Exact and Approximate Solutions for Fulfilment Metrics in Open Pipeline Order Fulfilment Systems

Bart L. MacCarthy[1][*], Tamer Ovutmen[1], S. Armagan Tarim[2] and Philip G. Brabazon[3]

1- **Nottingham University Business School, Nottingham, UK, NG8 1BB**

2- **Department of Management, Hacettepe University, Ankara, Turkey**

3- **Centre for Enterprise Research and Innovation, Portsmouth Business School, Portsmouth, UK, PO1 3DE**

Abstract

Open pipeline order fulfilment systems have emerged in sectors such as the automotive industry that offer a large number of product variants to the marketplace. In an open pipeline a customer can be fulfilled from anywhere in the system - by a stock product, by a product in the distribution system, by allocating a product in production or a product in the plan, or by supplying a product specially built to order (BTO). Here new results are presented to estimate the performance metrics for open pipeline systems. An exact expression is developed based on a Markov analysis to calculate the BTO proportion for any combination of the key system parameters - pipeline length, the variety level, and the initial stock level. Approximation schemes are developed to estimate the pipeline and stock fulfilment proportions, the mean lead time and the mean stock level. The approximation schemes provide accurate estimates for a large range of systems. Specific issues that affect the estimation of performance metrics in low variety open pipeline systems are discussed. The results presented provide generic insights for open pipeline systems design and management and provide a platform for further work in extending the applicability of open pipeline concepts.

Keywords: order fulfilment, automotive industry, Markov process, open pipeline.

*Corresponding author: bart.maccarthy@nottingham.ac.uk

1

1. Introduction

In many sectors high product variety presents significant operational challenges to producers seeking to satisfy as many customers as possible in a timely and cost effective way. Put simply, when the number of variants in a product range is low and demand reasonably predictable, the risks in holding sufficient stock to meet anticipated demand is low. However, in high variety contexts, holding stocks of an extensive range of product variants is much more costly, more risky and may be infeasible practically. In addition, customer populations for many products are heterogeneous in their requirements and desires. The desire of some customers to differentiate or customise products may further increase the operational challenges in fulfilling customers (McCutcheon et al. , 1994).

In large scale order fulfilment systems there is a need to develop approaches that can respond flexibly to meet the needs of a diverse customer population. The automotive sector is a case in point. Producers typically offer very high variety and the customer base is strongly heterogeneous (Holweg and Pil, 2004; Meyr, 2004; Fredrikkson and Gadde, 2005). These issues are exemplified in a recent major vehicle launch in the US market (WSJ, 2010).

A conventional fulfilment system might operate in a pure Make-to-Stock mode where customers are fulfilled only from stock, in a pure Make-to-Order mode where production is not initiated until an order is received, or might use a combination of both approaches with some customers being fulfilled from stock and some fulfilled by a product that is made- to-order. Here we consider an open pipeline fulfilment system that is emerging in some sectors. In such a system a customer may be satisfied from anywhere in the system - by a stock product held at a final stocking location, by a product in the distribution system, by allocating a product that is in production or in the plan, or by a product specifically built to order (BTO) for the customer. Figure 1 illustrates such a system conceptually with sequential phases of planning, production, distribution and finished stock (that may be in a single location or dispersed across a sales network).

Open pipeline order fulfilment systems seek to exploit the potential within the whole system to match each customer's requirements with an appropriate product. The approach seeks a balance between 'variety push' from the producer and 'variety pull' from the customer. This type of fulfilment strategy has been implemented to different degrees by many automotive producers (Agrawal et al., 2001; Holweg and Pil, 2004). It is particularly relevant for large scale systems with extensive variety and long planning pipelines, where the producer must build to forecast to some degree. Elements of the system are also used in other sectors and it has the potential for much wider use.

The open pipeline approach has had only a limited amount of analysis to date. Here we consider the estimation of the fulfilment performance metrics for open pipeline systems of different scale. Important metrics of interest are the proportions of customers fulfilled by different fulfilment mechanisms (from stock, from the pipeline and by build-to-order) and the average lead time and average stock levels in the system. In this paper, we develop models that enable these metrics to be estimated and that provide greater understanding and insights on the behaviour of open pipeline systems.

Antecedent research on related systems is first discussed and the origins of the work presented here are discussed. New analytical results based on a Markov analysis are presented to calculate exactly the build-to-order proportion in an open pipeline system of arbitrary scale. Approximation schemes are introduced to estimate the pipeline and stock fulfilment proportions, as well as the mean lead time and mean stock levels in the system. Specific issues that affect the estimation of fulfilment metrics in low variety open pipeline systems are illustrated. The new analytical results and approximation schemes presented advance the existing literature significantly. Opportunities for further development of models for open pipeline systems more generally are noted.
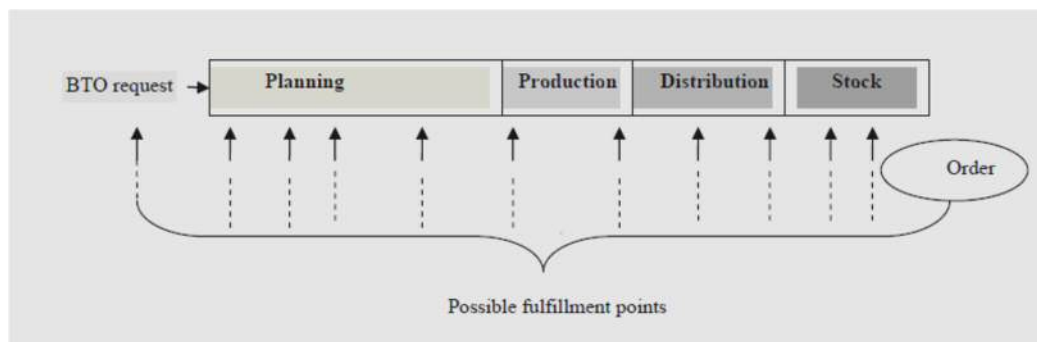

Figure 1. Open pipeline order fulfilment

2. Literature Review

McCutcheon et al. (1994) noted that firms in many industries were being squeezed by customer demands for both greater product variety and reduced delivery lead times. Standardization in mass production enables quick delivery of products. However, high product variety requires flexible and innovative organizations. Firms must therefore choose appropriate tactics to deal with the 'customization-responsiveness squeeze'. The operational challenges raised by offering high levels of product variety are well known (Ramdas, 2003; Er and MacCarthy, 2006).

Mass Customization has often been identified as an effective strategy to deliver relevant customer-led variety to markets. It refers to the production of customized products on a mass scale without

compromising on customer delivery times or quality and without incurring significantly increased producer costs (Tu et al., 2001; MacCarthy et al., 2003). However, the limitations of a Mass Customization strategy have been noted (Zipkin, 2000; Alford et al. 2000; Blecker and Abdelkafi, 2006; Squire et al. 2006). It is clear from the literature that implementing effective systems to deliver high variety and customization is difficult, particularly the design of effective flexible order fulfilment systems.

Different types of postponement strategies have been advocated to handle demands for product variety and customization (Zinn and Bowersox, 1988; van Hoek 2001; Yang and Burns, 2004; Graman and Magazine 2002; Jewkes and Alfa 2008; and Graman 2009). Although postponement provides a valuable strategy in many contexts, the systems studied do not provide strong analogues for the type of open pipeline order fulfilment system investigated in this paper, which has a 'floating' decoupling point. Relevant previous work on related flexible fulfilment systems is in fact quite sparse. Some work has been reported on multiple fixed point fulfilment strategies by Swaminathan and Tayur (1998) for computer assembly, by Denton et al. (2003) in the steel industry, and by Meredith and Akinc (2007) for long lead time large product assemblies but again the analogies are not strong.

In general, previous work (i) does not take account of the full potential extent of the fulfilment system (as in the open pipeline case) and (ii) typically considers only a small number of product variants. Order fulfilment in the volume automotive sector highlights the importance of addressing both of these issues. The very large number of product variants on offer in the volume automotive context is well known (Holweg and Pil, 2004; Meyr, 2004; Fredriksson and Gadde, 2005). For instance, Fredriksson and Gadde (2005) discuss the introduction of a number of new vehicle models over a four year period at Volvo based on the same product platform. The five models were offered in 14 colours, 9 engine types, 5 transmission alternatives, 22 types of interior design, and 9 wheel variants, resulting in potentially more than one million variants when all other optional choices are considered.

Holweg (2000) described multiple fulfilment mechanisms in the automotive sector by which a customer can receive a vehicle: from the dealer's stock; by transhipment from another dealer's stock; by a vehicle taken from a central stock holding centre; by a vehicle being submitted to the order bank as a build-to-order product; or by a vehicle that is in, or scheduled for, production being allocated to the customer, which may involve its specifications being amended. Agrawal et al. (2001) described this kind of approach in the automotive sector as a 'Virtual-Build-to-Order' system. They noted that fulfilment can be 'either via internet or in dealer's showrooms, to the vast, albeit far-flung, array of vehicles already in existence, including vehicles on dealer's lots, in transit, on

assembly line, and scheduled for production', with the expectation that 'customers are likely to find a vehicle with the colour and options they most want'. This kind of open pipeline system is the focus of the work here.

Brabazon and MacCarthy (2006) studied some of the fundamental behaviours of open pipeline systems using simulation modelling (they termed them Virtual-Build-to-Order or VBTO systems, after Agrawal's description). They identified relationships between the proportions of customers fulfilled by each fulfilment mechanism and the ratio of product variety level to pipeline length. Their work indicated many potential areas for detailed investigations of this type of order fulfilment system, both theoretically and practically. In particular it raised fundamental questions on how systems with open pipelines behave and how they should be designed and managed for effective performance. However, the model simulated did not include an initial stock position in the system.

Brabazon and MacCarthy (2010) in a theoretical study examined the feasibility of modelling a VBTO system as a Markov process. They presented 1st and 2nd order Markov approximations to estimate the fulfilment proportions for each fulfilment mechanism for a range of product variety levels in a very small VBTO system of pipeline length 3. However, the models could not be scaled up for open pipeline systems of arbitrary scale. In addition, the models assumed an initial stock position of zero in the system. They concluded that the approach was impractical for modelling larger systems, although the study showed valuable insights into the mechanics, operation and characteristics of VBTO systems. The work presented here builds on the previous studies to develop new approaches to calculate and estimate accurately all the fulfilment metrics for a generic open pipeline system of any scale.

