Stein–Chen Approximation and Error Bounds for Order Fill Rates in Assemble-to-Order Systems

Wenhui Zhou,[1] Xiuli Chao[2]

[1] *School of Business Administration, South China University of Technology, Guangzhou 510640, China*

[2] *Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan 48109*

Abstract: Assemble-to-order (ATO) is an important operational strategy for manufacturing firms to achieve quick response to customer orders while keeping low finished good inventories. This strategy has been successfully used not only by manufacturers (e.g., Dell, IBM) but also by retailers (e.g., Amazon.com). The evaluation of order-based performance is known to be an important but difficult task, and the existing literature has been mainly focused on stochastic comparison to obtain performance bounds. In this article, we develop an extremely simple Stein–Chen approximation as well as its error-bound for order-based fill rate for a multiproduct multicomponent ATO system with random leadtimes to replenish components. This approximation gives an expression for order-based fill rate in terms of component-based fill rates. The approximation has the property that the higher the component replenishment leadtime variability, the smaller the error bound. The result allows an operations manager to analyze the improvement in order-based fill rates when the base-stock level for any component changes. Numerical studies demonstrate that the approximation performs well, especially when the demand processes of different components are highly correlated; when the components have high base-stock levels; or when the component replenishment leadtimes have high variability. © 2012 Wiley Periodicals, Inc. Naval Research Logistics 59: 643–655, 2012

Keywords: assemble-to-order system; order fill rate; convex ordering; Stein–Chen approximation; error bounds; $M/G/\infty$ queue

1. INTRODUCTION

Assemble-to-order (ATO) is an important operational strategy for manufacturing firms to achieve quick response to customer orders while keeping low finished good inventories. This strategy has been successfully used not only by manufacturers but also by retailers. In an ATO system, components are produced in advance to stock, while the assembly of final products is delayed until detailed product specifications are available [16]. In case the time it takes to assemble a final product is relatively short compared with component replenishment times, the ATO production strategy is particularly attractive. Indeed, most of the mass customization companies adopt ATO systems; examples include Dell computer, Toyota's "build your Toyota," Amazon's "build your own ring," and Nike's "design your shoes," among others.

Despite its increasing popularity in reality, ATO systems with multiple products and multiple components are "notoriously difficult to analyze and manage" [3]. The main reason for that is that the demand processes for components in

such systems are highly intertwined and positively correlated. A key measure for such systems is the order-based fill rate, which is defined as the probability that a customer order, which typically consists of different types of components, is filled immediately. Song [12] is perhaps the first to investigate order-based fill rates of multiproduct and multicomponent ATO systems. As noted by Song [12] and Song and Yao [14], an exact analysis of order-based performance measures involves $2^N - 1$ factors (N being the total number of components), thus it is difficult to compute the performance for even a moderate-sized problem. The most relevant literature in this area uses the method of stochastic comparison to obtain bounds of order-based performance measures, such as order fill rate, mean backorders, and so forth. See, for example, Song [12], Xu [18], Xu and Li [19], and Song and Yao [14]. Another approach to this problem is asymptotic analysis based on large deviation theory, in which one is concerned with the asymptotic behavior of the fill rate when the promised due date becomes very long, see for example, Glasserman and Wang [6]. The latter approach is applicable only to systems where the component replenishment process is modeled as a single-server queue. Several authors have

Correspondence to: Wenhui Zhou (whzhou@scut.edu.cn)

studied ATO systems to minimize expected holding cost (of components) and expected back-order cost (of orders), see for example, Song [13], Song and Yao [14], Lu and Song [7], and Zhao and Simchi-Levi [20], among others. We refer the reader to Song and Zipkin [15] for an extensive discussion of examples and motivations for ATO systems.

In this article, we study a multiproduct and multicomponent ATO system similar to that of Song [12]. We first consider the case where components are managed using base-stock policies. The component leadtimes are random variables, and customer orders follow independent Poisson processes. We develop an extremely simple Stein–Chen approximation for the fill rates as well as the error bound for this approximation. Unlike that of Song [12], the Stein–Chen approximation works for arbitrarily distributed leadtimes for replenishing components, and as a matter of fact, we show that the approximation has a tighter error bound when the component replenishment leadtimes are more variable in the convex ordering. This approximation gives an explicit expression for the order-based fill rate for any customer in terms of component-based fill rates; it is easy to compute, and allows an operations manager to analyze the improvement in order-based fill rates when the base-stock level for any component changes. Our numerical studies demonstrate that the Stein–Chen approximation performs well, especially when demand processes for components are positively correlated, the components have high base-stock levels, or the component replenishment leadtimes are highly variable. When the components have deterministic leadtimes, we extend the results along several directions: first, we develop Stein–Chen approximation and error bounds for the order fill rate within any given time window; next, we obtain similar results for the case when components are managed using (r_i, Q_i) policies, and finally, we extend the results to the case when the customer orders follow compound (batch) Poisson processes. We also discuss extensions of these results to the case when the component leadtimes are stochastic and sequential (or noncrossing). We point out that our models are similar to those in Song [12, 13], Song and Yao [14], and Zhao and Simchi-Levi [20], in which the component suppliers have infinite production capacity, thus the component ordering process for each component is modeled as an $M/G/\infty$ queue, or a variation of the $M/G/\infty$ queue. This is different from Glasserman and Wang [6] who consider the case of finite production capacity of the component suppliers where the supply process for each component is modeled as an $M/G/1$ or a $G/G/1$ queue.

The Stein–Chen method is a powerful tool for approximating the sum of independent or dependent Bernoulli random variables with a Poisson random variable and evaluating the error bound of their difference, and it has been successfully used in probability and statistics (see e.g., Refs. 1, 9). Roughly, because a customer order is filled on arrival if and only if all the required components are available, we can construct a set of dependent indicator (Bernoulli) random variables, one for each component, such that the fill rate of the customer order is equal to the probability that the sum of these dependent Bernoulli random variables is zero. This allows us to obtain an approximation and both an upper and a lower bound for the fill rate of each class of customer orders. The detailed analysis is provided in Section 2. We present three sets of numerical examples to test the Stein–Chen approximation and the error bounds and compare them with those reported in the literature. We point out that, while the Stein–Chen approximation is very simple and performs well in general, it does not always outperform Song's approximation [12]. Our numerical examples show that the Stein–Chen approximation performs the best (and outperforms Song's approximation) when the superposed demand processes of components are highly correlated and/or when the component fill rates are high.

The rest of the article is organized as follows. In the next section, we study the Stein–Chen approximation for the case of stochastic component leadtimes, and outline the model description as well as the results under the base-stock component control policies. In Section 3, we focus on the case of deterministic component leadtimes and extend the results of Section 2 along three directions: fill rate within any given time window, batch component ordering policies, and batch Poisson customer order processes. Section 4 conducts some numerical studies and compares our results with several others reported in the literature; we also demonstrate how the approximation result depends on variability of component replenishment leadtimes. Concluding remarks and extensions to stochastic sequential component leadtimes are discussed in Section 5. Finally, all proofs are provided in the appendix. Throughout the article, for any two real numbers a, b and x, we use notation $a \vee b = \max\{a, b\}$ $a \wedge b = \min\{a, b\}$, and $x^+ = \max\{x, 0\}$.

2. RANDOM COMPONENT LEADTIMES

Consider an ATO system with N components and M final products. Let $\Omega = \{1, 2, \ldots, N\}$ be the set of all components and $\mathscr{D}_k \subseteq \Omega$ the set of components that are required to assemble one unit of product k, $k = 1, 2, \ldots, M$. For convenience, we shall also refer to the demand process for product k as class k order. We shall first assume that product k consists of exactly one unit of each component in $\mathscr{D}_k$, and later we discuss extension to arbitrary batch sizes. We follow the convention of using superscript to indicate order/product type and subscript to indicate component type. Class k orders arrive according to a Poisson process with rate λ^k, and the arrival processes of different classes of orders are independent of one another. As in Song [12], we assume that customer orders are satisfied on a first-come-first-serve basis and the

time to assemble the components into a final product is negligible and is assumed to be 0. That is, if all the components of a product are available, then the customer order for this product is filled immediately; otherwise, the order for the product is backlogged. This article is concerned with the evaluation of order-based fill rate, which is defined as the probability that a customer for a product is satisfied without delay upon arrival. We shall also discuss order-based fill rate within a time window, which is defined as the probability that a customer order is satisfied within a given time window after arrival.

For $i = 1, \ldots, N$, the inventory position, which is defined as the inventory on hand plus inventory on order subtract backorders, of component i is reviewed continuously. In this section, we consider the case where the component is controlled using a base-stock policy with base-stock level S_i for component i, and later we will extend the results to the case where component i is controlled using an (r_i, Q_i) policy. Under a base stock policy, as soon as the inventory position of component i drops to S_i, an order is placed immediately. The leadtime L_i for component i is independent and identically distributed with cdf $G_i(.)$, pdf $g_i(.)$, and mean $E[L_i]$. For convenience, we define some additional notations as follows.

