THE IMPACT OF DELIVERY LOT SPLITTING ON DELIVERY RELIABILITY IN A TWO-STAGE SUPPLY CHAIN

Inneke Van Nieuwenhuyse* and Nico Vandaele

University of Antwerp, Faculty of Applied Economics**

Abstract

Current supply chain policies are increasingly influenced by the lean philosophy, which calls for increased synchronization and smaller but more frequent shipments between the supply chain partners. One of the strategies used in this respect is *lot splitting*, i.e. the splitting of a single order quantity among multiple deliveries. Lot splitting is known to offer numerous advantages over a lot-for-lot policy, such as decreasing flow times and lower congestion levels. In this paper, we prove analytically that lot splitting improves the delivery reliability of the supplier, and hence the production schedule stability of the buyer. We also propose an approximation to estimate the delivery reliability in terms of the lot splitting policy and the system characteristics.

Keywords:

Lot Splitting, Production Control, Supply Chain Management, Lead Time Reduction

* Ms Van Nieuwenhuyse is a Postdoctoral Researcher of the Fund for Scientific Research-Flanders (Belgium).

** Corresponding author:

Inneke van Nieuwenhuyse, Prinsstraat 13 (B307), 2000 Antwerp, Belgium

Tel: +32 3 220 45 42

Fax: +32 3 220 47 99

E-mail: inneke.vannieuwenhuyse@ua.ac.be

1. Introduction and literature overview

Many of the practices that are currently advocated in lean supply chain management go back to the Just-in-Time (JIT) philosophy for production management, which was developed and promoted by Toyota Motor Co. Ltd. after World War II. The main principle underlying the lean philosophy is the elimination of waste, defined as anything that does not add value to the product (e.g. [43]). On the plant level, the lean philosophy is implemented by means of a set of techniques, such as group technology, reduction of setup times and kanban (see e.g. [34], [29], [36], [14]). On the supply chain level, one of the most important components of the philosophy is the relationship with the suppliers. Whereas in traditional purchasing practices, the relationship between supplier and buyer was mostly adversarial and based on price, the introduction of the lean philosophy in supply chain management calls for a long-term relationship between the supplier and the manufacturer, aimed at reducing delivery lot sizes, improving the quality of purchased parts, reducing the paperwork, and stabilizing production schedules. Many authors have acknowledged that this type of supplier relationships is primordial in the successful application of the lean strategy (see e.g. [12], [10], [26], [4], [1]).

The use of small lot sizes is the hallmark of the lean philosophy (see e.g. [35], [3], [28], [31]), in manufacturing as well as in purchasing. Although it is acknowledged that the ideal of "one-at-a-time" delivery is beyond reach ([34]), a relentless effort should be made to reduce delivery lot sizes.

According to the literature, one of the ways in which firms can achieve the objective of small but frequent deliveries is by *order splitting* ([34], [22]), which can be defined as the splitting of the order quantity among multiple deliveries ([30]). As shown in Figure 1, it

is the opposite of a lot-for-lot policy. The figure depicts a production order, consisting of 4 product units, going through a two-stage supply chain (consisting of a supplier and a buyer).

[Insert Figure 1 here]

The system in the top pane uses a lot-for-lot policy. Here, products are moved between the two stages in production orders: the products simply wait in the supplier's outbound buffer until the entire production order has been completed before they are forwarded to the buyer. The relationship between the order size and system performance in case of a lot-for-lot policy has been studied intensively in the operations management literature, and is already well understood (e.g. [21], [20], [39]).

The system in the lower pane makes use of lot splitting: as soon as the delivery lot size (which equals 1 product unit in the figure) is reached, the delivery lot is already sent to the buyer (this is also referred to in the literature as *batch availability*, e.g. [7], [33]). The use of delivery lot splitting offers numerous advantages, which fit within the framework of the lean philosophy. Firstly, it is widely acknowledged in the literature that lot splitting reduces product flow times by smoothing workflow and minimizing congestion levels (e.g. [33], [5], [11], [18], and [38]). This is due to the mechanism of *overlapping operations*: by allowing transportation of partial orders to a downstream stage in the production process, this stage can already start working on these partial orders while work proceeds at the upstream stage, thereby accelerating the progress of work through the production chain (e.g. [11], [13], [19], [25], and [40]). From Figure 1, it is indeed evident that the production order flow time F is shorter in case of the lot splitting policy.

Moreover, lot splitting may be useful when the buyer requires a small delivery lot size, while the setup costs of the supplier do not justify frequent setups. In that case, the supplier will be forced to produce in a lot size that is a multiple of the delivery lot size ([3]). Finally, it has been argued that lot splitting also fits within the framework of a "blanket order" strategy: this strategy consists of a long-term agreement between buyer and supplier for a specific quantity of products, against which the buyer can then generate short-term releases ([30], [24]).

Given these multiple advantages, the issue of lot splitting has already received considerable attention in the literature. The technique has been examined predominantly from a cost viewpoint, taking into account the impact of order splitting on cost components like setup costs for the supplier, transportation costs, holding costs of both buyer and supplier, and costs for order handling and receiving. Roughly, the literature can be split up in two separate substreams: those papers that consider the cost problem from the viewpoint of the buyer (the *independent optimization models*, e.g. [30], [24], [31], [9] and [8]), and those that approach the problem from the viewpoint of both buyer and supplier (the *joint optimization models*, e.g. [2], [16] and [22]).

