

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/74757/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Disney, Stephen M. , Maltz, Arnold, Wang, Xun and Warburton., Roger D. H. 2016. Inventory management for stochastic lead times with order crossovers. *European Journal of Operational Research* 248 (2) , pp. 473-486. 10.1016/j.ejor.2015.07.047

Publishers page: <http://dx.doi.org/10.1016/j.ejor.2015.07.047>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



# Inventory Management for Stochastic Lead Times with Order Crossovers

Stephen M. Disney <sup>\*1</sup>, Arnold Maltz<sup>2</sup>, Xun Wang<sup>1</sup>, and Roger D.H. Warburton<sup>3</sup>

<sup>1</sup>*Logistics Systems Dynamics Group, Cardiff Business School, Cardiff University, Aberconway Building,  
Colum Drive, Cardiff, CF10 3EU, UK.*

<sup>2</sup>*W.P. Carey School of Business, Arizona State University, PO Box 874706, Tempe, AZ 85287, USA.*

<sup>3</sup>*Metropolitan College, Boston University, 808 Commonwealth Avenue, Boston, MA 02215, USA.*

## Abstract

We study the impact of stochastic lead times with order crossover on inventory costs and safety stocks in the order-up-to (OUT) policy. To motivate our research we present global logistics data which violates the traditional assumption that lead time demand is normally distributed. We also observe that order crossover is a common and important phenomena in real supply chains. We present a new method for determining the distribution of the number of open orders. Using this method we identify the distribution of inventory levels when orders and the work-in-process are correlated. This correlation is present when demand is auto-correlated, demand forecasts are generated with non-optimal methods, or when certain ordering policies are present. Our method allows us to obtain exact safety stock requirements for the so-called proportional order-up-to (POUT) policy, a popular, implementable, linear generalization of the OUT policy. We highlight that the OUT replenishment policy is not cost optimal in global supply chains, as we are able to demonstrate the POUT policy always outperforms it under order crossovers. We show that unlike the constant lead-time case, minimum safety stocks and minimal inventory variance do not always lead to minimum costs under stochastic lead-times with order crossover. We also highlight an interesting side effect of minimizing inventory costs under stochastic lead times with order crossover with the POUT policy—an often significant reduction in the order variance.

Keywords: Stochastic lead-times; Safety stock; Order cross-over; Order-Up-To policy; Global supply chains.

---

\*Corresponding Author: DisneySM@cardiff.ac.uk

# 1 Introduction

Global sourcing often allows access to low-cost supply but is frequently associated with long and variable lead times. These longer and more variable lead times bring with them a number of complications and potential pitfalls, from both a cost and a service perspective (Stalk 2006). In particular, inventory planners must now account for uncertainty in both demand and lead time when determining safety stock levels (Warburton & Stratton 2002).

We add to the literature on planning with stochastic lead times by formulating and testing a calculation of safety stock that reflects these real-world complications. Our method allows for order crossover and correlation between pipeline inventory and replenishment orders, a factor often ignored. We introduce a novel approach to better understand the distribution of outstanding orders when lead times are uncertain.

This research was motivated by both practical and analytical issues. Practically, we have tracked and analyzed logistics data for global supply chains for both major forwarders and retailers, and were struck by the violations of the lead time normality assumption see Fig. 1. Furthermore, most inventory models do not allow for order crossover where shipments are received in a different sequence from which they were dispatched, yet variable shipment delays, clerical errors, and random custom inspections can easily delay a shipment long enough for others to pass it. Robinson et al. (2001) also provide real-world examples of order crossover. Another investigation we have conducted is summarised in Figs. 2 and 3. Fig. 2 shows the distribution lead time between a supplier in Colorado, USA and a customer in Shenzhen, China. Fig. 3 tracks how many queue positions each shipment gained or lost between the date-sorted list of dispatches and the date-sorted list of arrivals. There are clearly a significant amount of order crossovers nearly 40% of orders cross.

From the analytical perspective, two prescriptions for inventory management are widely disseminated. These approaches use either an average (or maximum) lead time in the constant lead time reorder point solution or assume that the demand during the lead time is normally distributed and then use the mean and variance of a random sum of random variables to determine the reorder point. Neither approach is well-suited to global supply chains with long transit times and multiple hand-offs. Rationalizations have been made for what is clearly a suspect assumption (Chopra et al. 2004, Eppen & Martin 1988, Tyworth & O'Neill 1997).

This paper develops an exact theoretical treatment of the impact of the stochastic lead times with order crossover on the probability density function (pdf) of the net stock levels. As we progressed in our investigations, we also began questioning the well-known assertion (Kaplan 1970) that the order-up-to (OUT) model is always a good fit for global supply chains. We find that, when there is order crossover, lower average inventories are possible when the ordering strategy follows the linear proportional order-up-to (POUT)

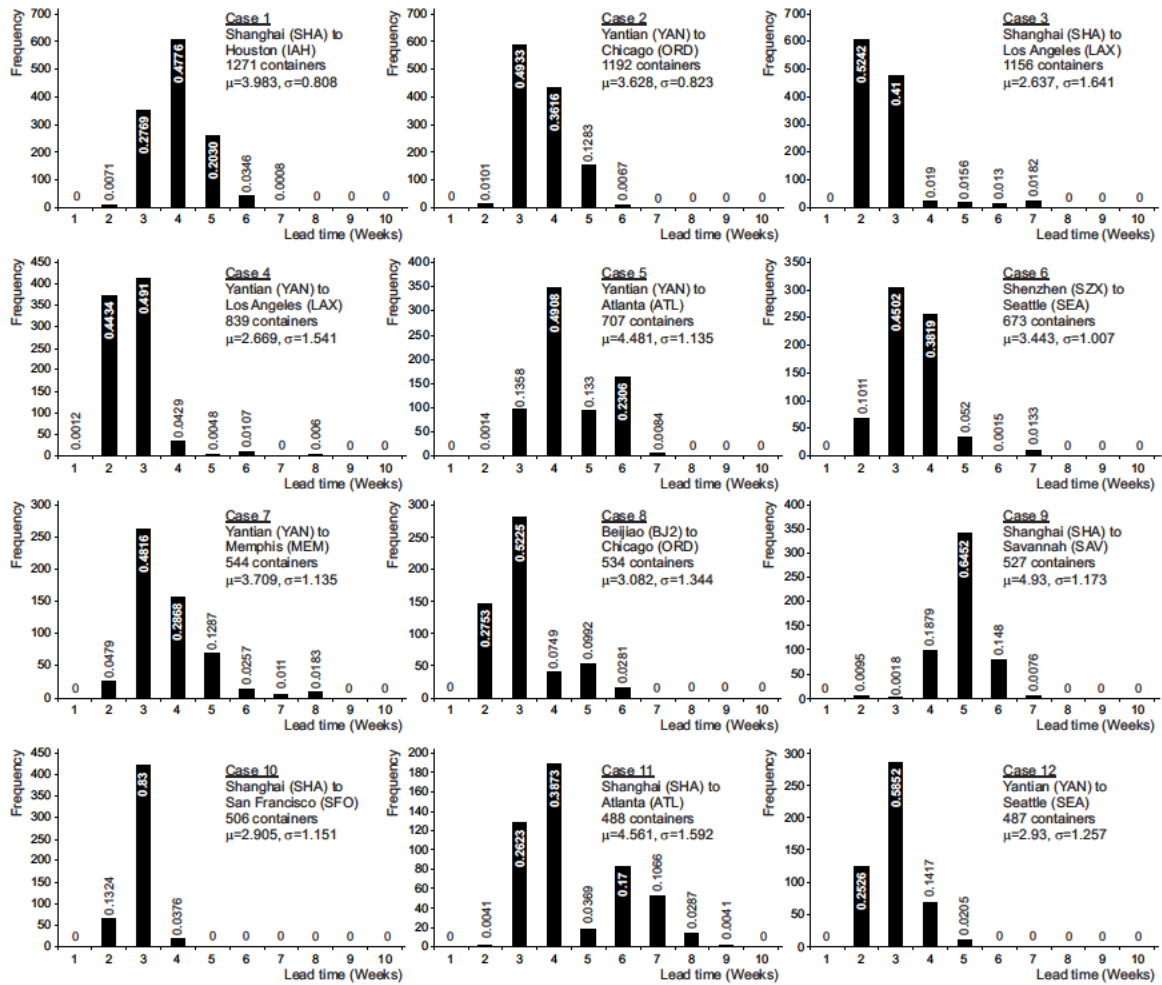


Figure 1: Real-world lead time distributions for port-to-port shipping from China to the USA during March-September, 2011.

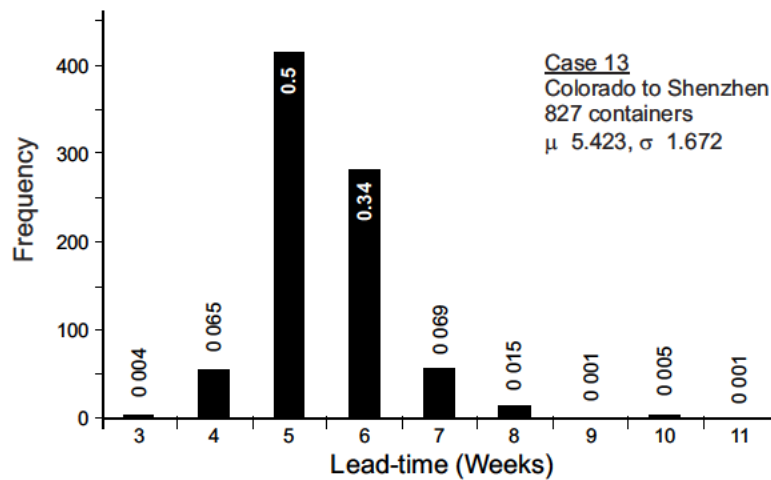


Figure 2: Door-to-door lead time distribution for Colorado-to-Shenzhen shipments.