$\lambda = \sum_{k=1}^{M} \lambda^k =$ total arrival rate of customer orders;

$S(i) = \{k \mid i \in \mathscr{D}_k\} =$ the set of customer orders that contain component i;

$\lambda_i = \sum_{k \in S(i)} \lambda^k =$ aggregate demand rate for component i;

$\lambda_{i \cap j} = \sum_{l \in S(i) \cap S(j)} \lambda^l =$ total demand rate of orders that contain both components i and j;

$\lambda_i^{-j} = \sum_{l \in S(i) \setminus S(j)} \lambda^l = \lambda_i - \lambda_{i \cap j} =$ total demand rate of orders that contain component i but not component j;

$f_i =$ fill rate of component i;

$F^k =$ fill rate of class k orders;

$F =$ aggregate fill rate. Clearly,

$$F = \sum_{k=1}^{M} \frac{\lambda^k}{\lambda} F^k. \tag{1}$$

Following base-stock policies for components, the arrival of a class k order with $i \in \mathscr{D}_k$ triggers a replenishment of component i, thus the aggregate demand process for component i is Poisson with rate λ_i. Let $IN_i(t)$ be the net inventory level of component i and let $IO_i(t) = S_i - IN_i(t)$ be the number of outstanding orders of component i. Because each order of component i is replenished after an independent and identically distributed amount of time, it is not hard to see that $IO_i(t)$ behaves exactly the same as an $M/G/\infty$ queue with arrival rate λ_i and service time being the component replenishment time, see for example, Song and Yao [14].

This observation will be used in computing the performance measure of component i.

For the ATO system operated under the base-stock policy S_i for component i, we additionally introduce the following notation.

$IN_i =$ steady-state net inventory of component i;

$IO_i =$ steady-state inventory on order of component i;

$\phi(\cdot \mid a) =$ the pmf of Poisson random variable with parameter a;

$\Phi(\cdot \mid a) =$ the cdf of Poisson random variable with parameter a;

$\bar{\Phi}(\cdot \mid a) = 1 - \Phi(\cdot \mid a)$.

In this article, we present an approximation for the fill rate of each class of customer orders based on the Stein–Chen method, as well as an error bound of this approximation. The Stein–Chen method evaluates the total variation distance between the distribution of the sum of a set of Bernoulli random variables X_i and the distribution of a Poisson random variable with the same mean. For more details of this method, the reader is referred to, for example, [1], and to the Appendix for a brief discussion of this method. The main idea of this approximation is the following: in the context of an ATO system, an order for product k is satisfied immediately only if all components $i \in \mathcal{D}_k$ are available when the order arrives. Thus, if we let X_i be the indicator function that component i is out of stock when needed by a class k order, then

$$W^k = \sum_{i \in \mathcal{D}_k} X_i$$

is a sum of Bernoulli random variables which, by the Stein–Chen method, can be approximated by a Poisson random variable with mean $E[\sum_{i \in \mathcal{D}_k} X_i]$. In particular, the probability $P\{W^k = 0\}$ is the fill rate of class k order and it can be approximated by the probability that the Poisson random variable is equal to 0.

The following is the first main result of this article.

THEOREM 1: For any k, the Stein–Chen approximation for the fill rate of class k order is $F_{\mathrm{SC}}^k = e^{-\Lambda_k}$ and it has error bound

$$\mid F^k - F_{\mathrm{SC}}^k \mid \ \leq (b_1^k + b_2^k)(1 - e^{-\Lambda_k})/\Lambda_k, \tag{2}$$

where

$$\Lambda_k = \sum_{i \in \mathscr{D}_k} p_i, \tag{3}$$

and

$$p_i = \bar{\Phi}(S_i - 1 \mid \lambda_i E[L_i]),$$

with b_1^k, b_2^k given by

$$b_1^k = \sum_{i,j \in \mathscr{D}_k} p_i p_j, \qquad (4)$$

$$b_2^k = \sum_{i,j \in \mathscr{D}_k} \sum_{i \neq j} p_{ij}, \qquad (5)$$

and

$$p_{ij} = \sum_{\ell=0}^{(S_i-1) \vee (S_j-1)} \phi(\ell \mid \lambda_{i \cap j} \theta_{ij}) \overline{\Phi}(S_i - 1 - \ell \mid \lambda_i^{-j} E[L_i]$$
$$+ \lambda_{i \cap j} \vartheta_{ij}) \overline{\Phi}(S_j - 1 - \ell \mid \lambda_j^{-i} E[L_j] + \lambda_{i \cap j} \delta_{ij})$$
$$+ \overline{\Phi}(S_i \vee S_j - 1) \mid \lambda_{i \cap j} \theta_{ij});$$

the parameters $\theta_{ij}, \vartheta_{ij},$ and δ_{ij} are given by

$$\theta_{ij} = \int_0^\infty (1 - G_i(x))(1 - G_j(x))dx, \qquad (6)$$

$$\vartheta_{ij} = \int_0^\infty (1 - G_i(x))G_j(x)dx, \qquad (7)$$

$$\delta_{ij} = \int_0^\infty G_i(x)(1 - G_j(x))dx. \qquad (8)$$

To illustrate the result, we consider the W-structure of ATO system, which was studied in Lu et al. [8], see Fig. 1. This ATO system has three components and two classes of orders, $\mathscr{D}_1 = \{1, 2\}$ and $\mathscr{D}_2 = \{2, 3\}$, with demand arrival rates λ^{12} and λ^{23}. The leadtimes for the three components have mean $E[L_1], E[L_2],$ and $E[L_3]$, respectively.

The Stein–Chen approximation and the error bound of fill rate F^{12} of order 12 are given by

$$F_{\text{SC}}^{12} = e^{-p_1 - p_2}$$

and

$$|F^{12} - F_{\text{SC}}^{12}|$$
$$\leq (p_1^2 + 2p_1 p_2 + p_2^2 + 2p_{12})(1 - e^{-p_1 - p_2})/(p_1 + p_2).$$

where

$$p_1 = \bar{\Phi}(S_1 - 1 \mid \lambda^{12} E[L_1]),$$
$$p_2 = \bar{\Phi}(S_2 - 1 \mid (\lambda^{12} + \lambda^{23}) E[L_2]),$$
$$p_{12} = \sum_{\ell=0}^{(S_1-1) \vee (S_2-1)} \phi(\ell \mid \lambda^{12} \theta_{12}) \overline{\Phi}(S_1 - 1 - \ell \mid \lambda^{12} \vartheta_{12})$$
$$\times \overline{\Phi}(S_2 - 1 - \ell \mid \lambda^{23} E[L_2] + \lambda^{12} \delta_{12})$$
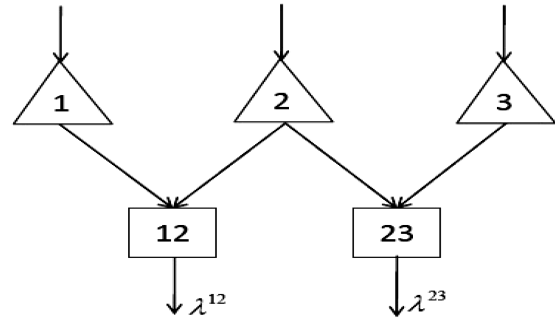$$+ \overline{\Phi}(S_1 \vee S_2 - 1) \mid \lambda^{12} \theta_{12});$$

Figure 1.   The W system.

Similar results are given for the fill rate of order 23. The approximation of the overall order fill rate is

$$F_{\text{SC}} = \frac{\lambda^{12} F_{\text{SC}}^{12} + \lambda^{23} F_{\text{SC}}^{23}}{\lambda^{12} + \lambda^{23}}.$$

The Stein–Chen approximation for order-based fill rates is very simple, and its computation $F_{\text{SC}}^k = e^{-\Lambda_k}$ boils down, by Eq. (3), to the evaluation of $p_i = \bar{\Phi}(S_i - 1 \mid \lambda_i E[L_i])$. Thus, if the probability distribution of Poisson is assumed to be known, then F_{SC}^k is obtained by simply adding them up, and the complexity to compute F_{SC}^k is linear in the number of components required by order k.

