. The error term is here given by $(1-\phi)\, G_t$. For $-1 < \phi < 0$, the demand process is negatively correlated and will exhibit period-to-period oscillatory behavior. For $0 < \phi < 1$, the demand process is positively correlated which will be reflected by a wandering or meandering sequence of observations.

As will become clear later in this paper, this notation reveals some elegant formulations and remarkable similarities between the demand pattern and the order pattern when demand is forecasted using the MSE forecasting technique. This considerably reduces the complexity of the queueing analysis, which is used to compute lead times.

It can be shown that when $0 \leq \phi < 1$, the minimum and maximum demand are given by the minimum and maximum values of G respectively, or $d_{min} = g_{min}$ and $d_{max} = g_{max}$. When $-1 < \phi \leq 0$, the minimum and maximum demand are given by, respectively, $d_{min} = (g_{min} + \phi\ g_{max})/(1+\phi)$ and $d_{max} = (g_{max} + \phi\ g_{min})/(1+\phi)$. This can be used to provide a condition on g_{min}, g_{max} and ϕ to avoid negative values for the observed consumer demand.

Whereas in the traditional notation (Eq. (1)) the error term ε has mean 0, and the average demand equals $\mu/(1-\phi)$, in our notation (Eq. (2)), the average demand $E(D) = E(G)$, and the variance of demand $Var(D) = \frac{1-\phi}{1+\phi}\ Var(G)$, implying that the demand decreases in variance as ϕ increases towards 1.

One can view ϕ as a marketing parameter related to the impact of promotion on demand. A negative value for ϕ means that the consumer's buying behavior is highly influenced by a promotion in the sense that consumers increase their purchases in the promotion week, and decelerate before and after the promotion. A positive ϕ value denotes a less aggressive reaction to the promotion: product demand is related to previous period's demand, rather than influenced by a price promotion.

3.2 Retailer's inventory policy and forecast method

The retailer controls his inventory with the standard periodic review base-stock policy, which is locally optimal when there is no fixed ordering cost and both holding and shortage costs are proportional to the volume of on-hand inventory or shortage (Nahmias, 1997; Zipkin, 2000). The base-stock level is determined to achieve a desired service level. Here, the service level is defined as the fraction of consumer demand that can be immediately satisfied from the inventory on hand, known as the *fill rate* (Zipkin, 2000).

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

Let O_t represent the order quantity in period t to be delivered in period $t + T_p + 1$, with T_p the stochastic lead time for the manufacturer to produce/fulfill an order, and let S_t be the base-stock level, which equals the inventory position after placing the order in period t. The timing of events (first receive goods from manufacturer, then satisfy demand and finally place order) and the conservation of flow imply that

$$O_t = S_t - S_{t-1} + D_t. \tag{3}$$

The base-stock level is the sum of the forecasted *lead time demand* and the safety stock. We define lead time demand as the total demand during the risk period, L, or $D_t^L = \sum_{i=1}^{L} D_{t+i}$, and let $\widehat{D}_t^L$ be its forecast. The risk period (the time between placing a replenishment order until receiving the subsequent replenishment order) is equal to the review period (= 1 period) plus the replenishment lead time (= T_p periods). Since lead time is stochastic, the lead time demand is a stochastic sum of a random number of random variables, or

$$S_t = \widehat{D}_t^L + SS, \tag{4}$$

with SS the safety stock required to achieve the desired service level. Due to the autocorrelation in demand, the demand forecast is updated when a new demand realization occurs. Hence lead time demand forecast changes every period, and the base-stock level S_t in Eq. (4) is *adaptive* over time (Graves, 1999). Combining (3) and (4), we obtain that the order quantity is equal to the observed demand plus the difference between the lead time demand forecast of the current period versus the previous period:

$$O_t = D_t + \left(\widehat{D}_t^L - \widehat{D}_{t-1}^L\right). \tag{5}$$

Several techniques are available to forecast lead time demand. The moving average (MA) and exponential smoothing (ES) forecast methods are widely employed because of their simplicity and ease of implementation. However, when demand follows an AR(1) process, the minimum Mean Squared Error (MSE) forecasting method is the preferred forecasting scheme. It is the conditional expectation of future demand, given current and previous demand observations (Box et al., 1994). Since it minimizes the expected mean squared forecast error, it is the preferred method when inventory cost is of primary concern (Zhang, 2004b). This forecasting technique assumes however that the underlying parameters of the demand model are known or that a suitable amount of demand data is available to estimate these parameters accurately.

For the AR(1) demand process given by Eq. (2), the MSE forecast of the i-period ahead demand

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

forecast is given by

$$\widehat{D}_{t+i} = \phi^i D_t + \left(1 - \phi^i\right) \cdot E(G). \tag{6}$$

By plugging the single period MSE forecast into the expression of the lead time demand forecast, $\widehat{D}_t^L = \sum_{i=1}^{L} \widehat{D}_{t+i}$, we obtain:

$$\widehat{D}_t^L = \left(\frac{\phi\left(1 - E\left(\phi^L\right)\right)}{1-\phi} \right) D_t + \left(E(L) - \frac{\phi\left(1 - E\left(\phi^L\right)\right)}{1-\phi} \right) E(G). \tag{7}$$

The MSE forecasting scheme clearly explicitly takes the autocorrelated demand structure into account, which is not the case in the non-optimal ES and MA forecasts. Moreover, instead of forecasting the one-period ahead demand and multiplying this with the lead time L, this technique calculates the forecast of the demand over the lead time horizon L (Kim et al., 2006).

3.3 Order process sent to production

Substituting (7) into (5) returns the order process generated by the retailer's base-stock policy:

$$O_t = \frac{1 - E\left(\phi^{L+1}\right)}{1-\phi} D_t - \frac{\phi\left(1 - E\left(\phi^L\right)\right)}{1-\phi} D_{t-1}. \tag{8}$$

The retailer's order quantity is a linear combination of the observed demand in the current period and the previous period. When we substitute (2) into (8) we obtain the following expression for the order process:

$$O_t = E\left(\phi^{L+1}\right) \cdot D_{t-1} + \left(1 - E\left(\phi^{L+1}\right)\right) \cdot G_t, \tag{9}$$

which is surprisingly similar to the expression of the observed consumer demand. Substituting ϕ by $E\left(\phi^{L+1}\right)$ in the expression of the demand process (Eq. (2)), results in the order process, given by Eq. (9). This equation actually has an ARMA(1,1) structure, similar to, but different to the AR(1) structure (Zhang, 2004a).

This order process is sent to the manufacturer's production queue. It is worthwhile analyzing some characteristics of this process. From Eq. (9) we can determine the order variance as

$$\begin{aligned} Var(O) &= \left(E\left(\phi^{L+1}\right)\right)^2 Var(D) + \frac{(1+\phi)\left(1 - E\left(\phi^{L+1}\right)\right)^2}{1-\phi} Var(D) \\ &= \left[1 + \frac{2\phi\left(1 - E\left(\phi^L\right)\right)\left(1 - E\left(\phi^{L+1}\right)\right)}{1-\phi} \right] Var(D). \end{aligned} \tag{10}$$

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

From Eq. (10) it can be shown that when consumer demand is positively correlated, the order variance is amplified with respect to the demand variance. This phenomenon is referred to as the bullwhip effect.

$$\begin{aligned} Var(O) > Var(D) &\Leftrightarrow 1 + \frac{2\phi\left(1 - E\left(\phi^{L}\right)\right)\left(1 - E\left(\phi^{L+1}\right)\right)}{1-\phi} > 1 \\ &\Leftrightarrow 2\phi\left(1 - E\left(\phi^{L}\right)\right)\left(1 - E\left(\phi^{L+1}\right)\right) > 0 \\ &\Leftrightarrow \phi > 0. \end{aligned} \tag{11}$$

Analogous to Eq. (11), it can be derived that when the autocorrelation coefficient is negative, there is an anti-bullwhip, or *de-whip* effect, which means that the orders are smoothed compared to the demand pattern.

$$Var(O) < Var(D) \Leftrightarrow \phi \leq 0. \tag{12}$$

This contrasts sharply with the traditional, non-optimal, MA and ES forecasting techniques, which always result in variance amplification, independent of the observed demand pattern (Dejonckheere et al., 2003). A similar conclusion was obtained by Alwan et al. (2003). If the autocorrelation coefficient is zero, we obtain an IID consumer demand, where orders equal the observed demand.