In this paper, we will investigate the impact of the lot splitting policy on the delivery reliability of the supplier. Indeed, the reliability of the deliveries is of utmost importance in a lean supply chain: as there are little or no safety stocks in the system, the inability of a supplier to stick to the delivery schedule could, just like bad quality, shut down a line (see e.g. [42], [15]).

We will study the reliability issue from the perspective of a two-stage stochastic supply chain, in which the buyer regularly places orders for a given product with the supplier. Clearly, it is desirable from the buyer's point of view that the arrival of the ordered

material can be predicted as accurately as possible, such that the start times of the production runs can be planned with a high level of certainty, and rescheduling efforts (due to material arriving too early or too late) can be restricted to a minimum.

In case of a lot-for-lot policy, the production run can only start when the entire order quantity has arrived in the buyer's inbound buffer; in case of the lot splitting policy, the run can start as soon as the first delivery lot (hereafter referred to as the *flag*¹) has arrived. Hence, we will measure delivery reliability by the predictability of the arrival times of the flags in case of a lot splitting policy, and the predictability of the arrival times of entire orders in case of a lot-for-lot policy.

In what follows, we will first analytically prove that the delivery reliability of a lot splitting policy outperforms the delivery reliability of a lot-for-lot policy. Next, we will develop an approximation for this delivery reliability, in terms of the policy used (i.e., the number of sublots) and the system characteristics (the levels of variability of setup and processing times, and the utilization of the supplier).

Despite the fact that we only focus on the start of the production runs, which simplifies the reliability issue when compared to a real life setting, we believe that this research makes a contribution: as the previous literature on lot splitting has focused on completely deterministic settings, the reliability issue has until now been overlooked. To our knowledge, this paper is the first one to develop a formal analytical model to study the issue, and provide analytical insight.

Section 2 gives an overview of the notation and the assumptions of the system under study. In Section 3, it is proven analytically that the delivery reliability of the supplier increases when a lot splitting policy is used. In Section 4, an approximation for the

delivery reliability is derived (based upon the lot splitting policy and the system characteristics), and tested versus simulation results. Finally, Section 5 summarizes the conclusions.

2. Notation and assumptions

We will consider a supply chain consisting of a single buyer and a single supplier, producing a single product type. We assume that the order quantity of the buyer equals the number of product units in a single production run (denoted by the symbol N). It is assumed that the product is processed in production runs of size N at both supplier and buyer.

The interarrival times of orders at the supplier are independently and identically distributed (IID), with a general probability distribution and a given mean and variance.

The queueing discipline between orders is supposed to be first-in first-out. At the start of every order, a setup time is to be performed, after which the product units in the order are individually processed. Setup times and unit processing times are assumed to be IID distributed with a general probability distribution and a given mean and variance.

The delivery lot size of the supplier (denoted by the symbol L) is assumed to be a common divisor of the order quantity N . Hence, the number of deliveries for a single order quantity is integer (denoted by the symbol T). As soon as the delivery lot size is reached at the supplier's outbound buffer, this delivery lot is moved to the buyer's facility. The transportation time between the two stages is assumed to be stochastic with a general distribution, and is denoted by D . The transportation times are supposed to be independent of the amount of material to be transported (as is usually the case), and are

¹ The term "flag" is taken from [37]. The purpose of the flag is to signal the upcoming arrival of subsequent sublots, and to authorize the start of the production run.

IID distributed. The inbound buffers of both buyer and supplier are assumed to have infinite capacity.

As mentioned above, the delivery reliability of the supplier is measured by the predictability of the production run start times in the buyer's production schedule. The predictability of these start times depends on the variance of the interarrival times of consecutive flags (in case of a lot splitting policy), and the variance of the interarrival times of entire orders (in case of a lot-for-lot policy) in the buyer's inbound buffer: the higher this variance, the lower the delivery reliability of the supplier, and the more uncertain the scheduled starting time of the production run at the buyer's plant will be.

[Insert Table 1 here]

Table 1 gives an overview of the notation that will be used in this section. An index n will refer to the number of the order (or, equivalently, the number of the production run) under consideration; e.g. $X_I(n)$ denotes the total service time at the supplier of the first delivery lot in the n th production order. $Y(n)$ refers to the interarrival time between the n th and the $(n+1)$ th order at the supplier, $\tau_f(n)$ denotes the interdeparture time between the flag of the n th order and the flag of the $(n+1)$ th order at the supplier, and $R(n)$ refers to the idle time occurring at the supplier between the processing of the n th and the $(n+1)$ th order.

3. Delivery reliability under a lot splitting policy: analysis

In this section, we will study the variance of the interarrival times of flags at the buyer's inbound buffer. To illustrate the analysis, Figure 2 shows a flowchart of two consecutive

orders (order n and order $n+1$), each consisting of 3 delivery lots, going through the supply and transportation stage. The vertical arrows at the top of the figure indicate the arrival times of the orders at the supplier; the vertical arrows at the bottom denote the arrival times of the consecutive flags at the buyer's stage. The dotted arrows indicate the times at which material is handed over to the transportation resource.