The result allows us to compute the order-based fill rate for any class of order in terms of component-based fill rates. Let f_i be the fill rate for component i, meaning that $100 f_i\%$ of the time component i is available when needed. As p_i is the probability, under base-stock policy S_i for component i, that component has no on-hand inventory when needed, we have $f_i = 1 - p_i$. This gives

$$F_{\text{SC}}^k = e^{-\sum_{i \in \mathscr{D}_k}(1-f_i)} = \prod_{i \in \mathscr{D}_k} e^{-(1-f_i)}.$$

This simple expression allows us to answer various questions of interest. For example, when the base-stock level for component i is increased from S_i to $S_i + 1$, the fill rate for any order of class k with $i \in \mathscr{D}_k$ will be improved, and we estimate that the percentage of improvement is

$$e^{-(\Phi(S_i)-\Phi(S_i-1))} - 1 = e^{-\phi(S_i)} - 1 = e^{-e^{-\lambda_i E[L_i]} \frac{(\lambda_i E[L_i])^{S_i}}{S_i!}} - 1.$$

The Stein–Chen approximation also suggests an efficient method for setting base-stock levels for managing components to achieve a desired customer service level, specified in terms of order-based fill rates. For instance, suppose that the desired fill rate for class k order is $F^k, k = 1, \ldots, M$. Using the Stein–Chen approximation, we want to achieve, for all k, that

$$F_{\text{SC}}^k = e^{-\sum_{i \in \mathscr{D}_k}(1-f_i)} \geq F^k.$$

As $1 - f_i = \bar{\Phi}(S_i - 1 \mid \lambda_i E[L_i])$, the inequalities above are equivalent to

$$\sum_{i \in \mathscr{D}_k} \bar{\Phi}(S_i - 1 \mid \lambda_i E[L_i]) \leq - \log F^k, \quad k = 1, \ldots, M. \quad (9)$$

These inequalities determine the required base-stock levels for each component to achieve the desired order-based fill rates. As an example, consider the W-system described earlier with three components and two classes of orders. Suppose the desired fill rates for the two classes of orders are F^{12} and F^{23}, respectively. Then Eq. (9) can be written as

$$\bar{\Phi}(S_1 - 1 \mid \lambda^1 E[L_1]) + \bar{\Phi}(S_2 - 1 \mid (\lambda^1 + \lambda^2) E[L_2])$$
$$\leq - \log F^{12},$$
$$\bar{\Phi}(S_2 - 1 \mid (\lambda^1 + \lambda^2) E[L_2]) + \bar{\Phi}(S_3 - 1 \mid \lambda^2 E[L_3])$$
$$\leq - \log F^{23}.$$

These constraints clearly show the trade-offs between the component base-stock levels and the order-based fill rates for different classes of customers.

As $e^{-x} \geq 1 - x$ for all $x \geq 0$, the approximation of fill rate satisfies

$$e^{-\Lambda_k} = \prod_{i \in \mathscr{D}_k} e^{-p_i} \geq \prod_{i \in \mathscr{D}_k} (1 - p_i)$$
$$= \prod_{i \in \mathscr{D}_k} \Phi(S_i - 1 \mid \lambda_i E[L_i]). \quad (10)$$

We note that right hand side of (10) is precisely the lower bound of fill rate reported in Proposition 5.1 of Song [12].

Numerical studies on the Stein–Chen approximation and error bounds are reported in Section 4. It is interesting to observe, from these numerical studies, that the fill rate approximation tends to be more accurate as the leadtimes for replenishing components are larger and more variable. Thus, we next investigate the effect of the variability of component replenishment leadtimes on the error bound of order fill rate approximation.

Consider two ATO systems that differ only in component replenishment leadtimes. Let the leadtimes of the two systems be denoted by L_i and $\tilde{L}_i$, respectively, with corresponding distribution functions G_i and $\tilde{G}_i$. Suppose L_i and $\tilde{L}_i$ are ordered in convex ordering. Recall that $L_i \geq_{cx} \tilde{L}_i$ if $E[h(L_i)] \geq E[h(\tilde{L}_i)]$ for all convex function h for which the expectations exist (see Ref. 10). In particular, the above inequality implies that $E[L_i] = E[\tilde{L}_i]$ and $Var(L_i) \geq Var(\tilde{L}_i)$. Therefore, convex order measures the variability of leadtimes.

THEOREM 2: The more variable the component replenishment leadtimes, the tighter the error bound of the Stein–Chen approximation for fill rates.

One possible explanation for this result is the following: the Stein–Chen method approximates the sum of random variables by a Poisson random variable with the same mean, the more random these random variables, the better the approximation.

Although the Stein–Chen approximation gives an explicit and extremely simple estimate for fill rate and, as will be seen in Section 5, it performs well in general, but we remark that we cannot guarantee that this approximation always outperforms the existing ones, even in simple systems such as W, N, and M systems. Therefore, it is recommended that the approximation is used jointly with other estimates and approximations reported in the literature.

3. DETERMINISTIC COMPONENT LEADTIMES

If the component replenish times are deterministic, then we can extend the results in the previous section along several important directions. Parts of these extensions hold true for stochastic but sequential (no-crossing) component leadtimes, and we will discuss that in Section 5.

To proceed, we first note that the error bound for the Stein–Chen approximation for order-based fill-rates can be simplified for deterministic leadtimes. Suppose l_i is the deterministic leadtime of component i, $i = 1, \ldots, M$, then the error bound is obtain by Eq. (2) with p_{ij} given by, if we assume without loss of generality that $l_i \leq l_j$,

$$p_{ij} = \sum_{\ell=0}^{S_i \vee S_j - 1} \phi(\ell \mid \lambda_{i \cap j} l_i) \bar{\Phi}(S_i - 1 - \ell \mid \lambda_i^{-j} l_i)$$
$$\times \bar{\Phi}(S_j - 1 - \ell \mid \lambda_j(l_j - l_i) + \lambda_j^{-i} l_j)$$
$$+ \bar{\Phi}(S_i \vee S_j - 1 \mid \lambda_{i \cap j} l_i).$$

Our first extension is for the order-based fill-rate within any given time window, which is defined as the probability an order is filled within any given time window, say $\tau \geq 0$. As the fill rate of class k order is 1 if $\tau \geq l_i$ for all $i \in \mathcal{D}_k$, here we only consider nonnegative τ that satisfies $\tau < \max_{i \in \mathcal{D}_k} l_i$.

THEOREM 3: Suppose component leadtimes are deterministic. For any k, the Stein–Chen approximation for fill rate $F_k(\tau)$ of class k order within a time window τ is $F_{SC}^k(\tau) = e^{-\Lambda_k(\tau)}$, where

$$\Lambda_k(\tau) = \sum_{i \in \mathscr{D}_k} p_i(\tau),$$

and

$$p_i(\tau) = \bar{\Phi}(S_i - 1 \mid \lambda_i(l_i - \tau)^+).$$

Furthermore, the Stein–Chen approximation has error bound

$$| F^k(\tau) - F^k_{SC}(\tau) | \le (b_1^k(\tau) + b_2^k(\tau))(1 - e^{-\Lambda_k(\tau)})/\Lambda_k(\tau),$$

with $b_1^k(\tau)$ and $b_2^k(\tau)$ given by

$$b_1^k(\tau) = \sum_{i,j \in \mathscr{D}_k} p_i(\tau) p_j(\tau),$$

$$b_2^k(\tau) = \sum_{i,j \in \mathscr{D}_k} \sum_{i \ne j} p_{ij}(\tau),$$

and $p_{ij}(\tau)$ given by

$$p_{ij}(\tau) = \sum_{\ell=0}^{S_i \vee S_j - 1} \phi(\ell | \lambda_{i \cap j}(l_i - \tau)^+)$$
$$\times \bar{\Phi}(S_i - 1 - \ell | \lambda_i^{-j}(l_i - \tau)^+)$$
$$\times \bar{\Phi}(S_j - 1 - \ell | \lambda_j((l_j - \tau)^+ - (l_i - \tau)^+) + \lambda_j^{-i}(l_j - \tau)^+)$$
$$+ \bar{\Phi}(S_i \vee S_j - 1 | \lambda_{i \cap j}(l_i - \tau)^+).$$

Note that, as we only consider the time window $\tau < \max_{i \in \mathcal{D}_k} l_i$, it holds that $p_i(\tau) > 0$ for at least some $i \in \mathcal{D}_k$, thus $\Lambda_k(\tau) > 0$.

Our second extension is on the component inventory control policies. In the previous section, we considered the scenario where the components are controlled using base-stock policies with base-stock level S_i for component i. In applications, it is often the case that, due to certain fixed ordering cost, the component i ordering follows an (r_i, Q_i) policy, $i = 1, \ldots, N$. That is, as soon as the inventory position of component i drops to r_i, an order of Q_i is placed to bring the inventory position to $r_i + Q_i$. In what follows, we extend the Stein–Chen approximation and error bound to the case when components inventories are managed using (r_i, Q_i) policy.

Let IP_i denote the steady state inventory position of component i. An important result that will be used in our analysis is that, when components ordering follows (r_i, Q_i) policies, IP_i is uniformly distributed on $\{r_i + 1, \ldots, r_i + Q_i\}$, and $IP_1(t), IP_2(t), \ldots, IP_N(t)$ are independent discrete uniform random variables [13]. It is rather easy to extend Song's result to that, for any time epochs $t_1, t_2, \ldots, t_N$, $IP_1(t_1), IP_2(t_2), \ldots, IP_N(t_N)$ are also independent and uniformly distributed. This latter result will be used in proving the following theorem.

