Spare parts management: linking distributional assumptions to demand classification

(Accepted for publication in the *European Journal of Operational Research*)

D. Lengu[1], A.A. Syntetos[2] and M.Z. Babai[3]

[1] University of Salford, UK; D.Lengu@salford.ac.uk

[2]Cardiff University, UK; SyntetosA@cardiff.ac.uk

[3]Kedge Business School, France; Mohamed-Zied.Babai@kedgebs.com

---

Spare parts are known to be associated with intermittent demand patterns and such patterns cause considerable problems with regards to forecasting and stock control due to their compound nature that renders the normality assumption invalid. Compound distributions have been used to model intermittent demand patterns; there is however a lack of theoretical analysis and little relevant empirical evidence in support of these distributions. In this paper, we conduct a detailed empirical investigation on the goodness of fit of various compound Poisson distributions and we develop a distribution-based demand classification scheme the validity of which is also assessed in empirical terms. Our empirical investigation provides evidence in support of certain demand distributions and the work described in this paper should facilitate the task of selecting such distributions in a real world spare parts inventory context. An extensive discussion on parameter estimation related difficulties in this area is also provided.

Keywords: Inventory; Demand distributions; Intermittent demand; Spare parts

---

1. Introduction

Many industries rely on the effective management of spare parts, including aerospace and defence, transportation, telecommunications and information technology, utilities and durable goods suppliers. Spare parts are held by firms for internal use in the maintenance of tools and equipment. They are also held by suppliers at the retail or wholesale supply chain level for sale to customers. The costs associated with the inventory management of spare parts can be substantial. According to US Bancorp, spare parts relate to a $700 billion annual expenditure that constitutes about 8 percent of the U.S. gross domestic product (Jasper, 2006).Given the very high level of inventory investments, it is clear that there is significant opportunity for cost-savings through better management.

The demand of spare part items is typically intermittent with demand orders arriving sporadically; the demand can also be highly variable as well as intermittent, in which case it is referred to as lumpy (Boylan and Syntetos, 2008). Kalchschmidtet al. (2006) have also defined lumpy demand as:

- variable, and therefore demand is characterized by fluctuations;
- sporadic, because the demand series are characterized by many periods of very low or no demand; and
- ‘nervous’, reflecting the low auto-correlation of the demand.

The area of inventory management has received a lot of attention in the Operations Research (OR) literature. Conventional inventory control approaches rely on a number of assumptions that are usually valid when demand is fast-moving. Demand over lead time is assumed to be normally distributed and standard forecasting methods are used to estimate the parameters of the normal distribution (see, for example, Strijbosch and Moors, 1996; Porras and Dekker, 2008). However, it has long been shown that such an assumption is invalid in a spare parts context where demand is usually intermittent (Mak and Hung, 1993; Botter and Fortuin, 2000). Moreover, the intermittent nature of the demand makes it very difficult to forecast future requirements with much accuracy (Fortuin and Martin, 1999). This problem is exacerbated when the replenishment lead times are long. Blumenfeldet al. (1999) have demonstrated, amongst others, that the longer the lead times are, the higher the levels of inventory required in order to accommodate the demand uncertainty. Forecasting is an integral part of inventory management systems. However, the challenges in forecasting intermittent demand have implications beyond inventory control; demand forecasts are also used in product development, production and supply chain planning.

Another important issue involved in inventory management is the categorisation of inventory items for the purpose of facilitating forecasting and stock control. When there is a large number of Stock Keeping Units (SKUs), it is not practical to evaluate them on an individual basis. In such cases, the SKUs will typically have to be categorised in order to facilitate decision-making and allow managers to focus their attention on the most important SKUs (however this is judged) (Teunter et al., 2010a). There have been a number of studies in the area of demand classification for inventory items with intermittent demand. A review of the studies in this area can be found in a number of papers including Bacchetti and Saccani (2012), Heinecke et al. (2012) and Van Kampen et al. (2012).

The main objective of this study is to advance the current state of knowledge in spare parts management by bringing together the issues of distributional assumptions and SKU classification. These issues will be linked together by using compound distributions to model demand during lead time. A number of authors (including Friend, 1960 and Kemp, 1967) have suggested that compound distributions (compound Poisson distributions in particular) may provide a good fit for the demand distributions of such SKUs. Compound distributions are appealing because their underlying structure is similar to the demand-generating process associated with intermittent demand.

A top down approach will be used in order to identify compound distributions that may accommodate the distributional properties observed among SKUs with intermittent demand. Firstly, we will consider the shapes that frequency distributions of order sizes will usually take in an intermittent demand context. We will then propose a number of probability distributions that could be used to model such order sizes. Finally, we will introduce the assumption that demand orders arrive according to a Poisson process and, by bringing together the proposed order size distributions and the Poisson arrival process, we will obtain compound distributions that may be used to model intermittent demand. As part of this process, we also develop a demand classification scheme. The categorisation[1] in this scheme will be motivated by a conceptual understanding of the distributional properties of the order sizes rather than a theoretically consistent match of every possible SKU in a particular category. This approach is different from the bottom up approaches that have previously been introduced in the area of intermittent demand management (for example by Syntetos et al., 2012). In the latter approaches, goodness-of-fit tests were first carried out for individual SKUs and the results of these tests were used towards the development of a possible classification scheme.

Our study also makes a number of further important contributions in the area including: (i) an empirical analysis in order to assess whether compound distributions provide a good fit for spare part SKUs; (ii) highlighting a number of challenges related to parameter estimation and goodness-of-fit testing in the area of intermittent demand management; (iii) the development of criteria that should be used when selecting distributions for modelling demand; (iv) deriving insights for practitioners and setting an agenda for further research.

[1] The words ‘classification’ and ‘categorisation’ are used interchangeably in this paper.

The remainder of this paper is structured as follows. In the next section, we shall provide a brief overview of the literature on inventory management related issues for SKUs with intermittent demand. Compound distributions that may model the distributional properties associated with intermittent demand are considered in Section 3. In that section, we will also propose a demand classification scheme that categorises SKUs based on the distributional properties of the order sizes. The empirical goodness of fit of the compound Poisson distributions discussed in this paper is then assessed on an extensive dataset of spare parts in Section 4. We will also compare the relative levels of fit achieved by the compound Poisson distributions in the different categories of the proposed scheme; this exercise allows us to assess the empirical validity of the proposed scheme with respect to the selection of demand distributions. The practical and theoretical implications of our study are discussed in Section 5. Finally, in Section 6, we will provide the conclusions of this study and also identify a number of areas of future research.

2. Research background

In the context of intermittent demand, the demand arrival can be reasonably modelled as a Bernoulli process if time is treated as a discrete variable. The Bernoulli process models whether or not an order arrives during any given unit time period. Demand orders arriving during each unit period are 'bucketed' and the aggregate demand over that period is known as the demand size. If demand arrives according to a Bernoulli process, then the inter-demand intervals will follow a geometric distribution. Croston (1972), Janssen et al. (1998), Syntetos et al. (2005), and Teunter et al. (2010b), among others, have modelled the demand arrival process as a Bernoulli one.

If time is treated as a continuous variable, then demand arrival can be modelled as a Poisson process. The Poisson process models the arrival of individual demand orders; the orders are therefore not 'bucketed'. As a result, the Poisson process captures more information about the demand occurrence than the Bernoulli one. Shale et al. (2008) have found that order arrival can be well represented by a Poisson process. Other studies have modelled order arrival as a Poisson process include Axsater (2006), Shale et al. (2005) and Larsen et al. (2008).

If orders arrive according to a Poisson process, then the intervals between order arrivals will have an exponential distribution. In this paper, we will assume that orders arrive according to a Poisson process; furthermore, we will assume that the order sizes (also known as 'transaction sizes') are distributed according to some arbitrary distribution. The distribution of demand during a fixed period of time will then have a compound Poisson distribution. Let us assume that demand has a compound Poisson distribution and let us denote sizes of the orders as X. In addition, let:

λ = the order arrival rate

$\mu = \mathrm{E}(X)$ = the mean of the order sizes

$\sigma^2 = \mathrm{Var}(X)$ = the variance of the order sizes

Y = the demand during a unit period of time.