This result is important for our analysis. The sign of the correlation coefficient determines whether orders are amplified in variability towards the manufacturer, or not. Since the manufacturer produces on a make-to-order basis, this has an impact on lead times. Positively correlated demand amplifies variability in orders, with increasing production/replenishment lead times as a consequence. Negatively correlated demand dampens the order variability, resulting in shorter lead times.

3.4 Production model

We characterize the manufacturer's production stage by a discrete time single server queue that sequentially processes single units with stochastic service times. The service times of a single item, denoted by M, are IID random variables. We make use of phase-type (PH) distributions, since they can approximate any general distribution and their Markovian structure greatly simplifies the queueing analysis. To ensure stability of the queue, we assume that the utilization of the production facility (average batch production time divided by average batch interarrival time) is strictly smaller than one.

The time from the instant the order is placed to the point that the production of the entire batch is finished, is the *production* or response time, denoted by T_r. This response time corresponds to

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

the sojourn time in a single server queueing system with batch arrivals (equal to the replenishment orders) and deterministic inter-arrival times (equal to one (review) period).

Note that the response time T_r is not necessarily an integer number of periods. In our inventory model, however, events occur on a discrete time basis with a time unit equal to one period; therefore the *replenishment* lead time, denoted by T_p, is expressed in terms of an integer number of periods. From the response time distribution T_r we obtain the replenishment lead time distribution T_p by relying on the sequence of events in a period. In our sequence of events, the demand need not be fulfilled until the end of the period, i.e., after the receipt of produced items in inventory, and a replenishment order is placed after demand is satisfied (see Fig. 1).

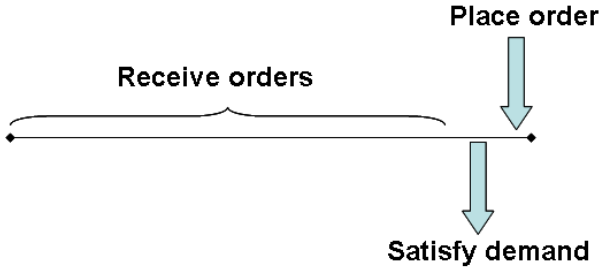


Figure 1: Sequence of events in a period: 1. receive produced orders, 2. satisfy demand, 3. place order

For instance, suppose that an order placed at the end of period t has a response time of 0.8 periods. This order quantity will be added to the inventory in the next period $t+1$, and can be used to satisfy demand in period $t+1$. Therefore the replenishment lead time is 0 periods, since the order can immediately be used to satisfy next period's demand. An order O_t with a production lead time of 1.4 periods is added to the inventory in period $t+2$ and can be used to satisfy demand D_{t+2}. Consequently we will treat the 1.4 period production lead time as an integer 1 period replenishment lead time. Hence, we round the response time T_r down to the nearest integer T_p (i.e., setting $T_p = \lfloor T_r \rfloor$) to obtain the (discrete) replenishment lead time. Note $T_p = 0$ implies the order arrives in the next period due to the sequence of events.

4. Determination of production lead times

4.1 Lead time dependency problem

The replenishment orders described by Eq. (9) load the production queue. The nature of this loading process relative to the available capacity and the variability it creates are the primary determinants of lead times (Karmarkar, 1993). According to the laws of factory physics, a more

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

variable arrival process at the production queue induces longer and more variable lead times (Hopp and Spearman, 2001). Consequently the order process determines the distribution of the lead times.

From Eq. (9) however, we see that the order process itself is dependent on the lead time distribution. In other words, we need the lead time distribution to determine order sizes. We consequently face a lead time dependency problem: the order process is dependent on the lead time distribution, while the lead time itself is dependent on the order process. In order to solve this mutual dependency problem, we develop an iterative procedure.

We start with an initial guess for the lead time distribution T_p^0. Typically, we select T_p^0 deterministically, equal to 0 periods. Next, for $n > 0$, we make use of T_p^{n-1} to determine the order pattern in Eq. (9). Given this expression for the order pattern, we determine the new lead time distribution T_p^n and repeat this procedure.

Notice, the only manner in which T_p^n affects the system behavior is through the term $E(\phi^{L_n+1})$ in Eq. (9), with $L_n = T_p^n + 1$. As soon as $\left|E(\phi^{L_n+1}) - E(\phi^{L_{(n-1)}+1})\right|$ drops below a predefined error value ϵ, e.g., $\epsilon = 10^{-14}$, we can consider T_p^n as sufficiently close to the actual lead time distribution T_p. We find that the lead time distribution converges towards the actual lead time distribution when $|\phi| < 1$. This assumption is not restrictive as $|\phi| < 1$ also assures that the demand process is covariance stationary.

4.2 Queueing model

To estimate the lead time distribution at iteration n, we develop a discrete time queueing model. By analyzing the characteristics of the replenishment orders, we implicitly analyze the characteristics of the production orders that arrive at the production queue. In a periodic review base-stock policy, the arrival process consists of batch arrivals with a fixed interarrival time (equal to the review period, i.e. 1 period) and with variable batch sizes, which are, in our model, correlated.

The service times of a single unit, denoted by M, are stochastic and IID according to a phase type (PH) distribution. The key idea behind PH distributions is to exploit the Markovian structure of the distribution to simplify the queueing analysis. Moreover, any general discrete distribution can be approximated in sufficient detail by means of a PH distribution (Horváth and Telek, 2002), since the class of discrete PH distributions is a versatile set that is dense within the set of all discrete distributions on the nonnegative integers (Bobbio et al., 2003; Latouche and Ramaswami, 1999; Neuts, 1989).

The computational complexity of our algorithm to compute the lead time distribution increases with the number of phases of the PH distributed service process. Therefore we want the service process to be PH-distributed with as few phases as possible. Since the lead time is expressed as an integer number of periods and the interarrival time is equal to one base period, we have the freedom to choose the time unit U of the queueing system in an appropriate manner (Bobbio et al., 2004).

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

When the time unit U is chosen as half the mean service time of a single item, i.e., $U = E(M)/2$, it is possible to match the first two moments of the single unit service times, $E(M)$ and $Var(M)$, by means of a PH distribution with only 2 phases (Boute et al., 2007). The PH distribution is then characterized by the pair (T, α), where T is a 2×2 substochastic matrix and α a 1×2 stochastic vector. Including more moments will lead to a higher number of phases.