THEOREM 4: Suppose component i is controlled using (r_i, Q_i) policy, $i = 1, \ldots, N$. For any k, the Stein–Chen approximation for fill rate of class k order F^k is still $F^k_{SC} = e^{-\Lambda_k}$, with error bound of the approximation given by (2), and $\Lambda_k = \sum_{i \in \mathscr{D}_k} p_i$, b_1^k, b_2^k are given by (4) and (5), but p_i and p_{ij} are replaced by

$$p_i = \frac{1}{Q_i} \sum_{q=1}^{Q_i} \bar{\Phi}(r_i + q - 1 | \lambda_i l_i),$$

$$p_{ij} = \frac{1}{Q_i Q_j} \sum_{q_i=1}^{Q_i} \sum_{q_j=1}^{Q_j} \sum_{\ell=0}^{(r_i+q_i-1) \vee (r_j+q_j-1)} \phi(\ell | \lambda_{i \cap j} l_i)$$
$$\times \bar{\Phi}((r_i + q_i - 1 - \ell | \lambda_i^{-j} l_i)$$
$$\times \bar{\Phi}(r_j + q_j - 1 - \ell | \lambda_j(l_j - l_i) + \lambda_j^{-i} l_j)$$
$$+ \bar{\Phi}((r_i + Q_i - 1) \vee (r_j + Q_j - 1) | \lambda_{i \cap j} l_i).$$

Using the same analysis as what follows Theorem 1, it can be seen that the Stein–Chen approximation for the fill rate of class k order, when components are controlled using (r_i, Q_i) policy, satisfies

$$e^{-\Lambda_k} \ge \prod_{i \in \mathcal{D}_k} \frac{1}{Q_i} \sum_{q_i=1}^{Q_i} \bar{\Phi}(r_i + q_i - 1 | \lambda_i l_i).$$

The right hand side is an extension of the lower bound for order-based fill rate obtained in Song [12] for the case of base-stock component control policies.

Our third extension is on the arrival processes of customer orders. In the previous section, the arrival process is assumed to be Poisson implying that each customer requires exactly one product. We now extend the result to the case that an order can be for an arbitrary number of products. This is done by assuming that the customer orders follow a batch Poisson process. Suppose the components are still managed using (r_i, nQ_i) policy, that is, as soon as the inventory position of type i component drops to or below r_i, an order of a minimum multiple of Q_i is placed to raise the inventory position to the interval $\{r_i + 1, \ldots, r_i + Q_i\}$.

Under very mild conditions, it has been shown in Song [12] that when component i is managed using (r_i, nQ_i) policy, $i = 1, \ldots, N$, the inventory positions of the different components $IP_1(t), \ldots, IP_N(t)$ are independent and uniformly distributed. And again, it can be shown that for any points in time $t_1, \ldots, t_N$, $IP_1(t_1), \ldots, IP_N(t_N)$ are also independent and uniformly distributed. This result will allow us to extend our results to the batch Poisson customer order processes.

For ease of exposition, we present the results for the case that each class k order requires a random number of product k, and each unit of product k consists of one item from each component in set $\mathcal{D}_k$, $k = 1, \ldots, M$. More specifically, class k orders arrive according to a Poisson process with rate λ^k, and each order requires ℓ units of product k with probability $z^{k,\ell}$, $\ell = 1, 2, \ldots$. Clearly, we have

$$\sum_{\ell=1}^{\infty} z^{k,\ell} = 1, \quad k = 1, \ldots, M.$$

Thus, the aggregate demand process for component i is also a batch Poisson process with rate $\lambda_i = \sum_{k \in S(i)} \lambda^k$ and the batch size of a random selected order for type i component, denoted by Z_i, has pmf

$$a_i(\ell) = P\{Z_i = \ell\} = \frac{1}{\lambda_i} \sum_{k \in S(i)} \lambda^k z^{z,\ell}, \quad \ell = 1, 2, \ldots.$$

To present our result, we need some additional notation. For any batch size distribution $\mathbf{a} = (a(\ell); \ell = 1, 2, \ldots)$, we let $\phi(\cdot \mid \mathbf{a}, \lambda)$ denote the pmf of the compound Poisson random variable $\sum_{l=1}^{N} Z_l$, where N has Poisson distribution with parameter λ and $Z_1, Z_2, \ldots$, are i.i.d. with pmf $\mathbf{a}$, and N is independent of the sequence $Z_1, Z_2, \ldots$. Moreover, we shall continue to use $\Phi(\cdot \mid \mathbf{a}, \lambda)$ and $\bar{\Phi}(\cdot \mid \mathbf{a}, \lambda)$ to denote the corresponding cdf and tail function of pmf $\phi(\cdot \mid \mathbf{a}, \lambda)$. Finally, for any components i and j, let $\mathbf{a}_i$ be the pmf of the batch size of a randomly selected order in $S(i)$, $\mathbf{a}_i^{-j}$ the pmf of the batch size of a randomly selected order in $S(i) \setminus S(j)$, $\mathbf{a}_{i \cap j}$ the pmf of the batch size of a randomly selected order in $S(i) \cap S(j)$, and

$$\tilde{\mathbf{a}}_j = \frac{\lambda_j (l_j - l_i) \mathbf{a}_j + \lambda_j^{-i} l_i \mathbf{a}_j^{-i}}{\lambda_j (l_j - l_i) + \lambda_j^{-i} l_i}.$$

The notation $\mathbf{a}_j$ and $\mathbf{a}_j^{-i}$ are similarly defined.

THEOREM 5: Suppose customer orders follow a batch Poisson process and component i is controlled using $(r_i, n Q_i)$ policy, $i = 1, \ldots, N$. For any k, the Stein–Chen approximation for fill rate of class k order F^k is still $F_{SC}^k = e^{-\Lambda_k}$, with error bound of the approximation given by (2), and $\Lambda_k = \sum_{i \in \mathscr{D}_k} p_i$, b_1^k, b_2^k are given by (4) and (5), but p_i and p_{ij} are replaced by

$$p_i = \frac{1}{Q_i} \sum_{q=1}^{Q_i} \bar{\Phi}(r_i + q - 1 \mid \mathbf{a}_i, \lambda_i l_i),$$

and

$$p_{ij} = \frac{1}{Q_i Q_j} \sum_{q_i=1}^{Q_i} \sum_{q_j=1}^{Q_j} \sum_{\ell=0}^{(r_i+q_i-1) \vee (r_j+q_j-1)} \phi(\ell \mid \mathbf{a}_{i \cap j}, \lambda_{i \cap j} l_i)$$
$$\times \bar{\Phi}((r_i + q_i - 1 - \ell \mid \mathbf{a}_i^{-j}, \lambda_i^{-j} l_i)$$
$$\times \bar{\Phi}(r_j + q_j - 1 - \ell \mid \tilde{\mathbf{a}}_j, \lambda_j (l_j - l_i) + \lambda_j^{-i} l_i)$$
$$+ \bar{\Phi}((r_i + Q_i - 1) \vee (r_j + Q_j - 1) \mid \mathbf{a}_{i \cap j}, \lambda_{i \cap j} l_i).$$

The result in Theorem 5 is based on that each class k order is for a random number of class k product, and each product k contains exactly one item of component i, $i \in \mathcal{D}_k$. We remark that a similar result can be obtained when each class

of order contains multiple and varying number of items for components in $\mathcal{D}_k$. Notation will be more involved in that case but the type of result is similar to what is reported in Theorem 5.

4. NUMERICAL STUDIES

We present three sets of numerical examples to test the effectiveness of the Stein–Chen approximation and its error bounds. As will be seen, the Stein–Chen approximation performs very well in general.

If we rewrite (2) as

$$F_{SC}^L = e^{-\Lambda_k} - (b_1^k + b_2^k)(1 - e^{-\Lambda_k})/\Lambda_k \le F^k \le e^{-\Lambda_k}$$
$$+ (b_1^k + b_2^k)(1 - e^{-\Lambda_k})/\Lambda_k = F_{SC}^U,$$

we then obtain, as by-products, a lower bound and an upper bound for fill rate F^k, F_{SC}^L, and F_{SC}^U, respectively, as defined above. We remark that these bounds are usually not very tight, and indeed, the main intention of the Stein–Chen approximation is to provide an approximation, and the error bound is a somewhat loose theoretical guarantee of the approximation. Our numerical studies show that F_{SC}^L is not as good as the lower bound (10); the upper bound F_{SC}^U, on the other hand, often performs better than the upper bound of Song [12], especially when the component fill rates are high.

The first example is taken from Song [12], which is on a personal computer ATO system. There are six types of components in the system: (1) interior tape backup; (2) standard hard drive; (3) high profile hard drive; (4) video memory card; (5) Pentium processor; (6) Pentium-Pro processor. And there are six classes of orders: $\{2, 5\}$, $\{3, 5\}$, $\{1, 2, 5\}$, $\{1, 3, 6\}$, $\{1, 3, 4, 5\}$, $\{1, 3, 4, 6\}$. The lead time is constant and $L = \{1, 1, 1, 1, 2, 2\}$, and the arrival rates are

$$\lambda^{25} = 0.1\lambda, \quad \lambda^{35} = 0.4\lambda, \quad \lambda^{125} = 0.15\lambda, \quad \lambda^{136} = 0.1\lambda,$$
$$\lambda^{1345} = 0.2\lambda, \quad \text{and} \quad \lambda^{1346} = 0.05\lambda.$$

In our numerical computations, we will use different values of λ.