Then the mean and variance of demand during a unit period of time are given respectively by (Satterthwaite, 1942):

$$E(Y) = \lambda\mu \tag{1}$$

$$Var(Y) = \lambda(\mu^2 + \sigma^2) \tag{2}$$

One of the appealing properties of compound Poisson distributions is that they are Lévy processes and, as such, they are infinitely divisible (Sato, 1999). Furthermore, a linear combination of a finite number of independent Lévy processes is again a Lévy process. The practical implication of this property is that, if the demand over a unit period of time (denoted as Y) is assumed to have a compound Poisson distribution, then the demand over a fixed period of length L (where L is a positive rational number) will also have a compound Poisson distribution.

A number of authors (e.g. Friend, 1960; Croston, 1972) have advocated the use of compound distributions to model intermittent demand patterns. The appeal of compound distributions stems from the fact that they can independently model the constituent elements of demand (order sizes and intervals between order arrivals). Feeney and Sherbrooke (1966) derived a simple analytic solution of the order-up-to level (under a base-stock policy) when demand follows a compound Poisson distribution. Ward (1978) proposed a regression model for calculating the reorder points of lumpy items. Watson (1987) examined the interactions between forecasting and inventory control in such a context. In the last two studies, demand was assumed to arrive according to a Poisson process and the order sizes were assumed to have a geometric distribution. Other researchers that have used a compound Poisson distribution to model intermittent demand include: Adelson (1966); Archibald and Silver (1978); Naddor (1978); Mitchell et al. (1983); Forsberg, (1995); Matheus and Gelders (2000); Hill and Johansen (2006); Zhao (2009); Babai et al. (2011).

There have been only a few empirical studies that consider goodness-of-fit related issues in the area of intermittent demand management. As far as we are aware, the only studies in this area are by Kwan (1991), Eaves (2002) and Syntetos et al. (2012). The negative binomial distribution (i.e. the Poisson-Logarithmic series distribution) performed well in Kwan's study but the sample used in that work was rather small (only 86 SKUs). Eaves (op. cit.) carried out goodness-of-fit tests on a larger set of spare/service parts SKUs (6,795 series). The demand orders were bucketed in monthly periods in that study. The goodness-of-fit tests were carried out separately for demand sizes (the total monthly demand) and the inter-demand intervals (the intervals between months with positive demand, again, measured in months). It is important to note that one cannot assume that a compound distribution will provide high levels of fit simply because the constituent distributions provide high levels of frequency of fit for the corresponding order sizes and inter-demand intervals (Katti and Gurland, 1962). (This issue is further discussed at the end of sub-section 4.2.) The results presented by Eaves (2002) therefore do not contribute much to our study. In this paper, we will take a direct approach. Goodness-of-fit tests will be carried out in order to assess the compound Poisson distributions and not just their constituent parts (order sizes and inter-order intervals).

Syntetos et al. (2012) have also assessed the goodness of fit of a number of distributions, including two of the distributions examined in this study (the Poisson-Geometric and Poisson-Logarithmic series distributions). The goodness-of-fit test used in that study was the Kolmogorov-Smirnov (K-S) test and the empirical database was made up of approximately 13,000 SKUs. They found that the Poisson-Geometric distribution outperformed the Poisson-Logarithmic series distributions. However, in the process of carrying out the tests, Syntetos et al. (op. cit.) derived the number of categories based on the hypothesised distribution and not the empirical one. As a result, the number of categories was too high and, consequently, the critical values were too low and the test was excessively 'liberal' (Syntetos et al., 2013). A goodness-of-fit test is referred to as liberal if the test incorrectly rejects the null hypothesis (that the distribution in question provides good fit) more often than is suggested by the specified significance level.

One topic that has not received sufficient attention in the area of intermittent demand is the classification of SKUs. An extensive review of the classification schemes that have been proposed in literature can be found in a recent paper by van Kampen et al. (2012). One of the main observations that came out of this review was that most of the classification schemes that have been proposed in literature are inspired by the specific context of the relevant studies. The characteristics used to classify the SKUs, the number of classes and the boundaries between the categories are often chosen in order to address the concerns associated with a particular application. It is therefore not always clear whether such schemes would have wider applicability. Moreover, in some of the proposed schemes, the boundaries between different classes may not even have any intrinsic meaning (D'Alessandro and Baveja, 2000).

In this paper, we propose a demand classification scheme that attempts to overcome these shortcomings. A deliberate approach was taken in the development of the scheme to ensure that the SKUs would be classified by general and not context-specific factors. As will be explained in more detail in the next section, the proposed scheme classifies SKUs solely based on the distributional patterns of demand. Contextual factors such as the product, customer or industry characteristics are not considered in the scheme. In addition, the boundaries between the different classes have some meaning in terms of the order size distributions. The only assumption that was made in the development of this scheme is that the demand follows a compound Poisson distribution. The validity of this assumption will be tested in section 4 by carrying out goodness-of-fit tests on empirical data.

3. Order size distributions and a demand classification scheme

In this section, we shall consider a number of distributions that could be used to model the order sizes. The term “order size” refers in this paper to the number of units in a distinct customer order. The term should not be confused with “demand size”, which is the total numbers of units ordered during a given period of time. The distributions used to model order sizes should ideally provide good empirical fit but they should not be computationally demanding for use in practical settings. For SKUs with intermittent demand, the order size frequency distributions observed in practice are usually either monotonically decreasing or more centred (i.e. “mounded”) but with a significant right skew. Boylan (1997) proposed three criteria for assessing the suitability of hypothesised demand distributions (regardless of the context of application): (a) A priori grounds for modelling demand, (b) The flexibility of the distribution to represent different types of demand, (c) Empirical evidence available in support of the distribution. The same criteria were adopted in this paper when selecting order sizes distributions. The three criteria are discussed in more detail below.

The first criterion (a priori grounds for modelling demand) relates to the intuitive appeal that a distribution may (or may not) have for representing the data under consideration. The hypothesised distribution has to match the underlying structure of the order sizes, as understood by inventory managers. By their nature, order sizes are discrete and they have to be greater than zero. These properties would suggest that the hypothesised order size distributions should ideally be discrete distributions that are defined in the positive domain.

Flexibility (the second criterion) refers to robustness in terms of the ability of the distribution to cope with diverse order size profiles. For practical purposes, it would be more convenient to have a manageably small number of distributions that are collectively robust enough to cover a great majority of possible empirical scenarios. The third criterion requires that there should be corroborative empirical evidence, where possible, in support of the selected distributions. Unfortunately, there have only been a few empirical studies on the goodness of fit of distributions for intermittent demand items (specifically, Kwan, 1991; Eaves, 2002; Syntetoset al., 2012). The findings of those studies will be used to inform our selection of the order size distributions.

In this paper, we add a fourth criterion – the selected distribution should have a probability distribution function that is computationally easy to work with in practice. The moments and parameter estimates of the selected distributions should take functional forms that can be computed easily and quickly. The distributions should also have as few parameters as possible (ideally, one or two); otherwise it becomes harder for practitioners to get a good grasp of the relationship between the parameters and the probabilities or any statistics of interest.

The four order size distributions considered in this paper are the Geometric, Logarithmic series, Poisson and Pascal distributions. These distributions were selected for two reasons. Firstly, the four distributions satisfy, for the most part, the four criteria discussed above. All four distributions are discrete, and the Geometric and Logarithmic series distributions take only positive values. The distributions are also flexible in the sense that different levels of skewness may be obtained for all four of them by adjusting the parameters accordingly. As far as we are aware, as yet, there have been no goodness-of-fit studies carried out specifically for order sizes. However, the compound representations associated with the Geometric and Logarithmic series distributions have been found to provide good fit for demand during lead time (Syntetos et al., 2012). The probability functions of all four distributions can be easily computed in practice. All of the distributions have one parameter except of the Pascal distribution which has two parameters.