The approaches developed in this paper differ significantly from the approaches presented in Brabazon and MacCarthy (2010), both in their construction and in what they achieve. Firstly, a new 1ST order Markov model is developed that provides an exact calculation of BTO fulfilment in an open pipeline system of any pipeline length, any level of variety, and any initial stock position – it is fully scalable. Secondly, the new approximation schemes developed for pipeline and stock fulfilment also apply to a system of any scale. In contrast, the previous models approximated all three fulfilment metrics for a very small system with 1st and 2nd order Markov models but were not scalable to larger systems and assumed a zero initial stock in the system. In addition, the new approaches have been extended to estimate mean lead time and mean stock level in an open pipeline system of any scale.

3. Exact solutions using Markov modelling

A basic open pipeline system is modelled here as a system with three segments: a stock of unsold products, a production plan, which is the sequence of products to be manufactured and the products in production (the pipeline), and a Build-to-Order (BTO) cycle. The structure is shown in Figure 2. Note that actual production is assumed to occur at the end of the pipeline and once a product is produced it enters stock unless it has entered the pipeline as a BTO product allocated to a customer or has been allocated previously to a customer from the pipeline. As with any modelling work, assumptions are made here to enable analysis of the system. The structure, logic and assumptions are the same as in Brabazon and MacCarthy (2010), the implications of which are discussed below.

Much of the work in this area by the authors has been carried out in the context of large scale automotive order fulfilment systems and the assumptions will be discussed mainly in that context. The principal purpose of these models is to understand the ability of the fulfilment system at a macro level to fulfil demand at different levels of variety over an extended period of time such as a selling season. The models are intended to shed light on macro system performance, in particular the overall likelihood of fulfilment by each mode of fulfilment.

The models developed assume that all the stock is centrally pooled in a single stocking holding. This is realistic in the context of macro level open pipeline system performance as the objective in such a system is to expose as much as possible of the stock in the system to incoming demand, irrespective of its location in vehicle holding centres, in the distribution channels and across dealership networks.

It is assumed that each arriving customer knows the specification of the product they are looking for and that a search is made for an exact match for that specification. When a customer arrives the highest priority is given to fulfilment from stock. If the requested product is not available in stock then the pipeline is searched. If a suitable unallocated product is not found in the pipeline, then a BTO product with the required specification is fed into the pipeline. This fulfilment logic is realistic as, in an open pipeline system, the producer will actively discourage (and may penalise) BTO requests when an appropriate product is available somewhere in stock or in the pipeline and similarly for pipeline sales when a product is available in stock. When a BTO variant is required in a demand cycle, it is assumed to take precedence and is fed into the system in the next simulation cycle, which ensures all BTO sales are handled without holding a BTO queue. This assumption is realistic as in a real system actual sold items typically get the highest priority. When a product is allocated to a

6

customer, either as a pipeline sale or a BTO, it is made unavailable to subsequent customers, which again is a realistic assumption.
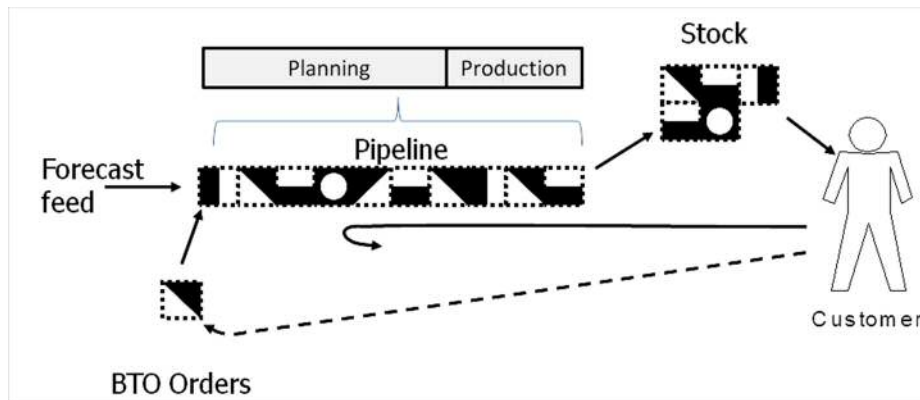

Figure 2 Open pipeline fulfilment system, adapted from Brabazon & MacCarthy (2006)

Product variety is captured simply in the basic model as a specific number of distinct product variants. In the models presented here the sequence of products fed into the pipeline is assumed to be uniformly distributed, random and uncorrelated. The production rate and customer arrival rate are assumed constant and equal. The assumption is that the system is in balance in terms of production and demand. Products enter the pipeline and customers arrive at a fixed constant rate. Once products enter the pipeline they cannot be re-sequenced. As with the pipeline feed, it is assumed the sequence of variants requested by customers is uniform and randomly distributed but is independent of the feed distribution

The assumption of a constant production rate in a balanced supply and demand system is appropriate at the macro level considered here The producer's goal is to match the level of production with demand (e.g. over a selling season, where all vehicles produced are sold). In the automotive sector, the main volume producers can claim that every vehicle is sold eventually. Thus, from a macro level perspective a balanced system is a reasonable modelling assumption. Clearly, to investigate micro level queuing and waiting in such a system, the arrival pattern and pipeline feed patterns would need to be modelled in detail explicitly but this is not the objective of the work presented here.

Essentially the assumption here is that the producer knows the overall shape of the demand distribution and achieves it in terms of feed over the long term. In practice a volume automotive producer will have good knowledge of the overall shape of the demand function but prediction of specific demands at any point will be uncertain, as is the case with the models here. Indeed volume producers can strongly influence the demand distribution through promotion and pricing. Thus,

having the same demand and feed distribution is an appropriate modelling assumption for the perspective considered here.

The assumption of uniform random feed and demand does facilitate the development of the models presented here. Assuming uniform random feed initially provides a baseline from which to judge systems with other characteristics. We discuss the extension and use of the models presented here to the skewed demand case in section 6. Further potential enhancements of the modelling approach are discussed in sections 6 and 7.

Given these assumptions, three system design parameters are identified with positive integer values - a fixed pipeline length (P), a fixed level of variety (V), and a fixed initial stock level (S) in the system. For a general open pipeline fulfilment system the primary performance metrics of interest are the proportions of customers fulfilled by each of the three fulfilment mechanisms – from stock, from the pipeline and by build-to-order (BTO). The average level of stock held in the system and the average lead time experienced by customers are also critical performance metrics. Stock level is an important metric for the producer. Although the initial stock position is fixed, the actual stock level varies dynamically over time as unallocated products from the pipeline go into stock (i.e. those not allocated through pipeline fulfilment as they progress though the pipeline) and fulfilment from stock takes place.

Lead time is an important metric for both the customer and the producer as excessive lead times may result in lost sales. In the model here, lead time is defined as a normalised metric between 0 and 1. The maximum lead time is experienced when a BTO fulfilment occurs and the minimum lead time occurs when a stock fulfilment occurs (logistics times in fulfilling from stock are assumed negligible or constant). Thus, the following conventions are used in measuring lead time (LT) in the model:

i.   LT equals 0 if a customer is fulfilled from stock
ii. LT equals 1, if a customer is fulfilled by BTO
iii. LT equals $(P-p)/P$, if a customer is fulfilled from the pth position in the pipeline.

3.1 Calculating the BTO proportion for an open pipeline system

An important consequence of the assumptions made in the model is that the total number of unallocated products in the system (i.e. in stock and in the pipeline) is always fixed at S+P, irrespective of the mode of fulfilment. A BTO occurs if and only if a match for the demanded

product cannot be found in the S+P unallocated products in the system. As V is known, there is a finite number of combinations for the number of available products for each variant in the system. This logic forms the basis for the exact calculation model for the BTO fulfilment proportion. The approach is first described here with an illustrative example using a small system.

Let $s = (I_1, I_2, I_3)$ be the inventory level vector for a system with $S + P = 3$ and $V = 3$, where I_i is the total inventory of variant i, i = 1, 2,3. An inventory level vector such as (1,2,0) represents a system with one product of variant 1, two products of variant 2, and zero products of variant 3 available in the system (stock and pipeline). As $I_1 + I_2 + I_3 = 3$, there are 10 possible combinations for s : (0,0,3), (0,1,2), (0,2,1), (0,3,0), (1,0,2), (1,1,1), (1,2,0), (2,0,1), (2,1,0), and (3,0,0).

In addition to the fixed unallocated inventory level property, as the system has one demand and one feed per cycle, all the transitions that can occur between different inventory level vectors can be defined.

- If a BTO fulfilment occurs, no product in the unallocated inventory leaves the system, an allocated product is fed into the system and so the inventory vector of unallocated products remains unchanged.
- If a stock or pipeline fulfilment occurs, one of the unallocated products in the system will leave the system and a randomly generated unallocated product will be added to the system. The unallocated inventory in the system remains the same if the variant fed into the system is the same as the demanded product. If a variant different to the demanded product is fed into the system, then the inventory level of the demanded variant decreases by one, and the inventory level of the fed in variant increases by one.

Based on this information, all the transitions between different inventory vectors can be defined. The possible transitions from an inventory vector (1,0,2) are shown in Figure 3.
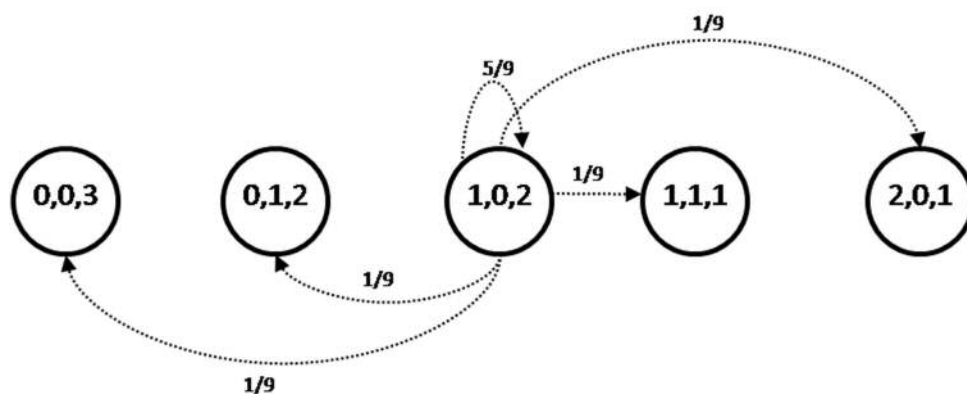

Figure 3 Example for possible transitions from an inventory vector

Given the assumptions about the demand distribution, the demand for each variant is equally likely irrespective of the current inventory state of the system. Below, the possible demand and feed combinations from state (1,0,2) are discussed:

1. If the demand is for product 1, either a pipeline or stock fulfilment occurs and the next state of the system is:
   a. (1,0,2), if product 1 is fed into the system,
   b. (0,1,2) if product 2 is fed into the system,
   c. (0,0,3) if product 3 is fed into the system.
2. If the demand is for product 2, then a BTO fulfilment occurs. An allocated product will enter the system, which will not change the inventory levels of unallocated products in the system.
3. If the demand is for product 3, then either a pipeline or stock fulfilment occurs, and the next state of the system is:
   a. (2,0,1) if product 1 is fed into the system,
   b. (1,1,1) if product 2 is fed into the system,
   c. (1,0,2) if product 3 is fed in to the system.