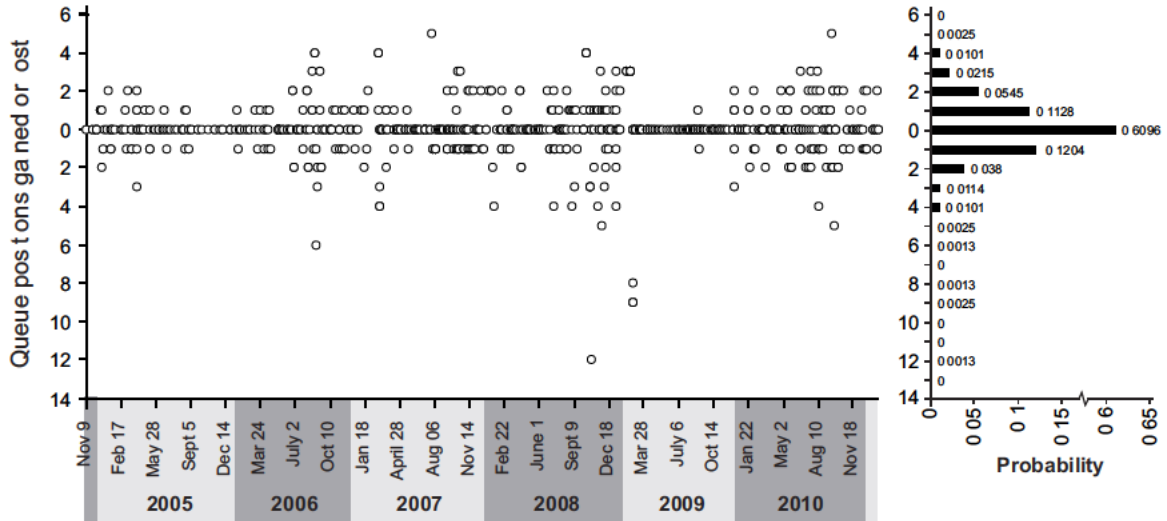


Figure 3: Order crossovers between Colorado and Shenzhen.

policy (Disney & Lambrecht 2008). We also became aware that the orders and work-in-progress (WIP) are cross-correlated, an effect that does not happen under the assumptions of independently and identically distributed (i.i.d.) demand, minimum mean square error forecasting, and OUT replenishment. However, whenever any one of these assumptions is relaxed, there is a co-variance term that must be accounted for. We believe this is the first treatment of this effect in the literature and this innovation allows one to relax a wide range of modeling assumptions in the study of order crossovers.

In the constant lead time case it is known that piecewise linear convex inventory holding and backlog costs are minimised when the safety stock is set to the critical newsvendor fractile. In this case, the replenishment policies should be selected and designed to minimise inventory variance and the optimal policy is the OUT policy. The minimal cost design also leads to minimal safety stock requirements. We will show that in the stochastic lead time case with order crossovers these insights do not always hold. While it remains true that optimal availability is the critical fractile, we find a linear policy that is always more economical than the OUT policy which sometimes requires less, sometimes more, safety stock. The inventory levels under our policy have less variance than the OUT policy. As an additional side benefit, our solution significantly reduces the variance of the replenishment orders even though our objective function is only concerned with inventory related costs.

The structure of the paper is as follows. After a review of the relevant literature (Section 2), we discuss some of the major assumptions and choices that we make in Section 3. We discuss the POUT/OUT policy with a constant lead time in Section 4.1. This establishes the notation for the stochastic lead time analysis. Developing further the approach of Robinson et al. (2001), we provide a new method to determine the distribution of the number of open orders for the stochastic lead time in Section 4.2. We

then study the POUT policy and present a new method for determining the pdf of the inventory levels in Section 4.3. This method allows for order crossover and is required as the orders and the WIP become correlated in the POUT policy. Section 5 presents a numerical example. Section 6 applies our theory to the empirical lead time distributions in Fig. 1 and Fig. 2. Section 7 concludes.

## 2 Literature Review

As with other recent treatments of these issues, we work in the periodic review, base stock inventory management framework (Bischak et al. 2014, Muharremoglu & Yang 2010, Srinivasan et al. 2011). Several streams of research are particularly relevant to our goal of better understanding the effects of stochastic lead times that are virtually certain to come up in global supply chains.

1. *Determining order quantities*: Simon (1952) outlined a mechanism for determining order quantities based on information about demand, inventory, WIP, lead times, and demand variability. Later, Kaplan (1970) proved that the so-called base stock or OUT inventory procedures were cost-optimal for stochastic lead times with no order crossover and independence between lead times and the number and size of open orders. This result is still widely cited and used today. In this paper, we suggest an alternative procedure for sizing orders based on a linear generalization of the OUT policy (Dejonckheere et al. 2003).
2. *Determining safety stocks*: Replenishment systems require safety stock targets,  $T$ , to be specified;  $T$  is the average inventory level. This is usually achieved with  $T = \Phi^{-1}[\alpha]\sigma\sqrt{1+k}$ , where  $\sigma$  is the standard deviation of demand,  $k$  is the replenishment lead time, and  $\Phi^{-1}[\alpha]$  is the inverse of the normal cumulative distribution function the ‘safety factor’ that achieves an availability of  $\alpha$ . This mainstream formula explicitly incorporates demand variability, but does not do the same for lead time variability. As lead times lengthen and process complexity increases with globalization, ignoring lead time variability seems ill-advised at best. Order crossover and correlation between orders and WIP, in particular, are real possibilities (Bradley & Robinson 2005, Muharremoglu & Yang 2010, Robinson et al. 2001, 2008, Srinivasan et al. 2011).
3. *Characterizing variability*: Reorder points and safety stock levels in a variable lead time setting are often based on the variance of a random sum of random variables. This second order moment is then used to set safety stock requirements via  $T = \Phi^{-1}[\alpha]\sqrt{k\sigma^2 + \mu^2\sigma_k^2}$  where  $\{\bar{k}, \sigma_k^2\}$  is the mean and variance of the lead time and  $\{\mu, \sigma^2\}$  is the mean and variance of the demand. This approach assumes that

the inventory levels are normally distributed. It is also a popular approach, despite having been shown to result in clear errors for even simple systems (Chopra et al. 2004, Eppen & Martin 1988, Tadikamalla 1984, Tyworth & O'Neill 1997) and somewhat more sophisticated treatments being available (Cachon & Terwiesch 2009, Silver et al. 1998).

Textbooks treatments of inventory planning and ordering policies generally start with purely deterministic models, such as the economic order quantity model, that posit constant demand and lead times. Typically, they then progress to 'probabilistic' models, such as Reorder Point,  $(Q, R)$  and OUT models, that account for demand variability and forecast errors by incorporating safety stock calculations into the setting of reorder points.

A few textbooks then try to add provisions for understanding lead time variability, usually through the well-known formula for computing the variance of a random sum of random numbers (Brown 1963, Feller 1966). Textbooks often then emphasize the need to have sufficient inventory on hand and on order, to cover customer demands until the next order arrives. Thus, they call out the critical issue of the demand during the lead time. Unfortunately, many texts simply assume that lead times are stable enough to be considered constant or that an uncertain demand combined with an uncertain lead time will result in a normal distribution of demand during the lead time. This has the advantage of resulting in a fairly simple safety stock calculation, but its reliability is in question. Others, (Axsäter 2000, for example) separate the problem into two classes: those without order crossover, which can be modeled using queuing theory, and those with crossover which can be approximated by the random sum of random variables approach.

With longer global supply chains, these assumptions become less tenable. Robinson et al. (2001) claimed that order crossover is actually fairly common, and our experience is consistent with their assertion (see Fig. 3). They found that using the shortfall distribution for planning purposes, rather than the distribution of lead time demand, results in better inventory performance. Although in later papers they acknowledge the usefulness of approximations (Robinson et al. 2008, Bradley & Robinson 2005), their fundamental insight is one that we build on in this research. Another important stream of research on order crossover involves the concept of an *effective lead time* (Hayya et al. 2011). This refers to the fact that the order crossover has the effect of reducing the average lead time.

There are practical issues that introduce uncertainty and variability into lead times from multiple sources. These include: ocean issues (Saldanha et al. 2009), import issues (Leachman & Jula 2012, Jula & Leachman 2011), transit reliability (Caplice & Kalkanci 2012, Kalkanci & Caplice 2012), containers (Fransoo & Lee 2012), and general issues (Stalk 2006).

There are also a number of common inventory approaches to studying the stochastic lead time problem. They can be classified as follows:

1. Lead times described by distribution functions (Bagchi et al. 1986, Bischak et al. 2014, Chopra et al. 2004, Eppen & Martin 1988, Hayya et al. 2008, 2011, Kim et al. 2006, Mentzer & Krishnan 1985).
2. OUT and base stock policies and procedures (Bischak et al. 2014, Hayya et al. 2011, Kaplan 1970).
3. Correlated and non-i.i.d. lead times (Bischak et al. 2014, Muharremoglu & Yang 2010).
4. Order crossovers (Bischak et al. 2014, Hayya et al. 2011, Muharremoglu & Yang 2010, Robinson et al. 2001, 2008, Srinivasan et al. 2011).
5. Effective lead times with order crossover (Bischak et al. 2014, Hayya et al. 2011).
6. Stochastic lead times and the bullwhip effect (Chatfield et al. 2004, Kim et al. 2006).
7. Endogenous lead times, where there may also be correlation between the orders and the WIP (Boute et al. 2014, So & Zheng 2003).

### 3 Safety Stocks, Lead Times and Demand

Kaplan (1970) found that the OUT policy will result in an optimal inventory cost if incoming orders do not cross. A natural question then arises: Is this still the case when the OUT policy is used in the presence of stochastic lead times with crossovers? We show that it is not as we are able to find a linear policy that outperforms the OUT policy. Whilst our policy may not be the optimal policy itself (see Srinivasan et al. 2011, where characteristics of the optimal non-linear policy are discussed), it is a linear policy that is well understood and implementable.

Our objective is always to minimise the sum of the unit inventory holding cost ( $h$ ) and the unit backlog cost ( $b$ ) as given  $J$ ,

$$J = h(I_t)^+ + b(-I_t)^+, \quad (1)$$

where  $I_t$  is the inventory levels at time  $t$ . It is well known that this cost is minimised when the safety stock is set so availability equals  $b/(h + b)$  (Brown 1963). The term *availability* is defined as:

$$\text{Availability} = \frac{\text{Number of periods without a stock-out}}{\text{Total number of periods}}. \quad (2)$$



Availability targets vary widely by product type. Basic products with no inventory risk can be sold over long periods and tend to have high availability. On the other hand, products with short life cycles may have stock-out rates of 50% or more (Warburton & Stratton 2002). Low availability does not necessarily imply a low fill rate, which directly measures the proportion of customer demand fulfilled from stock (Chopra et al. 2004). The fill rate is probably a more popular metric, especially in high-volume settings. However, due to its analytical complexity resulting from the double accounting of backlogs and the correlation between demand and the net stock (Disney et al. 2015), we have not pursued this approach herein.

### 3.1 Lead Times

From a practical perspective, the definition of the *lead time* deserves some thought. While it is straightforward to define lead time as the time from order to receipt, in the real-world this includes many factors. The lead time may include manufacturing, shipment to port, ship transit time, unloading (possibly with transfer to another ship), transfer to rail and/or truck, and unloading. Within that process are often customs clearances, both at outbound export and inbound import. Any of these factors can introduce variation into the lead time. If the product is not too bulky, many companies order some fraction to be delivered quickly via air, with the remainder to be delivered by sea. It is hoped that the availability of air transport for emergency shipments can reduce the risk of a stock-out.