Secondly, as it will be shown below, the four distributions are also associated with varied and distinct properties that naturally suggest a scheme for classifying SKUs. Table 1 shows the properties of the distributions with respect to modality and variability (as measured by the squared coefficient of variation). The mode ($\tilde{m}(X)$) and the squared coefficient of variation ($CV^2(X)$) are arguably two statistics that may collectively best describe the shape of an order size distribution. The mode will help us determine whether the order size distribution is monotonically decreasing or more 'mounded'. The squared coefficient of variation will give us an idea about the relative spread of the distribution

Table 1. Summary of the properties of the order size distributions

Order size distribution			**Compound Poisson distribution**
Name	$\tilde{m}(X)$	$CV^2(X)$	**Name**
Geometric	$\tilde{m}(X)=1$	$0<CV^2(X)<1$	Poisson-Geometric distribution
Log series	$\tilde{m}(X)=1$	$0<CV^2(X)<\infty$	Poisson–Logarithmic series distribution
Poisson	$\tilde{m}(X)\geq 1$	$0<CV^2(X)\leq 1$	Poisson-Poisson distribution
Pascal	$\tilde{m}(X)\geq 1$	$0<CV^2(X)<\infty$	Poisson-Pascal distribution

For each of the four distributions, the corresponding compound Poisson distribution is given in the final column of Table 1. (The probability mass functions of these compound distributions are presented separately as part of the supplementary material in an electronic companion to this paper). Based on these differences, we developed a classification scheme (illustrated in Figure 1 below) according to which SKUs are categorised based on $\tilde{m}(X)$ and $CV^2(X)$. The distributions were assigned to the various categories sequentially. The definitions of the categories are discussed immediately after Figure 1.

Figure 1. Demand classification based on the properties of the order sizes

Category B Order size mode: $\tilde{m}(X)=1$ Order size variability: $CV^2(X)\geq 1$ Proposed demand distribution: *Poisson-Logarithmic series*	**Category D** Order size mode: $\tilde{m}(X)\geq 2$ Order size variability: $CV^2(X)\geq 1$ Proposed demand distribution: *Poisson-Pascal*
Category A Order size mode: $\tilde{m}(X)=1$ Order size variability: $0<CV^2(X)<1$ Proposed demand distribution: *Poisson-Geometric*	**Category C** Order size mode: $\tilde{m}(X)\geq 2$ Order size variability: $0<CV^2(X)<1$ Proposed demand distribution: *Poisson-Poisson*

a) Category A – Contains all SKUs with $\tilde{m}(X)=1$ and $CV^2(X)<1$. The demand for each of these SKUs may be assumed to follow a Poisson-Geometric distribution (also known as the Pólya-Aeppli or Stuttering Poisson distribution). All four distributions could be used to model the order sizes in this category. However, the Geometric distribution was preferred to all the alternatives because it fully meets the criteria set out above. Unlike the Poisson and Pascal distributions which can take a value of zero, the Geometric distribution only takes strictly positive values. Syntetos et al. (2012) also found that the Poisson-Geometric distribution provided higher levels of frequency of fit than the Poisson-Logarithmic series distribution.

b) Category B – Contains all SKUs with $\tilde{m}(X)=1$ and $CV^2(X)\geq 1$. The demand for each of these SKUs may be assumed to follow a Poisson-Logarithmic series distribution (also known as the Negative Binomial distribution). While the Pascal distribution could also have been used to model the order sizes in this category, the

Logarithmic series distribution was preferred because it has only one parameter (unlike the Pascal distribution which has two) and there is empirical evidence in its support (Kwan, 1991; Syntetos et al., 2012).

c) Category C– Contains all SKUs with $\widetilde{m}(X) \geq 2$ and $CV^2(X) < 1$. The demand for each of these SKUs may be assumed to follow a Poisson-Poisson distribution (also known as the Neyman type A distribution). The Pascal distribution could also be used to model the order sizes in this category but the Poisson distribution was preferred instead because it has only one parameter.

d) Category D– Contains the SKUs that have not been assigned to the three other categories. For each of the SKUs in this category, the demand will be assumed to follow a Poisson-Pascal distribution.

Our selection of order size distribution is based, amongst other things, also on practical convenience. The four distributions considered are examined in most standard introductory textbooks on probability and statistics (e.g. Upton and Cook, 1996; Wackerly et al., 2002). As such, practitioners with little background in statistics can easily find out more about these distributions if they feel the need to do so.

The Pascal distribution will be used to model the order sizes of the SKUs falling in category D because it could (at least in theory) be able to perform as well as any of the other three alternatives. The Pascal distribution provides a good approximation for each of the three other order size distributions. If the Pascal distribution is denoted by $Ne(r,p)$, where r is the number of successes and p is the probability of success, then the Geometric, Logarithmic series and Poisson distributions are all limiting forms of the Pascal distribution given the right choice of the parameter r (Katti and Gurland, 1961). However, the compound distribution associated with the Pascal distribution (i.e. the Poisson-Pascal distribution) is comparatively more demanding in terms of computational effort. The Pascal distribution will therefore only be used when the three other distributions are not appropriate (i.e. in Category D).

It should be noted that the Pascal distribution cannot have, simultaneously, a mode greater than 1 and a squared coefficient of variation also greater than 1. There are however very few distributions that can meet those conditions (among them, the Lognormal, Inverse-Gaussian and the five-parameter Bi-Weibull continuous distributions as well as the Beta-Binomial and the Beta-Negative Binomial discrete distributions). These distributions however pose their own challenges. The continuous distributions are obviously not well suited for modelling discrete order size distributions. As for the discrete distributions, both the Beta-Binomial and the Beta-Negative Binomial distributions have three parameters, one more parameter than the Pascal distribution which has two. As will be demonstrated in sub-section 4.2, parameter estimation become more challenging as the number of distributional parameters increases. These challenges are bound to outweigh any improvements in modelling accuracy that may be obtained by using them in a real world context. To summarise, the Pascal distribution fails to meet simultaneously the modality and variability requirements in category D. However, it represents a compromise that should in theory perform as well as the three other distributions considered without introducing other undesirable challenges associated with distributions that meet these requirements.

Finally, it is worth pointing out that the proposed scheme was developed based on a fixed set of criteria relating to the distributional properties of order sizes. Unlike in other studies (Williams, 1984; Eaves, 2002; Syntetoset al., 2009), the scheme was not developed based on characteristics of a particular sample under concern. The scheme is therefore bound to be more widely applicable. There are also a priori theoretical grounds (stipulated by the criteria) justifying our selection of the proposed order size distributions. The boundaries between the different categories have a clear meaning that follows from the theoretical properties of the proposed order size distributions.

4.Empirical analysis

4.1 Empirical datasets

In this section, we shall carry out an empirical analysis to assess the validity of the theoretical propositions made in Section 3. Goodness-of-fit tests will be carried out to assess whether the compound Poisson distributions proposed in Section 3 provide a good fit for the demand distributions of empirical SKUs. We will also assess the effectiveness of the proposed demand classification scheme. The empirical datasets are made up of individual demand histories of nearly 15,000 spare part SKUs from two different industries. Table 2 below provides a summary description of the datasets.

Table 2. Summary description of the empirical datasets

Dataset	Industry	No. of SKUs	Time buckets	History length (months)
1	Domestic Appliances	14,874	Order level	60
2	Commercial airlines	496	Order level	28

Detailed demand information, at order level, was available for both datasets. Additional statistics on the characteristics of the demand series in each of the datasets are presented in Table 3. This table provides information on the average order arrival rate (λ) which is expressed in terms of the number of orders per month. The average order arrival rate is calculated by dividing the total number of orders over the demand history by the length of the demand history (measured in months). The order sizes of the SKUs were also examined and the table provides information on the mode ($\tilde{m}(X)$) and the squared coefficient of variation ($CV^2(X)$) of the order sizes. The percentages indicated represent the proportions of SKUs (in the corresponding dataset) that fall within each category. For example, 62.13% of the SKUs in dataset 1 have a mode of order sizes equal to 1.