When we choose U to be the time unit of our queueing system, this implies that orders placed every period arrive at the queue at time epochs $0, e, 2e, \ldots$, where $e \times U = 1$ period. The order size at these time epochs is driven by an underlying Markov process with state space $\{d_{min}, d_{min}+1, ..., d_{max}\}$, where d_{min} and d_{max} are respectively the minimum and maximum value the random variable D can attain (as defined in section 3.1). Indeed, according to Eq. (9), the order size generated at time te is determined as

$$O_{te} = E\left(\phi^{L+1}\right) D_{(t-1)e} + \left(1 - E\left(\phi^{L+1}\right)\right) G_{te}. \tag{13}$$

By keeping track of the state of the demand at time $(t-1)e$ we can determine the order size at time te. The state of the demand itself evolves as

$$D_{te} = \phi D_{(t-1)e} + (1-\phi) G_{te}, \tag{14}$$

which has an obvious Markovian nature. Using induction on t we easily establish that $E(O) = E(D) = E(G)$. From Eqs. (13-14) we see that if we know the value of $D_{(t-1)e}$, we can define the transition to the values of D_{te} and O_{te} (and their respective probabilities) from the value of G_{te} (and its probability function). This of course reduces the complexity of our Markov analysis, since we only need the value of $D_{(t-1)e}$ to determine the transition probabilities to both D_{te} and O_{te}.

The demand and order size resulting from (14) and (13), respectively, can be a real number. As it is more natural to have demands of integer size, the actual demand (determining the order size) is stochastically rounded to have size D^*_{te}:

$$D^*_{te} = \begin{cases} D_{te} & \text{if } D_{te} \in \mathbb{N}, \\ \lceil D_{te} \rceil & \text{with probability } D_{te} - \lfloor D_{te} \rfloor \text{ if } D_{te} \notin \mathbb{N}, \\ \lfloor D_{te} \rfloor & \text{with probability } \lceil D_{te} \rceil - D_{te} \text{ if } D_{te} \notin \mathbb{N}. \end{cases} \tag{15}$$

Suppose for instance that the arrival process generates a demand quantity of 5.8, then we round this to 5 units with a probability of 0.20 and to 6 units with a probability of 0.80. This (integer) number of units constitutes the demand which determines the batch size that has to be produced by the manufacturer. Notice, the expected value $E(D^*_{te}) = E(D_{te}) = E(D)$, meaning the expected

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

value is not affected.

Moreover, because only an integer number of items can be produced, the batch size passed to the manufacturer at time t has size O^*_{te}:

$$O^*_{te} = \begin{cases} O^+_{te} & \text{if } O^+_{te} \in \mathbb{N}, \\ \lceil O^+_{te} \rceil & \text{with probability } O^+_{te} - \lfloor O^+_{te} \rfloor \text{ if } O^+_{te} \notin \mathbb{N}, \\ \lfloor O^+_{te} \rfloor & \text{with probability } \lceil O^+_{te} \rceil - O^+_{te} \text{ if } O^+_{te} \notin \mathbb{N}, \end{cases} \tag{16}$$

where O^+_{te} is found by Eq. (13) when replacing $D_{(t-1)e}$ by $D^*_{(t-1)e}$. In order to simplify the notation, however, we will use respectively D_{te} and O_{te} instead of D^*_{te} and O^*_{te}, and assume in the remainder of this section that D_{te} and O_{te} are rounded according to Eqs. (15) and (16) respectively.

Discretizing the range of the demand and order sizes on the integer values is not only more natural, but also helps in computing the lead time distribution in an efficient manner. That is, it allows us to construct a Markov chain that has a considerably smaller state space, leading to a less stringent time and memory complexity for the numerical procedure involved.

4.3 Markov chain analysis

In order to set up the Markov chain to find the lead time distribution, we define the following random variables:

- t_n : the time of the n-th observation point, which we define as the n-th time epoch during which the server is busy,
- $a(n)$: the arrival time of the order in service at time t_n,
- V_n : the age of the order in service at time t_n, defined as the duration (expressed in the time unit of the queueing model, i.e., U) of the time interval $[a_n, t_n)$,
- C_n : the number of items part of the order in service that still need to either start or complete service at time t_n,
- S_n : the service phase at time t_n.

All events, such as arrivals, transfers from the waiting line to the server and service completions are assumed to occur at instants immediately after the discrete time epochs. This implies that the age of an order in service at some time epoch t_n is at least 1.

Thus, $(V_n, D_{a(n)}, C_n, S_n)$ forms a Markov chain on the state space $\mathbb{N}_0 \times \{x : x = d_{min}, d_{min} + 1, ..., d_{max}\} \times \{c \in \{1, 2, \ldots, d_{max}\}\} \times \{1, 2\}$, because V_n is a positive integer, $D_{a(n)}$ (the demand at

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

the time the order in service was placed) is an integer between d_{min} and d_{max}, C_n an integer between 1 and d_{max} and the PH service has two phases. In order to characterize its transition matrix, we start by deriving an expression for the probabilities $\Pr(G_{te} = g, D_{te} = k'|D_{(t-1)e} = k)$ for k, k' in $\{d_{min}, d_{min}+1, ..., d_{max}\}$ and g in $\{1, \ldots, g_{max}\}$. As a result from the stochastic rounding to integer demand values given by Eq. (15), these conditional probabilities, which we denote as $p^{(g)}(k, k')$, can be computed as:

$$p^{(g)}(k,k') = \Pr(G=g) \cdot \Big\{ 1_{\{k'-1<\phi k+\bar{\phi}g<k'\}} \left((\phi k + \bar{\phi} g) - \lfloor \phi k + \bar{\phi} g \rfloor \right) + 1_{\{\phi k+\bar{\phi}g=k'\}} + 1_{\{k'<\phi k+\bar{\phi}g<k'+1\}} \left(\lceil (\phi k + \bar{\phi} g) \rceil - (\phi k + \bar{\phi} g) \right) \Big\}, \quad (17)$$

where $\bar{\phi} = (1 - \phi)$ and the indicator function $1_{\{A\}}$ is 1 if the event A is true and 0 otherwise. Combining these probabilities with Eq. (16), we are now in a position to set up an expression for $\Pr(O_{te} = q, D_{te} = k'|D_{(t-1)e} = k)$, which we denote as $p_{[q]}(k, k')$:

$$p_{[q]}(k,k') = \sum_{g=1}^{g_{max}} p^{(g)}(k,k') \cdot \Big\{ 1_{q-1<\gamma k+\bar{\gamma}g<q\}} \left((\gamma k + \bar{\gamma} g) - \lfloor \gamma k + \bar{\gamma} g \rfloor \right) + 1_{\{\gamma k+\bar{\gamma}g=q\}} + 1_{\{q<\gamma k+\bar{\gamma}g<q+1\}} \left(\lceil (\gamma k + \bar{\gamma} g) \rceil - (\gamma k + \bar{\gamma} g) \right) \Big\}, \quad (18)$$

where $\gamma = E(\phi^{L+1})$ and $\bar{\gamma} = (1 - \gamma)$.

Let us now have a look at the evolution of the Markov chain $(V_n, D_{a(n)}, C_n, S_n)$. At each transition step, there are three possibilities. First, the current item in production stays in service and the phase of the service process may change. Second, the current item in service finishes production, and a new item of the same batch enters production. Third, the current item in service finishes production and when this is the last item of the batch, the complete batch is produced. The order quantity of the new batch that is taken in production is given by $p_{[q]}(k, k')$ according to Eq. (18). Let $(P)_{(a,k,r,s),(a',k',r',s')}$ be the transition probabilities of the Markov chain $(V_n, D_{a(n)}, C_n, S_n)$. These probabilities are then given by

$$(P)_{(a,k,r,s),(a',k',r',s')} = \begin{cases} T_{s,s'} & a' = a+1, k' = k, r' = r, \\ t_s \alpha_{s'} & a' = a+1, k' = k, r' = r-1 \geq 1, \\ t_s \alpha_{s'} p_{[q]}(k,k') & a' = \max(a-e+1, 1), r' = q, r = 1, \\ 0 & \text{else}, \end{cases} \quad (19)$$

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

with $t = (e - Te)$ denoting the probability that the current unit in service finishes production. As a consequence, we have the following form for the transition matrix P of $(V_n, D_{a(n)}, C_n, S_n)$:

$$P = \begin{bmatrix} A_e & A_0 & 0 & \dots & 0 & 0 & \dots \\ A_e & 0 & A_0 & \dots & 0 & 0 & \dots \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \ddots \\ A_e & 0 & 0 & \dots & A_0 & 0 & \dots \\ 0 & A_e & 0 & \dots & 0 & A_0 & \ddots \\ \vdots & \vdots & \ddots & \ddots & \vdots & \ddots & \ddots \end{bmatrix}, \tag{20}$$

where A_0 and A_e are square matrices of dimension $m_{tot} = 2(d_{max} - d_{min} + 1)d_{max}$. The matrix A_0 represents the probabilities that the service of the batch continues and is given by the first two equations of (19), while the matrix A_e represents the probabilities that the service of the batch finishes and is given by the 3rd equation of (19).