The probabilities of moving from state (1,0,2) to each of the other possible states can therefore be calculated, as shown below:

i. (0,0,3) if product 1 is demanded, and product 3 is fed, the probability is: $\frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$

ii. (0,1,2) if product 1 is demanded, and product 2 is fed, the probability is: $\frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$

iii. (1,0,2) if product 1 is demanded, and product 1 is fed; or if product 2 is demanded (it is a BTO fulfilment); or product 3 is demanded and fed. The probability is: $\left(\frac{1}{3} \times \frac{1}{3}\right) + \frac{1}{3} + \left(\frac{1}{3} \times \frac{1}{3}\right) = \frac{5}{9}$

iv. (1,1,1) if product 3 is demanded and product 2 is fed, the probability is: $\frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$

v. (2,0,1) if product 3 is demanded and product 1 is fed, the probability is: $\frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$

These probabilities only depend on the current state of the system and are not dependent on previous states of the system. Because the system has a finite number of mutually exclusive states, and the probabilities of transition from one state to another depends only on the current state of the system (i.e. the system is memoryless), then the process can be modelled as a Markov Chain.
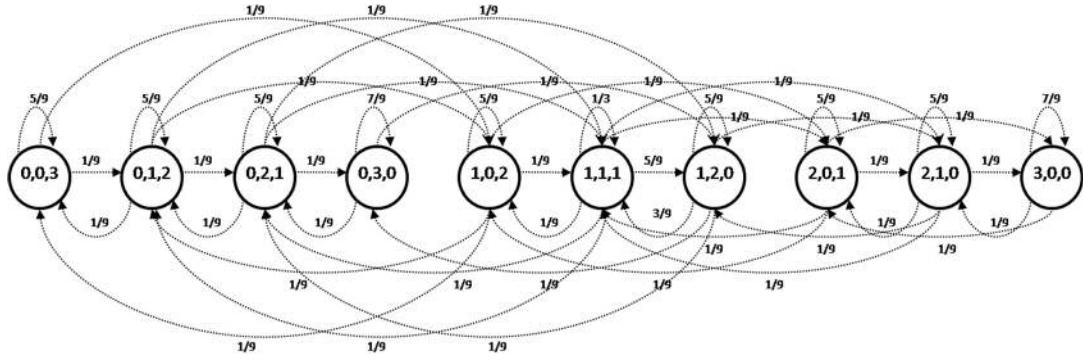
Figure 4 All states and transition probabilities for a system with S+P=3, and V=3

All the possible states and the transition probabilities of the system with S+P = 3 and V = 3 are shown in Figure 4. The probabilities are also shown in matrix form in Table 1 in terms of the variety level V.

Table 1 The states and the transition probabilities for a system with S+P=3 and V=3

	(0,0,3)	(0,1,2)	(0,2,1)	(0,3,0)	(1,0,2)	(1,1,1)	(1,2,0)	(2,0,1)	(2,1,0)	(3,0,0)
(0,0,3)	$\dfrac{2V+1}{V^2}$	$\dfrac{1}{V^2}$	0	0	$\dfrac{1}{V^2}$	0	0	0	0	0
(0,1,2)	$\dfrac{1}{V^2}$	$\dfrac{V+2}{V^2}$	$\dfrac{1}{V^2}$	0	$\dfrac{1}{V^2}$	$\dfrac{1}{V^2}$	0	0	0	0
(0,2,1)	0	$\dfrac{1}{V^2}$	$\dfrac{V+2}{V^2}$	$\dfrac{1}{V^2}$	0	$\dfrac{1}{V^2}$	$\dfrac{1}{V^2}$	0	0	0
(0,3,0)	0	0	$\dfrac{1}{V^2}$	$\dfrac{2V+1}{V^2}$	0	0	$\dfrac{1}{V^2}$	0	0	0
(1,0,2)	$\dfrac{1}{V^2}$	$\dfrac{1}{V^2}$	0	0	$\dfrac{V+2}{V^2}$	$\dfrac{1}{V^2}$	0	$\dfrac{1}{V^2}$	0	0
(1,1,1)	0	$\dfrac{1}{V^2}$	$\dfrac{1}{V^2}$	0	$\dfrac{1}{V^2}$	$\dfrac{V}{V^2}$	$\dfrac{1}{V^2}$	$\dfrac{1}{V^2}$	$\dfrac{1}{V^2}$	0
(1,2,0)	0	0	$\dfrac{1}{V^2}$	$\dfrac{1}{V^2}$	0	$\dfrac{1}{V^2}$	$\dfrac{V+2}{V^2}$	0	$\dfrac{1}{V^2}$	0
(2,0,1)	0	0	0	0	$\dfrac{1}{V^2}$	$\dfrac{1}{V^2}$	0	$\dfrac{V+2}{V^2}$	$\dfrac{1}{V^2}$	$\dfrac{1}{V^2}$
(2,1,0)	0	0	0	0	0	$\dfrac{1}{V^2}$	$\dfrac{1}{V^2}$	$\dfrac{1}{V^2}$	$\dfrac{V+2}{V^2}$	$\dfrac{1}{V^2}$
(3,0,0)	0	0	0	0	0	0	0	$\dfrac{1}{V^2}$	$\dfrac{1}{V^2}$	$\dfrac{2V+1}{V^2}$

Once the system is defined as a Markov chain, it is possible to calculate the BTO fulfilment probability of the system as follows:

  i.   Calculate the steady state probabilities of the system for each state,
 ii.   Calculate the probability of generating a BTO, given the state of the system,
iii.   Combine i. and ii. and calculate the steady state probability of a BTO.

When the transition probability matrix is analysed, it is observed that for every two states *i* and *j,* the probability of transitioning from *i* to *j* is the same as transitioning from *j* to *i.* The steady state probabilities in this type of symmetric Markov chain are uniform (Kannan, 1994). Hence for the system above, the steady state probabilities for each state is: $^1/_{10}$.

For any given system state the probability of BTO fulfilment depends on the number of variants that have zero stock in the system. Based on the assumption of uniformly distributed demand, the probability of demand for each product is equally likely. For example, at state (0,1,2) a BTO fulfilment occurs only if the demand is for product 1 (probability is $^1/_3$); at state (0,0,3) a BTO fulfilment occurs if the demand is for product 1 or 2 (probability is $^1/_3 + ^1/_3 = ^2/_3$). In general, given the assumption of a uniform distribution, the probability of BTO fulfilment for any given state is the ratio: (number of stock out products)/(variety level).

As both the steady state probability and the BTO fulfilment probability for each state are known, the steady state BTO fulfilment probability of this system can be calculated, which gives the steady state BTO fulfilment proportion for this system.

Let s_t be the state of the system where $t = 1, 2, \cdots, 10$, and p_t be the probability of BTO fulfilment for state t, then the BTO% of this system is:

$$BTO\% = \sum_{t=1}^{10} s_t \times p_t = ^1/_{10}\left(^2/_3 + ^1/_3 + ^1/_3 + ^2/_3 + ^1/_3 + 0 + ^1/_3 + ^1/_3 + ^1/_3 + ^1/_3\right) = ^2/_5 \, .$$

The next section will generalise this analytical expression to calculate the BTO proportion for arbitrary positive values of S, P, and V.

3.2 A general Markov Chain model to calculate the BTO proportion

Given the assumptions in the basic open pipeline system, the total number of available products in the system is fixed and equals $S + P$. Consider each state in a Markov model defined as a specific permutation of the available number of products in the system (combination of stock and pipeline). A state *s* is represented by its inventory vector $= [I_1, I_2, \cdots, I_V]$, where I_i denotes the total unallocated inventory for variant *i* for $i = 1, 2, \cdots, V$. As the total number of products in the system is equal to

S+P and is fixed, it is clear that $[I_1 + I_2 + \cdots + I_v] = S + P$. Hence, for any given $S + P$ and V values, the number of states to uniquely represent the system is finite and is equal to $V - 1$ combinations of a set $S + P + V - 1$.

For the transition probabilities from one system inventory state to another, there are three possible cases:

i. Infeasible, if $s_t = [\cdots, x, \cdots]$ and $s_{t+1} = [\cdots, x + k, \cdots]$ where $k \geq 2$.
    As there is only one feed and one demand at each fulfilment cycle, the stock level cannot increase or decrease by more than 1 for any variant in the system. The transition probability is zero.

ii. New state, if $s_t = [\cdots, x, \cdots, y, \cdots]$ and $s_{t+1} = [\cdots, x + 1, \cdots, y - 1, \cdots]$ where $x \geq 0$ and $y > 0$. The transition probability is $1/V^2$.

iii. Same state, if $s_t = [0, \cdots, 0, +, \cdots, +]$ and $s_{t+1} = [0, \cdots, 0, +, \cdots, +]$ where "+" denotes positive integer values. Assume that there are m zeros, and $V - m$ positive values. The transition probability is $m/V + (V - m)/V^2$.

Using the above conditions, the transition matrix for any S+P and V values can be generated. By using the symmetric Markov chain property (Kannan, 1994), it can be concluded that the stationary distribution is uniform. As explained above the number of states the system can have is the $V - 1$ combinations of a set $S + P + V - 1$. Since the stationary state probabilities are equally likely, each possible state appears with a probability of:

$$\Pr\{s_t\} = \frac{1}{\binom{S+P+V-1}{V-1}} \tag{1}$$

To calculate the BTO proportion, all possible states that can lead to a BTO request and the probability of observing a BTO in each of those states need to be identified. A BTO can only occur if there is at least one zero entry in the inventory vector. The numbers of states that have at least one zero entry are:

$$1\text{-zero: } \binom{V}{1}, 2\text{-zeros: } \binom{V}{2}, \cdots, k\text{-zeros: } \binom{V}{k}, \cdots, (V\text{-}1)\text{-zeros: } \binom{V}{V-1}$$

In addition, when a BTO occurs in the system, an allocated product enters the pipeline, and hence the state of the inventory vector does not change ($s_t = s_{t+1}$). As the demand distribution is assumed to be uniform, the probability of generating a BTO order when there are k zero entries in the stock is k/V.