[Insert Figure 2 here]

From the figure, we can see that:

$$W(n+1) = \max[0, W(n) + SU(n) + \sum_{i=1}^T X_i(n) - Y(n)] \quad (1)$$

If we then define $U(n) = SU(n) + \sum_{i=1}^T X_i(n) - Y(n)$, we can write:

$$W(n+1) = \max[0, W(n) + U(n)] \quad (2)$$

The idle time $R(n)$ can then be written as:

$$\begin{aligned} R(n) &= -\min[0, W(n) + U(n)] \\ &= \max[0, -W(n) - U(n)] \end{aligned} \quad (3)$$

Looking at the arrival process of consecutive flags in the buyer's inbound buffer, we can write:

$$\tau_f(n) = \sum_{i=2}^T X_i(n) + SU(n+1) + X_1(n+1) + R(n) + D(n+1) - D(n) \quad (4)$$

such that

$$\begin{aligned} \text{Var}[\tau_f(n)] &= \text{Var}\left[\sum_{i=2}^T X_i(n)\right] + \text{Var}[SU(n+1)] + \text{Var}[X_1(n+1)] + \text{Var}[R(n)] \\ &\quad + 2\text{Cov}\left[\sum_{i=2}^T X_i(n), R(n)\right] + \text{Var}[D(n+1)] + \text{Var}[D(n)] \end{aligned} \quad (5)$$

As the setup times, unit processing times and transportation times are IID distributed, we can write:

$$Var[\tau_f(n)] = T * Var(X) + Var[SU] + 2Var[D] + Var[R(n)] + 2Cov[\sum_{i=2}^T X_i(n), R(n)] \quad (6)$$

If a lot-for-lot policy were used to move products from the supplier to the buyer, the interarrival times of entire orders in the buyer's inbound buffer would equal:

$$\tau_o(n) = R(n) + SU(n+1) + \sum_{i=1}^T X_i(n+1) + D(n+1) - D(n) \quad (7)$$

such that the variance of the interarrival times of orders is given by:

$$Var[\tau_o(n)] = Var[R(n)] + Var[SU] + T * Var(X) + 2Var[D] \quad (8)$$

Using expressions (6) and (8), the relationship between the delivery reliability of a lot-for-lot policy ($Var[\tau_o(n)]$) and the delivery reliability of a lot splitting policy ($Var[\tau_f(n)]$) is given immediately by:

$$Var[\tau_f(n)] = Var[\tau_o(n)] + 2Cov[\sum_{i=2}^T X_i(n), R(n)] \quad (9)$$

The covariance term in expression (9) does not equal zero: indeed, we know from

expression (3) that $R(n) = \max[0, Y(n) - W(n) - SU(n) - \sum_{i=1}^T X_i(n)]$, such that $\sum_{i=2}^T X_i(n)$ and

$R(n)$ are clearly not independent. However, it is possible to analytically derive the

following upper and lower bound for $Cov[\sum_{i=2}^T X_i(n), R(n)]$:

$$-(T-1) * Var(X) \leq Cov[\sum_{i=2}^T X_i(n), R(n)] \leq 0 \quad (10)$$

The proof for these bounds is given in Appendices A and B. Combining expressions (9) and (10), we can conclude that:

$$Var[\tau_f(n)] \leq Var[\tau_o(n)] \quad (11)$$

In other words, the use of a lot splitting policy improves the delivery reliability of the supplier when compared to the lot-for-lot policy.

This result provides additional rationale for preferring the lot splitting policy. Indeed, from the buyer's point of view, an increase in delivery reliability implies that production plans can be made with a higher level of certainty, such that the need for rescheduling diminishes. From the supplier's point of view, it implies that there is a higher probability of meeting the delivery schedule, such that penalties for late deliveries can more likely be avoided.

4. Delivery reliability under a lot splitting policy: an approximation

4.1. Derivation of the approximation

From expression (9), it is clear that an approximation for $Var[\tau_f(n)]$ can be obtained by approximating its two main components:

$$Var[\tau_f(n)]_{app} = Var[\tau_o(n)]_{app} + 2Cov[\sum_{i=2}^T X_i(n), R(n)]_{app} \quad (12)$$

An expression for $Var[\tau_o(n)]_{app}$ is directly obtained by using the following result derived by Marshall ([27]):

$$Var[R(n)] + Var[SU] + T * Var(X) = 2(Var[SU] + T * Var[X]) + Var[Y] - 2E[W](E[Y] - E[SU] - T * E[X]) \quad (13)$$

In this expression, $E[W]$ refers to the average waiting time of a production order in the input buffer of the supplier. It can be approximated using well-known results from queueing theory, e.g. the formula of Kramer-Lagenbach Belz ([23]) (see Appendix C).

Using expressions (8) and (13), an expression for $Var[\tau_o(n)]_{app}$ is readily obtained as:

$$Var[\tau_o(n)]_{app} = 2Var[D] + 2(Var[SU] + T * Var[X]) + Var[Y] - 2E[W]_{app}(E[Y] - E[SU] - T * E[X]) \quad (14)$$

Consequently, we can restrict ourselves to developing an expression for

$Cov[\sum_{i=2}^T X_i(n), R(n)]_{app}$ in expression (12). From expression (3), we know that:

$$Cov[\sum_{i=2}^T X_i(n), R(n)] = Cov[\sum_{i=2}^T X_i(n), \max[0, -W(n) - SU(n) - \sum_{i=1}^T X_i(n) + Y(n)] \quad (15)$$

An approximation can now be developed by looking at two limiting cases (the case where the utilization of the supplier is extremely high, and the case in which the utilization is extremely low), and then interpolating between the two results.