Real-world logistics data typically include the time taken by all of the above activities, and any analysis is complicated by additional issues. Companies typically ship products in individual containers, but a ship transports many such containers, which then all experience the same transit time and, possibly, delays. Also, customs clearance depends on the type of cargo, random inspections, and the port at which it occurs. For example, clearance in a busy port (e.g., Los Angeles) can take a week, while in a small port (e.g., Providence) it may take less than a day. This is further exacerbated by the extremely large (and growing) container liner ships that frequent the busier ports. In addition, industrial action can have a major impact on port performance (O'Marsh 2014).

Fig. 1 presents a collection of data on lead times for shipments of product in containers from a port in China to a port in the United States. The figure plots the time from the COB to COT dates, highlighting the variability of the shipping delay. The COB is the date a container was *confirmed on-board* a ship at the port of origin. The COT refers to the *container out date*, which is the time that a container leaves its destination port. The key observation from Fig. 1 is that the lead time distribution is rarely normal, or even close to normal, and it often has a long tail. Also, this lead time actually represents only a part of the total lead time between a factory in China and the customer's warehouse in the U.S. The time required to get the products from the factory, into a container, delivered

to the port, through port operations, and onto a ship is not included. Also not included are the activities needed to get the container from the destination port to the customer's warehouse and to unload the container. These in-bound and out-bound activities may take a considerable amount of time, and may also be quite variable. As there is little reason to suspect that this variability is in any way correlated to the variability in the COB to COT lead times, we can only suppose that this variability will add to the shipping variability.

Another complication in the definition of the lead time that occurs in global supply chains is the inclusion of the time to manufacture products 'to order'. For example, many Chinese manufacturers will not schedule the production of an order until it has been paid for. In that case, the replenishment lead time may include the time to schedule and manufacture the product.

Rather than just the port-to-port lead times shown in Fig. 1, Fig. 2 gives the door-to-door lead times between a supplier in Colorado and a manufacturer in Shenzhen, China. This is the complete lead time from the moment the product leaves the factory in Colorado to the time it arrives and is booked in at the Shenzhen factory. This gives a more complete view of the lead time than that given by Fig. 1. We were also able to carefully investigate the timing of the dispatches and arrivals, and determined that order crossovers were actually quite frequent. In Fig. 3, we have illustrated how many queue positions each shipment gained or lost in sequences of dispatches and arrivals. Note that when one order gains (or loses) one position in the queue another must have lost (or gained) a position. However, if one order gains (or loses) several positions in the queue, one or more orders may have lost (or gained) one or more positions in the queue.

## 3.2 Demand

While some question the use of the normal distribution to represent demand (see Strijbosch et al. (2002)) the assumption of normally distributed demand has been adopted frequently in inventory management literature (Schneeweiss 1974, Disney & Towill 2003, Sobel 2004). We later show that under stochastic lead times the inventory distribution is made up of a weighted sum of sub-processes. The normally distributed demand, together with a linear system assumption means that these sub-processes are also normally distributed. If we further assume that demand is i.i.d., then the equations describing the first and second moments of the inventory levels are relatively simple.

Although most real demand patterns are likely to be autocorrelated over time, i.i.d. demands are observed in practice. Fig. 4 shows a demand series that we have collected from an industrial equipment manufacturer and it is both normally distributed and temporally independent. This series of data has passed the K-S test for normality hypothesis at significance of 0.1 ( $p = 0.1326$ ). The property of independence can be veri-

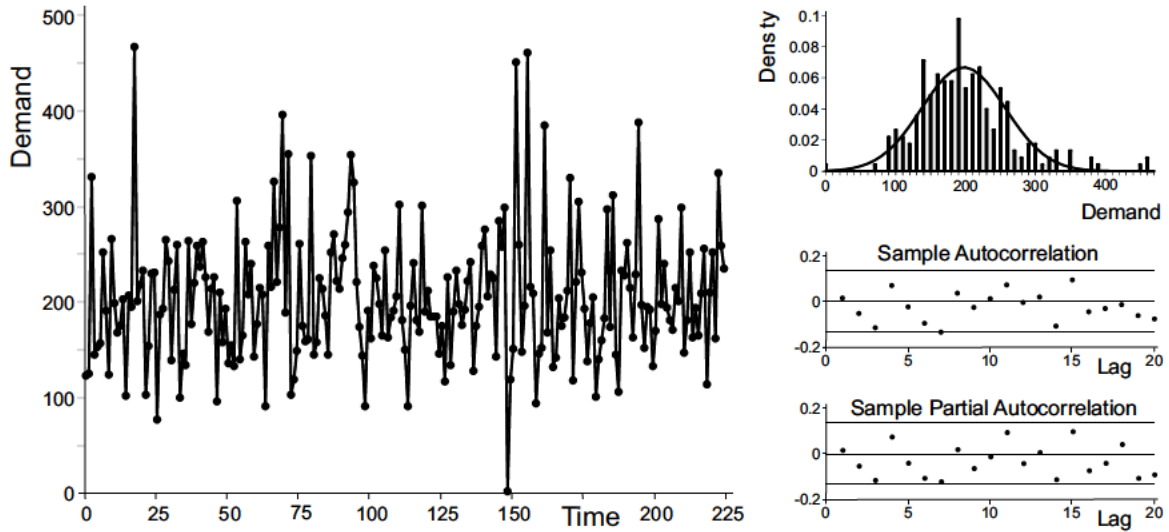


Figure 4: A real-world normally distributed i.i.d. demand pattern for industrial equipment.

fied through the autocorrelation function (ACF) and the partial autocorrelation function analysis (PACF), see Fig. 4. The two horizontal lines in the ACF and PACF diagrams show the critical values for the null hypothesis of independence. For a visual verification, Fig. 4 plots a normal pdf with a mean 197.1 and a standard deviation of 59.9. These parameters were chosen as they minimized the mean absolute error to the real data.

This particular demand pattern consists of integer units of demand as it is a discrete product. However given the large demand numbers, the relative error introduced by the continuous normal distribution is rather small. We also note that many other products can be sold in continuous units, especially those sold by weight, length or volume.

## 4 The Proportional Order-Up-To Policy

We first specify the discrete time POUT policy with constant lead times in Section 4.1 as it defines notation and introduces the policy. The POUT policy is a linear generalisation of the OUT policy, so studying the POUT policy allows us to also understand the OUT policy ‘for free’. We then move on to study the POUT policy with stochastic lead times and order crossover in Sections 4.2–4.3.

### 4.1 The Proportional Order-Up-To Policy with Constant Lead Times

We assume a sequence of events as shown in Fig. 5. We satisfy the demand and receive previously ordered product during the period. We observe the inventory and place a new

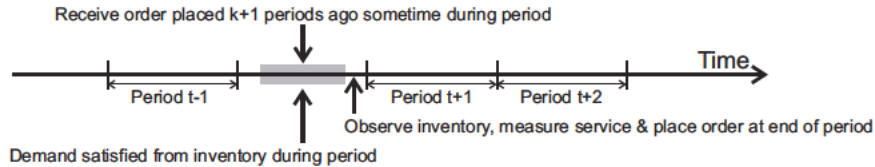


Figure 5: Sequence of events in the discrete time OUT policy.

order at the end of the period. At the end of the period we also determine the service level achieved with the availability metric. Note that if the physical lead time  $k = 0$ , then the order will be received in the next period and information about this receipt will be used in the next period's replenishment decision.

As the POUT policy operates on a discrete time basis, the lead time,  $k$ , must be an integer. Thus, an order with a physical lead time of 0.5 time periods, placed at the end of period  $t = 0$  will influence the inventory that will be observed at the end of period  $t = 1$ . That is, if the lead time is a decimal number then it must be rounded down to the integer. It is for this reason that we presented the real-world lead time data in Section 3.1 as discrete lead times (we will also make the assumption that the order lead times are uncorrelated and independent of the orders and all other system states). Furthermore, note the POUT policy has a sequence-of-events delay (the  $+1$  in the ' $k + 1$ ' of Fig. 5). The inventory balance equation is given by

$$I_t = I_{t-1} + R_t - D_t. \quad (3)$$

Here  $I_t$  is the inventory at time  $t$ ,  $D_t$  is the demand at time  $t$  and the variable  $R_t$  is used to keep track of the receipts. The receipts are merely the orders placed in period  $(t - k - 1)$ , so  $R_t = O_{t-k-1}$ , where  $k$  is a discrete (possibly random) variable that represents the time delay in the completion of orders ( $O_t$ ), which we leave undefined at the moment (see Fig. 5).

We assume that the demand is an i.i.d. random variable drawn from a normal distribution with a mean of  $\mu$  and standard deviation of  $\sigma$ . The minimum mean square error forecast of the demand for all periods ahead is given by its conditional expectation. For our i.i.d. demand process, this is

$$\hat{D}_{t:t+k} = \mu. \quad (4)$$

The POUT replenishment policy can be formulated (Disney & Lambrecht 2008) as

$$O_t = \mu + \beta(T + \bar{k}\mu - (I_t + W_t)). \quad (5)$$

Here,  $\beta$  is the proportional feedback controller that we can tune to regulate the speed at which deviations of the inventory position ( $I_t + W_t$ ) from the target ( $T + \bar{k}\mu$ ) are

corrected. This approach is common in hardware control and has a long history in both physical systems and inventory control systems (Simon 1952, Nise 2011).  $0 \leq \beta < 2$  is required for stability. When  $\beta = 1$  the POUT policy degenerates into the OUT policy. The POUT policy has been shown to attenuate the bullwhip effect (Disney & Towill 2003, Dejonckheere et al. 2003, Disney et al. 2004), and is relatively easy to implement in real supply chains; see Potter & Disney (2010) and Disney et al. (2013) for two case studies reporting how it has been implemented in practice. We can arrange (5) into the following form

$$O_t = \beta(D_t - \mu) + (1 - \beta)(O_{t-1} - \mu) + \mu. \quad (6)$$

showing that the POUT policy generates orders that are a convex combination of the stochastic component of the demand and the previous order (Balakrishnan et al. 2004, Boute & Van Mieghem 2014).