So, the resulting expression for calculating the steady state BTO proportion for an arbitrary system is:

$$BTO\% = \frac{\sum_{k=1}^{V-1} \frac{k}{V} \binom{V}{k}\binom{S+P-1}{V-k-1}}{\binom{S+P+V-1}{V-1}} \qquad (2)$$

This can be simplified to:

$$\frac{V-1}{S+P+V-1} \qquad (3)$$

Some illustrative calculations are presented in Table 2 showing the dramatic changes in the proportion fulfilled by build-to-order products for systems of different scale. For a short pipeline, low variety, low initial stock system, a third of the orders are fulfilled by BTO. When the initial stock level is ten times higher, the BTO proportion decreases to approximately 8%. However, for the same short pipeline and high stock configuration but with a variety level ten times higher, the BTO proportion increases to approximately 48%. In contrast, a long pipeline high stock system with low variety has a BTO proportion under 5%. The complementary metric – the proportion fulfilled from within the system (stock +pipeline) is also of interest to the producer. In the latter case, knowing that all but a small proportion of orders can be fulfilled from within the system may provide valuable information to system designers and managers.

Table 2 Illustrative BTO calculations for different system configurations

Configuration	P	V	S	BTO proportion
Short pipe/Low variety/ Low initial stock	256	256	256	0.332
Short pipe/Low variety/ High initial stock	256	256	2560	0.083
Short pipe/High variety/ High initial stock	256	2560	2560	0.476
Long pipe/Low variety/ High initial stock	2560	256	2560	0.047

4. Estimating the stock and pipeline fulfilment metrics using approximation schemes

Once the BTO proportion is known, the next step is to calculate the fractional share of fulfilment between the pipeline and stock. An essential property used in the exact calculation of the BTO proportion in section 3.2 above is that the number of unallocated products for a given system is a fixed value. This enabled the use of a Markov chain approach. However, the same assumption cannot be made for the stock level or the number of unallocated products in the pipeline, neither of

which is fixed. Changes in the stock level depend on the fulfilment mode and the specific product variant that occupies the last position in the pipeline:

• The stock level increases by 1 if a stock fulfilment does not occur and the last product in the pipeline (shown as P in Figure 2) is unallocated (and will therefore enter into stock at the end of the cycle).

• The stock level stays the same if (i) a stock fulfilment occurs and the last product in the pipeline is unallocated or (ii) a stock fulfilment does not occur and the last product in the pipeline is allocated.

• The stock level decreases by one if a stock fulfilment occurs and the last product in the pipeline is an allocated product.

Thus, the change in the stock level in the system depends on both the fulfilment mode and the last product in the pipeline. The allocation status of the last product in the pipeline depends on the history of the process. The last product in the pipeline is an allocated product if a BTO fulfilment occurred a pipeline time length previously or if a pipeline fulfilment has occurred previously that has allocated that specific variant from the pipeline. The next state in the system not only depends on the current state but also on the previous states of the system. Thus, the Markov analysis used above cannot be used to calculate the proportion fulfilled from the pipeline and/or from stock as the required assumptions are violated. It is for this reason that approximation schemes are developed here to estimate the pipeline and stock fulfilment metrics and the mean lead time and mean stock level in the system. As with the analytical model for BTO, the approximation schemes are developed for a system of arbitrary scale. These are presented in section 4.3 and their accuracy is discussed and summarised in section 4.4. The estimation of the two other metrics of interest – mean lead time (MLT) and means stock level (MST) - is discussed in section 4.5.

In order to test the accuracy of the approximation schemes a simulation model has been developed, which is outlined in section 4.1 below. The simulation model can be used to estimate the metrics accurately for most systems. However, systems with low variety give rise to particular issues that affect the estimation process. These are illustrated in section 4.2 below and their consequences noted.

4.1 A simulation model of a basic open pipeline system

A model has been developed in Matlab (www.mathworks.com/ products/matlab/) to simulate the basic open pipeline systems shown in Figure 2 using precisely the same logic and making the same assumptions explained in the introduction to section 3 above. The model can take as input any combination of system parameters (P, V and S) and generate each of the performance metrics

discussed in section 3 above after a simulation run. The model can use any input feed distribution and any demand distribution.

In conducting simulation experiments to estimate the steady state performance metrics for a given set of parameters, a simulation protocol was developed using a warm-up period and a 'batch means' approach with dead periods to reduce the variance in output metrics (Law and Kelton, 1999). A long warm-up period of 200,000 cycles was chosen. Law and Kelton (2000) note that many methods have been proposed to decide an appropriate length for the warm-up period but none have been found to perform well consistently under all conditions. In this research graphical methods were used for simplicity. Hoad et al.(2009) define a graphical method as a truncation method that involves visual inspection of time-series output and human judgment. For different combinations of pipeline length, variety and initial stock levels, the outputs for each of the fulfillment metrics, mean stock levels and mean lead times were plotted and it was observed that a warm-up period longer than 20 times pipeline length was adequate to observe steady-state behavior for the open pipeline systems considered. For simplicity in the simulation coding, a single warm-up period of 200, 000 was used for all experiments, giving over 20 pipeline lengths even for the longest pipeline.

After the warm-up period, statistics were gathered using 5 batches of length 10,000. Each period of data gathering was followed by a dead period of 10,000 cycles in which statistics were not gathered. A batch length of 10,000 cycles was selected so that even for the longest pipeline length of 8192, the whole pipeline was replenished in each batch. While the average of the batches was used as output metrics in the simulation, the individual batch results were also analysed and it was observed that individual batch averages were very similar to each other for the open pipeline scenarios tested in the approximation study below.

The total time required to run the simulations varies significantly and is mainly dependent on the variety level. The simulation algorithm has approximately quadratic complexity, $O(V^2)$. For low variety level small systems the simulations take less than a minute but at the highest variety level the simulations take up to 10 hours.

Figures 5a and 5b show similar outputs using the Matlab simulation and the expression in (3) above, respectively, to estimate the proportion of orders fulfilled by the BTO fulfilment mode for a given variety level (V = 1024) and a range of S and P values.
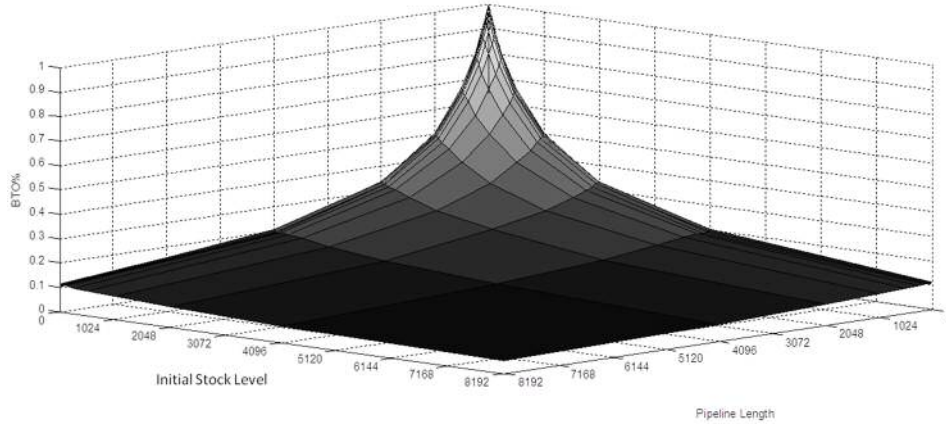
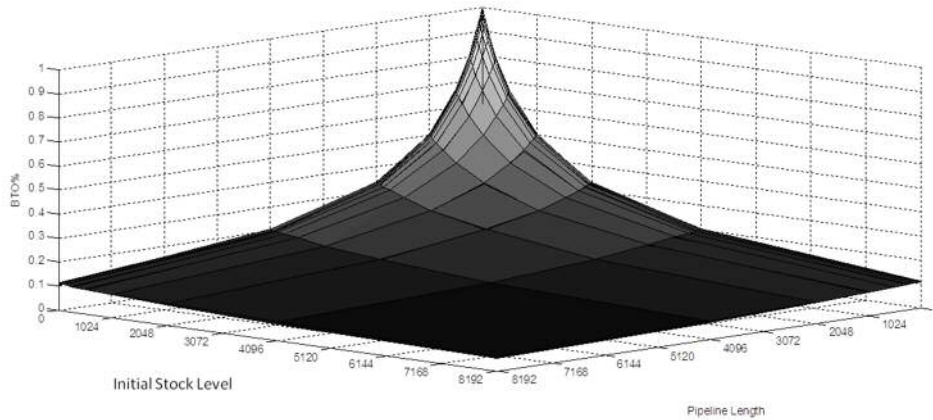**Figure 5a. 3D plot for BTO fulfilment using simulation**

Figure 5b. 3D plot for BTO fulfilment using the analytical formula

4.2 Fulfilment metrics in open pipeline systems with low variety

Experience with the simulation model has shown that it can be used to estimate average metrics accurately for all the open pipeline systems considered in this study except those with low variety. Specifically, the variance in the estimates of average performance metrics is larger for low variety systems than for high variety systems. This rather counter-intuitive behaviour results from an intrinsic property of any open pipeline system using the fulfilment logic explained in section 3. We illustrate the issue here using an extreme example with minimal variety.

Table 3 below shows the results for five independent simulation runs for a system with a long very pipeline (P = 8192), a low initial stock setting (S = 256) and a minimal level of variety (V =2). Each run uses a different random number seed. This is a system in which no BTO fulfilment would be expected and this is observed in each simulation run. However, the results show a large range for each of the other metrics (pipeline fulfilment, stock fulfilment, mean lead time and mean stock level)

17

across the five simulation runs, even though the same simulation protocols described above have been used.

Table 3 Estimates from five independent runs for V=2, P=8192, and S=256

Run	PIPE	STOCK	MLT	MSL
1	0.21576	0.78424	0.005	296.5976
2	0.15854	0.84146	0.001	264.7572
3	0.49536	0.50464	0.069	823.2719
4	0.29128	0.70872	0.003	279.3066
5	0.34576	0.65424	0.007	310.9149

Figures 6 and 7 illustrate what can occur to cause the large variance in estimates in low variety systems. The plots show the stock level of each variant for two independent simulation runs with different random seeds for the above system for a run length of 100, 000 cycles. It can be observed that once the stock level of one of the variants reaches zero, it is hard for the system to recover from the unfavourable stock mix. In Figure 6 the stock level of variant 2 reaches zero around the 30,000th simulation cycle and the system struggles to recover. In Figure 7 the system recovers from the stock-out of variant 2 around the 18,000th and 25,000th simulation cycles, but does not recover after 76,000th cycle. The estimates of stock fulfilment are very different in these systems with consequential effects on both the observed mean stock level and the observed mean lead times.