On the one hand, when the utilization of the supplier is extremely low, we expect that

$W(n)$ will be almost constant and close to zero. Hence, as $\rho \rightarrow 0$, $Cov[\sum_{i=2}^T X_i(n), R(n)]$ will

approach:

$$Cov[\sum_{i=2}^T X_i(n), R(n)]_{\rho \rightarrow 0} = Cov[\sum_{i=2}^T X_i(n), \max[0, -SU(n) - \sum_{i=1}^T X_i(n) + Y(n)]] \quad (16)$$

Using the definition of covariance (see e.g. [6]), this can be rewritten as:

$$\begin{aligned} & Cov[\sum_{i=2}^T X_i(n), R(n)]_{\rho \rightarrow 0} \\ &= E[\sum_{i=2}^T X_i(n) * \max[0, -SU(n) - \sum_{i=1}^T X_i(n) + Y(n)]] \\ & - E[\sum_{i=2}^T X_i(n)] * E[\max[0, -SU(n) - \sum_{i=1}^T X_i(n) + Y(n)]] \end{aligned} \quad (17)$$

When we assume that both $V = \sum_{i=2}^T X_i(n)$ and $U = -SU(n) - \sum_{i=1}^T X_i(n) + Y(n)$ are

normally distributed, we can approximate this covariance as follows:

$$\begin{aligned}
& Cov[\sum_{i=2}^T X_i(n), R(n)]_{\rho \rightarrow 0} \\
& \approx \int_{-\infty}^{+\infty} \int_0^{+\infty} u * v * f_{U,V}(u, v) * dudv - E[V] * E[\max[0, U]]
\end{aligned} \tag{18}$$

where the means and variances of U and V are known and given by:

$$\begin{aligned}
E[U] &= m_U = E[Y] - E[SU] - T * E[X] \\
E[V] &= m_V = (T - 1) * E[X] \\
Var[U] &= \sigma_U^2 = Var[Y] + Var[SU] + T * Var[X] \\
Var[V] &= \sigma_V^2 = (T - 1) * Var[X]
\end{aligned} \tag{19}$$

and where $f_{U,V}(u, v)$ refers to the bivariate normal distribution of U and V :

$$f_{U,V}(u, v) = \frac{1}{2\pi\sigma_V\sigma_U\sqrt{1-\rho_{U,V}^2}} e^{\frac{-1}{2(1-\rho_{U,V}^2)}[\frac{(v-m_V)^2}{\sigma_V^2} - 2\rho_{U,V}\frac{(v-m_V)(u-m_U)}{\sigma_V\sigma_U} + \frac{(u-m_U)^2}{\sigma_U^2}]}$$

The notation $\rho_{U,V}$ denotes the correlation between U and V , given by:

$$\rho_{U,V} = \frac{-\sigma_V^2}{\sigma_U\sigma_V}$$

Given these assumptions, the double integral in expression (18) can be evaluated numerically. In expression (18), $E[\max[0, U]]$ can be approximated by ([6]):

$$E[\max[0, U]] = \frac{m_U}{2} (1 + Erf\{\frac{m_U}{\sqrt{2}\sigma_U}\}) + e^{\frac{-m_U^2}{2\sigma_U^2}} \frac{\sigma_U^2}{\sqrt{2\pi}\sigma_U} \tag{20}$$

where $Erf\{z\}$ refers to the error function in z :

$$Erf\{z\} = \frac{2}{\sqrt{\pi}} \int_0^z Exp\{-t^2\} dt \tag{21}$$

Using expressions (19) and (20), expression (18) yields an approximation for

$$Cov[\sum_{i=2}^T X_i(n), R(n)]_{\rho \rightarrow 0}.$$

On the other hand, when the utilization of the first stage is high ($\rho \rightarrow 1$), we expect that

$R(n)$ will be almost constant and close to zero. Hence, at high utilization,

$Cov[\sum_{i=2}^T X_i(n), R(n)]$ will approach zero:

$$Cov[\sum_{i=2}^T X_i(n), R(n)]_{\rho \rightarrow 1} \approx 0 \quad (22)$$

We propose to develop the final approximation for $Cov[\sum_{i=2}^T X_i(n), R(n)]$ by interpolating

between expression (18) and expression (22) in terms of ρ :²

$$\begin{aligned} Cov[\sum_{i=2}^T X_i(n), R(n)]_{app} &= \rho^2 \cdot 0 + (1 - \rho^2) \cdot Cov[\sum_{i=2}^T X_i(n), R(n)]_{\rho \rightarrow 0} \\ &= (1 - \rho^2) \cdot Cov[\sum_{i=2}^T X_i(n), R(n)]_{\rho \rightarrow 0} \end{aligned} \quad (23)$$

Using expressions (12), (14) and (23), we now dispose of an approximation for the delivery reliability of a supplier using a lot splitting policy:

$$\begin{aligned} Var[\tau_f(n)]_{app} &= 2Var[D] + 2(Var[SU] + T * Var[X]) + Var[Y] \\ &\quad - 2E[W]_{app} (E[Y] - E[SU] - T * E[X]) + 2Cov[\sum_{i=2}^T X_i(n), R(n)]_{app} \end{aligned} \quad (24)$$

In the next subsection, we will test the performance of this approximation by means of simulation.

4.2. Validation of the approximation versus simulation

In the experiment, we simulated production orders going through the supply stage of a two-stage supply chain. Table 2 gives an overview of the input data used in the experiment. As we know from the analysis in Section 4.1 that the transportation times do

² The use of ρ^2 as an interpolation factor in expression (23) is supported by earlier results on the approximation of variability in departure processes (see e.g. [17]).

not play a role in either $Cov[\sum_{i=2}^T X_i(n), R(n)]_{app}$ or $E[W]_{app}$, they were left out of the

validation experiment (i.e., D was assumed to be constant and equal to zero).