The first seven columns of Table 1 are the total arrival rate λ and the base-stock levels for controlling the six components. Among the other columns, F_{SC} is the Stein–Chen approximation of the aggregate fill rate, computed using (1), F is the exact order fill rate, F^L and F^U are the lower and upper bounds of Song [12], F_{DSX}^L is the setwise-bonferroni combination lower bounds reported in Dayanik et al. [5], and $\%Err$ is the relative error of the Stein–Chen approximation, defined by

$$\%Err = 100 \times \frac{|F_{SC} - F|}{F}\%.$$

Table 1.   Personal computer example.

λ	S_1	S_2	S_3	S_4	S_5	S_6	F^L	F_{DSX}^L	F_{SC}^L	F_{SC}	F	F_{SC}^U	F^U	$\%Err$
4	5	2	7	2	17	3	0.809	0.817	0.717	0.823	0.830	0.929	0.842	0.84
4	6	3	9	3	20	3	0.932	0.933	0.913	0.934	0.938	0.955	0.943	0.43
4	7	3	10	3	23	4	0.952	0.953	0.946	0.953	0.955	0.960	0.960	0.21
4	7	4	10	4	23	5	0.986	0.986	0.984	0.986	0.986	0.988	0.989	0.14
8	6	3	9	3	20	3	0.595	0.622	0.149	0.631	0.660	1.000	0.728	4.39
8	8	4	12	4	27	4	0.862	0.867	0.792	0.869	0.876	0.946	0.894	0.80
8	10	5	15	5	34	6	0.964	0.964	0.957	0.964	0.968	0.971	0.970	0.41
8	10	6	15	6	34	7	0.985	0.985	0.981	0.985	0.986	0.989	0.990	0.17
16	12	6	18	6	40	7	0.774	0.788	0.602	0.786	0.807	0.970	0.848	2.60
16	14	7	21	7	47	8	0.904	0.909	0.867	0.908	0.913	0.949	0.929	0.55
16	16	8	24	8	54	9	0.961	0.954	0.954	0.962	0.960	0.970	0.968	0.21
16	16	10	24	10	54	11	0.989	0.989	0.986	0.989	0.989	0.992	0.994	0.12

We observe from Table 1 that the approximation performs very well in general. The highest percentage of error is 4.39%.

Next, we conduct a numerical study for a more complex scenario. Consider an ATO system which contains 12 components and all possible combinations of order classes. In other words, any subset of $\{1, 2, \ldots, 12\}$ is an order class. There are a total of $2^{12} - 1$ order classes, and the computation of the exact fill rates becomes impossible. We randomly generate a set of data and test the performance of the approximation result. Specifically, we randomly generate arrival rate of orders from a uniform distribution on $[0, 0.1]$; the deterministic leadtimes are randomly generated from a uniform distribution on $[0, 2]$; the component base-stock level for component i, S_i, is determined using a guaranteed component-based fill rate, that is, S_i is set to guarantee that the component-based fill rate is α (here, we set α to 95, 97.5, and 99%, respectively). Note that due to problem complexity, the order fill rates are not possible to compute exactly, hence we define $\%Err_B$ as the ratio of the error bound with the Stein–Chen approximation of fill rate, again measured in percentage, as

$$\%Err_B = 100 \times \sum_{k=1}^{2^{12}-1} \frac{\lambda^k (b_1^k + b_2^k)(1 - e^{-\Lambda_k})}{\lambda \Lambda_k F_{\mathrm{SC}}^k} \%.$$

The results of this numerical study are rather consistent and here we only report the results for three random instances, in Table 2.

As mentioned earlier, in our numerical examples F_{SC}^L is not as good as the lower bound (10), but F_{SC}^U often performs better than the upper bound of Song [12], especially when the component fill rates are high. For instance, for the ATO system above with component-based fill rates $\alpha = 95\%$, Song's upper bounds for the fill rates are better than F_{SC}^U, but when $\alpha = 97.5\%$ or 99%, F_{SC}^U is better than Song's upper bounds. Indeed, in all our numerical examples, we noted that the larger α and the correlation among order classes, the better the upper bounds F_{SC}^U for the order fill rates. We also

compared the Stein–Chen approximation with the approximation of Song [12], which is the average of the upper and lower bounds developed in Song [12], and we found that the Stein–Chen approximation performs very well in general, and it outperforms Song's approximation when the demand processes of the components are highly correlated, and/or when the component fill rates are high.

Finally, we present an example to demonstrate the effects of variability of leadtimes on the Stein–Chen approximation and its error bound. To do this, we consider three ATO systems which differ only in leadtime distributions. The ATO system consists of 12 components and all possible classes of orders. As in the previous set of numerical examples, all order classes have the same arrival rate 0.1, and the component base-stock levels S_i are set based on component-based fill rate $\alpha = 99\%$. Let L_i^j denote the leadtime for replenishing component i in system j. For $i \in \{1, 2, \ldots, 12\}$, $L_i^1 = 1$, $L_i^2 \sim U[0.7, 1.3]$, $L_i^2 \sim U[0.2, 1.8]$, where $U[a, b]$ is the uniform random variable on $[a, b]$. It is obvious that $L_i^1 \leq_{cx} L_i^2 \leq_{cx} L_i^3$. Table 3, Figs. 2, and 3 reveal the impact of leadtime variability on the errors of the Stein–Chen approximation. In Table 3, F_{SC} is the Stein–Chen approximation of fill rate, F_U^i and F_L^i, $i = 1, 2, 3$, represent the upper and lower bounds F_{SC}^U and F_{SC}^L of the fill rates for three different component leadtimes, respectively,

Table 2.   ATO with randomly generated data.

No.	α	F^L	F_{SC}^L	F_{SC}	F_{SC}^U	F^U	$\%Err_B$
	95%	0.7604	0.5619	0.7651	0.9684	0.9130	26.57
1	97.5%	0.8729	0.8035	0.8742	0.9450	0.9548	8.10
	99%	0.9476	0.9305	0.9478	0.9651	0.9813	1.80
	95%	0.7649	0.5876	0.7695	0.9513	0.9123	23.63
2	97.5%	0.8664	0.7996	0.8679	0.9361	0.9518	7.86
	99%	0.9470	0.9320	0.9473	0.9626	0.9815	1.62
	95%	0.7566	0.5554	0.7615	0.9675	0.9131	27.05
3	97.5%	0.8707	0.8006	0.8721	0.9435	0.9541	8.19
	99%	0.9480	0.9315	0.9482	0.9650	0.9820	1.77

Table 3. Comparison of ATO systems with different leadtimes.

	D_3	D_4	D_5	D_6	D_7	D_8	D_9
F_{SC}	0.978	0.971	0.964	0.957	0.950	0.943	0.936
F_U^1	0.983	0.980	0.977	0.976	0.975	0.976	0.978
F_U^2	0.982	0.978	0.975	0.973	0.971	0.970	0.970
F_U^3	0.981	0.976	0.972	0.968	0.965	0.963	0.961
F_L^1	0.973	0.962	0.950	0.938	0.924	0.909	0.894
F_L^2	0.974	0.964	0.953	0.941	0.928	0.915	0.901
F_L^3	0.975	0.966	0.956	0.945	0.934	0.922	0.910
$\%Err_B^1$	0.51	0.93	1.35	1.99	2.63	3.50	4.49
$\%Err_B^2$	0.41	0.72	1.14	1.67	2.21	2.86	3.63
$\%Err_B^3$	0.31	0.51	0.83	1.15	1.58	2.12	2.67

and $\%Err_B^j$ is the ratio of error bound for system j and F_{SC}, measured in percentages. In Figure 2, we also display the ratio of the error bound and the Stein–Chen approximation of fill rate, for order classes D_1 to D_9, measured in percentage. In Figure 3, we display both the upper and lower bounds F_{SC}^L and F_{SC}^U for three different component leadtimes. The curve in the middle of Figure 3 is the result from the Stein–Chen approximation.

It is interesting to note that leadtime variability has positive effect on the error bounds of the Stein–Chen approximation. In other words, when the mean leadtimes are the same, the larger the variances of leadtimes, the smaller the error of the approximations.