This behaviour is an inherent property of open pipeline order fulfilment systems of this type. It occurs because the pipeline is used both to replenish stock and to fulfil customers. Once a variant runs out of stock, when a new demand for that variant occurs in the system, the demand is fulfilled from the pipeline. Once a customer is fulfilled from the pipeline the relevant product is allocated and leaves the system at the end of the pipeline without going into the stock. Hence, once the stock level of a variant reduces to zero, not only is the coverage of variants in stock affected but also the variant mix in the downstream end of the pipeline is affected. The variants that could have replenished the stock instead leave the system because of the pipeline fulfilment mode. Once this begins to occur it may be difficult for stock composition to recover as the relevant variants are taken from the pipeline.
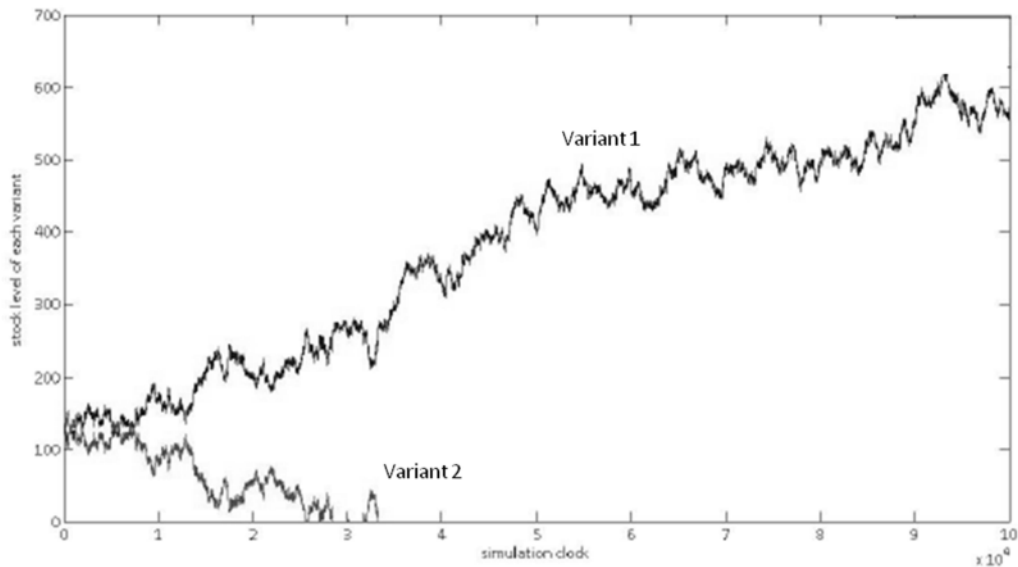
Figure 6 Stock levels plot for a low stock fulfilment proportion
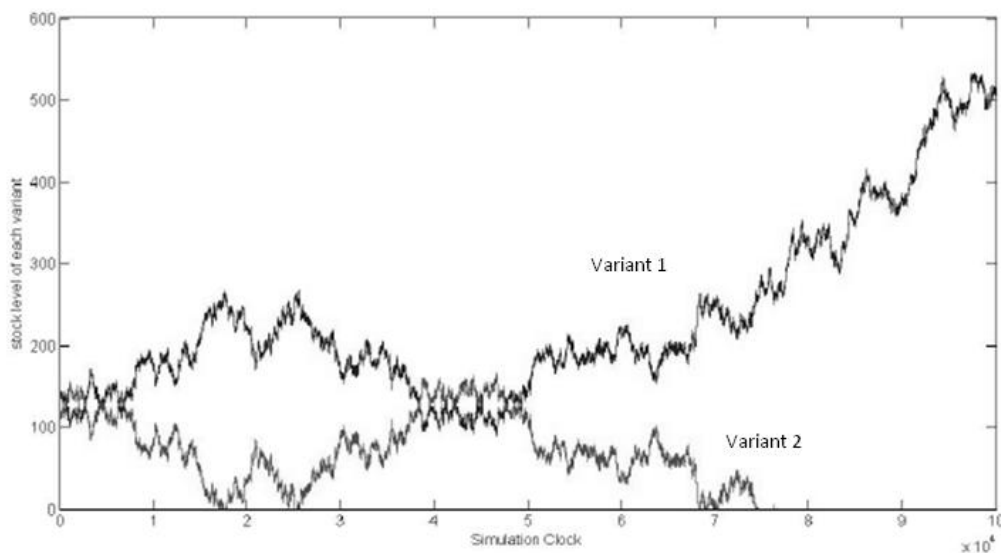
Figure 7 Stock levels plot for a high stock fulfilment proportion

As noted, this behaviour is an inherent property of an open pipeline system, including systems with high variety. In a high variety system, when a stockout of one of the variants occurs, the pipeline will be searched to fulfil incoming customer orders for that variant. As with the low variety case, the stock of that particular variant will not be replenished by the pipeline and the stockout may persist. However, the effect of such a stockout on average fulfilment metrics is less in a high variety system compared to a low variety system as these are aggregate metrics across all variants. The value of the metric is not therefore strongly affected in a high variety system when a variant, or a small proportion of the variants, are stocked out and not replenished from the pipeline for a period of

19

time. In a low variety system, having longer simulation runs or using more elaborate simulation variance reduction techniques do not provide more precise simulation estimates as the effect is a property of the system.

Given the issue noted above, the approximation schemes developed below to estimate the stock and pipeline fulfilment proportions exclude low variety systems. For the open pipeline systems considered in the study below, the analysis of experimental results has indicated that simulation estimates are stable for variety levels of 256 or larger. Hence, variety levels at or above this level have been used in evaluating the accuracy of the approximation schemes developed.

Importantly, it should be noted that this is not a significant issue in practice. Allowing a pipeline fulfilment mechanism is less likely in low variety systems as conventional fulfilment modes (i.e. fulfilment from Stock and by BTO) are likely to be used and to be sufficient in practice.

4.3 Approximation schemes to estimate stock and pipeline fulfilment proportions

Here we use the terms BTO%, PIPE%, STOCK% for the proportion of orders fulfilled by BTO, from the pipeline and from stock, respectively. Once the BTO% is calculated, its complement (1-BTO%) needs to be split between PIPE% and STOCK%. A naïve approximation scheme could assume that PIPE% and STOCK% are proportional to the relative sizes of P and S, i.e. $\frac{PIPE\%}{STOCK\%} = \frac{P}{S}$. Hence an estimate of PIPE% and STOCK% might be calculated:

$$PIPE\% = (1 - BTO\%) \times P/(S + P) \tag{4}$$

$$STOCK\% = (1 - BTO\%) \times S/(S + P) \tag{5}$$

This scheme results in large errors because it fails to capture the dynamics in the system. It implies that the stock level (SL) and unallocated pipeline length (UL) stay constant. However, every BTO fulfilment product enters the pipeline as an allocated product. Hence, on average, a BTO% of the pipeline is allocated with products arising from BTO fulfilment. Thus, the effective number of available products in the pipeline is (1-BTO%) x P. As the number of available products in the system is fixed, the effective stock level must be larger than the initial stock level. Thus, for the first approximation (Scheme-1) the logic in the naïve approach is used but with better approximations for both the unallocated pipeline (UP) and the stock level (SL) (i.e. $\frac{PIPE\%}{STOCK\%} = \frac{UP}{SL}$)

$$UP = (1 - BTO\%) \times P \tag{6}$$

$$SL = BTO\% \times P + S \tag{7}$$

So the resulting approximation formulas are:

$$PIPE\% = (1 - BTO\%) \times UP/(S + P) \qquad (8)$$

$$STOCK\% = (1 - BTO\%) \times SL/(S + P) \qquad (9)$$

The accuracy of the approximation scheme has been evaluated on 2002 different system configurations capturing a wide range of variety levels (≥ 256) and initial positions for three pipeline length representing a small, medium and large sized pipeline (P=256, 2048, 8192; V = 256, 384, ..., 2048; S = 2,4,8, ..., 8192). The accuracy is determined by the differences between the approximation scheme output and the Matlab simulation output for any given set of conditions. The mean absolute deviation (MAD) error metric was chosen as (i) some error values are very small and an error metric like mean absolute percentage error (MAPE) would not be efficient for small values; and (ii) the fulfilment metrics are proportions/percentages and taking the percentage error of percentages may lead to misleading results especially for small values. For conciseness, the full tabulation of results is presented only for the most accurate scheme.

Scheme-1

The average MAD error for Scheme-1 for the medium length pipeline (P = 2048) across the range of variety levels and initial stock conditions is 11.1%.   The error decreases for shorter pipelines and increases for longer pipelines. The error decreases as S increases, and for a fixed S, the error decreases as V increases. Thus, in general, Scheme-1 is more accurate for shorter pipelines, higher stock levels and higher variety levels.

Scheme-2

Equation (6) above is an upper-bound for UP and therefore equation (7) is a lower-bound for SL. These approximations ignore the number of products allocated in the pipeline based on PIPE%. In approximation Scheme-2 it is assumed that, in addition to the products allocated in the pipeline because of BTO%, half of the pipeline is also added to the stock. Thus, the approximation scheme formulas are:

$$PIPE\% = (1 - BTO\%) \times \frac{UP}{(S + UP + P/2)} \qquad (10)$$

$$STOCK\% = (1 - BTO\%) \times \frac{S + P/2}{(S + UP + P/2)} \qquad (11)$$

For the medium pipeline length scenario (P = 2048) the overall average error is 3.8% for Scheme-2 compared to 11.1% with Scheme-1. As with Scheme-1, accuracy is better for shorter pipelines and

worse for longer pipelines and for a fixed V the error decreases as S increases. For a fixed S the error is more variable but generally reduces as V increases.