The transfer batch size L is equal to 1, with T equal to 4, 10, 20 or 30. The average setup time $E[SU]$ for a production order at the supplier is equal to 2, the average subplot processing time $E[X]$ is equal to 1. The average interarrival time of production orders at the supplier ($E[Y]$) was varied in order to yield 4 different utilization levels ($\rho=30\%$, 50%, 70% and 90%).

[Insert Table 2 here]

The squared coefficients of variation of the setup time (c_{SU}^2), the subplot processing time (c_X^2) and the production order interarrival time (c_Y^2) were taken as 3-level factors in the design of the experiment: each of these distributions could have low variability (SCV=0.25), medium variability (SCV=0.5) or high variability (SCV=1).

For each of the simulation runs, we kept track of the variance of the interarrival times of the flags ($Var[\tau_f]_{sim}$). In order to obtain proper estimates for $Var[\tau_f]_{sim}$ for a supplier with general probability distributions, the interarrival times of the process batches, the setup times and the subplot processing times were all simulated with either gamma, beta or lognormal distributions. As such, we obtained 3 independent estimates for each simulation setting. The final estimate for $Var[\tau_f]_{sim}$ was then obtained by calculating the average of these 3 independent estimates.

The design of the experiment was full factorial, resulting in 1296 simulation runs. In each run, 100,000 production orders were sent through the system; as the purpose is to

obtain a steady state estimate, the data for the first 20,000 orders in each run were considered as the warm-up period and left out of the analysis.

The results for $Var[\tau_f]_{sim}$ were compared to the approximation ($Var[\tau_f]_{app}$) in expression (24). The performance of the approximation was then judged by means of the relative errors (RE) of $Var[\tau_f]_{app}$ versus $Var[\tau_f]_{sim}$:

$$RE = \frac{Var[\tau_f]_{app} - Var[\tau_f]_{sim}}{Var[\tau_f]_{sim}}$$

and the absolute values of these errors (RE_{abs}). Figure 3 shows a histogram of the relative errors. Apparently, the approximation performs very well, yielding errors that stay between $\pm 3\%$. Table 3 gives an overview of the summary statistics of the relative errors (in percent), for the different values of ρ . The table confirms that the relative errors generated by the approximation are very small, irrespective of the utilization level.

Hence, we may conclude that expression (24) enables us to approximate very closely the delivery reliability of a supplier using a lot splitting policy.

[Insert Figure 3 here]

[Insert Table 3 here]

5. Conclusions and limitations

In this paper, we have investigated the delivery reliability of a supplier using a lot splitting policy, and compared it to the delivery reliability that would result from a lot-for-lot policy. We have been able to prove analytically that, in a general setting, the delivery reliability improves when a lot splitting policy is used. This is an interesting finding, as it implies that the scheduled start times for the production runs at the buyer's

facility can be planned with a higher level of certainty when a lot splitting policy is used. Hence, our findings provide additional rationale for the use of the lot splitting policy in a lean supply chain environment.

Furthermore, we have been able to develop an approximation, allowing us to estimate the delivery reliability of the supplier in terms of the system characteristics (e.g. level of variability of the setup and processing times, utilization of the supplier) and the lot splitting policy. The approximation has been tested versus simulation results, and proved to perform very well (the observed relative errors being between $\pm 3\%$). Hence, it provides us with a useful tool to predict the delivery reliability of any given lot splitting policy, such that this reliability can be taken into account in the design of production plans.

As we have only focused on the predictability of the start times of production runs, we have not taken into account the impact of the lot splitting policy on other supply chain performance measures, such as the transportation costs (which are bound to go up as the number of sublots increases) or the possibility of production disruptions at the buyer's facility (due to the late arrival of order sublots other than the flag; see [41]). When assessing the overall effectiveness of lot splitting in a particular context, these aspects should certainly be taken into account. However, as analytical insights into the impact of lot splitting in stochastic supply chains are currently scarce, we believe that our theoretical results may offer a useful first contribution to this field of research.

Appendix A: Proof for $-(T-1) * Var(X) \leq Cov[\sum_{i=2}^T X_i(n), R(n)]$

The proof is based on the concept of associated random variables. Therefore, we start by briefly stating its definition and two corollaries ([32]):

Definition: a set of random variables $\mathbf{X}=\{X_1, X_2, \dots, X_n\}$ is said to be associated if for all increasing functions f and g :

$$E[f(X)g(X)] \geq E[f(X)]E[g(X)]$$

A function h is said to be increasing if:

$$h(x_1, x_2, \dots, x_n) \leq h(y_1, y_2, \dots, y_n)$$

whenever $x_i \leq y_i$ for $i=1, \dots, n$.