Table 3. Summary statistics of the empirical datasets

λ	**$0.0 < \lambda \leq 0.25$**	**$0.25 < \lambda \leq 0.50$**	**$0.50 < \lambda \leq 1.00$**	**$1.00 < \lambda \leq 5.00$**	**$\lambda < 5.00$**
Dataset 1	*60.21%*	*12.09%*	*9.39%*	*11.95%*	*6.35%*
Dataset 2	*35.28%*	*34.88%*	*18.35%*	*11.29%*	*0.20%*
$\tilde{m}(X)$	**1**	**2-10**	**11-50**	**51-100**	**100+**
Dataset 1	*62.13%*	*34.64%*	*2.80%*	*0.28%*	*0.15%*
Dataset 2	*29.03%*	*60.48%*	*3.83%*	*6.65%*	*0.00%*
$CV^2(X)$	**$0.0 < CV^2 \leq 0.5$**	**$0.5 < CV^2 \leq 1.0$**	**$1.0 < CV^2 \leq 5.0$**	**$5.0 < CV^2 \leq 10.0$**	**$10.0 < CV^2$**
Dataset 1	*62.84%*	*13.57%*	*19.04%*	*2.78%*	*1.77%*
Dataset 2	*31.25%*	*33.27%*	*34.27%*	*1.21%*	*0.00%*

Key: $\tilde{m}(X)$ – The mode of order sizes; $CV^2(X)$ – The squared coefficient of variation of order sizes

The statistics provide some idea about diversity with respect to the distributional properties of the SKUs in the datasets. Most of the SKUs have very low order arrival rates (in both dataset 1 and 2, more than 70% of the SKUs have an average order arrival rate less than or equal to 0.50 orders per month). There is however less diversity in terms of the mode and the variability of the order sizes. Specifically, in the case of dataset 1, only 0.15% of the SKUs have order sizes with a mode greater than 100 and only 1.77% of the SKUs have order sizes with a squared coefficient of variation greater than 10.0. Such values may be attributed to the particular industries examined in our study; replication of our findings in more demand datasets is an avenue for further research and this issue is further discussed in the last section of the paper.

4.2 Goodness-of-fit

The goodness-of-fit test used in this paper is the Kolmogorov-Smirnov test (or K-S test, in short). Other goodness-of-fit tests were also considered. Pearson's χ^2 test is a well-known goodness-of-fit test that places observations in categories and compares the observed and expected frequencies in each of the categories. This test is easy to use but is associated with some requirements/'rules' (given in Cochran, 1952; Birnbaum, 1962; Roscoe and Byars, 1971; Kendallet al., 1987; Cramer, 1999) that specify the minimum and average expected frequencies for the categories. The data in our study mostly failed to meet these requirements. The intermittent nature of our data meant that the demand was zero in most periods and there were very few demand observations. In most of the cases, we could not create more than two viable categories and, as a result, we could not carry out a valid χ^2 test.

The Cramer von Mises and Anderson-Darling goodness-of-fit tests could potentially have also been used. However, whereas the K-S is distribution-free (i.e. the critical values are independent of the hypothesised distribution), the critical values of the Cramer von Mises and Anderson-Darling tests will depend on the hypothesised distribution. As a result, different tables of the critical values must be calculated for each of the proposed distributions. The computational effort involved in deriving the critical values (by using, for example, Monte Carlo methods) would be prohibitive.

The goodness-of-fit test used in this study is the K-S test with the significance level set at 5%. The distribution of the demand per month has been considered rather than the distribution of the lead-time demand; this is due to the lack of information on the actual lead times for the datasets. As was pointed out in section 2, compound Poisson distributions are Lévy processes; if demand arrivals follow a compound Poisson process, then the demand over any fixed period of time (in our case, a month) will also have a compound Poisson distribution. Monthly demand will be considered in the goodness-of-fit tests because the demand for the

SKUs in our sample is highly intermittent. According to the statistics in Table 3, at least 80% of the SKUs in each dataset have an order arrival rate (λ) of one order or less per month. With demand being so highly intermittent, daily or weekly demand figures are likely to be very small and this may present problems later on when the parameters are being estimated. Finally, unlike earlier studies such as Eaves (2002), the goodness-of-fit tests will be carried out on demand and not its constituent parts (order sizes and inter-demand intervals). This direct approach is taken up in this study because it may be incorrect to infer that a compound distribution provides high levels of fit simply because the constituent distributions do so. This discussion will be taken up further at the end of this sub-section after the parameter estimators of the compound Poisson distributions have been introduced.

In the K-S tests, the empirical distribution function (EDF) for each SKU was taken as the cumulative frequency distribution of the demand for the SKU under concern and the fitted distribution was the cumulative distribution function (CDF) of the hypothesised compound Poisson distribution. The parameters of the hypothesised distribution were estimated from the observed demand data using the following two methods:

a) The method of moments, using the first two moments (or MM, in short);
b) The method of mean and zero frequency (or M&Z, in short). With this method, the estimates are derived by equating: (i) the sample mean and the population mean; (ii) the observed and expected probabilities of zero observations. This method has been used in a number of studies including Katti and Gurland (1962), Bowman and Shenton (1967) and Shenton and Bowman (1977).

In the case of the Poisson-Pascal distribution, an additional moment is required under either method in order to obtain the estimate of the third parameter. The formulae for the parameter estimators under each of these methods are given in the *electronic companion*.

In this paper, a parameter estimator is referred to as *domain compliant* if the values of the estimator will always fall within the domain of the relevant parameter. Taking the well-known normal distribution $N\left(\mu,\sigma^2\right)$ with $-\infty < \mu < \infty$ and $0 \leq \sigma^2$ as an example, the sample mean is a domain compliant estimator of the parameter μ since this statistic will always fall within the domain of the parameter. The sample variance is also a domain compliant estimator of the parameter σ^2. For the purposes of this study, the value of the parameter estimate has to fall within the domain of the relevant parameter otherwise it is not possible to obtain a valid fitted CDF.

In the case of the compound Poisson distributions, the parameter estimation methods discussed above are not necessarily associated with domain compliant estimators. The MM estimators will provide meaningful estimates so long as the empirical data satisfies the theoretical relationship between the relevant moments. For the four compound Poisson distributions discussed in this paper, the variance is always greater than or equal to the mean (Keilson and Kubat, 1984; Johnson et al., 2005). The MM estimates will therefore fall outside the domain if the sample variance (denoted as s_y^2) is less than the sample mean (denoted as $\bar{y}$). This can be easily seen in the formulae given in the *electronic companion*. Most of the MM estimators fall outside the relevant domain whenever $s_y^2 < \bar{y}$. The method of mean and zero frequency also fails when there are no periods with zero demand (i.e. when the observed zero frequency, f_0, is equal to 0).

Parameter estimators that are not domain compliant present a practical challenge in goodness-of-fit tests; if any of the derived estimates falls outside the relevant domain, what conclusion do we draw with regards to the goodness-of-fit? For example, if the observed sample variance is less than the sample mean, then this might be a genuine reflection of the fact that the underlying demand distribution is under-dispersed. Alternatively, the underlying demand distribution might actually be over-dispersed, but the observed sample variance might be less than the sample mean simply as a result of sampling error. Without knowing the underlying distribution, it is not possible to know what the right conclusion should be.

Domain incompliance is a challenge not only in K-S goodness-of-fit tests, but also for every procedure that relies on parameter estimation, including parametric inventory management. Stock control parameters such as the reorder point and the order-up-to level are derived based on the distribution of demand during lead-time. If the parameter estimators that are used are domain incompliant, then the parameter estimates obtained may fall outside the relevant domain; such estimates are meaningless and they would not provide us with a valid distribution for the lead time demand.