The term  $T$  in (5) is the target net stock – the safety stock – the average inventory level.  $T$  is a decision variable to be optimized to minimize inventory holding and backlog costs via the newsvendor principle.  $\bar{k}$  is the average lead time (when a constant lead time exists,  $\bar{k} = k$ ). The WIP is the inventory on order – the orders placed but not yet received, the in-transit inventory – and is given by

$$W_t = \sum_{i=1}^k O_{t-i} = W_{t-1} + O_{t-1} - R_t. \quad (7)$$

When  $k = 0$ , there is no WIP, as orders are received before the next order is generated. With the POUT policy, with arbitrary but constant lead-times, i.i.d. demand and MMSE forecasting the following expressions hold for the variance of the orders,

$$\sigma_O^2 = \frac{\sigma^2 \beta}{2 - \beta} \quad (8)$$

and the net stock variance,

$$\sigma_I^2 = \sigma^2 \left( k + \frac{1}{\beta(2 - \beta)} \right), \quad (9)$$

Disney & Towill (2003). Fig. 6 plots the bullwhip ratio ( $\sigma_O^2/\sigma^2$ ), and the net stock variance ratio, NSamp ( $\sigma_I^2/\sigma^2$ ), minus  $k$  as the influence of  $k$  is independent of  $\beta$ . The bullwhip ratio is unity at  $\beta = 1$ , zero at  $\beta = 0$ ,  $\infty$  at  $\beta = 2$ , strictly increasing, and convex in  $\beta$ . Note that the bullwhip ratio and order variance are not affected by the (possibly stochastic) lead time. The inventory variance is minimal (and equal to  $1 + k$ ) when  $\beta = 1$ ,  $\infty$  at  $\beta = \{0, 2\}$ , and convex in  $\beta$ . As the lead time  $k$  influences the inventory variance, the stochastic lead time will have an effect on the inventory distribution. However, for

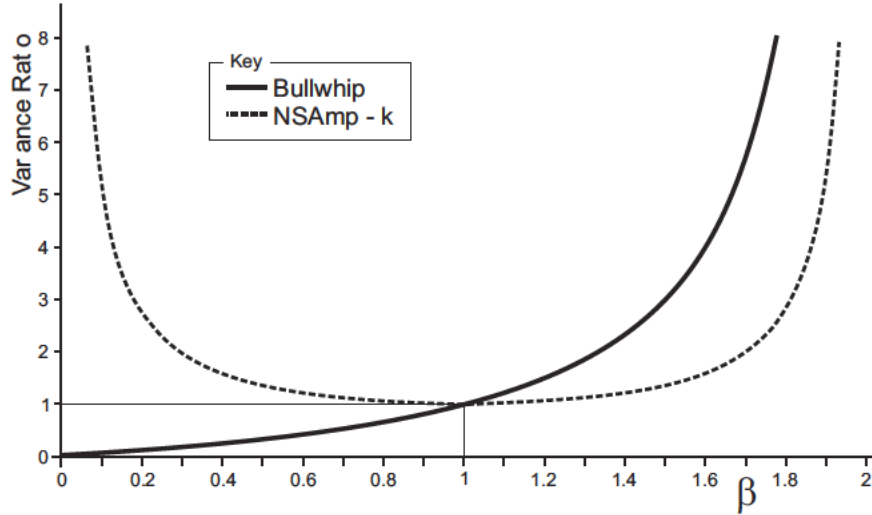


Figure 6: Bullwhip and net stock amplification ( $-k$ ) under a constant lead time of  $k$  for different  $\beta$ .

constant lead times,  $\beta = 1$  minimizes the net stock variance, safety stock requirements, and inventory costs. Also note that in this constant lead time setting, the influence of  $k$  on the inventory variance is independent of  $\beta$ .

When  $\beta = 1$ , the MMSE forecasting assumption (4) has a number of consequences. The most important consequence is that the orders in each period are simply the demand in that period,  $O_t = D_t$ . That is, the OUT policy acts as a ‘one-for-one’ or ‘pass-on-orders’ policy, a strategy often advocated by the lean production community.

In the constant lead time case the standard newsvendor critical fractile means that linear, convex inventory holding costs,  $h$ , and backlog costs,  $b$ , are minimized when

$$T_{OUT}^* = \sigma_{I,OUT} \Phi^{-1} \left[ \frac{b}{b+h} \right] \quad (10)$$

for the OUT policy. In the constant lead time case of the POUT policy,

$$T_{POUT}^* = \sigma_{I,POUT} \Phi^{-1} \left[ \frac{b}{b+h} \right]. \quad (11)$$

For constant lead times, when  $\beta \neq 1$  then  $\sigma_{I,POUT} > \sigma_{I,OUT}$ . Also when  $b \geq h$ ,  $0.5 \leq \alpha = b/(b+h) \leq 1$ , so  $T_{OUT}^* < T_{POUT}^*$ . In other words, when the availability target,  $\alpha$ , is greater than 50% for a constant lead time, the OUT policy will require less safety stock, because

$$0 < (T_{OUT}^* |_{\alpha > 0.5}) < (T_{POUT}^* |_{\alpha > 0.5}). \quad (12)$$

However, when  $b < h$ ,  $0 \leq \alpha = b/(b+h) < 0.5$ , so  $T_{POUT}^* < T_{OUT}^*$ . That is, when the availability target is less than 50% then

$$(T_{POUT}^*|_{\alpha < 0.5}) < (T_{OUT}^*|_{\alpha < 0.5}) < 0. \quad (13)$$

Note that when  $\alpha > 0.5$  ( $\alpha < 0.5$ ), the safety stock is positive (negative). The consequences of (12) and (13) will mean that there will be levels of availability where tighter control of inventory variance *reduces* safety stocks and levels of availability where tighter control of inventory variance *increases* safety stock requirements. Reductions in safety stock result from minimising variance when the critical fractile lays on a *leading* (increasing) edge of the pdf; when the critical fractile lays on a *trailing* (falling) edge of the pdf, one should not minimise variance to reduce safety stock. Later we will show that the inventory pdf is multi-modal and this will lead to cases when the the critical fractile swaps from leading to trailing edges (and vice versa) of the pdf. This effect was also noticed by Chopra et al. (2004).

## 4.2 The State of the WIP Pipeline

In the stochastic lead time case the probability of a lead time of  $k$  periods is denoted by  $p_k$ . The minimum lead time is  $k = 0$  and the maximum lead time is  $k^+$ . The average lead time (Zalkind 1978) is given by

$$\bar{k} = \sum_{k=0}^{k^+} p_k k. \quad (14)$$

The key to understanding the impact of the stochastic lead time is to consider the number of *open* replenishment orders in the WIP pipeline. Open orders are those that have been placed but not yet received. Notice that we are not describing the *quantity* of products on order, but the *number* of open orders. All orders placed  $k^+$  or more periods ago are guaranteed to have been received. However, those placed later than  $k^+ - 1$  periods ago may either be open (not yet received, denoted by a '1') or closed (received, denoted by '0'). Since each of the  $k^+$  positions in the pipeline is either open or closed, this means that there are  $2^{k^+}$  possible *states* of the WIP pipeline.

The probability that the pipeline is in state  $i$  is denoted  $q_i$ . The relationship between  $p_k$  and  $q_i$  is rather complex. To explain the relationship, consider a case where the lead time possibilities are  $p_0 = 0$ ,  $p_1 = 1/3$ ,  $p_2 = 1/2$  and  $p_3 = 1/6$ . Note that the probabilities sum to unity and the maximum lead time is  $k^+ = 3$ .

Table 1 describes all of the eight possible states to the WIP pipeline. The first column lists the state index,  $i$ . The next 4 columns denote the probability that order placed in period  $t - x$  has been received. The final column is the product of the four probabilities and denotes the probability that the WIP pipeline is in a particular state,  $q_i$ .

Consider state 1. In this state all four positions in the WIP pipeline have been closed. This is denoted by  $m_{i,j}$  and made specific in the first row with binary digits. In the second

row we have enumerated  $q_{i,j}$  which is the probability of that binary state occurring. This means that the probability of the last order (placed in period  $t - 1$ ) being received is zero,  $q_{1,1} = 0$  as  $p_0 = 0$ . The probability of the order placed two periods ago (at  $t - 2$ ) being received is  $q_{1,2} = p_0 + p_1$ , the probability of the order placed three periods ago being received is  $q_{1,3} = p_0 + p_1 + p_2$ . The order placed four period ago, at  $t - 4$  are guaranteed to be closed as  $q_{1,4} = p_0 + p_1 + p_2 + p_3 = 1$ . These probabilities are listed in the second row of information about the state. As there is no chance that the lead time  $k = 0$  (as  $p_0 = 0$ ), then  $q_1$ , the probability of the pipeline being in state one is zero. This is also true for states two through four.

Now consider state 8. In this state the first three positions of the pipeline are open, the fourth is closed. Recall, an open order has yet to be received, a closed order has been received. The probability of the first position being open is  $q_{8,1} = 1 - p_0$ , the second position being open is  $q_{8,2} = 1 - (p_0 + p_1)$ , the third being open is  $q_{8,3} = 1 - (p_0 + p_1 + p_2)$ . The order placed four periods ago is still guaranteed to be closed,  $q_{8,4} = 1$  as before. The probability that the pipeline is in state 8 is  $q_8 = (1 + p_0)(1 - (p_0 + p_1))(1 - (p_0 + p_1 + p_2))(p_0 + p_1 + p_2 + p_3) = \frac{1}{9}$ .

The complete set of pipeline states are shown in Table 1. It should also be clear from the process that leads to Table 1, that the sequence in which the orders actually arrive does not affect the calculation of probabilities, revealing that  $q_i$  is independent of the state of the pipeline.

We now formalize our methodology described above. Recall,  $q_i$  is the probability that the pipeline is in state  $i$ . Let  $\mathbf{M}$  be a binary matrix with  $j = 1$  to  $k^+$  columns and  $i = 1$  to  $2^{k^+}$  rows. Assign the  $(i, j)$  element of  $\mathbf{M}$  a value according to

$$m_{i,j} = \frac{1 + (-1)^v}{2} \quad (15)$$

where  $v = \lfloor 2^{j-k^+} i \rfloor$ . Each row of the  $\mathbf{M}$  matrix represents a  $k^+$ -tuple of binary digits that describes the state of the WIP pipeline. A zero in element  $m_{i,j}$  of matrix  $\mathbf{M}$  indicates that for state  $i$ , the order placed  $j - 1$  periods ago has been received (the order is closed), unity indicates that the order placed  $j - 1$  periods ago has not yet been received (it is open). Note, the order placed  $k^+$  periods ago is always closed, thus  $j$  indexes through 1 to  $k^+$  to represent the lead times  $k = 0$  to  $k^+ - 1$ . There are  $2^{k^+}$  rows to  $\mathbf{M}$ , one for each possible state of the order pipeline. The probability that the WIP pipeline is in state  $i$  is given by

$$q_i = \frac{1}{2^{k^+}} \prod_{j=1}^{k^+} \left[ 1 + (-1)^v \left( 2 \sum_{k=j}^{k^+} p_k - 1 \right) \right] \quad (16)$$

One can derive (16) by observing that in the  $i$ th pipeline state, the probability that an order placed  $j$  periods ago is open/closed can be expressed universally as



Table 1: The combinations of open orders and their associated probabilities.