5. CONCLUDING REMARKS

In this article, we analyze the order-based fill rates of a multiproduct and multicomponent ATO system with stochastic leadtimes. We develop the Stein–Chen Poisson approximation for order fill rates as well as its error bound. Our approximation and bounds have an interesting property that the larger the leadtime variability, the smaller the error bound of the
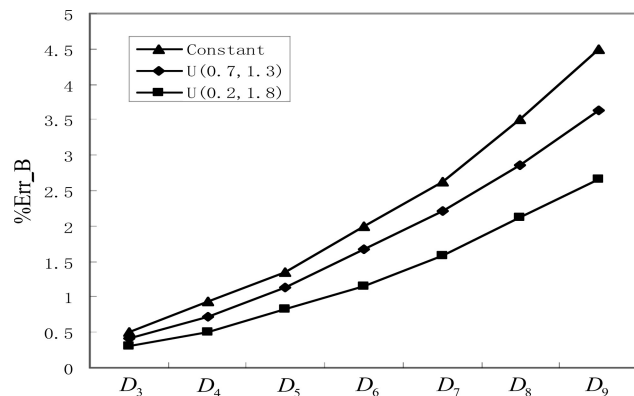
Figure 2. Effect of leadtime variability on approximation error bounds.
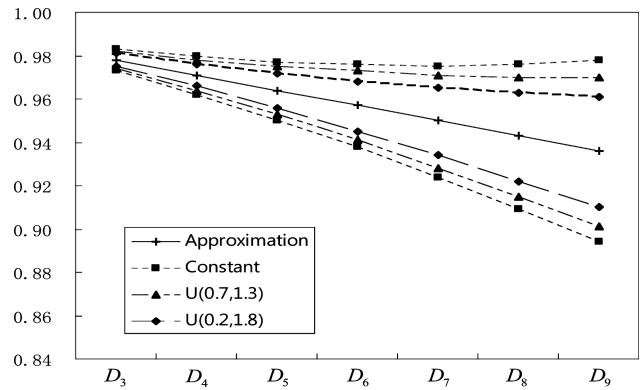
Figure 3. Effect of leadtime variability on the bounds of order fill rate.

approximation. Furthermore, our numerical studies suggest that the Stein–Chen approximation performs the best when the component-based fill rates are high, and that the accuracy of the approximation tends to improve as the leadtime variability increases. The Stein–Chen approximation presents an explicit and extremely simple expression for the order-based fill rate in terms of component-based fill rates. This expression allows the operations manager to analyze the benefits of increasing the component base-stock levels on the fill rates of each class of orders.

Some of the results in Section 3 can be further extended to stochastic component leadtimes, as long as the leadtimes for each component are sequential (or noncrossing). Stochastic sequential leadtimes have nice properties and they have been explored in Zhao and Simchi-Levi [20] for ATO systems. Suppose the leadtime of component i has cumulative distribution function $G_i(\cdot)$. Then, for the first model in Section 3 on order fill rate within a given time window, we can show that when the component leadtimes are stochastically sequential, then the Stein–Chen approximation for fill rate of class k orders, within a give time window τ, is given by $e^{-\Lambda_k}$, where $\Lambda_k = \sum_{i \in \mathcal{D}_k} p_i$ and p_i is computed as, borrowing some notations from Zhao and Simchi-Levi [20],

$$
\begin{aligned}
p_i &= 1 - f_i = 1 - P\{T(S_i) \geq (L_i - \tau)^+\} \\
&= 1 - \int_0^\tau P\{T(S_i) \geq 0\} dG_i(t) \\
&\quad + \int_\tau^{+\infty} P\{T(S_i) \geq t - \tau\} dG_i(t) \\
&= 1 - G_i(\tau) \\
&\quad - \int_\tau^{+\infty} \left(1 - \sum_{n=0}^{S_i-1} \frac{1}{n!} e^{-\lambda_i(t-\tau)} \lambda_i^n (t - \tau)^n \right) dG_i(t) \\
&= \sum_{n=0}^{S_i-1} \frac{\lambda_i^n}{n!} \int_\tau^{+\infty} e^{-\lambda_i(t-\tau)} (t - \tau)^n dG_i(t).
\end{aligned}
$$

Unfortunately, we cannot obtain an exact formula for p_{ij}, which means that we do not have the error bound for the Stein–Chen approximation of order fill rates for this case.

If the component leadtimes are stochastically sequential, we can also obtain the Stein–Chen approximation result for the case when component i is controlled by (r_i, Q_i) policy. To achieve that, we will need Proposition 5.2 of Zhao and Simchi-Levi [20] and a result that the uniform distribution of component inventory position extends to the case of component stochastic sequential leadtimes. In that case, the Stein–Chen approximation for order fill rates are still given by $e^{-\Lambda_k}$ with $\Lambda_k = \sum_{i \in \mathcal{D}_k} p_i$, but p_i is now computed as

$$
\begin{aligned}
p_i &= 1 - f_i \\
&= 1 - \frac{1}{Q_i} \sum_{q_i=r_i+1}^{r_i+Q_i} P\{T(q_i) \geq (L_i - \tau)^+\} \\
&= 1 - \frac{1}{Q_i} \sum_{q_i=r_i+1}^{r_i+Q_i} \left\{ \int_0^\tau P\{T(q_i) \geq 0\} \, dG_i(t) \right. \\
&\quad \left. - \int_\tau^{+\infty} Pr\{T(q_i) \geq t - \tau\} dG_i(t) \right\} \\
&= 1 - G_i(\tau) - \frac{1}{Q_i} \sum_{q_i=r_i+1}^{r_i+Q_i} \int_\tau^{+\infty} \\
&\quad \times \left(1 - \sum_{n=0}^{q_i-1} \frac{1}{n!} e^{-\lambda_i(t-\tau)} \lambda_i^n (t-\tau)^n \right) dG_i(t) \\
&= \frac{1}{Q_i} \sum_{q_i=r_i+1}^{r_i+Q_i} \sum_{n=0}^{q_i-1} \frac{\lambda_i^n}{n!} \int_\tau^{+\infty} e^{-\lambda_i(t-\tau)} (t-\tau)^n dG_i(t).
\end{aligned}
$$

Again, we are unable to obtain the error bound for the Stein–Chen approximation, as p_{ij} cannot be computed explicitly in this case.

The components suppliers in this article are assumed to have infinity capacities. That is, the order leadtime for a (or a batch of) component is independent of the number of out outstanding orders of this type of component (this is assumed in both stochastic sequential component leadtimes and i.i.d. component leadtimes). The model for the supply process of a component with infinite capacity is $M/G/\infty$ queue. If the supplier of a component has finite capacity, then it cannot be modeled as $M/G/\infty$ queue anymore. One model for the supply process with finite capacity is $M/G/1$ queue (and more generally, $M/G/m$ queue), which assumes that the component is produced one after another, see Glasserman and Wang [6]. Because there is no closed form solution for the stationary queueing length distribution of an $M/G/1$ queue, we cannot extend the result of this article to the finite supply capacity case. However, if the production times are exponentially distributed, we can still obtain the Stein–Chen approximation.

Suppose the production facility of component i is a single server with production times having exponential distribution and mean $E[L_i]$, then the Stein–Chen approximation of the order-based fill rate of class k order is $F_{SC}^k = e^{-\Lambda_k}$, with Λ_k given by

$$
\Lambda_k = \sum_{i \in \mathcal{D}_k} (\lambda_i E[L_i])^{S_i}.
$$

In this case, however, we will not be able to provide the error bound as in Theorem 1. This is because, when supply processes have finite capacity, the computation of p_{ij} relies on the inventory processes of components i and j, that are equivalent to two queues with a common Poisson arrival process in addition to their own arrival processes. The analysis of these two queues is a notoriously difficult problem in applied probability; and their joint queue length processes do not have an analytically tractable solution. The corresponding queueing problem is known as the Flatto-Hahn-Wright model (see e.g., Chapter 16 of Ref. 11, or 4), and only approximations are available for its analysis. As the p_{ij} values are not known in exact form, we cannot obtain the error bound for the Stein–Chen approximation of order fill rates for this case.

APPENDIX: PROOFS

In this appendix, we provide all the proofs of the technical results.

The Stein–Chen method bounds the total variation distance between the distribution of the sum of a set of Bernoulli random variables X_i and the distribution of a Poisson random variable with parameter $\lambda = \sum_i E[X_i]$. For two random variables X and Y, the total variation distance between X and Y is defined as

$$
d(X, Y) = \sup_A | P\{X \in A\} - P\{Y \in A\} |.
$$

We briefly introduce the Stein–Chen method. Details can be found, for example, in Arratia, et al. [1]. Let I be a finite or countable index set. For any $i \in I$, let X_i be a Bernoulli random variable. For each $i \in I$, B_i is called a dependent neighborhood of i, if $B_i \subset I$, and for any $i, j \in B_i$, X_i and X_j are dependent. Let b_1 and b_2 be defined by (4) and (5) but with $p_i = P\{X_i = 1\}$ and p_{ij} given by

$$
p_{ij} = E[X_i X_j] = P\{X_i X_j = 1\} = P\{X_i = 1, X_j = 1\},
$$

and let

$$
b_3 = \sum_{i \in I} E[E[X_i - p_i \mid X_j : j \notin B_i]].
$$

The following result can be found, for example, in Arratia et al. [1].