Scheme-3

Scheme-3 is an improved version of scheme-2. It is based on searching for the optimal pipeline fraction which goes into stock. In scheme-2 a value of P/2 was used. In approximation Scheme-3, instead of using P/2, an error minimization problem has been solved using Excel in order to find the factor for the proportion of the pipeline products added to stock that minimizes the MAD of the approximation scheme across all the experimental space studied. The minimum error for these approximations is given by P/1.51. The resulting equations are:

$$PIPE\% = (1 - BTO\%) \times \frac{UP}{(S + UP + P/1.51)} \tag{12}$$

$$STOCK\% = (1 - BTO\%) \times \frac{S + P/1.51}{(S + UP + P/1.51)} \tag{13}$$

Table 4, 5 and 6 show the MAD errors for Scheme-3 for small, medium and large pipeline lengths (P = 256, P = 2048, and P = 8192), across the same ranges of initial stock positions and variety levels. The overall average errors for each scenario are 0.6%, 1.2% and 2.0%, respectively. For conciseness the results are tabulated using ratios of variety and stock level to pipeline length (V/P and S/P). As with the previous schemes, for a fixed V/P ratio, the error decreases as S/P increases, and for a fixed S/P ratio, the error reduces as V/P increases. Although the errors are much smaller, Scheme-3 still performs better for shorter pipelines and worse for longer pipelines.

Table 4 Approximation errors of scheme-3 for P=256

P=256		V/P										
		1	3/2	2	5/2	3	7/2	4	5	6	7	8
S/P	0	0.8%	1.2%	1.4%	0.7%	0.8%	0.4%	0.5%	0.6%	0.3%	0.4%	0.3%
	1/128	1.1%	1.2%	1.4%	0.6%	0.8%	0.4%	0.8%	0.6%	0.5%	0.3%	0.2%
	1/64	1.1%	1.1%	1.4%	0.6%	0.8%	0.4%	0.9%	0.4%	0.4%	0.4%	0.3%
	1/32	0.8%	1.3%	1.1%	1.5%	1.3%	1.1%	1.2%	0.5%	0.4%	0.3%	0.4%
	1/16	1.0%	1.2%	1.1%	1.4%	1.2%	1.0%	0.8%	0.5%	0.5%	0.3%	0.4%
	1/8	0.8%	0.8%	1.2%	0.4%	0.7%	0.3%	0.6%	0.6%	0.2%	0.4%	0.6%
	1/4	0.8%	1.1%	0.9%	1.2%	1.2%	0.8%	0.7%	0.2%	0.5%	0.6%	0.4%
	1/2	0.5%	0.9%	1.1%	1.0%	0.9%	0.7%	0.7%	0.4%	0.7%	0.6%	0.0%
	1	0.6%	0.6%	0.5%	0.5%	0.0%	0.1%	0.3%	0.4%	0.2%	0.1%	0.5%
	2	0.1%	0.9%	0.1%	0.6%	0.5%	0.4%	0.1%	0.7%	0.3%	0.0%	0.0%
	4	0.1%	0.5%	0.0%	0.5%	0.3%	0.2%	0.4%	0.2%	0.2%	0.0%	0.3%
	8	0.8%	0.5%	1.1%	0.5%	0.2%	0.1%	0.1%	0.1%	0.3%	0.5%	0.1%
	16	0.0%	0.5%	0.8%	0.0%	0.1%	0.2%	0.3%	0.2%	0.0%	0.1%	0.1%
	32	0.5%	0.4%	0.2%	0.3%	0.3%	0.8%	0.2%	0.6%	0.4%	0.3%	0.2%

Table 5 Approximation errors of scheme-3 for P=2048

P=2048		V/P										
		1/8	3/16	1/4	5/16	3/8	7/16	1/2	5/8	3/4	7/8	1
S/P	0	2.8%	2.5%	1.8%	1.5%	0.9%	0.3%	0.0%	0.7%	0.7%	1.3%	1.2%
	1/1024	2.9%	2.6%	1.4%	1.5%	0.9%	0.3%	0.1%	0.1%	0.9%	1.4%	1.2%
	1/512	3.2%	2.5%	1.9%	1.5%	0.9%	0.3%	0.1%	0.6%	0.3%	1.1%	0.9%
	1/256	3.0%	2.6%	1.6%	1.5%	0.9%	0.3%	0.4%	0.1%	0.7%	0.6%	1.2%
	1/128	3.1%	3.2%	0.4%	1.6%	1.1%	0.6%	0.5%	0.8%	1.0%	1.1%	1.1%
	1/64	2.5%	2.7%	0.8%	1.5%	0.9%	0.4%	0.4%	0.8%	0.8%	0.9%	1.0%
	1/32	3.1%	3.2%	1.0%	1.7%	1.3%	0.6%	0.4%	0.4%	0.6%	1.1%	1.1%
	1/16	3.1%	3.2%	2.7%	1.7%	1.5%	0.7%	0.2%	0.4%	0.7%	0.5%	1.1%
	1/8	4.0%	2.9%	1.6%	1.6%	0.8%	0.5%	0.4%	0.3%	0.8%	0.1%	0.5%
	1/4	3.1%	3.7%	0.7%	1.8%	2.2%	0.8%	0.7%	0.2%	0.2%	0.3%	0.4%
	1/2	0.9%	2.6%	2.5%	1.9%	0.9%	1.0%	0.6%	0.9%	0.3%	0.0%	0.6%
	1	1.3%	2.1%	1.9%	1.3%	2.3%	0.9%	1.2%	0.4%	0.5%	0.3%	0.9%
	2	0.5%	1.4%	0.7%	0.1%	1.5%	0.2%	0.2%	1.5%	0.2%	0.6%	0.4%
	4	1.7%	0.9%	1.9%	1.0%	0.7%	1.1%	0.6%	0.3%	0.7%	1.1%	0.7%

Table 6 Approximation errors of scheme-3 for P=8192

P=8192		V/P										
		1/32	3/64	1/16	5/64	3/32	7/64	1/8	5/32	3/16	7/32	1/4
S/P	0	0.1%	2.0%	2.1%	2.2%	3.2%	1.5%	1.8%	0.8%	2.4%	1.9%	1.1%
	1/4096	1.0%	2.0%	2.5%	2.3%	3.2%	1.5%	1.5%	2.0%	2.6%	1.5%	1.7%
	1/2048	0.3%	2.1%	2.8%	2.2%	3.2%	1.5%	1.9%	2.9%	1.5%	1.4%	1.0%
	1/1024	0.2%	2.0%	2.1%	2.2%	3.2%	1.5%	2.9%	2.9%	0.9%	1.5%	1.8%
	1/512	0.6%	2.0%	1.6%	2.2%	3.1%	1.5%	3.0%	1.0%	1.7%	1.7%	1.6%
	1/256	0.0%	2.0%	1.6%	2.2%	3.2%	1.5%	1.7%	1.8%	1.5%	2.2%	1.0%
	1/128	0.8%	3.7%	3.0%	3.5%	2.6%	2.4%	1.7%	1.4%	1.8%	1.4%	2.1%
	1/64	1.4%	1.9%	2.4%	2.1%	3.0%	1.5%	2.9%	2.2%	1.7%	2.5%	1.5%
	1/32	4.1%	1.8%	3.0%	2.1%	3.0%	1.6%	2.5%	1.9%	1.8%	1.7%	1.0%
	1/16	1.0%	1.6%	1.4%	1.8%	2.7%	1.3%	2.3%	2.4%	1.6%	1.9%	0.9%
	1/8	3.3%	3.0%	1.9%	2.8%	2.5%	2.2%	2.0%	0.9%	2.0%	1.4%	1.3%
	1/4	2.2%	0.2%	0.2%	0.3%	1.2%	1.1%	1.0%	1.4%	2.0%	0.9%	0.1%
	1/2	5.3%	1.0%	1.1%	0.4%	1.2%	1.3%	0.4%	0.8%	1.4%	0.4%	0.8%
	1	11.5%	7.1%	8.2%	3.3%	3.8%	3.4%	2.1%	2.1%	0.2%	0.0%	1.2%

4.4 Comparison of the approximation schemes

Each of the approximation schemes presented above performs better for short pipeline and high variety systems than for other combinations of parameters. In a short pipeline system, the pipeline has less effect on performance, the system is more stable and the initial stock level and average stock level metrics are close to each other. In a high variety system, the BTO% is higher than for a low variety systems (for fixed S and P), which leads to an exact calculation of a bigger proportion of the system.

The three approximation schemes are compared in Table 7, based on MAD metrics over the full experimental space (the 2002 system configurations studied). Scheme-3 outperforms all other approximation schemes for each of the error level boundaries across, providing approximations with errors less than 1% in 78.5% of the cases and less than 5% in 99.7% of cases.

Table 7 Comparison of approximation schemes

	Scheme – 1	Scheme - 2	Scheme -3
Error ≤ 1%	63.11%	59.62%	78.49%
Error ≤ 2%	67.71%	66.70%	92.16%
Error ≤ 3%	70.29%	71.96%	96.91%
Error ≤ 5%	73.63%	88.01%	99.65%
Error ≤ 10%	80.36%	99.70%	99.95%
Error ≤ 20%	89.73%	100.00%	100.00%

4.5 Approximations to estimate mean lead time and mean stock level

In addition to the approximation schemes developed for the fulfilment metrics, expressions have also been developed to approximate the mean lead time (MLT) and mean stock lead time (MSL). The MLT and MSL metrics use the outputs from the previous calculations as the input metrics for calculations.

By definition if a BTO fulfilment occurs in the system the lead time is 1 and if a stock fulfilment occurs the lead time is zero. For pipeline fulfilment the expected LT is half the pipeline length (LT = 1/2).The mean lead time in the system can thus be approximated by:

$$MLT = \left(BTO\% \times 1 + PIPE\% \times {}^{1}/_{2} + STOCK\% \times 0\right)\Big/_{P}$$

$$MLT = BTO\% + {}^{PIPE\%}/_{2} \qquad\qquad (14)$$

A number of observations can be made from the assumptions in the basic open pipeline model to estimate mean stock level (MSL). Firstly, as noted previously, S+P is constant and given the fulfilment logic in the model S is a lower bound for MSL. The increase in the stock level in the system from the initial stock level S equals the number of allocated products in the pipeline (i.e. if products are allocated in the pipeline, then stock level should increase by the same amount). Hence an approximation for MSL should seek to estimate the number of products fulfilled in the pipeline. Allocated products in the pipeline arise from BTO fulfilment and pipeline fulfilment. When a BTO

fulfilment occurs the allocated product passes through the whole pipeline and hence BTO%×P number of product are allocated in the pipeline in the long run.