The problem of domain incompliance is exacerbated in the case of intermittent demand. Intermittent demand is often characterised by having only a small number of demand observations; as the number of demand observations decreases, the standard error of the parameter estimates (and thus, the probability that the estimate will fall outside the domain) will usually increase. It is worth pointing out that domain incompliance will usually not be a problem for fast-moving items; the demand for such items is typically assumed to be normally distributed and as discussed above the parameter estimators for this distribution are domain compliant.

In this study, we have made some restrictions in order to ensure that the parameter estimates do not fall outside the domain. In the case of MM estimates,

a) whenever the sample variance (s_y^2) is less than or equal to the sample mean ($\bar{y}$), the sample variance is increased and made equal to $1.05\bar{y}$. This is similar to the approach adopted by Kwan (2002).
b) Furthermore, in the case of the Poisson-Pascal distribution, the sample variance has to fall within the following interval:

$$\frac{\bar{y}+\sqrt{\bar{y}^2+8m_3\bar{y}}}{4}<s_y^2<\frac{-\bar{y}+\sqrt{\bar{y}^2+4\bar{y}(m_3-\bar{y})}}{2}$$

(where $\bar{y}$, s_y^2 and m_3 are the sample mean, sample variance and the sample central third moment respectively). The derivation of this restriction is given in the *electronic companion*. If the sample variance falls outside this range, then the sample variance is increased (or, as the case may be, decreased) accordingly to move it just inside the interval.

As for the M&Z estimates,

a) the observed zero frequency, f_0, was bound within the range $1 \le f_0 \le N-1$, where N is the length of the demand series. Note that $f_0 = N$ represents the trivial case of a demand series that does not have any periods with positive demand; $f_0 = 0$ leads to computational problems (specifically, taking logarithms of zero).
b) Furthermore, in the case of the Poisson-Pascal distribution, whenever the estimated parameter p is less than 0, we have assumed that $s_y^2 = 1.05 \times -\bar{y}^2 / ln\left(f_0/N\right)$. The derivation of this restriction is also given in the *electronic companion*.

The additional restrictions arise in the case of the Poisson-Pascal distribution (for both estimators) necessarily as a consequence of the fact that the distribution has one more parameter. With these restrictions in place, the parameter estimates will always fall within the relevant domain and the probability distributions obtained from these estimates will be valid. The fitted CDF obtained under these restrictions is simply our best effort to obtain a valid hypothesised distribution that provides a close fit to the empirical data. The CDF fitted in this manner will still fail to provide significant fit if there is little agreement between the empirical data and the hypothesised compound Poisson distribution.

The empirical data used in our study is highly varied as indicated in Table 3 and it is not clear which of the two parameter estimation methods would perform best for our data. Goodness-of-fit tests have therefore been carried out using both methods. The results presented in Table 4 indicate, per dataset and under each parameter estimation method, the percentage of SKUs for which a distribution was found to provide a significant fit. The levels of frequency of fit achieved by the four compound Poisson distributions are quite high (for all four distributions and under each parameter estimation method, the level of frequency of fit was at least 70%). The proposed compound Poisson distributions therefore provided significant fit for most of the empirical demand data used in our study.

Table 4. Goodness-of-fit results for the compound Poisson distributions - % fit

	Po-Geo		**Po-Log**		**Po-Po**		**Po-Pa**	
	MM	M&Z	MM	M&Z	MM	M&Z	MM	M&Z
Dataset 1	*93.85%*	*94.47%*	*92.88%*	*93.77%*	*89.12%*	*91.02%*	*93.37%*	*87.11%*
Dataset 2	*84.07%*	*86.69%*	*82.66%*	*81.05%*	*72.18%*	*77.42%*	*85.08%*	*68.75%*

Key: MM – Method of moments; M&Z – Method of mean and zero frequency

With respect to the parameter estimation methods, a comparison of the levels of frequency of fit achieved reveals that, overall, there is little to choose between the MM and M&Z estimators. MM estimators performed better than M&Z estimators in some cases but they performed worse in others. The choice of parameter estimation method may therefore make a difference.

Finally, it is noteworthy that the MM estimators of the parameters of the compound Poisson distributions may not be the same as the MM estimator of the parameters of the constituent distributions that make up the compound Poisson distributions. Let us consider, for example, the Poisson-Geometric distribution. Suppose the demand series were n periods long and that there were m order arrivals during this period. Let us also denote the orders by $\{x_1,...,x_m\}$.

The MM estimators for the parameters of the geometric distribution $Geo(\theta)$ and the Poisson distribution $Po(\lambda)$ are $\hat{\theta} = 1/\bar{x}$ and $\hat{\lambda} = n/m$ respectively. These estimators are different from the corresponding estimators given for the compound Poisson distribution in equation A.2 in the *electronic companion*. The two sets of estimators may therefore give different parameter estimates. Thus, it would not be correct to infer the goodness-of-fit of a compound-Poisson distribution simply from the goodness-of-fit results of the distributions of the constituent elements of the demand.

4.3Validity of the proposed demand classification scheme

In this sub-section, we will assess the empirical validity of the proposed demand classification scheme. The SKUs in datasets 1 and 2 were first categorised according to the mode and variability of the order sizes as proposed by the relevant scheme. Goodness-of-fit tests were then carried out for each of the four compound Poisson distributions and the K-S statistics were calculated accordingly. Finally, sign tests were carried out - for each pair of distributions and in each category - to test the hypothesis that there is no difference in the goodness of fit achieved by the two distributions. The sign tests were carried out on the difference between the K-S statistics achieved by each pair of distributions calculated as follows:

> *K-S statistic achieved by the distribution given in the corresponding row minus (-) the K-S statistic achieved by the distribution given in the corresponding column.*

Suppose that there was no difference in the goodness of fit achieved by two distributions. In such cases, negative differences would be as likely as positive differences. Thus, if the null hypothesis that there is no difference is correct, then one would expect that roughly half of the differences would be negative differences. Let us denote the number of negative differences, when expressed as a proportion of total number of differences, by the term $\hat{d}$. The null hypothesis will thus be that $\hat{d} = 0.5$ and the alternative hypothesis is that $\hat{d} \neq 0.5$.

Hypothesis testing was done based on confidence intervals, where the confidence intervals for $\hat{d}$ were constructed under the assumption that $\hat{d}$ is normally distributed. Such an assumption is justified given the high number of SKUs in each category (Berry, 1941;

Esseen, 1956). The pair-wise comparisons between the distributions are performed simultaneously; a multiple-comparison correction is therefore required in order ensure that the overall confidence level (in this case, 95%) is maintained. If a multiple-comparison correction was not used, then the Type I error (i.e. the probability of incorrectly rejecting the null hypothesis) could be significantly higher than 0.05. The multiple-comparison correction used in this study was the Bonferroni correction (Benjamini and Hochberg, 1995).

Finally, to allow for easier interpretation, the confidence intervals are expressed in terms of the variable $\hat{\delta} = \hat{d} - 0.5$. This essentially corresponds to a translation of the confidence intervals so that, under this translation, the intervals are centred around the value $\hat{\delta} = 0$. Separate analysis was carried for the two data sets and the two parameter estimation methods and the results are given in Tables 5-8. For each of the datasets, the number of SKUs that fell in a given category is given by *N*. The top table in each category provides the results expressed in terms of confidence intervals. The results can be interpreted as follows:

a) If the lower limit of the confidence interval is positive, then we conclude that the distribution given in the row label outperformed the distribution given in the column label in the given category.
b) Alternatively, if the upper limit of the confidence interval is negative, we conclude that the distribution given in the column label outperformed the distribution given in the row label in the given category.
c) A conclusion that there is no difference between the two distributions is obtained if 0 falls within the confidence interval.

Corresponding to each of these ‘confidence intervals’ tables is an associated ‘Conclusions’ table. This latter table presents the conclusions drawn from the confidence intervals. The conclusions are expressed in terms of inequality signs with $X>Y$ indicating that distribution X outperformed distribution Y and $X<Y$ indicating the opposite. The results that are relevant for the purposes of assessing the validity of the scheme have been underlined. Results highlighted in bold indicate those instances in which the findings do not agree with the suggestions in the classification scheme.