PROPOSITION 1: Let $W = \sum_{i \in I} X_i$ be the number of occurrences of some dependent events, and let Z be a Poisson random variable with mean $E[Z] = E[W] = \lambda < \infty$. Then

$$
d(W, Z) \leq 2 \left[(b_1 + b_2) \frac{1 - e^{-\lambda}}{\lambda} + b_3 (1 \wedge 1.4 \lambda^{-0.5}) \right], \quad (11)
$$

and

$$| P\{W = 0\} - e^{-\lambda} | \le (b_1 + b_2 + b_3) \frac{1 - e^{-\lambda}}{\lambda}. \qquad (12)$$

REMARK 1: The Stein–Chen method provides an approximation and an error bound for the total variation distance between W and a Poisson random variable Z with the same mean. For the fill rate application, we only need an approximation for $P\{W = 0\}$ and the error bound between $P\{W = 0\}$ and $P\{Z = 0\}$, that is (12). However, the first result, (11), also offers useful insights; it provides an understanding on the probability for having any number of components out of stock when a particular class of order arrives.

PROOF OF THEOREM 1. To apply Proposition 1 to the multicomponent multiproduct ATO system, we need to construct a set of dependent Bernoulli random variables. Fix k and we consider class k order. Without loss of generality, suppose $\mathscr{D}_k = \{1, 2, \ldots, K\}$, where K is the number of components to assemble a class k order. We set $I = \mathscr{D}_k$, and define

$$X_i = \begin{cases} 0, & \text{if component } i \text{ is available when needed,} \\ 1, & \text{otherwise.} \end{cases}$$

The set of random variables, $\{X_i\}_{i=1}^K$, are dependent Bernoulli. Let

$$W^k = \sum_{i=1}^K X_i.$$

Then, we have

$$F^k = P\{W^k = 0\}.$$

By Stein–Chen method, F^k is approximated by $e^{-\Lambda_k}$, where $\Lambda_k = E[\sum_{i=1}^K X_i] = \sum_{i=1}^K p_i$ and $p_i = P\{X_i = 1\}$.

By Proposition 1, an error bound for the approximation is

$$|F^k - e^{-\Lambda_k}| \le (b_1 + b_2 + b_3)(1 - e^{-\Lambda_k})/\Lambda_k. \qquad (13)$$

In particular, when $B_i = I$, then $b_3 = 0$, and (13) is reduced to

$$e^{-\Lambda_k} - (b_1 + b_2)(1 - e^{-\Lambda_k})/\Lambda_k \le F^k$$
$$\le e^{-\Lambda_k} + (b_1 + b_2)(1 - e^{-\Lambda_k})/\Lambda_k,$$

where b_1 and b_2 are as given in (4) and (5), with $p_{ij} = E[X_i X_j] = P\{X_i = 1, X_j = 1\}$.

For any $i \in \mathscr{D}_k$, p_i is the probability that a class k order arrives during a backorder period of component i, which has been modeled as an $M/G/\infty$ queue. In particular, $IN_i = S_i - IO_i$ with IO_i having probability mass function $\phi(\cdot \mid \lambda_i L_i)$. Hence, it follows from Poisson arrivals see time averages (PASTA, see Ref. 17) that

$$p_i = Pr\{IN_i < 1\} = Pr\{IO_i > S_i - 1\} = 1 - \Phi(S_i - 1 \mid \lambda_i L_i)$$
$$= \bar{\Phi}(S_i - 1 \mid \lambda_i L_i).$$

It remains to compute p_{ij}. Note that for $i, j \in \mathscr{D}_k$ and $i \ne j$, we have

$$p_{ij} = P\{X_i = 1, X_j = 1\}$$
$$= P\{IN_i < 1, IN_j < 1\}$$
$$= P\{IO_i > S_i - 1, IO_j > S_j - 1\}. \qquad (14)$$

We derive the joint pmf of steady-state inventory on order (IO_i, IO_j) for arbitrary components i and j. It can be seen that (IO_i, IO_j) has a bivariate Poisson distribution [2], thus p_{ij} can be computed by conditioning on the common term in the order. Here, we follow Song and Yao [14] to compute p_{ij}.

For any i, j with $i \ne j$, denote

$$F_{i,j}(h, l) = P\{IO_i > h, IO_j > l\}.$$

Then, (14) can be recast as

$$p_{ij} = F_{i,j}(S_i - 1, S_j - 1).$$

For $i \ne j$ it holds that

$$IO_i = D_i^{-j} + D_i^{i,j} + +D^{i,j},$$

and

$$IO_j = D_j^{-i} + D_j^{i,j} + D^{i,j},$$

where D_i^{-j} is the steady state number of outstanding orders of component i that were triggered by order classes in $S(i) \setminus S(j)$, $D_i^{i,j}$ is the steady state number of outstanding orders of component i that were triggered by order classes $S(i) \cap S(j)$ but the corresponding orders in component j have been delivered, while $D^{i,j}$ is the number of outstanding orders that were triggered by orders classes in $S(i) \cap S(j)$ and that are outstanding for both components i and j. The other notations D_j^{-i} and $D_j^{i,j}$ are defined similarly.

Using the properties of Poisson processes, the dependency between IO_i and IO_j only relies on $D^{i,j}$. By Proposition 1 in Song and Yao [14], we have

$$P\{D_i^{i,j} = q, D_j^{i,j} = r, D^{i,j} = s\}$$
$$= \frac{(\lambda^{i \cap j} \theta_{ij})^s (\lambda^{i \cap j} \vartheta_{ij})^q (\lambda^{i \cap j} \delta_{ij})^r}{s! q! r!} e^{-\lambda^{i \cap j}(\theta_{ij} + \vartheta_{ij} + \delta_{ij})}$$
$$= \phi(q \mid \lambda^{i \cap j} \vartheta_{ij}) \phi(r \mid \lambda^{i \cap j} \delta_{ij}) \phi(s \mid \lambda^{i \cap j} \theta_{ij}),$$

where θ_{ij}, ϑ_{ij}, and δ_{ij} are given by (6), (7), and (8). Thus, we have

$$F_{i,j}(h, l) = P\{IO_i > h, IO_j > l\}$$
$$= P\{D_i^{-j} + D_i^{i,j} + D^{i,j} > h, D_j^{-i} + D_j^{i,j} + D^{i,j} > l\}$$
$$= \sum_{\ell=0}^{\infty} \phi(\ell \mid \lambda_{i \cap j} \theta_{ij})$$
$$\times P\{D_i^{-j} + D_i^{i,j} > h - \ell, D_j^{-i} + D_j^{i,j} > l - \ell\}$$
$$= \sum_{s=0}^{\infty} \phi(\ell \mid \lambda_{i \cap j} \theta_{ij}) \overline{\Phi}(\ell \mid \lambda_i^{-j} L_i + \lambda_{i \cap j} \vartheta_{ij})$$
$$\times \overline{\Phi}(l - \ell \mid \lambda_j^{-i} L_j + \lambda_{i \cap j} \delta_{ij})$$
$$= \sum_{\ell=0}^{h \vee l} \phi(\ell \mid \lambda_{i \cap j} \theta_{ij}) \overline{\Phi}(h - \ell \mid \lambda_i^{-j} L_i + \lambda_{i \cap j} \vartheta_{ij})$$
$$\times \overline{\Phi}(l - \ell \mid \lambda_j^{-i} L_j + \lambda_{i \cap j} \delta_{ij})$$
$$+ \overline{\Phi}(h \vee l \mid \lambda_{i \cap j} \theta_{ij}).$$

This completes the proof of Theorem 1. $\qquad \square$

If the component leadtimes are deterministic, then IO_i and IO_j can be expressed as, if we assume $l_i \leq l_j$,

$$IO_i = D_i^{-j} + D^{i,j}, \tag{15}$$

$$IO_j = D_j(l_j - l_i) + D_j^{-i} + D^{i,j}, \tag{16}$$

where $D^{i,j}$ is the number of orders of classes in $S(i) \setminus S(j)$ during leadtime L_i, D_i^{-j} is the number of orders of classes in $S(i) \setminus S(j)$ during leadtime L_i, D_j^{-i} is the number of orders of classes in $S(j) \setminus S(i)$ during leadtime L_j, and $D_j(l_j - l_i)$ is the number of orders of classes in S_j during time interval $l_j - l_i$, all independent Poisson random variables. Thus,

$$
\begin{aligned}
F_{i,j}(h,l) &= P\{IO_i > h, IO_j > l\} \\
&= P\{D_i^{-j} + D^{i,j} > h, D_j(l_j - l_i) + D_j^{-i} + D^{i,j} > l\} \\
&= \sum_{\ell=0}^{\infty} \phi(\ell \mid \lambda_{i \cap j} l_i) P\{D_i^{-j} > h - \ell\} \\
&\quad \times P\{D_j(l_j - l_i) + D_j^{-i} > l - \ell\} \\
&= \sum_{\ell=0}^{\infty} \phi(\ell \mid \lambda_{i \cap j} l_i) \\
&\quad \times \bar{\Phi}(h - \ell \mid \lambda_i^{-j} l_i) \bar{\Phi}(l - \ell \mid \lambda_j(l_j - l_i) + \lambda_j^{-i} l_j) \\
&= \sum_{\ell=0}^{h \vee l} \phi(\ell \mid \lambda_{i \cap j} l_i) \\
&\quad \times \bar{\Phi}(h - \ell \mid \lambda_i^{-j} l_i) \bar{\Phi}(l - \ell \mid \lambda_j(l_j - l_i) + \lambda_j^{-i} l_j) \\
&\quad + \bar{\Phi}(h \vee l \mid \lambda_{i \cap j} l_i). \tag{17}
\end{aligned}
$$