For pipeline fulfilment, the expected time a product will be allocated in the pipeline is P/2. Hence, PIPE%×P⁄2 products are allocated in the pipeline because of pipe fulfilment. Thus, the mean stock level in the system (MSL) can be approximated by:

$$MSL = S + BTO\% \times P + PIPE\% \times {P}/{2} \qquad (15)$$

The expressions to estimate MLT and MSL rely on the exact calculation of the BTO% and the approximation for PIPE%. Thus, their accuracy is dependent on the accuracy of PIPE%. These approximations work very well using approximation scheme-3 above. Tables 9 and 10 present the estimation errors for MLT and MSL for the system in Table 5 above (a medium pipeline length of 2048) using the mean absolute percentage error metric (MAPE). The MLT and MSL approximation schemes for this pipeline length scenario are 3.1% and 4.1%, respectively. The error educes for increased initial stock levels and increased variety levels. When compared to the simulation outputs across the whole experimental space, the MAPE for MLT and MSL are 1.6% and 1.2%, respectively.

Table 9 Approximation errors of MLT for P=2048

P = 2048	V/P										
	1/8	3/16	1/4	5/16	3/8	7/16	1/2	5/8	3/4	7/8	1
0	11.3%	9.9%	7.0%	6.3%	4.1%	3.5%	2.4%	1.2%	0.6%	0.3%	0.3%
1/1024	10.1%	9.9%	6.5%	6.3%	4.1%	3.6%	1.8%	1.2%	0.7%	0.3%	0.4%
1/512	11.9%	9.8%	7.8%	6.2%	4.1%	3.6%	2.1%	1.2%	0.7%	0.3%	0.3%
1/256	10.0%	9.9%	6.8%	6.2%	4.2%	3.5%	3.3%	1.2%	0.7%	0.4%	0.3%
1/128	11.4%	11.1%	3.7%	5.9%	5.1%	3.6%	3.4%	0.7%	0.3%	0.4%	0.5%
1/64	7.6%	10.0%	5.5%	6.2%	4.0%	3.6%	1.3%	1.4%	0.8%	0.3%	0.0%
1/32	10.6%	11.1%	4.8%	5.9%	5.2%	3.6%	3.0%	0.7%	0.4%	0.3%	0.3%
1/16	10.5%	11.0%	9.1%	6.0%	5.3%	3.6%	2.7%	0.8%	0.4%	0.4%	0.3%
1/8	13.6%	9.7%	7.4%	6.2%	3.7%	3.4%	3.5%	1.5%	0.9%	0.5%	0.5%
1/4	10.1%	11.5%	3.7%	5.8%	6.6%	3.5%	3.4%	1.4%	0.6%	0.3%	0.4%
1/2	8.1%	9.3%	7.4%	6.2%	3.4%	3.5%	2.9%	2.4%	1.8%	0.5%	2.0%
1	6.9%	7.9%	6.8%	4.9%	6.9%	3.7%	4.3%	3.0%	1.3%	0.9%	2.8%
2	6.3%	7.2%	3.9%	1.8%	5.8%	0.5%	1.5%	3.5%	0.8%	1.1%	1.2%
4	7.5%	0.9%	7.9%	4.1%	3.3%	5.5%	2.6%	0.6%	1.0%	1.5%	2.9%

Table 10 Approximation errors of MSL for P=2048

P = 2048	V/P										
	1/8	3/16	1/4	5/16	3/8	7/16	1/2	5/8	3/4	7/8	1
0	11.5%	10.3%	6.9%	6.5%	4.4%	4.0%	2.7%	1.5%	0.7%	0.6%	0.4%
1/1024	9.9%	10.2%	6.3%	6.5%	4.4%	4.0%	1.7%	1.5%	0.7%	0.6%	0.3%
1/512	12.1%	10.2%	7.6%	6.4%	4.4%	4.0%	2.2%	1.4%	0.7%	0.6%	0.2%
1/256	9.4%	10.2%	6.6%	6.5%	4.4%	3.9%	3.4%	1.5%	0.7%	0.7%	0.0%
1/128	10.8%	11.0%	3.9%	5.7%	5.0%	3.6%	3.4%	0.7%	0.4%	0.4%	0.4%
1/64	7.3%	10.0%	5.0%	6.3%	4.2%	3.9%	1.1%	1.6%	0.7%	0.6%	0.1%
1/32	10.0%	10.3%	4.6%	5.5%	4.8%	3.4%	2.9%	0.8%	0.5%	0.4%	0.2%
1/16	8.8%	9.6%	8.0%	5.2%	4.5%	3.3%	2.4%	0.7%	0.5%	0.3%	0.3%
1/8	10.2%	7.7%	5.8%	5.1%	3.3%	3.0%	2.7%	1.4%	0.8%	0.7%	0.4%
1/4	6.0%	7.0%	2.4%	3.6%	4.1%	2.2%	2.1%	1.0%	0.6%	0.1%	0.3%
1/2	3.0%	3.7%	3.1%	2.6%	1.4%	1.4%	1.4%	1.1%	0.8%	0.3%	1.0%
1	1.4%	1.7%	1.5%	1.2%	1.7%	0.8%	1.2%	0.9%	0.5%	0.2%	0.9%
2	0.5%	0.6%	0.4%	0.2%	0.6%	0.1%	0.2%	0.4%	0.1%	0.1%	0.1%
4	0.2%	0.0%	0.2%	0.1%	0.1%	0.2%	0.1%	0.0%	0.0%	0.1%	0.2%

6. Discussion

The paper has presented new results to estimate performance metrics for basic open pipeline configurations. An exact expression has been introduced to calculate the BTO proportion for any combination of the key system parameters - pipeline length (P), initial stock level (S) and variety level (V). Approximation schemes have been developed to estimate both the pipeline and stock fulfilment proportions. This is a significant advance on previously published work (Brabazon and MacCarthy, 2010). The new approaches not only provide an exact calculation for the BTO fulfilment metric compared to the approximations in a previous model but are also scalable for any open pipeline system of any pipeline length, any level of variety, and any initial stock position.

The best approximation scheme works very well for all but low variety systems with errors less than 1% MAD in almost 80% of cases studied (V/P ratios from 1/32 to 8 and S/P ratios from 0 to 32). Approximation schemes are also presented to estimate both the mean lead time (MLT) and mean stock level (MSL) metrics. These again work well for all but low variety systems generating errors less than 2% MAPE across the experimental space studied. Thus, the performance metrics for a large range of open pipeline systems can be predicted with good accuracy using the expressions developed.

The models presented here provide generic insights that are useful for systems design, management and control. The simplicity of the analytical and approximation results make them particularly relevant in the planning phases when considering the impact of design parameters such as the variety level to offer and the most appropriate stocking level for a given system. Two examples highlighting generic managerial insights are given here.

A producer that wants to satisfy a large majority of its customers from stock (i.e. have a stock dominant system) can, for a given pipeline length and variety level, estimate the initial stock level needed to prime the system to guarantee a required level of stock fulfilment. Use of the formulae presented here shows that for small and medium sized pipelines (relative to stock level), the effect of increasing initial stock level above a certain level results in a diminishing return on the increase in stock fulfilment proportion. Given that stock holding has associated costs, the producer may seek a balanced position to guarantee stock fulfilment above a certain level for a given cost.

In general, use of the formulae shows that the performance metrics are more reactive to changes in variety level and/or initial stock level for shorter pipelines than for longer pipelines. Thus, for a short pipeline a relatively small increase in initial stock level can generate a relatively large change in stock fulfilment compared to a longer pipeline, highlighting the benefits of shortening the pipeline in this

26

case. In contrast, a relatively small change in variety level can adversely affect performance metrics for a short pipeline, resulting in a relatively large increase in the BTO fulfilment proportion compared to a longer pipeline (although overall lead times may be shorter for a short pipeline system). The length of the pipeline is essentially determined by the planning and supply chain processes deployed by the producer. It is likely to be difficult to change for an existing system in many cases, particularly in the automotive sector. However, at the design stage, when considering the most appropriate fulfilment processes to put in place for a new product range, the results here enable understanding of the effects of pipeline length on likely performance.

An issue with the estimation of performance metrics for open pipeline systems with low variety has also been discussed. For any open pipeline system an unfavourable sequence of demands can result in a stockout of a particular variant and the stockout can persist as demand for that variant is satisfied from the pipeline. This is a feature of all open pipeline systems but its effects are more noticeable in low variety systems when average performance metrics are estimated across all variants. Thus, the estimation of steady state performance metrics at low variety is problematic. It is important to note however that in practice open pipelines are more likely to be deployed in conditions where there are a significant number of product variants than in low variety contexts where conventional fulfilment approaches are likely to suffice.

As with any modelling work refinements can be made to incorporate more features, consider a wider range of systems or to improve accuracy (see section 7 below). For instance, demand level may be skewed, violating the assumption of uniform demand made in the models presented here. Indeed in some cases such as the automotive sector the producer may be able to influence the shape of the demand function by its product offering, marketing and pricing strategies. Although it is not possible to use the calculation methods presented here directly with a skewed demand distribution, their performance can be approximated by using an approach based on the 'effective variety' in the system. We illustrate this in Figure 8 where we show the effect of increasing skew in demand and feed distributions on a particular metric for a particular system.

Figure 8 shows the results of experiments with the simulation model for a system with fixed S and P values of 1024, across 16 variety levels ranging from 128 to 2048 with equal increments of 128 (shown on the vertical axes). To control the shape in the feed and demand functions in the system, Beta distributions are used with a fixed α parameter of 1 and a β parameter varying between 1 and 30 (shown on the horizontal axis). As the β parameter increases, the distribution becomes progressively more skewed (the vertical lines indicate the approximate level of skew for some values

of β). In total 480 experimental settings are considered covering all the combinations of V and β (16*30).