The results given in Tables 5-8 largely agree with the suggestions in the proposed scheme. The Poisson-Geometric and Poisson-Logarithmic series distributions consistently performed as well as, or better than, all the alternatives in their assigned categories (i.e. categories A and B respectively). The results for the two other distributions were however ambiguous. In the case of dataset 1, the Poisson-Poisson distribution performed well in its assigned category, matching or outperforming the alternative distributions in category C. There were however instances in the case of dataset 2 where one of the alternative distributions performed better than the Poisson-Poisson distribution in category C (in particular, the Poisson-Pascal distribution in Table 7 and the Poisson-Geometric distribution in Table 8). The results for the Poisson-Pascal distribution in Category D are also mixed. The performance of this distribution however seems to depend on the parameter estimation method being used. Under the method of moments (Tables 5 and 7), the Poisson-Pascal distribution generally performed equally or better than the alternative distributions in Category D. The only exception was in the case of dataset 1 (Table 5) where the Poisson-Geometric distribution was found to perform better than the Poisson-Pascal distribution in Category D. The Poisson-Pascal distribution, however, consistently underperformed all the alternative distributions when the parameters were estimated using the method of mean and zero frequency.

The difference in the performance of the Poison-Pascal distribution under the two estimation methods is not surprising. Katti and Gurland (1962) compared the efficiency of the estimators under the two methods and they found that, for intermittent demand patterns, the method of moments estimators were more efficient than the method of mean and zero frequency estimators. For any pair of estimators, the more efficient estimator has a smaller variance and, in that sense, more accurate, than the less efficient estimator. The greater efficiency of the method of moments may explain its superior performance.

The results above suggest that, in the case of categories A and B, the proposed classification is effective in assigning the compound Poisson distribution that provides the best fit. However, the performance of the scheme in categories C and D is mixed and the scheme needs to be tested on more datasets in order to obtain more empirical evidence. In the case of the Poisson-Pascal distribution, the results above suggest that the parameters should be estimated using the method of moments and not the method of mean and zero frequency.

Table 5. Comparison of goodness-of-fit in the four categories – Dataset 1 (Method of moments)

Category B (*N* = 2,781)

Confidence intervals			
	PoLog	**PoPo**	**PoPa**
PoGeo	-0.12, -0.07	0.41, 0.46	-0.05, -0.01
PoLog		0.29, 0.34	0.19, 0.23
PoPo			-0.50, -0.45

Conclusions			
	PoLog	**PoPo**	**PoPa**
PoGeo	PoGeo<PoLog	PoGeo>PoPo	PoGeo<PoPa
PoLog		PoLog>PoPo	PoLog>PoPa
PoPo			PoPo<PoPa

Category D (*N* = 733)

Confidence intervals			
	PoLog	**PoPo**	**PoPa**
PoGeo	0.06, 0.15	0.42, 0.50	**0.03, 0.11**
PoLog		0.15, 0.24	-0.02, 0.07
PoPo			-0.54, -0.45

Conclusions			
	PoLog	**PoPo**	**PoPa**
PoGeo	PoGeo>PoLog	PoGeo>PoPo	**PoGeo>PoPa**
PoLog		PoLog>PoPo	No Difference
PoPo			PoPo<PoPa

Category A (*N* = 6,461)

Confidence intervals			
	PoLog	**PoPo**	**PoPa**
PoGeo	0.14, 0.17	-0.01, 0.02	-0.03, 0.00
PoLog		-0.08, -0.05	-0.08, -0.05
PoPo			-0.03, 0.00

Conclusions			
	PoLog	**PoPo**	**PoPa**
PoGeo	PoGeo>PoLog	No Difference	No Difference
PoLog		PoLog<PoPo	PoLog<PoPa
PoPo			PoPo<PoPa

Category C (*N* = 4,899)

Confidence intervals			
	PoLog	**PoPo**	**PoPa**
PoGeo	0.47, 0.50	-0.35, -0.31	-0.41, -0.38
PoLog		-0.47, -0.44	-0.49, -0.46
PoPo			0.13, 0.16

Conclusions			
	PoLog	**PoPo**	**PoPa**
PoGeo	PoGeo>PoLog	PoGeo<PoPo	PoGeo<PoPa
PoLog		PoLog<PoPo	PoLog<PoPa
PoPo			PoPo>PoPa

KEY: The comparative results that relate to the distribution that is theoretically expected to perform best in each quadrant are underlined.

Table 6. Comparison of goodness-of-fit in the four categories – Dataset 1 (Method of mean and zero frequency)

Category B (*N* = 2,781)

Confidence intervals	PoLog	PoPo	PoPa
PoGeo	-0.14, -0.10	0.42, 0.47	0.45, 0.49
PoLog		0.31, 0.35	0.32, 0.37
PoPo			0.31, 0.36

Conclusions	PoLog	PoPo	PoPa
PoGeo	PoGeo<PoLog	PoGeo>PoPo	PoGeo>PoPa
PoLog		PoLog>PoPo	PoLog>PoPa
PoPo			PoPo>PoPa

Category D (*N* = 733)

Confidence intervals	PoLog	PoPo	PoPa
PoGeo	0.05, 0.13	0.44, 0.53	0.43, 0.52
PoLog		0.32, 0.41	0.28, 0.37
PoPo			0.05, 0.13

Conclusions	PoLog	PoPo	PoPa
PoGeo	PoGeo>PoLog	PoGeo>PoPo	**PoGeo>PoPa**
PoLog		PoLog>PoPo	**PoLog>PoPa**
PoPo			**PoPo>PoPa**

Category A (*N* = 6,461)

Confidence intervals	PoLog	PoPo	PoPa
PoGeo	0.09, 0.12	0.05, 0.08	0.35, 0.38
PoLog		-0.02, 0.01	0.19, 0.22
PoPo			0.25, 0.28

Conclusions	PoLog	PoPo	PoPa
PoGeo	PoGeo>PoLog	PoGeo>PoPo	PoGeo>PoPa
PoLog		No Difference	PoLog>PoPa
PoPo			PoPo>PoPa

Category C (*N* = 4,899)

Confidence intervals	PoLog	PoPo	PoPa
PoGeo	0.46, 0.50	-0.23, -0.20	0.31, 0.34
PoLog		-0.40, -0.37	-0.35, -0.32
PoPo			0.37, 0.41

Conclusions	PoLog	PoPo	PoPa
PoGeo	PoGeo>PoLog	PoGeo<PoPo	PoGeo<PoPa
PoLog		PoLog<PoPo	PoLog<PoPa
PoPo			PoPo>PoPa

KEY: The comparative results that relate to the distribution that is theoretically expected to perform best in each quadrant are underlined.

Table 7. Comparison of goodness-of-fit in the four categories – Dataset 2 (Method of moments)

Category B ($N = 53$)

Confidence intervals

	PoLog	PoPo	PoPa
PoGeo	-0.27, 0.06	0.34, 0.66	-0.04, 0.29
PoLog		0.28, 0.61	0.09, 0.42
PoPo			-0.66, -0.34

Conclusions

	PoLog	PoPo	PoPa
PoGeo	No Difference	PoGeo>PoPo	No Difference
PoLog		PoLog>PoPo	PoLog>PoPa
PoPo			PoPo<PoPa

Category D ($N = 123$)

Confidence intervals

	PoLog	PoPo	PoPa
PoGeo	-0.11, 0.10	0.34, 0.56	-0.11, 0.10
PoLog		0.23, 0.45	-0.04, 0.18
PoPo			-0.59, -0.38

Conclusions

	PoLog	PoPo	PoPa
PoGeo	No Difference	PoGeo>PoPo	No Difference
PoLog		PoLog>PoPo	No Difference
PoPo			PoPo<PoPa

Category A ($N = 91$)

Confidence intervals

	PoLog	PoPo	PoPa
PoGeo	0.11, 0.36	0.03, 0.28	-0.10, 0.15
PoLog		-0.05, 0.20	-0.10, 0.15
PoPo			-0.24, 0.01

Conclusions

	PoLog	PoPo	PoPa
PoGeo	PoGeo>PoLog	PoGeo>PoPo	No Difference
PoLog		No Difference	No Difference
PoPo			No Difference

Category C ($N = 229$)

Confidence intervals

	PoLog	PoPo	PoPa
PoGeo	0.33, 0.49	0.00, 0.16	-0.31, -0.15
PoLog		-0.30, -0.15	-0.46, -0.30
PoPo			-0.29, -0.13

Conclusions

	PoLog	PoPo	PoPa
PoGeo	PoGeo>PoLog	No Difference	PoGeo<PoPa
PoLog		PoLog<PoPo	PoLog<PoPa
PoPo			**PoPo<PoPa**

KEY: The comparative results that relate to the distribution that is theoretically expected to perform best in each quadrant are underlined.