The MC characterized by Eq. (20) is of the GI/M/1 type (Neuts, 1981). From an operational point of view it is clear that the proposed queueing system is stable if and only if its utilization ρ is strictly smaller than one, or equivalently if the average production time of a batch order is strictly smaller than the average inter-arrival time of a batch order. Since we have chosen the time unit of our queueing model such that the average production time of a single unit is equal to 2, and the average batch order size is equal to the average demand $E(D)$, the average production time of a batch order is $2E(D)$. The inter-arrival time of an order is one (review) period, or, when we express it in the time unit of our queueing model, equal to e time units. Hence the stability condition can be rephrased as $2E(D) < e$. This condition is not restrictive as a system with a load $\rho > 1$ leads to infinite lead times as the demand is greater than the production capacity.

For an ergodic MC of the GI/M/1 type, one computes the steady state vector π of P, that is, $\pi P = \pi$ and $\pi \mathbf{1} = 1$, as follows:

$$\pi_1 = \pi_1 (I - R^e)(I - R)^{-1} A_e, \tag{21}$$

$$\pi_i = \pi_1 R^{i-1}, \tag{22}$$

where $\pi = (\pi_1, \pi_2, \dots)$ and π_i is a $1 \times m_{tot}$ vector, for all $i > 0$. The vector π_1 is normalized as $\pi_1 (I - R)^{-1} \mathbf{1} = 1$ and the $m_{tot} \times m_{tot}$ rate matrix R is the smallest nonnegative solution to the matrix equation $R = A_0 + R^e A_e$ and can be numerically solved with a variety of algorithms, e.g., Neuts (1981), Ramaswami (1988), Alfa et al. (2002).

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

Having obtained the steady state vector $\pi = (\pi_1, \pi_2, \ldots)$, we can obtain the response time using the following observation: the probability that an order has a response time of a time units can be calculated as the expected number of orders with an age of a time units that complete their service at an arbitrary time instant, divided by the expected number of orders that get completed during an arbitrary time instant (that is, $1/e$ for a queue with $\rho < 1$). As such, denoting T_r as the response time (expressed in the time unit U) we have

$$\Pr(T_r = a) = e\rho \sum_{k,s} \pi_a(k, 1, s)\ (t)_s, \tag{23}$$

where $\pi_a(k, r, s)$ represents the steady state probability of being in state (a, k, r, s). Notice, to make sure that an order completes its service, the number of remaining customers requiring service cannot be more than one.

We chose the time unit U of our queueing system as half of the mean production time of a single item (i.e., $E(M)/2$). Thus, if we want to express the lead time in terms of the number of periods needed to replenish an order, we still need to make the following conversion:

$$\Pr(T_p = i) = \sum_j \Pr(T_r = j) \cdot 1_{\{\lceil j/e \rceil = i\}}. \tag{24}$$

Note that this conversion at the same time rounds the (possibly fractional) response time T_r to the discrete replenishment lead time, T_p, expressed in an integer number of periods. This lead time distribution, T_p, is then used to start a new iteration.

5. Determination of safety stocks

Once the lead time distribution is known, we can analyze the retailer's inventory process and determine the safety stock requirements to provide a target service level. Since inventory is controlled by stochastic (endogenous) lead times, it is not necessarily replenished every period and we do not know exactly when a replenishment occurs. Moreover, the queueing analysis implies that it takes a longer time to produce (and consequently replenish) a larger order quantity, which involves that the order quantity and its replenishment lead time are correlated. This has an impact on the calculation of the inventory distribution. Therefore, if we want to determine the inventory distribution and the corresponding safety stock requirements in an exact way, we need to take this correlation into account. In this section we first define the evolution of the net stock over time, then we find its steady state distribution, and finally we determine the safety stock requirements to provide a target customer service.

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

5.1 Transient evolution of the net stock

We monitor the inventory on hand at the end of period t, after consumer demand D_t is observed and after a replenishment order O_t has been placed. At that time, there may be $l \geq 0$ orders waiting in the production queue and there is always 1 order in service (since the observation moment is immediately after an order placement) which is placed l periods ago (O_{t-l}). Note that l is a function of t, but we write l as opposed to $l(t)$ to simplify the notation.

The inventory on hand or net stock NS_t is equal to the initial inventory on hand plus all replenishment orders received so far minus total observed consumer demand. At the end of period t, the order O_{t-l} is in service, and orders placed more than l periods ago, i.e., $O_{t-i}, i \geq l+1$, are already received in inventory, while consumer demand is satisfied up to the current period t. For our purposes the initial inventory level is a control variable, equal to the safety stock SS, determining the retailer's customer service. Moreover we assume that $O_t = D_t = E(D)$ for $t \leq 0$, so that the net stock after satisfying demand in period t is equal to

$$NS_t = SS + (E(T_p) + 1) \cdot E(D) + \sum_{i=l+1}^{t-1} O_{t-i} - \sum_{i=0}^{t-1} D_{t-i}. \tag{25}$$

Since $D_t = E(D)$ for $t \leq 0$, the lead time demand forecast in period 0 is equal to $\widehat{D}_0^L = E(L) \cdot E(D)$, and the order quantity in period 1 is equal to

$$\begin{aligned} O_1 &= \left(\widehat{D}_1^L - \widehat{D}_0^L\right) + D_1 \\ &= \frac{1 - E\left(\phi^{L+1}\right)}{1-\phi} D_1 - \frac{\phi\left(1 - E\left(\phi^L\right)\right)}{1-\phi} E(D). \end{aligned} \tag{26}$$

For $t > 1$, the order quantity is given by Eq. (8):

$$O_t = \frac{1 - E\left(\phi^{L+1}\right)}{1-\phi} D_t - \frac{\phi\left(1 - E\left(\phi^L\right)\right)}{1-\phi} D_{t-1}.$$

Hence,

$$\begin{aligned} \sum_{i=l+1}^{t-1} O_{t-i} &= \frac{1 - E\left(\phi^{L+1}\right)}{1-\phi} D_1 - \frac{\phi\left(1 - E\left(\phi^L\right)\right)}{1-\phi} E(D) + \sum_{i=l+1}^{t-2} \left(\frac{1 - E\left(\phi^{L+1}\right)}{1-\phi} D_{t-i} - \frac{\phi\left(1 - E\left(\phi^L\right)\right)}{1-\phi} D_{t-i-1} \right) \\ &= \sum_{i=l+1}^{t-1} \left(\frac{1 - E\left(\phi^{L+1}\right)}{1-\phi} - \frac{\phi\left(1 - E\left(\phi^L\right)\right)}{1-\phi} \right) D_{t-i} + \frac{\phi\left(1 - E\left(\phi^L\right)\right)}{1-\phi} \left(D_{t-l-1} - E\left(D\right)\right) \\ &= \sum_{i=l+1}^{t-1} D_{t-i} + \frac{\phi\left(1 - E\left(\phi^L\right)\right)}{1-\phi} \left(D_{t-l-1} - E\left(D\right)\right). \end{aligned} \tag{27}$$