Corollaries:

- Independent random variables are associated
- Increasing functions of associated random variables are associated

According to expression (3), we have:

$$\begin{aligned} & Cov\left[\sum_{i=2}^T X_i(n), R(n)\right] \\ &= Cov\left[\sum_{i=2}^T X_i(n), \max[0, Y(n) - W(n) - SU(n) - \sum_{i=1}^T X_i(n), 0]\right] \\ &= Cov\left[\sum_{i=2}^T X_i(n), \max[0, Y(n) - W(n) - SU(n) - X_1(n) - \sum_{i=2}^T X_i(n), 0]\right] \end{aligned}$$

For ease of notation, let

$$Z = Y(n) - W(n) - SU(n) - X_1(n)$$

$$V1 = \sum_{i=2}^T X_i(n)$$

Then we have to prove that:

$$-Var[V1] \leq Cov[V1, \max[0, Z - V1]]$$

We can write:

$$\max[0, Z - V1] = \gamma * (Z - V1)$$

with $\gamma = 1$ if $Z > V1$, and $\gamma = 0$ if $Z \leq V1$. Then:

$$\text{Cov}[V1, \max[0, Z - V1]] = \text{Cov}[V1, \gamma Z] - \text{Cov}[V1, \gamma V1] \quad (\text{A.1.})$$

As Z and $V1$ are independent, and γ^*Z is an increasing function of Z , we can use the concept of associated random variables ([35]) which yields:

$$\text{Cov}[V1, \gamma^*Z] \geq 0$$

such that from (A.1.), it follows that:

$$\text{Cov}[V1, \max[0, Z - V1]] \geq -\text{Cov}[V1, \gamma V1] \quad (\text{A.2.})$$

Now it remains to be proved that:

$$-\text{Cov}[V1, \gamma V1] \geq -\text{Var}[V1]$$

or, equivalently,

$$\text{Var}[V1] \geq \text{Cov}[V1, \gamma V1]$$

we know that

$$\begin{aligned} & \text{Var}[V1] - \text{Cov}[V1, \gamma V1] \\ &= \text{Cov}[V1, (1 - \gamma)V1] \end{aligned}$$

As $\gamma = 1$ if $Z \geq V1$, and $\gamma = 0$ otherwise, we know that $(1 - \gamma) * V1$ is an increasing function of $V1$, and it follows from the concept of associated random variables that:

$$\begin{aligned} & \text{Var}[V1] - \text{Cov}[V1, \gamma V1] \\ &= \text{Cov}[V1, (1 - \gamma)V1] \\ &\geq 0 \end{aligned}$$

Consequently,

$$\text{Var}[V1] \geq \text{Cov}[V1, \gamma V1]$$

and

$$-\text{Var}[V1] \leq -\text{Cov}[V1, \gamma V1]$$

such that, in combination with (A.2.), we have:

$$\text{Cov}[V1, \max[0, Z - V1]] \geq -\text{Cov}[V1, \gamma V1] \geq -\text{Var}[V1]$$

which is what we had to prove.

Appendix B: Proof for $\text{Cov}[\sum_{i=2}^T X_i(n), R(n)] \leq 0$

This proof is also based on the concept of associated random variables (see Appendix A).

According to expression (3), we have:

$$\begin{aligned}
& \text{Cov}[\sum_{i=2}^T X_i(n), R(n)] \\
&= \text{Cov}[\sum_{i=2}^T X_i(n), \max[0, Y(n) - W(n) - SU(n) - \sum_{i=1}^T X_i(n)]] \\
&= \text{Cov}[\sum_{i=2}^T X_i(n), -\min[0, -Y(n) + W(n) + SU(n) + X_1(n) + \sum_{i=2}^T X_i(n)]]
\end{aligned}$$

We can then define a set of random variables V :

$$V = \{SU(n) + X_1(n), W(n), Y(n), \sum_{i=2}^T X_i(n)\}$$

These random variables are independent, so they are associated ([32]).

Let's then define:

$$f(V) = \sum_{i=2}^T X_i(n)$$

$$g(V) = \min[0, W(n) + SU(n) + X_1(n) + \sum_{i=2}^T X_i(n) - Y(n)]$$

We can then write:

$$\begin{aligned}
& E[f(V) * g(V) | Y(n) = c] \\
&= E[\sum_{i=2}^T X_i(n) * \min[0, W(n) + SU(n) + X_1(n) + \sum_{i=2}^T X_i(n) - c]] \\
&\geq E[\sum_{i=2}^T X_i(n)] * E[\min[0, W(n) + SU(n) + X_1(n) + \sum_{i=2}^T X_i(n) - c]]
\end{aligned}$$

The last inequality follows from the concept of associated random variables. Since $W(n)$,

$SU(n) + X_1(n)$ and $\sum_{i=2}^T X_i(n)$ are independent, they are associated. The minimum

function is an increasing function in $W(n)$, $SU(n) + X_1(n)$ and $\sum_{i=2}^T X_i(n)$. Moreover, the

function $f(V) = \sum_{i=2}^T X_i(n)$ is also increasing in $W(n)$, $SU(n) + X_1(n)$ and $\sum_{i=2}^T X_i(n)$. As

increasing functions of independent random variables are associated ([32]), the inequality follows.

We then have:

$$E[f(V) * g(V) | Y(n) = c] \geq E[f(V)] * E[g(V) | Y(n) = c]$$

Taking expectations, we have:

$$E[E[f(V) * g(V) | Y(n) = c]] \geq E[E[f(V)] * E[g(V) | Y(n) = c]]$$

which yields ([32]):

$$E[f(V) * g(V)] \geq E[E[f(V)] * E[g(V) | Y(n) = c]]$$

As $E[f(V)]$ is independent of $Y(n)$, we can write:

$$E[f(V) * g(V)] \geq E[f(V)] * E[E[g(V) | Y(n) = c]]$$

As $E[E[g(V) | Y(n) = c]] = E[g(V)]$, we have:

$$E[f(V) * g(V)] \geq E[f(V)] * E[g(V)]$$

which proves that

$$\begin{aligned} & \text{Cov}[f(V), g(V)] \\ &= \text{Cov}\left[\sum_{i=2}^T X_i(n), \min[0, W(n) + SU(n) + X_1(n) + \sum_{i=2}^T X_i(n) - Y(n)]\right] \\ &\geq 0 \end{aligned}$$

so

$$\text{Cov}[\sum_{i=2}^T X_i(n), -\min[0, W(n) + SU(n) + X_1(n) + \sum_{i=2}^T X_i(n) - Y(n)]] \leq 0$$

which is what we had to prove.