The simulation results are shown here for one metric - the BTO%. Using shading, the contour plots show the changes in BTO% fulfilment for different levels of variety and different levels of skew. In the plot each shade represents a different BTO% interval. Nine intervals are shown, from a 0% – 7% interval to a 56% – 63% interval.
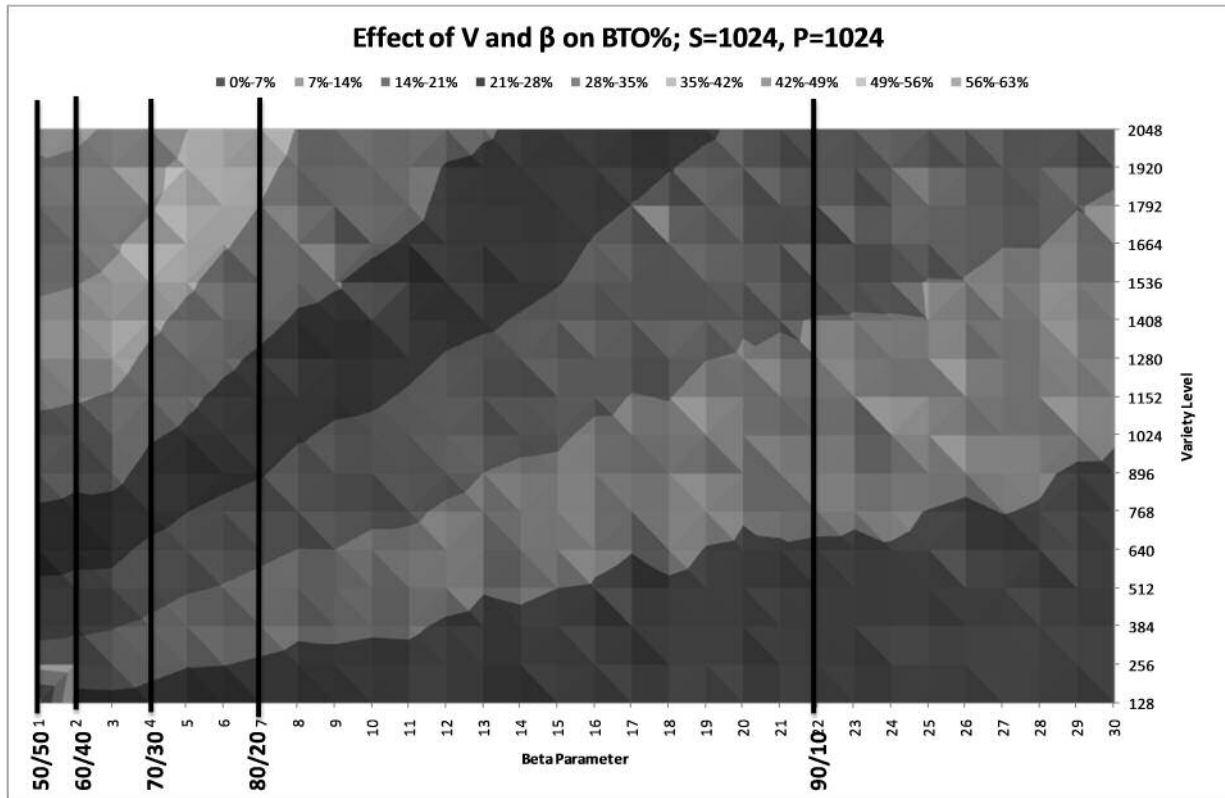

Figure 8 Effect of product variety, and beta value on BTO% for S=1028, P=1024

The level of skew affects the behavior of the open pipeline system. However, for the same pipeline length (P) and initial stock level setting (S), systems with different variety levels perform similarly depending on their relative skew levels. For example, a system with V = 512 and a skew level of β = 2 performs similarly to a system with four times the level of variety (V = 2048) but a much higher skew level (β = 22). In general, a system with a large number of variants but highly skewed demand across the variants performs similarly to a system with a much fewer number of variants but with a more uniform demand. This gives rise to the concept of the 'effective variety level', where for a fixed pipeline length and a fixed initial stock level, systems are considered to have the same effective variety if both systems perform similarly. Further work is on-going to estimate effective variety level

28

for different systems, taking into account the degree of skew in the demand distribution, and thereby extending the applicability of the estimation methods presented here.

7. Direction for future work

The models presented here can be refined, developed or augmented in many ways. One example has been given at the end of the previous section above. Some of the many potentially fruitful directions are noted here.

Where there is uncertainty in the timing of arrivals the demand arrival process could be modelled as a stochastic process. This is not a major issue in the automotive sector at the macro level considered here as incoming demands are usually batched in an unscheduled order bank. However, this approach could be worth considering in order fulfilment processes in other sectors with different operating practices. The loss of customers resulting from failure to obtain desired variants in a timely manner could also be considered. Differences between the producer's knowledge of demand and actual demand could also be studied. The basic model could be further developed to incorporate some of the features and the flexibilities sometimes seen in real systems with multiple demand points (e.g. dealers and distributors), such as the trading of physical and planned products and the ability to amend the specification of products in the planning pipeline. These have been studied for real automotive order fulfilment systems (Brabazon et al. 2010). Further theoretical analysis can add to the insights gained.

The complexity and tractabilty of any model will of course influence the modelling approach adopted, whether analytical modelling, approximation schemes or simulation. The Matlab model used here for the approximation study can simulate a wider range of systems. The approach taken to developing the approximation schemes here may also be further developed to approximate the performance of open pipelines operating with different assumptions.

As noted in the introduction, the factors driving the need for open pipeline principles are not confined to the automotive industry. In many sectors the characteristics for considering open pipelines are present to some degree - a high number of variants, different customer needs within the customer base and significant planning pipelines from which customers can be allocated products. Elements of the approach are evident in sectors such as instrumentation, machinery and electronic equipment (Meredith and Akinc, 2007; Closs et al., 2010). Opening the pipeline can potentially satisfy more customers, more closely and more quickly than a conventional fixed mode fulfilment system. However, open pipeline systems are more dynamic than conventional systems and it is clear from the studies here that such systems have to be designed with care. The study

identifies some of the system characteristics that are most relevant in designing such systems for high performance and provides a platform for further work in extending their applicability.

Acknowledgements

References

Agrawal M, Kumaresh T V, Mercer G (2001), 'The false promise of Mass Customization', *The McKinsey Quarterly*, Vol. 3, pp 62-71.

Alford D, Sackett P, Nelder G (2000). 'Mass customisation — an automotive perspective'. *International Journal of Production Economics*, Vol. 65(1), pp.99-110.

Blecker T, Abdelkafi N (2006). 'Complexity and variety in mass customization systems: analysis and recommendations', *Management Decision*, Vol. 44, pp 908 – 929

Brabazon P G, MacCarthy (2006). 'Fundamental behaviour of virtual-build-to-order systems'. International Journal of Production Economics , Vol. 104, pp 514-524.

Brabazon P G, MacCarthy B L (2010). 'On Markovian approximations for virtual-build-to-order systems'. *Journal of the Operational Research Society*, Vol. 61, pp 1471–1484.

Brabazon P G, MacCarthy B L, Woodcock A, Hawkins R, (2010). 'Mass Customization in the automotive industry: comparing inter-dealing trading and reconfiguration flexibility in order fulfilment'. *Production and Operations Management*, Vol. 19 (5), pp 489 - 502.

Closs D J, Nyaga G N, Voss M D, (2010). 'The differential impact of product complexity, inventory level, and configuration capacity on unit and order fill rate performance'. *Journal of Operations Management*, Vol. 28 (1), pp 47-57

Denton B, Gupta D, Jawahir K (2003). 'Managing increasing product variety at integrated steel mills'. *Interfaces*, Vol. 33 (2), pp 41-53

Er M, MacCarthy B L (2006). 'Managing product variety in multinational corporation supply chains: a simulation study'. *International Journal of Manufacturing Technology Management*, Vol 17(8), pp 117-1138.

Fredriksson P, Gadde, L.-E (2005). 'Flexibility and rigidity in customization and build-to-order production'. *Industrial Marketing Management*, 34(7), pp 695–705.

Graman G A (2010). 'A partial-postponement decision cost model'. *European Journal of Operational Research*, Vol. 201(1), pp 34-44.

Graman G A, Magazine M J, (2002). 'A numerical analysis of capacitated postponement'. *Production and Operations Management*, Vol.11 (3), pp 340-357.

Hoad K, Robinson S, Davies R (2009). 'Automating discrete event simulation output analysis automatic estimation of number of replications, warm-up period and run length.' In *INFORMS Simulation Society Research Workshop* (Eds. Lee LH, Kuhl M E, Fowler J W, Robinson S), INFORMS Simulation Society, pp 75–79.

Holweg, M. (2000). 'The order fulfilment process in the automotive industry'. *Cardiff Business School, Lean Enterprise Research Centre,* Report Reference: S1 – 7/00.

Holweg M, Pil F (2004). *The Second Century: Reconnecting Customer and Value Chain through Build-to-Order, Moving beyond Mass and Lean Production in the Auto Industry.* MIT Press.

Jewkes E M, Alfa A S (2009). 'A queuing model of delayed product differentiation'. *European Journal of Operational Research*, Vol. 199(3), pp.734-743.

MacCarthy B L, Brabazon P G, Bramham J (2003). 'Fundamental modes of operation for Mass Customization'. *International Journal of Production Economics*, Vol. 85 (3), pp 289-304.

McCutcheon D M, Raturi A S, Meredith J R, (1994). 'The Customization-Responsiveness Squeeze', *Sloan Management Review*, 35(2) , pp 89-99.

Meredith J, Akinc U (2007). 'Characterizing and structuring a new make-to-forecast production strategy'. *Journal of Operations Management*, Vol. 25 , pp 623-642.

Meyr, H. (2004). 'Supply chain planning in the German automotive industry'. *OR Spectrum*, Vol. 26(4), pp 447-470.

Ramdas K, (2003). 'Managing Product Variety: An Integrative Review and Research Directions'. *Production and Operations Management,* Vol. 12 (1), pp 79–101.

Squire B, Brown S, Redman J, Bessant J, 2006. 'The impact of mass customisation on manufacturing trade-offs', *Production and Operations Management*, Vol. 15(1), pp 10–21.

Swaminathan J M, Tayur S R (1998), 'Managing broader product lines through delayed differentiation using vanilla boxes'. *Management Science*, Vol. 44 (12), pp S161-S172.

Tu Q, Vonderembse M A, Ragu-Nathan T S (2001). 'The impact of time-based manufacturing practices on mass customization and value to customer'. *Journal of Operations Management*, Vol. 19(2), pp 201-217.

Van Hoek R I (2001), 'The rediscovery of postponement: a literature review and directions for research', *Journal of Operations Management*, Vol. 19, pp 161-184.

WSJ. (2010). 'Options Overload for Fiats 500'. *The Wall Street Journal*, 21st of November, 2010 (by Jeff Bennett). http://online.wsj.com/ article/ SB10001424052748704443045756284936173638 72. html1- 88.

Yang B, Burns N D, Backhouse C J (2004). 'Postponement: a review and integrated framework'. *International Journal of Production Research*, Vol. 24, pp 468 – 487.

Zinn W, Bowersox D J, (1988). 'Planning physical distribution with the principle of postponement'. *Journal of Business Logistics*, Vol. 9(2), pp.117 - 136.

Zipkin P (2001). 'The limits of Mass Customization'*. MIT Sloan Management Review*, Vol. 42(3), pp 81-88.