Table 8. Comparison of goodness-of-fit in the four categories – Dataset 2 (Method of mean and zero frequency)

Category B (*N* = 53)

Confidence intervals			
	PoLog	**PoPo**	**PoPa**
PoGeo	-0.17, 0.15	0.28, 0.61	0.34, 0.66
PoLog		0.11, 0.44	0.28, 0.61
PoPo			0.26, 0.59

Conclusions			
	PoLog	**PoPo**	**PoPa**
PoGeo	No Difference	PoGeo>PoPo	PoGeo>PoPa
PoLog		PoLog>PoPo	PoLog>PoPa
PoPo			PoPo>PoPa

Category D (*N* = 123)

Confidence intervals			
	PoLog	**PoPo**	**PoPa**
PoGeo	-0.04, 0.18	0.31, 0.53	0.38, 0.59
PoLog		0.13, 0.35	0.23, 0.45
PoPo			0.21, 0.43

Conclusions			
	PoLog	**PoPo**	**PoPa**
PoGeo	No Difference	PoGeo>PoPo	**PoGeo>PoPa**
PoLog		PoLog>PoPo	**PoLog>PoPa**
PoPo			**PoPo>PoPa**

Category A (*N* = 91)

Confidence intervals			
	PoLog	**PoPo**	**PoPa**
PoGeo	0.01, 0.25	0.08, 0.33	0.31, 0.56
PoLog		-0.07, 0.19	0.14, 0.39
PoPo			0.33, 0.58

Conclusions			
	PoLog	**PoPo**	**PoPa**
PoGeo	PoGeo>PoLog	PoGeo>PoPo	PoGeo>PoPa
PoLog		No Difference	PoLog>PoPa
PoPo			PoPo>PoPa

Category C (*N* = 229)

Confidence intervals			
	PoLog	**PoPo**	**PoPa**
PoGeo	0.35, 0.51	0.05, 0.21	0.32, 0.48
PoLog		-0.28, -0.12	-0.17, -0.01
PoPo			0.21, 0.37

Conclusions			
	PoLog	**PoPo**	**PoPa**
PoGeo	PoGeo>PoLog	**PoGeo>PoPo**	PoGeo<PoPa
PoLog		PoLog<PoPo	PoLog<PoPa
PoPo			PoPo>PoPa

KEY: The comparative results that relate to the distribution that is theoretically expected to perform best in each quadrant are underlined.

5. Implications for the OR theory and practice

As others (e.g. Fortuin and Martin, 1999; Botter and Fortuin, 2000; Syntetos et al., 2009) have already pointed out, the management of spare parts and other inventory items with intermittent demand is a difficult task. A number of authors have argued that compound distributions could be used to model such intermittent demand patterns. However, there have been very few empirical studies in this area. The main contribution of this paper relates to a detailed empirical investigation on the viability of using compound Poisson distributions to model intermittent demand. Goodness-of-fit tests were carried out for various compound Poisson distributions and the challenges involved in using such distributions were explored.

Compound Poisson processes have a structure that is similar to the demand-generating process associated with intermittent demand – events (in this case demand orders) arrive sporadically and the size of the events is variable. The likeness between compound Poisson processes and the order arrival processes typically observed among spare parts will have an intuitive appeal to inventory managers. Goodness-of-fit tests were carried out in this study for four different compound Poisson distributions: (i) Poisson-Pascal; (ii) Poisson-Poisson; (iii) Poisson-Log Series and (iv) Poisson-Geometric. The empirical demand data used in these tests was extensive and consisted of the demand histories of more than 15,000 spare parts SKUs. All four distributions were found to provide high levels of frequency of fit.

Compound Poisson distributions also model the order sizes independently of the order arrival process. Orders are assumed to arrive according to a Poisson process but different distributions could be used to model the order sizes. Different compound Poisson distributions could therefore be used to model SKUs with differing order size profiles. In the area of inventory management, there is wide agreement that effective classification can lead to substantial improvements in performance. In this paper, we proposed a scheme that assigns different compound Poisson distributions to SKUs with differing order size properties. The scheme classifies SKUs based on the modality and variability of the observed orders sizes and it can greatly facilitate the process of selecting distributional models for items with intermittent demand. The scheme has been assessed for its empirical validity in terms of the goodness of fit. The results suggest that the scheme is very effective in assigning the best-fitting distribution to SKUs falling in two of the four identified categories. Ambiguous results were obtained in the case of the other two categories and further empirical tests need to be carried in order to ascertain the effectiveness of the scheme in these categories. The scheme was developed, not based on empirical findings from individual studies, but rather on a fixed set of criteria relating to order size distributions of intermittent demand items. As such, the scheme is generally applicable and we recommend that practitioners and researchers may adopt this solution (after a simulation of its performance on real data related to the cases under concern).

However, it is true to say that further tests are required in order to assess the effectiveness of the scheme in terms of its stock control performance. The goodness-of-fit results in this study will be instructive for practitioners concerned with performance targets (such as the fill rate) which are derived based on the entire demand distribution. In such cases, the scheme identifies a distribution that in theory is likely to provide practitioners with a good fit for the observed demand data. The scheme is however less useful for performance targets that are concerned with only a single point in the demand distribution. For example, the cycle service level is defined as the appropriate percentile of the demand distribution. There is therefore less benefit in such cases in identifying a distribution that provides good fit across the entire demand distribution. A distribution might perform well under the cycle service level definition if it provides good fit at the specified percentile but very poor fit across the rest of the distribution. For such performance targets, practitioners are bound to find the classification scheme proposed in this study less effective.

One of the issues considered in this study is the need for a hierarchical list of criteria that should be used when selecting distributions for modelling demand. The most important criterion is that the hypothesised distribution has to match the underlying structure of demand as understood by the inventory managers. But based on the challenges encountered in this study, it seems that the next most important criterion should be the mathematical tractability of the distribution. If the distribution is to be useful in practical settings, then it needs to have a probability function that is easy to compute using readily available software packages such as Microsoft Excel ®. In the context of intermittent demand, distributions with large number of parameters should be avoided as much as possible. For a given demand pattern, as the number of parameters increases, the degrees of freedom (the number of independent observations in a sample that are available to estimate parameters) decrease. The accuracy of the parameter estimates will therefore deteriorate as the number of parameters increases. This is particularly a problem in the case of intermittent demand. In general, the accuracy of the parameter estimates will improve as the samples becomes more diverse. When demand is intermittent, there is little diversity in the observations (most of the observations are zeroes). Finally, mathematical tractability in terms of the domain compliance of the parameter estimators is also an issue worth considering. Some distributions might seem appealing in theory but, if they have domain incompliant estimators, they might not perform as well (for example, the Poisson-Pascal distribution in this study).

After mathematical tractability, the next most important criterion is corroborative empirical evidence. However, relevant empirical evidence might be hard to come by and occasionally the findings in different studies might contradict one another. The final criterion should be the flexibility of the distribution. While flexibility might be desirable, this is an issue that can be easily resolved by simply increasing the number of distributions in order to ensure that there is a distribution to accommodate each of the possible demand profiles. While this might seem inconvenient, the challenges encountered in this study suggest that it might be worthwhile to sacrifice flexibility for mathematical tractability.