State $i$	$t - 1, j = 1$	$t - 2, j = 2$	$t - 3, j = 3$	$t - 4, j = 4$	Probability, $q_i = \prod_{j=1}^4 q_{i,j}$
1	0	0	0	0	$q_1 = 0$
	0	$0 + \frac{1}{3}$	$0 + \frac{1}{3} + \frac{1}{2}$	$0 + \frac{1}{3} + \frac{1}{2} + \frac{1}{6}$	
2	0	0	1	0	$q_2 = 0$
	0	$0 + \frac{1}{3}$	$1 - (0 + \frac{1}{3} + \frac{1}{2})$	$0 + \frac{1}{3} + \frac{1}{2} + \frac{1}{6}$	
3	0	1	0	0	$q_3 = 0$
	0	$1 - (0 + \frac{1}{3})$	$0 + \frac{1}{3} + \frac{1}{2}$	$0 + \frac{1}{3} + \frac{1}{2} + \frac{1}{6}$	
4	0	1	1	0	$q_4 = 0$
	0	$1 - (0 + \frac{1}{3})$	$1 - (0 + \frac{1}{3} + \frac{1}{2})$	$0 + \frac{1}{3} + \frac{1}{2} + \frac{1}{6}$	
5	1	0	0	0	$q_5 = \frac{5}{18}$
	$1 - 0$	$0 + \frac{1}{3}$	$0 + \frac{1}{3} + \frac{1}{2}$	$0 + \frac{1}{3} + \frac{1}{2} + \frac{1}{6}$	
6	1	0	1	0	$q_6 = \frac{1}{18}$
	$1 - 0$	$0 + \frac{1}{3}$	$1 - (0 + \frac{1}{3} + \frac{1}{2})$	$0 + \frac{1}{3} + \frac{1}{2} + \frac{1}{6}$	
7	1	1	0	0	$q_7 = \frac{5}{9}$
	$1 - 0$	$1 - (0 + \frac{1}{3})$	$0 + \frac{1}{3} + \frac{1}{2}$	$0 + \frac{1}{3} + \frac{1}{2} + \frac{1}{6}$	
8	1	1	1	0	$q_8 = \frac{1}{9}$
	$1 - 0$	$1 - (0 + \frac{1}{3})$	$1 - (0 + \frac{1}{3} + \frac{1}{2})$	$0 + \frac{1}{3} + \frac{1}{2} + \frac{1}{6}$	

$$q_{i,j} = m_{i,j} \sum_{k=j}^{k^+} p_k + (1 - m_{i,j}) \sum_{k=1}^{j-1} p_k = \frac{1}{2} \left[ 1 + (-1)^v \left( 2 \sum_{k=j}^{k^+} p_k - 1 \right) \right], \quad (17)$$

and that  $q_i$  is the product of  $q_{i,j}$  over  $j$ .

Robinson et al. (2001) provided an iterative algorithm for determining the distribution of the number of open orders. It produces exactly the same results as (16) for finitely dimensioned discrete distributions.

### 4.3 The pdf of the Inventory Levels with Stochastic Lead Times and Order Crossover.

We define a *process* as the sequence of a variable over time ( $\{I_t\}, \{W_t\}$  etc). A *sub-process* is a subset of the process where the pipeline states (the completion status of previous orders, the rows in the  $\mathbf{M}$  matrix) are the same. Each sub-process is normally distributed (as the demand is normally distributed and each sub-process is the output of a linear system) and the distribution of the entire process can be multi-modal. We now require the mean and variance of the inventory levels in each of the sub-processes. We obtain this by first determining the distribution of the WIP in each sub-process and then each WIP sub-process is combined with a scaled replenishment order to obtain something we call the *scaled shortfall distribution*. A weighted sum of the scaled shortfall distributions in each sub-process then forms the complete inventory distribution.

We can rearrange (5) to obtain

$$I_t = T + \mu (\bar{k} + 1/\beta) - (W_t + O_t/\beta). \quad (18)$$

For OUT policy (that is, when  $\beta = 1$ ) we can see that the inventory distribution is a reflected *shortfall distribution*,  $(W_t + O_t)$ , translated by  $T + \mu(\bar{k} + 1/\beta)$  (Zalkind 1978; Robinson et al. 2001). When  $\beta \neq 1$  the  $O_t$  component has become scaled by  $O_t/\beta$ , in which case we call the distribution of  $(W_t + O_t/\beta)$  the *scaled shortfall distribution*. We now require the mean and the variance of the scaled shortfall distribution for each sub-process. The complicating factors are that  $O_t$  is auto-correlated and that the distributions of  $W_t$  and  $O_t/\beta$  are correlated with each other. As the system is linear the simplest way to proceed is to exploit the z-transform, which is defined by

$$F(z) = Z \{f[t]\} = \sum_{t=0}^{\infty} f[t] z^{-t}. \quad (19)$$

To determine the variance of the WIP in sub-process  $i$ , we first note that the variance of the orders maintained by the POUT policy is independent of the lead-time, as

$$\frac{\sigma_O^2}{\sigma^2} = \sum_{t=0}^{\infty} \left( Z^{-1} \left\{ \frac{z\beta}{z + \beta - 1} \right\} \right)^2 = \sum_{t=0}^{\infty} ((1 - \beta)^t \beta)^2 = \frac{\beta}{2 - \beta}, \quad (20)$$

Disney & Towill (2003). Here,  $z$  is the  $z$ -transform operator,

$$Z^{-1} \{F(z)\} = \frac{1}{2\pi j} \oint_C F(z) z^{t-1} dz = f[t], \quad (21)$$

$f[t]$  is the inverse  $z$ -transform of the transfer function,  $F(z)$ .  $z\beta(z + \beta - 1)^{-1}$  is the transfer function of the orders maintained by the POUT policy under i.i.d. demand and minimum mean squared error forecasting (Disney and Towill 2003). The relationship between the variance ratio and the sum of the squared impulse response is known as Tsyppkin's (1964) relationship.

The pdf of the normal distribution with an argument of  $x$ , a mean of  $\mu$ , and a standard deviation of  $\sigma$ , is defined by

$$\phi[x|\mu, \sigma] = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma}. \quad (22)$$

Using this notation, (9) leads to an order process described by the pdf,

$$\phi_O = \phi \left[ x | \mu, \sqrt{\sigma^2 \beta / (2 - \beta)} \right]. \quad (23)$$

The variance of WIP sub-process  $i$ , is given by the variance of the sum of the impulse responses of the open orders,

$$\frac{\sigma_{W,i}^2}{\sigma^2} = \sum_{t=0}^{\infty} \sum_{j=1}^{k^+} \left( m_{i,j} Z^{-1} \left\{ \frac{\beta z^{1+j}}{z + \beta - 1} \right\} \right)^2. \quad (24)$$

where  $m_{i,j}$  is an element of the binary matrix  $\mathbf{M}$  that captures whether an order is open or closed. The distribution of the scaled orders,  $O_t/\beta$ , for all sub-processes, can be obtained using

$$\frac{\sigma_{O/\beta}^2}{\sigma^2} = \sum_{t=0}^{\infty} (Z^{-1} \{z/(z + \beta - 1)\})^2 = \sum_{t=0}^{\infty} ((1 - \beta)^t)^2 = (2\beta - \beta^2)^{-1} = \frac{\sigma_O^2}{\sigma^2 \beta^2}, \quad (25)$$

which leads to the following expression for its pdf,

$$\phi_{O/\beta} = \phi \left[ x | \mu/\beta, \sqrt{\sigma^2 / (2\beta - \beta^2)} \right]. \quad (26)$$

The covariance between the WIP sub-process and the scaled orders sub-process is given by

$$\text{cov}(W_i, O/\beta) = \sum_{t=0}^{\infty} \left( Z^{-1} \left\{ \sum_{j=1}^{k^+} m_{i,j} \frac{\beta z^{1-j}}{z + \beta - 1} \right\} Z^{-1} \left\{ \frac{z}{z + \beta - 1} \right\} \right) = \text{cov}(W_i, O) / \beta. \quad (27)$$

$\sigma_{I,i}^2$ , the variance of sub-process  $i$  in the inventory distribution, is equal to the variance of the shortfall distribution,

$$\sigma_{I,i}^2 = \sigma_{W,i}^2 + \sigma_O^2 / \beta^2 + 2\text{cov}(W_i, O) / \beta. \quad (28)$$

The mean of the each of the sub-processes of the inventory distribution can be shown to be

$$\mu_{I,i} = T + \mu \left( \bar{k} - \sum_{j=1}^{k^+} m_{i,j} \right). \quad (29)$$

The complete pdf inventory distribution is then given by

$$\phi_I = \sum_{i=1}^{2^{k^+}} q_i \phi \left[ x \mid \mu_{I,i}, \sqrt{\sigma^2 \sigma_{I,i}^2} \right]. \quad (30)$$

We emphasize that  $\phi_I$  is a multi-modal pdf as it is a combination of the normally distributed pdfs with different means and variances weighted by  $q_i$ . The average inventory is given by  $T$  and this can be set arbitrarily. However, if one wishes to minimise inventory holding and backlog costs,  $T$  becomes a function of  $\beta$ . The variance of the complete, multi-modal, inventory is given by

$$\sigma_I^2 = \int_{-\infty}^{\infty} \phi_I (T - x)^2 dx = \sum_{i=1}^{2^{k^+}} q_i \left( \left( \sum_{j=1}^{k^+} m_{i,j} \right)^2 \mu^2 + \sigma^2 \sigma_{I,i}^2 \right). \quad (31)$$

Equation (31) shows that the inventory variance contains a weighted sum of the variances of individual sub-processes. Equation (31) also shows that the mean demand has an influence on the variance of the inventory levels, an effect that does not happen with constant lead times. An influence of the mean demand can be also seen in the standard random sum of random variables variance equation,  $\bar{k}\sigma^2 + \mu^2\sigma_k^2$ . However, we note that (31) has a different structure from this formula, and will produce significantly different guidance for the inventory variance and safety stock. One should not be surprised at this because the random variables in the random sum are drawn from different distributions in the (P)OUT policy.

When simulating this scenario in a spreadsheet for verification of our analytical approach, we found that it is convenient to generate alongside the order,  $O_t$ , a random number used to determine the lead time for that order. We can then determine with a simple logical test the time at which the order is received. This ensures that complete

orders are received, i.e. they are not split – each individual order is received all at once. As we allow for orders to cross some periods receive more than one order; in other periods there could be no receipts. This is consistent with our practical experience discussed earlier. Notice that we also assumed that the random lead times are independent of all the other system states.

## 5 A numerical example when $k^+ = 4$

Consider the situation when  $k^+ = 4$ . Table 2 details the pipeline states ( $\mathbf{M}$ ), the variance of the net stock, and the mean of each of the  $2^{k^+} = 16$  individual sub-processes to the inventory distribution. It can be easily shown that each of the expressions for the variance (and the standard deviations) of the inventory sub-processes is infinite at  $\beta = \{0, 2\}$ . Furthermore, each sub-process has a single unique minimum,  $\beta_i^*$ , which is also detailed in Table 2. We can see that  $\beta_i^* = 1$  exists only in the sub-processes that do not contain order cross-overs. All of the sub-processes that contain order-crossover have  $\beta_i^* < 1$ .