Substituting (27) into (25), we find that the net stock is equal to the safety stock plus the difference

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

between the average lead time demand and the realized lead time demand plus a fraction of the difference between the last observed demand before the order in service was placed, and the average demand. This last term intends to incorporate the autocorrelation in demand into the replenishment orders:

$$NS_t = SS + \left[(E(T_p)+1) \cdot E(D) - \sum_{i=0}^{l} D_{t-i} \right] + \frac{\phi \left(1 - E\left(\phi^L\right)\right)}{1-\phi} \left(D_{t-l-1} - E\left(D\right)\right). \quad (28)$$

Using Eq. (2), the realized lead time demand can be written as

$$\sum_{i=0}^{l} D_{t-i} = \sum_{i=1}^{l+1} \phi^i D_{t-l-1} + \sum_{i=0}^{l} \left(1 - \phi^{i+1}\right) G_{t-i}, \quad (29)$$

so that substituting (29) into (28) provides the following expression for the evolution of the net stock:

$$NS_t = SS + \left((E(T_p)+1) - \frac{\phi \left(1 - E\left(\phi^L\right)\right)}{1-\phi} \right) \cdot E(D) - \sum_{i=0}^{l} \left(1 - \phi^{i+1}\right) G_{t-i} - \frac{\phi}{1-\phi} \left(E\left(\phi^L\right) - \phi^l \right) D_{t-l-1}. \quad (30)$$

5.2 Steady state distribution of the net stock

We need to determine the steady state distribution NS of the net stock evolution NS_t, characterized by Eq. (30). To do so, we focus on the steady state distribution of Z_t, defined as:

$$Z_t = \sum_{i=0}^{l} \left(1 - \phi^{i+1}\right) G_{t-i} + \frac{\phi}{1-\phi} \left(E\left(\phi^L\right) - \phi^l \right) D_{t-(l+1)}. \quad (31)$$

Some care must be taken when evaluating (31), since there is correlation between the terms that make up Z_t. The values of G_{t-l} and $D_{t-(l+1)}$ influence the number of orders l in the queue. According to Eq. (9), the values of G_{t-l} and $D_{t-(l+1)}$ determine the order quantity O_{t-l}. It is intuitively clear that if O_{t-l} is large, it takes a longer time until production is completed, so that l increases.

It is possible to include this correlation in our analysis, making use of the Markov analysis described in section 4.3. Since this analysis is done in the time unit U of our queueing system, where 1 period corresponds to e time units U, we will work in the remainder of this section in the

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

time unit of the queueing system. Rewriting Eq. (31) in time unit U then gives

$$Z_t = \sum_{i=0}^{l} \left(1 - \phi^{i+1}\right) G_{t-ei} + \frac{\phi}{1-\phi} \left(E\left(\phi^L\right) - \phi^l\right) D_{t-e(l+1)}. \tag{32}$$

To obtain the distribution of $Z = \lim_{t\to\infty} Z_t$ we need to find the joint probabilities of having l ongoing orders at the end of a period (i.e., immediately after placing a new order), while $G_{t-el} = g$ and $D_{t-e(l+1)} = k$. We denote these probabilities as $\Pr\left(\hat{B} = l, \hat{G} = g, \hat{D} = k\right)$. Notice, $\hat{B}$ is the limiting distribution of $l(t)$ as t goes to infinity. In order to determine these joint probabilities, we could extend the 4-dimensional Markov chain $(V_n, D_{a(n)}, C_n, S_n)$, set up to find the lead time distribution, to a 6-dimensional Markov chain $(V_n, D_{a(n)-e}, D_{a(n)}, G_{a(n)}, C_n, S_n)$, which tracks the error term $G_{a(n)}$ and the demand $D_{a(n)-e}$. However, doing so will increase the dimensions of the block matrices of the transition matrix (20) with a factor $g_{max}(d_{max} - d_{min} + 1)$, which increases the time and memory complexity of the numerical procedure to find the steady state probabilities of the corresponding Markov chain.

Instead, we compute the required joint probabilities from the (known) steady state vector π of the previously used Markov chain $(V_n, D_{a(n)}, C_n, S_n)$ in a number of steps. First, we observe this Markov chain just before the service completion of the n'th replenishment order, and we determine the system state probabilities at the start of service of the next replenishment order $n+1$. In this transition step, we retain the error term $G(n+1)$, the order quantity $O(n+1)$ and the value of the previous consumer demand $D(n)$. Then, we observe the system at an arbitrary busy moment and derive its steady state vector. This is nearly identical to the steady state vector π, but additionally contains the values of $G(n+1)$ and $D(n)$. In the last step, we restrict to arrival instants only, i.e., we observe the system just after an order arrives at the queue, which corresponds to the end of a period. This allows us to determine the joint probabilities $\Pr\left(\hat{B} = l, \hat{G} = g, \hat{D} = k\right)$, which enables to find the end-of-period inventory distribution in an exact way.

Step 1

We start by determining the system state probabilities at the start of service. To describe this state we define the probability that given the demand value $D_{(t-1)e} = k$, the error term in the next period G_{te} equals g and the order quantity O_{te} equals q, denoted by $p_{[q]}^{(g)}(k) = \Pr\left(O_{te} = q, G_{te} = g | D_{(t-1)e} = k\right)$. These probabilities are similar to Eq. (17), but in this case we are not interested in the next period's demand size k', but in the next period's order quantity q. These probabilities can be found

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

as follows:

$$p_{[q]}^{(g)}(k) = \Pr(G=g) \cdot \left\{ 1_{\{q-1<\gamma k+\bar{\gamma} g<q\}} \left((\gamma k+\bar{\gamma} g) - \lfloor \gamma k+\bar{\gamma} g \rfloor \right) + 1_{\{\gamma k+\bar{\gamma} g=q\}} + 1_{\{q<\gamma k+\bar{\gamma} g<q+1\}} \left(\lceil (\gamma k+\bar{\gamma} g) \rceil - (\gamma k+\bar{\gamma} g) \right) \right\}. \quad (33)$$

Denote $\bar{\pi}_{a'}(g,k,r)$ as the probability that immediately after we start serving an order (say at time t), we observe an order with age a', an order size equal to r, while $G_{t-ea'} = g$ and $D_{t-(a'+1)e} = k$. Similar to Eq. (23), we can express $\bar{\pi}_{a'}(g,k,r)$ as

$$\bar{\pi}_{a'}(g,k,r) = e\rho \sum_{a,s} \pi_a(k,1,s)\ (t)_s\ p_{[r]}^{(g)}(k)\ 1_{\{a'=[a-e]^+\}}, \quad (34)$$

where $[x]^+ = \max(0,x)$. This can be explained as follows. First we divide the expected number of orders with age a and demand size k that complete service, given by $\pi_a(k,1,s)\ (t)_s$, by the expected number of orders that start service during an arbitrary time instant (that is, $1/e$ for a queue with $\rho < 1$). After service completion, the subsequent order that starts service has age a'. Since the Markov chain is only defined at time slots when the server is busy, the age $a' = 0$ when the previous order completes service before the next order arrives at the queue, or, equivalently, when the order's age a is smaller than the interarrival time e. When $a > e$ however, this implies that the next order waits in the queue for $a - e$ time instants until it starts service, and consequently $a' = a - e$. The probability $p_{[r]}^{(g)}(k)$ then defines that the next order in service has size r and the error term equals g. Finally, we multiply these probabilities by the average load ρ, in order to shift from busy time slots to all time slots.