Appendix C

The formula of Kramer-Lagenbach Belz ([23]) for the average waiting time in queue is given by:

$$\begin{aligned} E[W]_{app} &= \frac{\rho^2(ca^2 + cs^2)}{2l(1-\rho)} \text{Exp}\left\{\frac{-2(1-\rho)(1-ca^2)^2}{3\rho(ca^2 + cs^2)}\right\} & \text{if } ca^2 \leq 1 \\ &= \frac{\rho^2(ca^2 + cs^2)}{2l(1-\rho)} & \text{if } ca^2 > 1 \end{aligned} \quad (\text{C.1.})$$

In this expression, ρ refers to the utilization of the stage, l refers to the average arrival rate of production orders at the stage ($l = \frac{1}{E[Y]}$), and ca^2 and cs^2 refer to the squared coefficient of variation of the interarrival times of production orders, and the service time of production orders, respectively. For a single-product system, they are given by:

$$\begin{aligned} ca^2 &= \frac{\text{Var}[Y]}{(E[Y])^2} \\ cs^2 &= \frac{\text{Var}[SU] + N * \text{Var}[x]}{(E[SU] + N * E[x])^2} \end{aligned}$$

Acknowledgement

The authors wish to thank Prof. Dr. David D. Yao (Columbia University) and the two referees for their useful suggestions.

References

- [1] Ansari, A. and B. Modaress, 1987, The potential benefits of Just-in-Time purchasing for US manufacturing, *Production and inventory management*, 28, 30-36.
- [2] Banerjee, A., 1986, A joint economic-lot-size model for purchaser and vendor, *Decision Sciences*, 17, 292-311.
- [3] Banerjee, A. and S.-L. Kim, 1995, An integrated JIT inventory model, *International Journal of Operations and Production Management*, 15, 237-244.
- [4] Bartholomew, D., 1984, The vendor-customer relationship today, *Production and Inventory Management*, 25, 106-121.
- [5] Benjaafar, S., 1996, On production batches, transfer batches, and lead times, *IIE Transactions*, 28, 357-362.
- [6] Blumenfeld, D., 2001, *Operations research calculations handbook* (CRC Press).
- [7] Bukchin, J., M. Tzur and M. Jaffe, 2002, Lot splitting to minimize average flow time in a two-machine flow shop, *IIE Transactions*, 34, 953-970.
- [8] Chiang, C., 2001, Order splitting under periodic review inventory systems, *International Journal of Production Economics*, 70, 67-76.

- [9] Chiang, C. and W.-C. Chiang, 1996, Reducing inventory costs by order splitting in the sole sourcing environment, *Journal of the Operational Research Society*, 47, 446-456.
- [10] Ellram, L.M., 1991, A managerial guideline for the development and implementation of purchasing partnerships, *International Journal of Purchasing and Materials Management*, 27, 2-8.
- [11] Goldratt, E.M. and J. Cox, 1984, *The Goal: a process of ongoing improvement* (North River Press).
- [12] Goyal, S.K. and Y.P. Gupta, 1989, Integrated inventory models: the buyer-vendor coordination, *European Journal of Operational Research*, 41, 261-269.
- [13] Graves, S.C. and M.M. Kostreva, 1986, Overlapping operations in materials requirements planning, *Journal of Operations Management*, 6, 283-294.
- [14] Harber, D., D.A. Samson, A.S. Sohal and A. Wirth, 1990, Just-in-time: the issue of implementation, *International Journal of Operations and Production Management*, 10, 21-30.
- [15] Hay, E.J., 1990, Implementing JIT purchasing, *Production and Inventory Management Review*, 10, 38-40.
- [16] Hill, R.M., 1997, The single-vendor single-buyer integrated production-inventory model with a generalised policy, *European Journal of Operational Research*, 97, 493-499.

- [18] Hopp, W. J., M. L. Spearman and D.L. Woodruff, 1990, Practical strategies for lead time reduction, *Manufacturing Review*, 3, 78-84.
- [19] Jacobs, F.R. and D.J. Bragg, 1988, Repetitive lots: flow-time reductions through sequencing and dynamic batch sizing, *Decision Sciences*, 19, 281-294.
- [20] Karmarkar, U., 1987, Lot sizes, lead times and in-process inventories, *Management Science*, 33, 409-417.
- [21] Karmarkar, U., S. Kekre and S. Kekre, 1985, Lot sizing in multi-item multi-machine job shops, *IIE Transactions*, 17, 290-298.
- [22] Kim, S.-L. and D. Ha, 2003, A JIT lot-splitting model for supply chain management: enhancing buyer-supplier linkage, *International Journal of Production Economics*, 86, 1-10.
- [23] Kramer, W. and M. Lagenbach-Belz, 1976, Approximate formulae for the delay in the queueing system GI/GI/1, in: *Congressbook of the Eight International Teletraffic Congress (Melbourne)*, 235-1/8.
- [24] Larson, P., 1989, An inventory model which assumes the problem away: a note on Pan and Liao, *Production and Inventory Management*, 4, 73-74.