PROOF OF THEOREM 2. We next prove Theorem 2. Because $|F^k - e^{-\Lambda_k}| \leq (b_1 + b_2)(1 - e^{-\Lambda_k})/\Lambda_k$, and b_1 and Λ_k only depend on the mean of leadtimes, they are the same for the two systems. Now we compare b_2 with $\tilde{b}_2$. Let IO_i and $\tilde{IO}_i$ denote the steady state inventory on order of component i for the two systems with leadtimes L_i and $\tilde{L}_i$, respectively. It follows from Proposition 2 of Song and Yao [14], that for all $i \neq j$ and $h > 0, l > 0$,

$$P\{IO_i \geq h, IO_j \geq l\} \leq Pr\{\tilde{IO}_i \geq h, \tilde{IO}_j \geq l\}.$$

This shows that

$$p_{ij} = Pr\{IO_i \geq S_i, IO_j \geq S_j\} \leq Pr\{\tilde{IO}_i \geq S_i, \tilde{IO}_j \geq S_j\} = \tilde{p}_{ij}.$$

Consequently, we have $b_2 \leq \tilde{b}_2$, and the proof of Theorem 2 is complete. ☐

PROOF OF THEOREM 3. This result follows from Theorem 1 above and Proposition 1.1 of Song [12], which states that when the leadtimes are deterministic, the order-based fill rate with a time window x can be transformed into another order-based fill rate problem with a time window 0. ☐

PROOF OF THEOREM 4. According to Song [12], we know that for $q_i = 1, 2, \ldots, Q_i$,

$$P\{IP_i = r_i + q_i\} = \frac{1}{Q_i},$$

and $IP_1(t), IP_2(t), \ldots, IP_N(t)$ are independent of one another. It is easily shown that $IP_1(t_1), IP_2(t_2), \ldots, IP_N(t_N)$ are also independent. We illustrate this using $N = 2$. Suppose $(IP_1(t), IP_2(t))$ has reached steady state.

Let $t_1 \leq t_2$, then $IP_1(t_1)$ and $IP_2(t_1)$ are independent and uniformly distributed. Given $IP(t_1)$ and $IP_2(t_1)$, $IP_2(t_2)$ only depends on $IP_2(t_1)$ and some other events that are independent of $IP_1(t_1)$, thus $IP_2(t_2)$ is also independent of $IP_1(t_1)$. Because it is known that $IP_2(t_2)$ is uniformly distributed, this shows that $IP_1(t_1)$ and $IP_2(t_2)$ are independent and uniformly distributed random variables.

Using the fundamental identity of inventory, we have $IN_i = IP_i - IO_i$, where IO_i is the leadtime demand for class i item and it has probability mass function $\phi(\cdot \mid \lambda_i l_i)$. Hence, it follows PASTA, and by conditioning on the inventory position IP_i, we obtain

$$
\begin{aligned}
p_i &= P\{IN_i < 1\} \\
&= \sum_{q_i=1}^{Q_i} P\{IO_i > r_i + q_i - 1\} P\{IP_i = r_i + q_i\} \\
&= \frac{1}{Q_i} \sum_{q_i=1}^{Q_i} (1 - \Phi(r_i + q_i - 1 \mid \lambda_i l_i)) \\
&= \frac{1}{Q_i} \sum_{q_i=1}^{Q_i} \bar{\Phi}(r_i + q_i - 1 \mid \lambda_i l_i).
\end{aligned}
$$

It remains to compute p_{ij}. For $i, j \in \mathscr{D}_k$ and $i \neq j$. We first observe that, the inventory positions $IP_i(t - l_i)$ and $IP_j(t - l_j)$ are independent uniform random variables. Thus, by conditioning on IP_i and IP_j, at times $t - l_i$ and $t - l_j$, respectively, we obtain

$$
\begin{aligned}
p_{ij} &= P\{X_i = 1, X_j = 1\} \\
&= P\{IN_i < 1, IN_j < 1\} \\
&= \sum_{q_i=1}^{Q_i} \sum_{q_j=1}^{Q_j} P\{IO_i > r_i + q_i - 1, IO_j > r_j + q_j - 1\} \\
&\quad \times P\{IP_i = r_i + q_i, IP_j = r_j + q_j\} \\
&= \frac{1}{Q_i Q_j} \sum_{q_i=1}^{Q_i} \sum_{q_j=1}^{Q_j} F_{i,j}(r_i + q_i - 1, r_j + q_j - 1),
\end{aligned}
$$

where $F_{i,j}(h,l)$ is computed the same as before, and it is given by (17). ☐

PROOF OF THEOREM 5. The proof of Theorem 5 is similar to that of Theorem 4, except that $D_i^{-j}, D^{i,j}, D_j^{-i}$, and D_j are now independent compound Poisson random variables instead of Poisson random variables. Therefore, in computing the joint distribution of IO_i and IO_j, we need to condition on the pdf of $D^{i,j}$ which is compound Poisson. The rest of the proof is similar to that of Theorem 4 and the analysis following the proof of Theorem 1 on deterministic component leadtimes. ☐

ACKNOWLEDGMENTS

REFERENCES

[1] R. Arratia, L. Goldstein, and L. Gordon, Poisson approximation and the Chen-Stein method, Stat Sci 5 (1990), 403–434.

[2] R. Barlow and F. Proschan, Statistical theory of reliability and life testing, Holt Rinehart and Winston, New York, 1975.

[3] S. Benjaafar and M. Elhafsi, Production and inventory control of a single product assemble-to-order systems with multiple customer classes, Manag Sci 52 (2006), 1896–1912.

[4] X. Chao, M. Miyazawa, and M. Pinedo, Queueing networks: customers, signals, and product form solutions, John Wiley & Sons, 1999.

[5] S. Dayanik, J.S. Song, and S.H. Xu, The effectiveness of several performance bounds for capacitated production, partial-order-service assemble-to-order systems, Manuf Serv Oper Manag 5, 230–251.

[6] P. Glasserman and Y. Wang, Leadtime-inventory tradeoffs in assemble-to-order systems, Oper Res 46 (1998), 858–871.

[7] Y. Lu and J.S. Song, Order-based optimization in assemble-to-order systems, Oper Res 53 (2005), 151–169.

[8] Y. Lu, J.S. Song, and Y. Zhao, No-holdback allocation rules for continuous-time assemble-to-order systems, Oper Res 58 (2008), 691–705.

[9] S.M. Ross and E.A. Pekoz, A second course in probability, Publisher: www.ProbabilityBookstore.com, Boston, MA, 2007.

[10] M. Shaked and J.G. Shanthikumar, Stochastic orders and their applications, Academic Press, New York, 1994.

[11] A. Shwartz and A. Weiss, Large deviations for performance analysis, Chapman & Hall, New York, 1995.

[12] J.S. Song, On the order fill rate in a multi-item, base-stock inventory system, Oper Res 46 (1998), 831–845.

[13] J.S. Song, A note on assemble-to-order systems with batch ordering, Manag Sci 46 (2000), 739–743.

[14] J.S. Song and D.D. Yao, Performance analysis and optimization of assemble-to oreder systems with random lead time, Oper Res 50 (2002), 889–903.

[15] J.S. Song and P. Zipkin, "Supply chain operations: Assemble-to-order systems," Chapter 11: in Handbooks in Operations Research and Management Science, Volume 11, Supply Chain Management, Elsevier, Amsterdam, The Netherlands, 2003, pp. 561–596.

[16] U. Wemmerlov, Assemble-to-order manufacturing: Implications for materials management, J Oper Manag 4 (1984), 347–368.

[17] R. Wolff, Stochastic modeling and the theory of queues, Prentice-Hall, Englewood Cliffs, NJ, 1989.

[18] S.H. Xu, Structural analysis of a queueing system with multi-classes of correlated arrivals and blocking, Oper Res 47 (1999), 264–276.

[19] S.H. Xu and H. Li, Majorization of weighted trees: a new tool to study correlated stochastic systems, Math Oper Res 25 (2000), 298–323.

[20] Y. Zhao and D. Simchi-Levi, Performance analysis and evaluation of assemble-to-order systems with stochastic sequential lead times, Oper Res 54 (2006), 706–724.

[21] P.H. Zipkin, Foundations of inventory management, McGraw Hill, New York, 2000.