Step 2

Given the system state probabilities at the start of service, we establish an expression for the probability vector of the system at an arbitrary busy moment. We denote $\tilde{\pi}_a(g,k,r',s)$ as the probability of having an order in service with age a, r' remaining items that still need to be produced, and service phase equal to s, provided that the system is busy (say at time t), while $G_{t-ea} = g$ and $D_{t-(a+1)e} = k$. Notice, $\tilde{\pi}_a(g,k,r',s)$ and $\pi_a(k',r',s)$ have a nearly identical interpretation, except that k' is the demand D at time $t-ea$, while k is the demand D at time $t-(a+1)e$ and g reflects the outcome of G at time $t-ea$.

If we observe the system at an arbitrary busy moment t_b, then the probability that t_b falls within the service of an order of size r, is given by

$$\frac{\sum_v \Pr(O=r)\Pr(M^{r*}=v)\,v}{E(O)\,E(M)} = \frac{\Pr(O=r)\,r}{E(G)}, \quad (35)$$

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

since $E(O) = E(G)$ and $\sum_v \Pr(M^{r*} = v)\, v$ defines the expected service time of a batch of size r, equal to $E(M) \cdot r$. Thus, the probability that we observe the system during the u-th time slot after starting the service of an order of size r, is then given by

$$\frac{\Pr(O = r)\, r}{E(G)} \left(\sum_{v \geq u} \frac{\Pr(M^{r*} = v)\, v}{E(M^{r*})}\ (1/v) \right) = \frac{\Pr(O = r) \Pr(M^{r*} \geq u)}{2E(G)}, \tag{36}$$

as the service has to last for at least u time slots and we have a probability $1/v$ that t_b is located in the u-th time epoch of a length v interval.

This observation allows us to write $\tilde{\pi}_a(g, k, r', s)$. Let $p_{\langle s \rangle}(u, r, r')$ denote the probability that an order of size r requires at least u time slots to complete, r' equals the number of remaining items that require completion and s is the service phase after u time units. These probabilities are computed from the matrix T and the vector α. Then,

$$\tilde{\pi}_{a'}(g, k, r', s) = \frac{1}{2E(G)} \sum_{u,a,r} \bar{\pi}_a(g, k, r)\ p_{\langle s \rangle}(u, r, r') 1_{\{a' = a+u\}}. \tag{37}$$

Step 3

We are now in a position to compute the probabilities at arrival instants by observing that all time epochs where the age of the customer is a multiple of e correspond to an arrival instant. Hence, these probabilities are given by

$$\Pr\left(\hat{B} = l, \hat{G} = g, \hat{D} = k\right) = \rho e \sum_{r,s} \tilde{\pi}_{el}(g, k, r, s), \tag{38}$$

$$\Pr\left(\hat{B} = 0, \hat{G} = g, \hat{D} = k\right) = \rho e \left(\sum_{s} \sum_{a=1}^{e-1} \pi_a(k, 1, s)(t)_s \right) \Pr(G = g), \tag{39}$$

for $l > 0$.

Observe that when $\hat{B} = 0$, we use the steady state vector π of our original Markov chain instead of $\tilde{\pi}$. This can be seen as follows. When an order arrives at an empty queue, then the value of the consumer demand that corresponds to the previous order is in fact the demand corresponding to the order that just finished service. This demand value can easily be found from the steady state vector π.

We are now able to compute the steady state probabilities Z of Z_t. Making use of the proba-

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

bilities $\Pr\left(\hat{B}=b,\hat{G}=g,\hat{D}=k\right)$ we readily find

$$\Pr\left(Z=s\right)=\lim_{t\to\infty}\Pr\left(Z_t=s\right)=\sum_{b=0}^{\infty}\sum_{g_b,k}\Pr\left(\hat{B}=b,\hat{G}=g_b,\hat{D}=k\right)\cdot \sum_{g_0,g_1,\ldots,g_{b-1}}\left(\prod_{j=0}^{b-1}\Pr(G=g_j)\right)\cdot 1_{\{\sum_{i=0}^{b}(1-\phi^{i+1})g_i+\phi k(E(\phi^L)-\phi^b)/(1-\phi)=s\}}, \tag{40}$$

as the random variables G_{t-ej} form an independent set for $j=0,\ldots,b$.

Let

$$S=SS+\left((E(T_p)+1)-\frac{\phi\left(1-E\left(\phi^L\right)\right)}{1-\phi}\right)\cdot E(D), \tag{41}$$

then, we find from Eq. (30) that the steady state probabilities of the net stock $\Pr(NS=k)=\lim_{t\to\infty}\Pr(NS_t=k)$ can then be computed from Eq. (40):

$$\Pr(NS=s)=\Pr(Z=S-s). \tag{42}$$

5.3 Safety stock determination

Given the inventory distribution, we can find the safety stock requirements to provide a target customer service. To measure customer service, we use the fill rate, which measures the proportion of demand that can be immediately fulfilled from the inventory on hand (Zipkin, 2000). The probability of a stock-out can be found from the inventory distribution, or, $\Pr(NS<0)=\Pr\left(Z>S\right)$, and the average number of shortages when a stock-out occurs is given by $E\left(NS^{-}\right)=E\left(\left[Z-S\right]^{+}\right)$, where $x^{+}:=\max\left\{0,x\right\}$. Hence, the fill rate can then be calculated as

$$\text{Fill rate}=1-\frac{E\left(\left[Z-S\right]^{+}\right)}{E(D)}. \tag{43}$$

In practice, decision makers often have to find the minimal safety stock that is required to achieve a given fill rate. From (43) we can compute the minimal value of S that is required such that an imposed fill rate is met, the corresponding safety stock is then found using Eq. (41).

6. Numerical experiment

In this section we numerically illustrate our procedure and investigate the impact of autocorrelation in consumer demand on lead times and safety stocks. To do so, we compare the supply chain

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

performance when demand is *uncorrelated* (i.e., an IID demand) with the performance when there is (positive or negative) autocorrelation in demand.

It is a well known result that when consumer demand is an IID process, the base-stock policy with MSE forecasting generates orders equal to the observed consumer demand, or $O_t = D_t$. In other words, a chase sales strategy. We use the procedure developed by Boute et al. (2007) to analyze this type of periodic review production/inventory system with endogenous lead times.

We set up a small numerical experiment. We assume an autoregressive demand given by Eq. (2). We consider a corresponding IID demand of the following form:

$$D_t = \left[1 - \sqrt{\frac{1-\phi}{1+\phi}}\right] \cdot E(D) + \sqrt{\frac{1-\phi}{1+\phi}} \cdot G_t, \tag{44}$$

where G_t is the same error term and ϕ has the same value as in the correlated demand pattern (Eq. (2)). As such, both demand processes have the same average, $E(D) = E(G)$, and variance, $Var(D) = \frac{1-\phi}{1+\phi}\, Var(G)$; they only vary in the autocorrelated structure of the demand process.

We assume G is uniformly distributed between 6 and 15, so that $\Pr(G = g) = 0.1$ for $g \in \{6, 7, ..., 15\}$ and $\Pr(G = g) = 0$, else. The average demand and average order quantity are then given by $E(D) \equiv E(O) \equiv E(G) = 10.5$ units per day. The manufacturer's production is available 10 hours per day and it takes on average 48 minutes to produce a single unit, with a coefficient of variation equal to 1. This results in an average production load equal to $(10.5 \times 48)/(10 \times 60) = 0.84$.

Then, for a given value of ϕ, we determine the lead time distribution and safety stock requirements for both the AR(1) demand (through the procedure described in this paper) and the corresponding IID process (with a chase sales policy, through the procedure described in Boute et al. (2007)). We consider values $-0.3 \leq \phi \leq 0.7$, which avoids negative demand and order sizes.