[25] Litchfield, J.L. and R. Narasimhan, 2000, Improving job shop performance through process queue management under transfer batching, *Production and Operations Management*, 9, 336-348.

[26] Manoochehri, G.H., 1984, Supplier and the just-in-time concept, *Journal of Purchasing and Materials Management*, 20, 16-21.

[27] Marshall, K.T., 1968, Some inequalities in queueing, *Operations Research*, 16, 651-665.

[28] Monden, Y., 1981, Smoothed production lets Toyota adapt to demand changes and reduce inventory, *Industrial Engineering*, 13, 42-51.

[29] Monden, Y., 1983, *Toyota production system: a practical approach to production management* (Industrial Engineers and Management Press).

[30] Pan, A. and C.-J. Liao, 1989, An inventory model under just-in-time purchasing agreements, *Production and Inventory Management Journal*, 30, 49-52.

[31] Ramasesh, R.V., 1991, Recasting the traditional inventory model to implement just-in-time purchasing, *Management Science*, 37, 428-443.

[32] Ross, S. M., 1996, *Stochastic Processes* (John Wiley and Sons).

- [33] Santos, C. and M. Magazine, 1985, Batching in single-operations manufacturing systems, *Operations Research Letters*, 4, 99-103.
- [34] Schonberger, R.J., 1982, *Japanese Manufacturing Techniques: nine hidden lessons in simplicity* (The Free Press).
- [35] Schonberger, R.J. and A. Ansari, 1984, Just-in-time purchasing can improve quality, *Journal of Purchasing and Materials Management*, 20, 2-7.
- [36] Shingo, S., 1985, *A revolution in Manufacturing: the SMED system* (Productivity Press).
- [37] Smunt, T.L., A.H. Buss and D.H. Kropp, 1996, Lot Splitting in stochastic flow shop and job shop environments, *Decision Sciences*, 27, 215-238.
- [38] Umble, M. and M. L. Srikanth, 1995, *Synchronous Manufacturing: principles for world-class excellence* (The Spectrum Publishing Company).
- [39] Vandaele, N., I. Van Nieuwenhuyse and S. Cupers, 2003, Optimal grouping for a nuclear magnetic resonance scanner, *European Journal of Operational Research*, 151, 181-192.
- [40] Van Nieuwenhuyse, I. and N. Vandaele, 2004, Determining the optimal number of sublots in a single-product, deterministic flow shop with overlapping operations, *International Journal of Production Economics*, 92, 221-239.

[41] Van Nieuwenhuyse, I., 2004, Lot splitting in single-product flow shops: issues of delivery reliability, production disruptions and flow times, PhD dissertation, University of Antwerp (Belgium)

[42] Waters-Fuller, N., 1995, Just in time purchasing and supply: a review of the literature, *International Journal of Operations and Production Management*, 15, 220-236.

[43] White, R.E. and W.A. Ruch, 1990, The composition and scope of JIT, *Operations Management Review*, 7, 9-18.

Figure 1: Flowchart of a production order consisting of 4 product units through a two-stage chain, lot-for-lot policy versus lot splitting policy

N	= size of the production run
L	= delivery lot size
T	= total number of deliveries in a single production run
ρ	= utilization of the supplier
Y	= interarrival time of orders at the supplier
SU	= setup time required at the supplier
x	= unit processing time at the supplier
W	= waiting time of the orders in the input queue of the supplier
τ_f	= interarrival time of lags at the buyer
τ_o	= interarrival time of complete orders at the buyer
X_i	= processing time of delivery lot i at the supplier ($i=1, \dots, T$)
R	= length of the idle period between busy periods at the supplier ≥ 0
D	= transportation time between supplier and buyer

Table 1: Overview of symbols and notation

Figure 2: Possible realization of interarrival times of orders at the supplier, service times of sublots and the resulting interarrival times of flags at the buyer

Factor	Levels	Number of levels
T	4, 10, 20, 30	4
L	1	1
$E[X]$	1	1
$E[SU]$	2	1
c_x^2	0.25, 0.5, 1	3
c_{SU}^2	0.25, 0.5, 1	3
c_Y^2	0.25, 0.5, 1	3
Probability distribution	Gamma, beta, lognormal	3
ρ	0.3, 0.5, 0.7, 0.9	4

Table 2: Input factors and levels for the simulation experiment

Figure 3: Histogram of the relative errors of $Var[\tau_f]_{app}$ versus $Var[\tau_f]_{sim}$

ρ	0.3	0.5	0.7	0.9
average RE	-0.57 %	-0.94 %	-0.78 %	-0.16 %
stdev RE	0.41 %	0.55 %	0.93 %	1.02 %
median RE	-0.50 %	-0.76 %	-0.64 %	-0.30 %
5% percentile RE	-1.51 %	-1.93 %	-2.53 %	-1.82 %
95% percentile RE	-0.16 %	-0.34 %	0.56 %	1.58 %
max RE	-0.13 %	-0.24 %	0.67 %	1.72 %
min RE	-1.66 %	-1.93 %	-2.83 %	-1.91 %
average RE_{abs}	0.57 %	0.94 %	0.95 %	0.87 %
stdev RE_{abs}	0.41 %	0.55 %	0.75 %	0.52 %

Table 3: Summary statistics for the relative errors (in percent) of $Var[\tau_f]_{app}$ versus

$$Var[\tau_f]_{sim}$$