An intuitive explanation of this is as follows (to avoid unnecessary notation assume, for this paragraph only, that  $\sigma = 1$ ). Recall, the variance of each of the inventory sub-processes is given by (28) and  $\sigma_O^2/\beta^2$  is infinite at  $\beta = \{0, 2\}$ , minimised to unity at  $\beta = 1$ , and convex in  $\beta$ . For the sub-processes without order crossover then  $\sigma_{W_i}^2 + 2\text{cov}(W_i, O)/\beta = \sum_{j=1}^{k^+} m_{i,j}$ , a constant. The variance of the sub-processes without order crossover are then clearly minimised at  $\beta = 1$ . However, for sub-processes with order cross-over then  $\sigma_{W_i}^2 + 2\text{cov}(W_i, O)/\beta$  is convex in  $\beta$  between  $\beta = 0$  and  $\beta = 1$  and equal to  $\sum_{j=1}^{k^+} m_{i,j}$  at  $\beta = \{0, 1\}$ . This implies there will be a minimum in  $\sigma_{I_i}^2$  between  $0 \leq \beta < 1$ . As the complete inventory pdf is a weighted sum of independent variances, some minimised with  $\beta = 1$ , some minimised with  $\beta < 1$ , then the  $\beta$  that minimises the variance of the complete inventory distribution is  $\beta_\sigma^* < 1$ . Also, the proportion of states with order crossover increases in  $k^+$  as the number of states with order crossovers is given by  $2^{k^+} - (k^+ + 1)$ . This suggests that the role of  $\beta$  becomes more important as  $k^+$  increases.

Table 2 details the first and second order moments of each of the subprocesses for a maximum lead time of  $k^+ = 4$ . To identify the probability of the pipeline being in state  $i$ , we now need to make the results in Table 2 specific by enumerating the lead time probabilities. Assume  $\{p_0 = \frac{1}{2}, p_1 = p_2 = p_3 = 0, p_4 = \frac{1}{2}\}$  which we have chosen as it allows us to better reveal the impact of the order crossovers. It may also be representative of a supply chain where 50% of orders are sent via ship with a lead time of four and 50% are sent by air with a lead time of zero. That is, the air shipments arrive before the next order is made. The maximum lead time is  $k^+ = 4$  and the average lead time is  $\bar{k} = 2$ . Using (5) we are then able to determine the probability that the pipeline is in state  $i$ , is  $\forall i, q_i = 0.0625$ . Note that in general, the probability that the pipeline is in a particular

Table 2: The combinations of open orders and their associated variance, mean and variance minimizing feedback controller.

M	j				$\frac{\sigma_{I,i}^2}{\sigma^2}$	$\mu_{I,i}$	$\beta_i^*$
	i	1	2	3			
1	0	0	0	0	$\frac{1}{\beta(2-\beta)}$	$T + \mu\bar{k}$	1
2	0	0	0	1	$\frac{-2\beta^5+8\beta^4-12\beta^3+7\beta^2-2\beta-1}{\beta(2-\beta)}$	$T + \mu(\bar{k} - 1)$	0.656633
3	0	0	1	0	$\frac{2\beta^4-6\beta^3+5\beta^2-2\beta-1}{\beta(2-\beta)}$	$T + \mu(\bar{k} - 1)$	0.689845
4	0	0	1	1	$\frac{2\beta^5-10\beta^4+16\beta^3-10\beta^2+4\beta+1}{\beta(2-\beta)}$	$T + \mu(\bar{k} - 2)$	0.60974
5	0	1	0	0	$\frac{2\beta^3-3\beta^2+2\beta+1}{\beta(2-\beta)}$	$T + \mu(\bar{k} - 1)$	0.751274
6	0	1	0	1	$\frac{2\beta^5-6\beta^4+10\beta^3-8\beta^2+4\beta+1}{\beta(2-\beta)}$	$T + \mu(\bar{k} - 2)$	0.676129
7	0	1	1	0	$\frac{-2\beta^4+6\beta^3-6\beta^2+4\beta+1}{\beta(2-\beta)}$	$T + \mu(\bar{k} - 2)$	0.689845
8	0	1	1	1	$\frac{-2\beta^5+8\beta^4-12\beta^3+9\beta^2-6\beta-1}{\beta(2-\beta)}$	$T + \mu(\bar{k} - 3)$	0.656633
9	1	0	0	0	$\frac{\beta^2-\beta-1}{\beta(2-\beta)}$	$T + \mu(\bar{k} - 1)$	1
10	1	0	0	1	$\frac{-2\beta^4+6\beta^3-6\beta^2+4\beta+1}{\beta(2-\beta)}$	$T + \mu(\bar{k} - 2)$	0.689845
11	1	0	1	0	$\frac{2\beta^3-4\beta^2+4\beta+1}{\beta(2-\beta)}$	$T + \mu(\bar{k} - 2)$	0.751274
12	1	0	1	1	$\frac{2\beta^4-6\beta^3+7\beta^2-6\beta-1}{\beta(2-\beta)}$	$T + \mu(\bar{k} - 3)$	0.689845
13	1	1	0	0	$\frac{-2\beta^2+4\beta-1}{\beta(2-\beta)}$	$T + \mu(\bar{k} - 2)$	1
14	1	1	0	1	$\frac{2\beta^3-5\beta^2+6\beta+1}{\beta(2-\beta)}$	$T + \mu(\bar{k} - 3)$	0.751274
15	1	1	1	0	$\frac{-3\beta^2+6\beta+1}{\beta(2-\beta)}$	$T + \mu(\bar{k} - 3)$	1
16	1	1	1	1	$\frac{-4\beta^2+8\beta+1}{\beta(2-\beta)}$	$T + \mu(\bar{k} - 4)$	1
Overall	-	-	-	-	Eq (31)	T	$\beta_\sigma^* = 0.73$

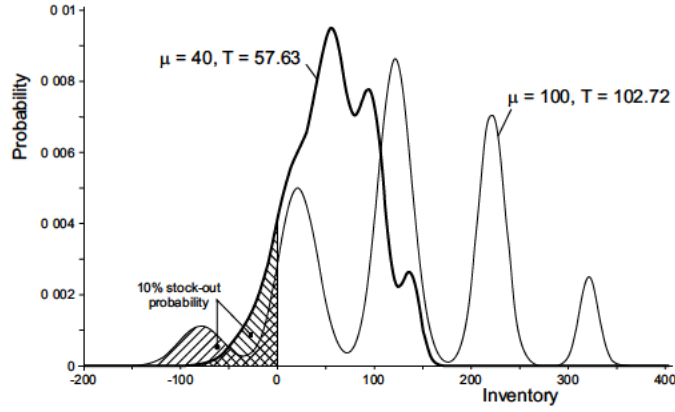


Figure 7: Distribution of the inventory levels maintained by the OUT policy for 90% availability

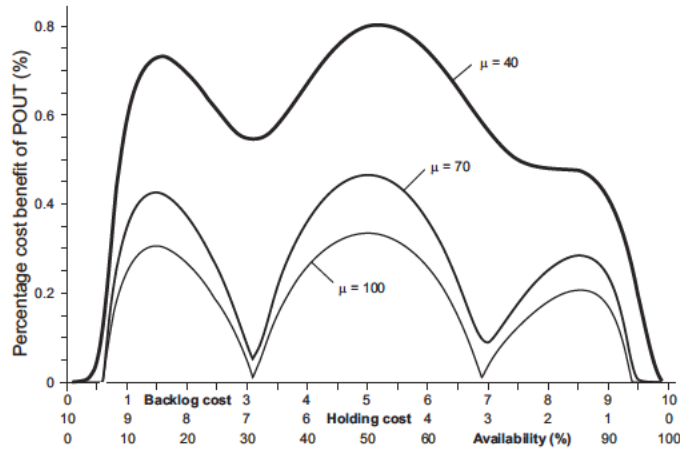


Figure 8: The percentage economic gain from using the POUT policy

state need not be, and almost never is, the same as the probability that the pipeline is another state.

Assigning costs to the inventory holding and backlog positions with (1) it is possible to show that

$$\left. \frac{dJ}{d\beta} \right|_{\beta=1} = \frac{\sqrt{\sigma^2} (b + h) e^{-(T+\mu)^2/4\sigma^2} \left( 6 + 3\sqrt{2}e^{(T^2+6T\mu+\mu^2)/8\sigma^2} + 4\sqrt{6}e^{(T^2+6T\mu+3\mu^2)/12\sigma^2} \right)}{64\sqrt{\pi}} > 0, \quad (32)$$

implying that, for all cost combinations, the OUT policy is never optimal as there always exists a  $\beta < 1$  which is more economical.

Together with the variances of the individual sub-processes detailed in Table 2, (4) and (18) we are now able to obtain an expression of the pdf of the inventory levels, which we plot in Fig. 7. Fig. 7 illustrates two cases of the OUT policy (that is when  $\beta = 1$ ,

and in both we have set the safety stock,  $T$ , to minimize  $J$  when  $h=1$ ,  $b=9$ . In the first case  $\mu = 100$  and we can clearly see that there are five modes in the inventory pdf. Although the probability of being in each of the 16 states is equal in this case, the 16 states have only five unique means. This leads to the differences in the size of each mode but, ultimately, they are all functions of  $\mu$ ,  $\sigma$ , and the lead time probabilities. In the other case,  $\mu = 40$  and the distributions of the sub-processes overlap more. Furthermore, the complete pdf of the  $\mu = 40$  case has less variance, and requires less safety stock, than the  $\mu = 100$  case.

When  $\mu = 100$ , the inventory levels have a variance of 10,300 for the OUT policy. Numerical experiments reveal that there is a single minimum inventory variance (or standard deviation) at  $\beta_\sigma^* = 0.73$  and the net stock variance is 10,280 – 0.2% less than the OUT variance. For the  $\mu = 40$  case, the inventory variance maintained by the OUT policy is 1900, the numerically optimised feedback parameter is the same,  $\beta_\sigma^* = 0.73$ , and the net stock variance is 1879 – a 1% reduction.

Using numerical techniques we can find the optimal proportional feedback controller  $\beta^*$ , and safety stock  $T$ , that minimizes the inventory cost. When we have set  $\{\beta^*, T\}$  optimally, Fig. 8 describes the percentage economic gain  $((J_{OUT} - J_{POUT}) J_{OUT}^{-1} \times 100\%)$ , from using the POUT policy. While the improvement is rather small (note that 0.8 means 0.8% not 80%), the POUT is always more economical than the OUT policy. These cost reductions are induced by a reduction in the inventory variance. The variance consists of two parts, one with different modes which depend on  $\mu$ , but independent of  $\beta$ ; the other one is a function of the variances of each sub-process which does depend on  $\beta$ . The former part is dominant, hence the inventory cost benefit of the POUT policy is limited and decreases in  $\mu$ .

Fig. 9 plots  $\beta^*$  for different cost ratios and different mean demands. We see that  $\beta^*$  is near unity when the availability target is (very) near 0% or 100%, but for most availability targets  $\beta^* \approx 0.725$ . Interestingly, almost always,  $\beta^* \neq \beta_\sigma^*$  implying that the tightest inventory control does not always lead to the minimal cost. The abrupt changes in  $\beta^*$  that we see in Fig. 9 are a result of the critical fractile moving from leading to trailing edges of the modes in the inventory pdf. When the cost ratios are such that the critical fractile lies on the leading edge,  $\beta^*$  tends towards unity; when they are on a falling edge  $\beta^*$  reduces. The sharp changes occur when the critical fractile occurs at a peak of a mode.