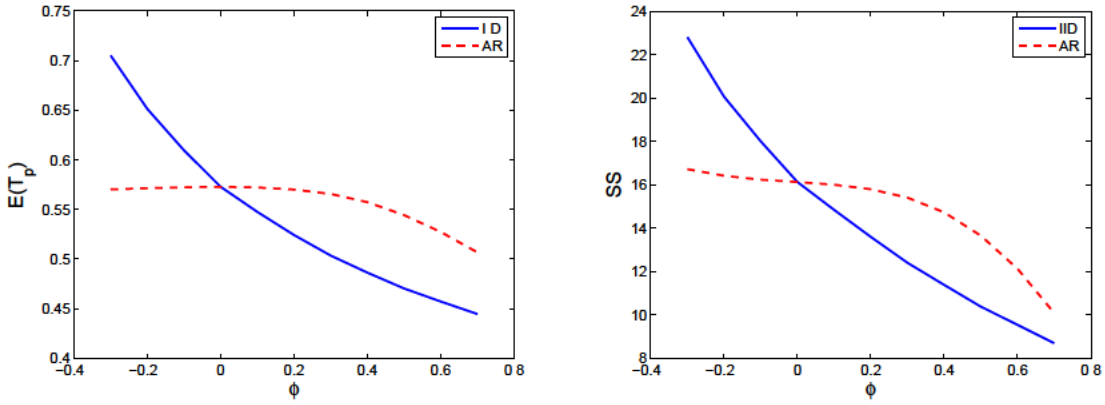
Let us first discuss the dynamics when demand is an IID process.

- As ϕ increases towards one, demand variance decreases towards zero, and hence, in a chase sales policy, order variance goes down. As a consequence, when ϕ increases, lead times go down (see Fig. 2(a), where the solid line represents the chase sales policy for an IID demand).

- The lead times have an impact on safety stocks: longer lead times inflate safety stocks. The same holds for demand variability: a more variable (or more *uncertain*) demand inflates safety stocks as well. Since both lead time and demand variance decrease as ϕ goes up, safety stock requirements decrease with a higher ϕ. This is observed in Fig. 2(b), where the safety stock is plotted to meet a 98% fill rate (solid line represents the chase sales policy for an IID demand).

We now compare these observations with the performance when there is autocorrelation in demand.

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.



(a) Impact of ϕ on average lead times

(b) Impact of ϕ on safety stocks

Figure 2: Comparison of P/I systems with an AR and IID demand process

- Fig. 3 illustrates the convergence of the lead time distribution in our iterative procedure for both $\phi = 0.2$ and $\phi = -0.2$ when we start with an initial lead time distribution $T_p^0 = 0$. The abscissa shows the iteration step n and the ordinate represents the difference between the average lead time at iteration step n and the actual average lead time: $E(T_p) - E(T_p^n)$. We observe a convergence for both a positive and negative ϕ.

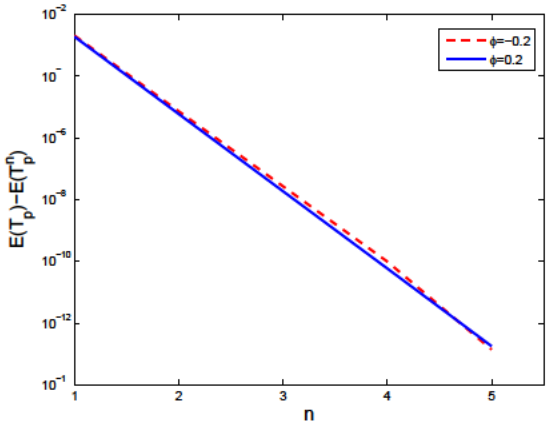


Figure 3: Convergence of lead time distribution

- When autocorrelation in demand is positive, we find that lead times are longer compared to the IID demand. This is illustrated in Fig. 2(a) for values $\phi > 0$, (the dotted line represents the case where demand is an AR process). This can be explained by the fact that a positively correlated demand amplifies the variability in the orders (see Eq. (11)), implying more variability in the arrival process at the queue, resulting in longer lead times compared to the

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

chase sales strategy for an IID demand.

- The inverse is true for a negatively correlated demand. When $\phi < 0$, the variability in orders is dampened with respect to the observed demand (see Eq. (12)) and consequently lead times are shorter compared to the chase sales strategy. In our specific example, this dampening effect in orders is so strong that, although the demand variability increases as ϕ decreases, the corresponding lead times do not increase.

- We find a similar conclusion for the safety stocks. When demand is positively correlated, safety stocks are higher compared to the safety stocks when demand is IID (see Fig. 2(b)). When demand is negatively correlated, safety stocks are much lower compared to the case with IID demand. This is explained by their impact on lead times.

What are the implications of these results in the supply chain? In the first place, one has to realize that when there is positive autocorrelation in demand, the order variance is amplified compared to consumer demand, even if the optimal forecasting scheme is used. When the manufacturer produces on a make-to-order basis, this increased order variance will result in longer production lead times, and consequently longer replenishment lead times. This in turn inflates the safety stock requirements at the retailer. The inverse is true when demand is negatively autocorrelated. The optimal MSE forecasting scheme dampens the variability in the replenishment orders, with shorter lead times as a consequence, decreasing the safety stock requirements at the retailer.

This sheds new light to the Sales & Operations Planning (S&OP) meetings, where sales and marketing managers decide, amongst others, on pricing their products, and link it with required inventories and production lead times, which is the responsibility of operations managers. Traditionally, operations managers tend to constrain the pricing flexibility for sales managers since they may create vexing ripple effects in operations. However, as we show in this paper, we need to consider both the variability and the autocorrelation in demand caused by promotions, since they both have an impact on the operational performance of the supply chain. Given the same variability, a price promotion policy leading to negatively autocorrelated demand provides better performance.

7. Concluding remarks

Much of the management science literature separates the questions of production and inventory control. However, inventory influences production by initiating orders, and production influences inventory by completing and delivering orders to inventory. Modeling a two-echelon supply chain (retailer-manufacturer) as a production/inventory system complies with this research question and explicitly analyzes the interaction between the retailer's inventory and the manufacturer's produc-

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

tion management. This results in new insights. For instance, an increased demand variability has a double impact on supply chain performance: it not only increases inventory variability (thereby inflating safety stocks), lead times go up as well due to the increased order variability, which reinforces the increase in safety stocks. Boute et al. (2007) show that decoupling the inventory and production systems, thereby treating lead times as (exogenous) IID variables, drastically underestimates the required safety stocks and consequently results in lower fill rates.

In this paper we studied the autocorrelation in demand, rather than its variability. Autocorrelated demand behavior can be impacted by marketing promotions. A negative autocorrelation involves that consumers increase their purchases in the promotion period, and strongly decrease their demand in the periods preceding and subsequent to the promotion period, resulting in erratic sales. Positive autocorrelation, on the other hand, denotes a consistent, meandering sales pattern. When we consider the demand variance to be the same, we find that the erratic pattern results in an improved supply chain performance compared to the meandering sales. In the former case (negative correlation), there is a natural smoothing in the replenishment orders. This dampening effect decreases lead times at the manufacturer, which has a compensating effect on the corresponding safety stocks. In the latter case (positive correlation), order variance is amplified towards the manufacturer, even if the optimal forecasting scheme is employed. This results in higher lead times and higher safety stocks.

We have developed an exact and stable solution for this problem, modeled as a periodic review base-stock controlled production-inventory system with autoregressive demand. Since the order decision depends on the lead time distribution and lead times depends on the replenishment order process, we encountered a lead time dependency problem, which we solved through an iterative procedure. An exact solution to the inventory distribution was developed, taking the correlation between consumer demand and lead times into account.