Fig. 10 shows the safety stock requirements when  $\beta^*$  is used for different cost ratios. The multi-modal nature of the  $\mu = 100$  case results in rapid increases in the safety stock requirement at predictable points on the availability scales. These are also related to the multi-modal pdf of the inventory levels as the safety stock requirements are a function of the cdf of inventory. Furthermore, between 40-60% availability, the two demand settings require very similar amounts of safety stock. As it is not possible to visually distinguish



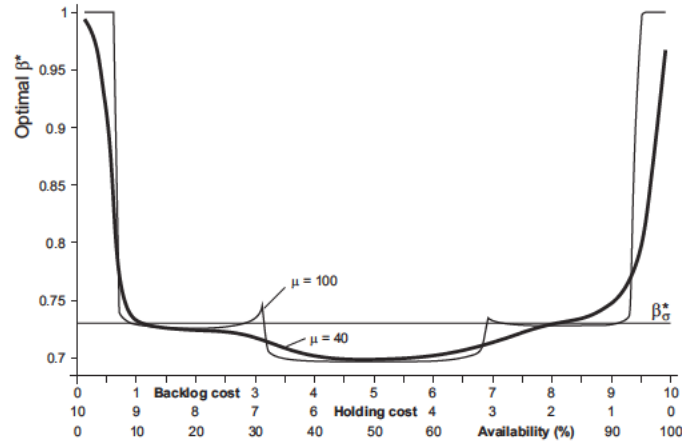


Figure 9: The inventory cost optimal  $\beta^*$  in the POUT policy

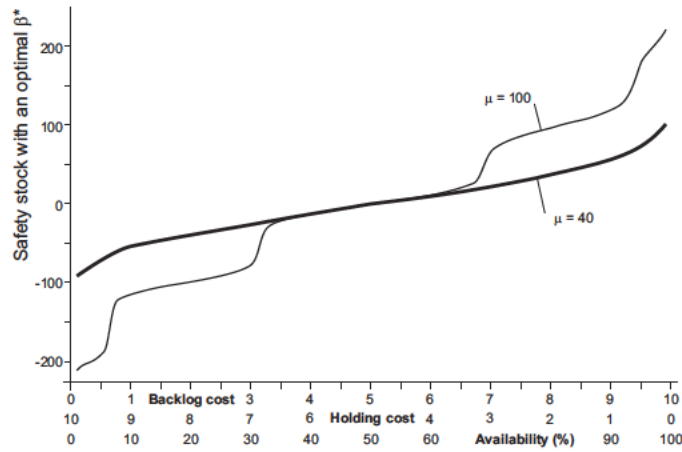


Figure 10: Safety stock with the inventory cost optimal POUT policy

between the optimal POUT safety stock,  $T_{POUT}$ , and the optimal safety stock for the OUT policy,  $T_{OUT}$ , we have plotted the difference,  $T_{POUT} - T_{OUT}$ , in Fig. 11. Here we can see that, although the POUT policy is economically superior, minimum safety stock does not always coincide with least costs, even when the service level is high. This is an insight that is contrary to the constant lead time case where the least cost solution always has the smallest safety stocks.

An intuitive explanation of this is as follows. Consider a constant lead time where the cost ratio is such that the optimal availability is less than 50%. In order to minimise the safety stock requirements then we should set  $\beta = 0$  to make  $\sigma_I = \infty$  and the safety stock required is then  $T^* = -\infty$ . However, in this case the inventory costs  $J = \infty$ . Under a stochastic lead time, due to the multi-modality of the inventory distribution, these *trailing edge* effects can happen in the whole range of availability targets (over  $0 < \alpha < 1$ , not just for  $\alpha < 0.5$ ).

Fig. 12 highlights the bullwhip ratio achieved when  $\beta^*$  is used in the POUT policy. We

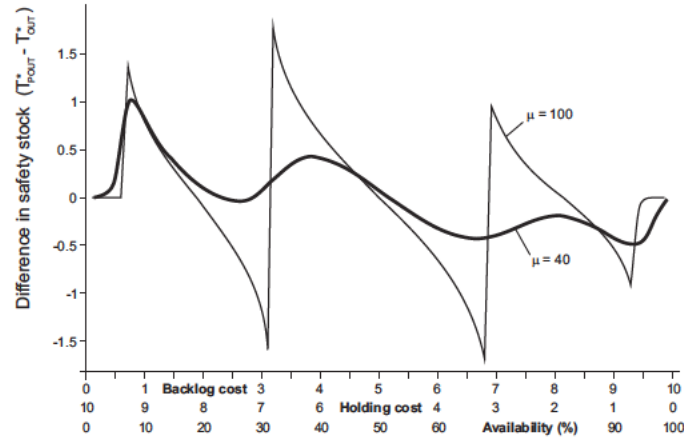


Figure 11: Safety stock difference between the OUT policy and the inventory cost optimal POUT policy ( $T_{POUT}^* - T_{OUT}^*$ )

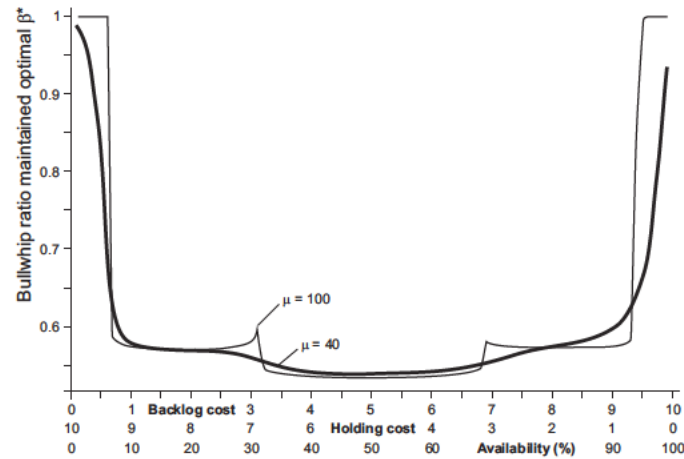


Figure 12: Bullwhip with the inventory cost optimal  $\beta^*$

can see that a 40% reduction in order variance is possible between 8% to 92% availability. This is interesting as the objective function consists only of inventory related costs. If costs are associated with order variability are also present, these will also be reduced. This is an important result as order variance costs are somewhat harder to quantify, even though the bullwhip effect is widely recognized as having a negative effect on supply chain performance (Lee et al. 2000).

## 6 Revisiting the empirical lead time distributions

Table 3 revisits the 12 empirical lead time distributions in Fig. 1 (case 1 to 12) and the case in Fig. 2 (case 13). A demand with a mean of  $\mu = 40$  and a standard deviation of  $\sigma = 10$  was assumed. The availability target is held constant at 90% with a unit backlog cost per period of  $b = 9$  and a unit holding cost of  $h = 1$ . We take note of the minimised

Table 3: Policy comparison for the real-world lead time distributions between China and the USA.

Case	OUT Policy		POUT Policy			% benefit in $J$	$\sigma_O^2/\sigma^2$
	$T^*$	$J$	$T^*$	$\beta^*$	$J$		
1	44.8623	62.8224	44.8442	0.9156	62.7996	0.0363	0.8443
2	42.8846	60.3743	42.8547	0.9147	60.348	0.0436	0.8428
3	41.3303	60.0697	41.3313	0.845	59.968	0.1693	0.7316
4	39.1185	56.1142	39.1223	0.8962	56.0718	0.0756	0.8119
5	49.5645	69.04	49.4522	0.8524	68.9504	0.1298	0.7427
6	43.8723	60.8159	43.8569	0.9027	60.781	0.0574	0.8226
7	48.0291	68.6473	47.9623	0.8223	68.5191	0.1868	0.6982
8	46.0299	65.3138	45.9663	0.8197	65.1635	0.2301	0.6944
9	42.1678	59.6323	42.1636	0.9605	59.6278	0.0075	0.924
10	30.4934	43.1683	30.4933	0.9946	43.1682	0.0002	0.9892
11	56.3319	79.0999	56.1169	0.7722	78.8626	0.3	0.6289
12	39.1313	56.4232	39.1244	0.9356	56.41	0.0234	0.8789
13	47.793	66.7626	47.7706	0.8837	66.7453	0.0259	0.7916

costs ( $J$ ) from optimising the safety stock ( $T^*$ ) in the OUT policy. Recall, the variance of the orders in the OUT policy is always equal to the demand variance ( $\sigma_O = \sigma = 10$ ). For the POUT policy we note the minimised costs ( $J$ ) from optimising both the safety stock ( $T^*$ ) and the feedback controller ( $\beta^*$ ). We also calculate the variance of the orders and the percentage reduction in the inventory costs from using the POUT policy.

The POUT policy is always more economical than the OUT policy, and the optimal  $\beta^* < 1$ . Usually, the safety stock requirements of the POUT policy is less than the OUT policy, but not for case 3 and 4. In case 10, the probability of order crossovers is very small and the  $\beta$  is very close to unity. However in these practical examples, the optimal  $\beta$  is generally around 0.8 to 0.9. Whilst the inventory cost benefit is very small, the reduction in the order variance is more significant and comes without cost under stochastic lead times. The order was reduced by 37% in case 11, and just 1% in case 10, but the average reduction in order variance is 20%.

## 7 Conclusions

As globalization accelerates, lengthening supply chains bring complications and pitfalls associated with increasing uncertainty in both demand and lead time. We determined how these characteristics actually impact the safety stock required to minimise inventory holding and backlog costs. We began by presenting global logistics data that violated the traditional normality assumption about lead times. We also provided a real-world example of order crossover.

We introduced a new approach for calculating the distribution of open orders. This allowed us to formulate and test a method that resulted in an exact solution for the safety stock calculation. Since the enumeration of open orders is combinatoric in nature, it explicitly allowed for order crossover. We showed that using the POUT strategy always results in lower inventory costs when stochastic lead times with order crossover are present. Our model settings were motivated by real-world settings and consisted of a discretely distributed lead time and a continuously distributed demand process.

Our novel contribution is a new method to obtain the distribution of the inventory levels in the presence of correlation between the WIP and orders, via the so-called scaled shortfall distribution. This builds upon another unique contribution – the M-matrix and the associated method to determine the probability of the pipeline being in each of its possible  $2^{k+}$  states. Furthermore, we considered the impact of correlation in orders and the covariance between orders and WIP in a stochastic lead time setting with order crossovers. Our methodology can be used to investigate the impact of auto-correlated demand, non-MMSE forecasting methods, more sophisticated replenishment policies, and information sharing strategies.