References

A. Alfa, B. Sengupta, T. Takine, and J. Xue. A new algorithm for computing the rate matrix of GI/M/1 type Markov chains. In *Proc. of the 4th Int. Conf. on Matrix Analytic Methods*, pp 1–16, Adelaide, Australia, 2002.

L. C. Alwan, J. J. Liu, and D. G. Yao. Stochastic characterization of upstream demand processes in a supply chain. *IIE Transactions*, 35, pp 207–219, 2003.

A. Balakrishnan, J. Geunes, and M. Pangburn. Coordinating supply chains by controlling upstream variability propagation. *Manufacturing & Service Operations Management*, 6(2), pp 163–183, 2004.

F. Bernstein, F. Chen, and A. Federgruen. Coordinating supply chains with simple pricing schemes. *Management Science*, 52(10), pp 1483–1492, 2006.

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

D. Bini, G. Latouche, and B. Meini. *Numerical methods for structured Markov chains*. Oxford University Press, Oxford and New York, 2005.

D. Bini, B. Meini, B. Van Houdt, and S. Steffe. Structured markov chains solver: software tools. In *Proceedings of SMCtools'06*, Pisa (Italy), 2006. ACM Press.

R. C. Blattberg and S. A. Neslin. Sales promotion models. In E. J. and L. G. L., editors, *Handbooks in Operations Research and Management Science: Marketing*, pp 553–609. North Holland, Amsterdam, 1993.

A. Bobbio, A. Horváth, M. Scarpa, and M. Telek. Acyclic discrete phase type distributions: Properties and a parameter estimation algorithm. *Performance Evaluation*, 54(1), pp 1–32, 2003.

A. Bobbio, A. Horváth, and M. Telek. The scale factor: a new degree of freedom in phase type approximation. *Performance Evaluation*, 56(1-4), pp 121–144, 2004.

R. N. Boute, M. R. Lambrecht, and B. Van Houdt. Performance evaluation of a production/inventory system with periodic review and endogenous lead times. *Naval Research Logistics*, 54(4), pp 462–473, 2007.

G. E. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis: Forecasting and Control*. Prentice Hall, Englewood Cliffs, NJ, 3rd edition, 1994.

G. Cachon. Planning supply chain operations: definition and comparison of planning concepts. In A. G. De Kok and S. C. Graves, editors, *Supply chain management: design, coordination and operation*, chapter 6. North Holland, Amsterdam, 2003.

A. G. De Kok and J. C. Fransoo. Planning supply chain operations: definition and comparison of planning concepts. In A. G. De Kok and S. C. Graves, editors, *Supply chain management: design, coordination and operation*, chapter 12. North Holland, Amsterdam, 2003.

J. Dejonckheere, S. M. Disney, M. R. Lambrecht, and D. R. Towill. Measuring and avoiding the bullwhip effect: A control theoretic approach. *European Journal of Operational Research*, 147(3), pp 567–590, 2003.

S. M. Disney, I. Farasyn, M. R. Lambrecht, D. R. Towill, and W. Van de Velde. Taming bullwhip whilst watching customer service in a single supply chain echelon. *European Journal of Operational Research*, 173(1), pp 151–172, 2006.

S. C. Graves. Safety stocks in manufacturing systems. *Journal of Manufacturing and Operations Management*, 1(1), pp 67–101, 1988.

S. C. Graves. A single-item inventory model for a nonstationary demand process. *Manufacturing & Service Operations Management*, 1(1), pp 50–61, 1999.

I. Hendel and A. Nevo. Measuring the implications of sales and consumer inventory behavior. *Econometrica*, 74(6), pp 1637–1673, 2006.

W. J. Hopp and M. L. Spearman. *Factory Physics*. Irwin, McGraw-Hill, 2nd edition, 2001.

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

A. Horváth and M. Telek. PhFit: A general phase-type fitting tool. In *Proceedings of Performance TOOLS 2002*. London, UK, April 2002.

Z. Jemaï and F. Karaesmen. The influence of demand variability on the performance of a make-to-stock queue. *European Journal of Operational Research*, 164(1), pp 195–205, 2005.

J. A. Kahn. Inventories and the volatility of production. *The American Economic Review*, 77(4), pp 667–679, 1987.

U. S. Karmarkar. Manufacturing lead times, order release and capacity loading. In S. C. Graves, A. H. G. Rinnooy Kan, and P. H. Zipkin, editors, *Logistics of Production and Inventory*, volume 4 of *Handbooks in Operations Research and Management Science*, pp 287–329. Elsevier Science Publishers B.V, 1993.

J. G. Kim, D. C. Chatfield, T. P. Harrison, and J. C. Hayya. Quantifying the bullwhip effect in a supply chain with stochastic lead time. *European Journal of Operational Research*, 173(2), pp 617–636, 2006.

G. Latouche and V. Ramaswami. *Introduction to Matrix Analytic Methods and stochastic modeling*. SIAM, Philadelphia, 1999.

H. L. Lee, V. Padmanabhan, and S. Whang. Information distortion in a supply chain: The bullwhip effect. *Management Science*, 43(4), pp 546–558, 1997.

S. Macé and S. A. Neslin. The determinants of pre- and postpromotion dips in sales of frequently purchased goods. *Journal of Marketing Research*, 16, pp 339–350, 2004.

C. F. Mela, K. Jedidi, and D. Bowman. The long-term impact of promotions on consumer stockpiling behavior. *Journal of Marketing Research*, 35, pp 250–262, 1998.

S. Nahmias. *Production and Operation Analysis*. McGraw-Hill, 3rd edition, 1997.

S. A. Neslin, C. Henderson, and J. Quelch. Consumer promotions and the acceleration of product purchases. *Marketing Science*, 4(2), pp 147–165, 1985.

M. Neuts. *Matrix-Geometric Solutions in Stochastic Models, An Algorithmic Approach*. John Hopkins University Press, 1981.

M. Neuts. *Structured Stochastic Matrices of M/G/1 type and their applications*. Marcel Dekker, Inc, New York and Basel, 1989.

J. Pahl, S. Voss, and D. L. Woodruff. Production planning with load dependent lead times. *4OR, A quarterly journal of operations research*, 3(4), pp 257–302, 2005.

J. S. Raju. The effect of price promotions on variability in product category sales. *Marketing Science*, 11(3), pp 207–220, 1992.

V. Ramaswami. Nonlinear matrix equations in applied probability - solution techniques and open problems. *SIAM review*, 30(2), pp 256–263, June 1988.

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.

B. Selcuk, I. J. Adan, A. G. De Kok, and J. C. Fransoo. An explicit analysis of the lead time syndrome: stability condition and performance evaluation. *International Journal of Production Research*, 47(9), pp 2507–2529, 2009.

J.-S. Song and P. Zipkin. The joint effect on leadtime variance and lot size in a parallel processing environment. *Management Science*, 42(9), pp 1352–1363, 1996.

C. R. Sox, L. J. Thomas, and J. O. McClain. Coordinating production and inventory to improve service. *Management Science*, 43(9), pp 1189–1197, 1997.

X. Zhang. Evolution of arma demand in supply chains. *Manufacturing & Service Operations Management*, 6, pp 195–198, 2004a.

X. Zhang. The impact of forecasting methods on the bullwhip effect. *International Journal of Production Economics*, 88, pp 15–27, 2004b.

P. H. Zipkin. *Foundations of Inventory Management*. McGraw-Hill, New York, 2000.

Boute, R.N., Disney, S.M., Lambrecht, M.R. and Van Houdt, B., (2014), "Coordinating lead-times and safety stocks under auto-correlated demand", European Journal of Operational Research, Vol. 232, No. 1, pp52–63. DOI: 10.1016/j.ejor.2013.06.036.