In the constant lead time case, or the non-crossover stochastic lead time case,  $\beta = 1$  will minimize the variance (or equivalently the standard deviation) of the inventory levels and result in the minimum inventory costs when the safety stock is set to the critical fractile (Brown 1963). However, in the stochastic lead time with order crossover case, minimizing the variance of the inventory levels, by tuning  $\beta$ , will not always result in minimal costs. While the optimal  $\beta^*$ , may be near unity, it is never unity and changes significantly with the availability target, see Fig. 9.

The stochastic lead time case with order crossover results in a surprising paradox. Minimizing inventory costs does not always lead to minimum safety stocks. However, the relationship between holding and backlog costs and the availability achieved at the most economical solution does still hold. This leads to an important insight: Costs should be used to design the system because focusing on minimizing inventory variance, or safety stocks, can lead to an incorrectly specified system. We conclude that care must be taken when determining safety stock requirements under a stochastic lead time with order crossovers. One cannot simply use the first two moments of the inventory distribution;

one needs to use whole pdf because the distribution of the inventory levels is multi-modal.

We have demonstrated that the OUT policy is not the optimal policy when order crossover exists, as the linear POUT economically outperforms it. We have not proven the optimality of the POUT policy itself. Indeed it is known that the optimal policy is non-linear, see Srinivasan et al. (2011). However, the POUT policy has a long history and has been successfully implemented in practice. See Potter & Disney (2010) for details of an implementation at the UK grocery retailer, Tesco and Disney et al. (2013) for an implementation in a global printer manufacturer.

We note that our model / approach takes no account of state dependent or autocorrelated lead times. Seasonal congestion in ports is a well observed phenomena and this would lead one to suspect that lead times are positively auto-correlated. Furthermore, in order to avoid an imminent stock-out, companies may air-freight cargo leading to a lead time that is a function of the state of the supply chain. We have also not considered the consequences of non-normal demands. If demand is not normally distributed, then the complete pdf of each sub-process has to be obtained. Presumably this could be achieved with convolution but this it is beyond the scope of the paper. These issues remain interesting areas for future work.

## References

- Axsäter, S. (2000), *Inventory Control*, 1st edn, Kluwer Academic Publishers, Boston, MA.
- Bagchi, U., Hayya, J. C. & Chu, C.-H. (1986), 'The effect of lead-time variability: The case of independent demand', *Journal of Operations Management* **6**(2), 159–177.
- Balakrishnan, A., Geunes, J. & Pangburn, M. (2004), 'Coordinating supply chains by controlling upstream variability propagation', *Manufacturing and Service Operations Management* **6**(2), 163–183.
- Bischak, D. P., Robb, D. J., Silver, E. A. & Blackburn, J. D. (2014), 'Analysis and management of periodic review order-up-to level inventory systems with order crossover', *Production and Operations Management* **23**(5), 762–772.
- Boute, R. N., Disney, S. M., Lambrecht, M. R. & Houdt, B. V. (2014), 'Coordinating lead times and safety stocks under autocorrelated demand', *European Journal of Operational Research* **232**(1), 52–63.
- Boute, R. & Van Mieghem, J. A. (2014), 'Global dual sourcing and order smoothing: The impact of capacity and lead times', *Management Science* (In Press, <http://dx.doi.org/10.1287/mnsc.2014.1992>).

- Bradley, J. R. & Robinson, L. W. (2005), 'Improved base-stock approximations for independent stochastic lead times with order crossover', *Manufacturing and Service Operations Management* **7**(4), 319–329.
- Brown, R. G. (1963), *Smoothing Forecasting and Prediction of Discrete Time Series*, 1st edn, Prentice-Hall, Michigan.
- Cachon, G. & Terwiesch, C. (2009), *Matching Supply with Demand: An Introduction to Operations Management*, 1st edn, McGraw-Hill/Irwin, New York, NY.
- Caplice, C. & Kalkanici, B. (2012), Ocean transportation reliability: Myth vs. reality, in 'Presentation to Annual Meeting of Council of Supply Chain Management Professional', Atlanta, GA.
- Chatfield, D. C., Kim, J. G., Harrison, T. P. & Hayya\*, J. C. (2004), 'The bullwhip effect: Impact of stochastic lead time, information quality, and information sharing: A simulation study', *Production and Operations Management* **13**(4), 340–353.
- Chopra, S., Reinhardt, G. & Dada, M. (2004), 'The effect of lead time uncertainty on safety stocks', *Decision Sciences* **35**(1), 1–24.
- Dejonckheere, J., Disney, S. M., Lambrecht, M. R. & Towill, D. R. (2003), 'Measuring and avoiding the bullwhip effect: A control theoretic approach', *European Journal of Operational Research* **147**(3), 567–590.
- Disney, S. M., Gaalman, G., Hedenstierna, C. P. T. & Hosoda, T. (2015), 'Fill rate in a periodic review order-up-to policy under auto-correlated normally distributed, possibly negative, demand', *International Journal of Production Economics: Under review*.
- Disney, S. M., Hoshiko, L., Polley, L. & Weigel, C. (2013), Removing bullwhip from Lexmark's toner operations, in 'Production and Operations Management Society Annual Conference', Denver, CO.
- Disney, S. M. & Lambrecht, M. R. (2008), 'On replenishment rules, forecasting and the Bullwhip Effect in supply chains', *Foundations and Trends in Technology, Information and Operations Management* **2**(1), 1–80.
- Disney, S. M. & Towill, D. R. (2003), 'On the Bullwhip and Inventory variance produced by an ordering policy', *The International Journal of Management Science* **31**(3), 157–167.
- Disney, S. M., Towill, D. R. & Van de Velde, W. (2004), 'Variance amplification and the golden ratio in production and inventory control', *International Journal of Production Economics* **90**(3), 295–309.

- Eppen, G. D. & Martin, R. K. (1988), 'Determining safety stock in the presence of stochastic lead time and demand', *Management Science* **34**(11), 1380–1390.
- Feller, W. (1966), *An Introduction to Probability Theory and its Applications*, Vol. I, 2nd edn, John Wiley & Sons, Inc., New York, NY.
- Fransoo, J. C. & Lee, C.-Y. (2012), 'The critical role of ocean container transport in global supply chain performance', *Production and Operations Management* **22**(2), 253–268.
- Hayya, J. C., Bagchi, U., Kim, J. G. & Sun, D. (2008), 'On static stochastic order crossover', *International Journal of Production Economics* **114**(1), 404–413.
- Hayya, J. C., Harrison, T. P. & He, X. J. (2011), 'The impact of stochastic lead time reduction on inventory cost under order crossover', *European Journal of Operational Research* **211**(2), 274–281.
- Jula, P. & Leachman, R. C. (2011), 'Long and short run supply chain optimization models for the allocation and congestion management of containerized imports from Asia to the United States', *Transportation Research, Part E* **47**(5), 593–608.
- Kalkanci, B. & Caplice, C. (2012), Empirical investigation of global transit reliability, Technical report, MIT, Cambridge, MA.
- Kaplan, R. S. (1970), 'A dynamic inventory model with stochastic lead times', *Management Science* **16**(7), 491–507.
- Kim, J. G., Chatfield, D., Harrison, T. P. & Hayya, J. C. (2006), 'Quantifying the bullwhip effect in a supply chain with stochastic lead time', *European Journal of Operational Research* **173**(2), 617–636.
- Leachman, R. C. & Jula, P. (2012), 'Estimating flow times for containerized imports to the United States through the Western Rail Network', *Transportation Research, Part E* **48**(1), 296–309.
- Mentzer, J. T. & Krishnan, R. (1985), 'The effect of the assumption of normality on inventory control/customer service', *Journal of Business Logistics* **6**(1), 101–120.
- Muharremoglu, A. & Yang, N. (2010), 'Inventory management with an exogenous supply process', *Operations Research* **58**(1), 111–129.
- Nise, N. S. (2011), *Control Systems Engineering*, 6th edn, John Wiley and Sons, Inc, USA.
- O'Marsh, K. (2014), 'Us port congestion is a warning shot for supply chain strategists', <http://www.scmworld.com/columns/beyond-supply-chain/us-port-congestion-is-a-warning-shot-for-supply-chain-strategists/>. Accessed Feb 1, 2015.

- Potter, A. & Disney, S. M. (2010), Removing bullwhip from the Tesco supply chain, *in* 'Production and Operations Management Society Annual Conference', Vancouver, Canada.
- Robinson, L. W., Bradley, J. R. & Thomas, L. J. (2001), 'Consequences of order crossover under order-up-to inventory policies', *Manufacturing and Service Operations Management* **3**(3), 175–188.
- Robinson, L. W., Bradley, J. R. & Thomas, L. J. (2008), 'Further improvements on base-stock approximations for independent stochastic lead times with order crossover', *Manufacturing and Service Operations Management* **10**(2), 325–327.
- Saldanha, J. P., Tyworth, J. E., Swan, P. F. & Russell, D. M. (2009), 'Cutting logistics costs with ocean carrier selection', *Journal of Business Logistics* **30**(2), 175–191.
- Schneeweiss, C. A. (1974), 'Optimal production smoothing and safety inventory', *Management Science* **20**(7), 1122–1130.
- Silver, E. A., Pyke, D. F. & Peterson, R. (1998), *Inventory Management and Production Planning and Scheduling*, 3rd edn, John Wiley, New York, NY.
- Simon, H. A. (1952), 'On the application of servomechanism theory to the study of production control', *Econometrica* **20**(2), 247–268.
- So, K. & Zheng, X. (2003), 'Impact of supplier's lead time and forecast demand updating on retailer's order quantity variability in a two-level supply chain', *International Journal of Production Economics* **86**, 169–179.
- Sobel, M. (2004), 'Fill rates of single stage and multistage supply systems', *Manufacturing and Service Operations Management* **6**(1), 41–52.
- Srinivasan, M., Novack, R. & Thomas, D. (2011), 'Optimal and approximate policies for inventory systems with order crossover', *Journal of Business Logistics* **32**(2), 180–193.
- Stalk, G. (2006), 'Surviving the China riptide', *Supply Chain Management Review* **10**(4), 18–26.
- Strijbosch, L. W. G., Heuts, R. M. J. & Luijten, M. L. J. (2002), 'Cyclical packaging planning at a pharmaceutical company', *International Journal of Operations and Production Management* **22**(5), 549–564.
- Tadikamalla, P. R. (1984), 'A comparison of several approximations to the lead time distribution', *Omega: International Journal of Management Science* **12**(6), 575–581.



Tyworth, J. E. & O'Neill, L. (1997), 'Robustness of the normal approximation of lead-time demand in a distribution setting', *Naval Research Logistics* **44**(2), 165–186.

Warburton, R. D. H. & Stratton, R. (2002), 'Questioning the relentless shift to offshore manufacturing', *Supply Chain Management* **7**(2), 101–108.

Zalkind, D. (1978), 'Order-level inventory systems with independent stochastic leadtimes', *Management Science* **24**(13), 1384–1392.