6.Conclusions and further work

Demand classification is an important operational issue in the management of spare part inventory items. Demand classification facilitates decision-making with respect to forecasting and stock control and enables managers to focus their attention on the SKUs considered most important. In this paper, we carried out goodness-of-fit tests to assess whether compound Poisson distributions provide a good fit to SKUs with intermittent demand. An empirical dataset of nearly 15,000 spare part SKUs from two different industries was used in these tests. The compound Poisson distributions were found to provide good fit for most of the SKUs in the empirical dataset. These results suggest that managers should consider using the compound distributions discussed in this work to model the demand of intermittently moving inventory items. We have also proposed a demand classification scheme that categorises SKUs based on the mode and variability of the observed order sizes. The scheme facilitates the process of selecting distributional models for items with intermittent demand. The scheme was also tested for its empirical validity and the results suggest that it is mostly effective in the sense that the proposed compound distribution often provided the highest levels of frequency of fit for SKUs falling within the associated category. A comprehensive list of criteria to be used when selecting demand distributions has also been proposed. Finally, an extensive discussion has been provided on parameter estimation related difficulties in this area. As such, we feel that our work should enable further theoretical developments in the area of spare parts management and should successfully inform relevant real world practices.

In the next steps of our research, we plan to replicate our findings on more demand datasets and assess the empirical validity of the classification scheme in terms of its implications for forecast accuracy. The scheme will also need to be assessed for effectiveness in terms of stock control performance. Further work and empirical studies on the performance of non-parametric approaches (like Bootstrapping for example) and the way such approaches compare to the more 'traditional' distribution-based inventory control considered in this paper should also contribute significantly towards extending the current state of knowledge in this area. Finally, an attempt will be made to link the quantitative measures in the scheme (i.e. the mode and squared coefficient of variation of the order sizes) with the qualitative aspects of SKUs. The linkage between the technical attributes of the classification scheme and the qualitative attributes of the SKUs assigned to the various categories may be of great value to practitioners operating in this area.

References

Adelson, R. (1966). Compound Poisson Distributions. *Journal of the Operational Research Society, 17*, 73-75.

Axsẽter, S. (2006). *Inventory control* (Vol. 90). New York, NY: Springer

Archibald, B. C., & Silver, E. A. (1978). (s, S) Policies Under Continuous Review and Discrete Compound Poisson Demand. *Management Science, 24*, 899.

Babai, M., Jemai, Z., & Dallery, Y. (2011). Analysis of order-up-to-level inventory systems with compound Poisson demand. *European Journal of Operational Research, 210*, 552-558.

Bacchetti, A., Saccani, N. (2012). Spare parts classification and demand forecasting for stock control: Investigating the gap between research and practice. *OMEGA:International Journal of Management Science*, *40 (6)*, 722-737.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*, 289-300.

Berry, A. C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates. *Transaction Orders of the American Mathematical Society, 49*, 122-139.

Birnbaum, Z. W. (1962). *Introduction to probability and mathematical statistics*. New York: Harper.

Blumenfeld, D. E., Daganzo, C. F., Frick, M. C., & Gonsalvez, D. J. A. (1999). Impact of manufacturing response time on retailer inventory. *International Journal of Operations & Production Management, 19*, 797-811.

Botter, R., & Fortuin, L. (2000). Stocking strategy for service parts–a case study. *International Journal of Operations & Production Management, 20*, 656-674.

Bowman, K. O., & Shenton, L. (1967). *Remarks on estimation problems for the parameters of the Neyman Type A distribution*, Union Carbide Corporation, Report ORNL-4102, Oak Ridge, TN.

Boylan, J. E. (1997). *The centralisation of inventory and the modelling of demand.* Unpublished unpublished Ph.D. Thesis, University of Warwick, Coventry, UK.

Boylan, J. E., & Syntetos, A. A. (2008). Forecasting for inventory management of service parts. In K. A. H. Kobbacy & D. N. P. Murthy (Eds.), *Complex System Maintenance Handbook* (pp. 479-506). New York: Springer Verlag.

Cochran, W. G. (1952). The Chi-Squared test of goodness of fit. *The Annals of Mathematical Statistics, 23*, 315-345.

Cramer, H. (1999). *Mathematical methods of statistics* (19th ed.). Princeton, New Jersey: PrincetonUniversity Press.
Croston, J. (1972). Forecasting and Stock Control for Intermittent Demands. *Journal of the Operational Research Society, 23*, 289-303.
D'Alessandro, A. J. & Baveja, A. (2000). Divide and conquer: Rohm and Haas' response to a changing specialty chemicals market, *Interfaces,30,*1-16.
Eaves, A. H. C. (2002). *Forecasting for the ordering and stock-holding of consumable spare parts.* Unpublished PhD thesis, LancasterUniversity, Lancaster, UK.
Esseen, C. G. (1956). A moment inequality with an application to the central limit theorem. *Scandinavian aktuarmæssige tidsskrift, 39*, 160-170.
Feeney, G. J., &Sherbrooke, C. C. (1966). The (s-1, s) inventory policy under compound Poisson demand. *Management Science, 12*, 391-411.
Forsberg, R. (1995). Optimization of order-up-to-S policies for two-level inventory systems with compound Poisson demand. *European Journal of Operational Research, 81*, 143-153.
Fortuin, L., & Martin, H. (1999). Control of service parts. *International Journal of Operations & Production Management, 19*, 950-971.
Friend, J. (1960). Stock control with random opportunities for replenishment. *Journal of the Operational Research Society, 11*, 130-136.
Gutierrez, R. S., Solis, A. O., & Mukhopadhyay, S. (2008). Lumpy demand forecasting using neural networks. *International Journal of Production Economics, 111*, 409-420.
Heinecke, G., Syntetos, A., & Wang, W. (2011). Forecasting-based SKU classification. *International Journal of Production Economics*. Advance online publication, 10.1016/j.ijpe.2011.11.020
Hill, R. M., & Johansen, S. G. (2006). Optimal and near-optimal policies for lost sales inventory models with at most one replenishment order outstanding. *European Journal of Operational Research, 169*, 111-132.
Janssen, F., Heuts, R., & De Kok, T. (1998). On the (R, s, Q) inventory model when demand is modelled as a compound Bernoulli process. *European Journal of Operational Research, 104*, 423-436.
Jasper, J. B. (2006). *Quick Response Solutions, FedEx Critical Inventory Logistics Revitalized*, FedEx, Memphis, TN.
Johnson, N. L., Kemp, A. W., & Kotz, S. (2005). *Univariate discrete distributions*. Hoboken, New Jersey: Wiley-Interscience.
Kalchschmidt, M., Verganti, R., & Zotteri, G. (2006). Forecasting demand from heterogeneous customers. *International Journal of Operations & Production Management, 26*, 619-638.
Katti, S., & Gurland, J. (1961). The Poisson Pascal distribution. *Biometrics, 17*, 527-538.
Katti, S., & Gurland, J. (1962). Efficiency of certain methods of estimation for the negative binomial and the Neyman type A distributions. *Biometrika, 49*, 215-226.
Keilson, J., & Kubat, P. (1984). Parts and service demand distribution generated by primary production growth. *European Journal of Operational Research, 17*, 257-265.
Kemp, C. D. (1967). 'Stuttering - Poisson' distributions. *Journal of the Statistical and Social Inquiry Society of Ireland, 21*, 151-157.
Kendall, M. G., Stuart, A., Ord, J. K., & O'Hagan, A. (1987). *Kendall's advanced theory of statistics* (Vol. 1). New York: OxfordUniversity Press.
Kwan, H. W. (1991). *On the demand distributions of slow moving items.* Unpublished unpublished Ph.D. thesis, LancasterUniversity, Lancaster, UK.
Larsen, C., Seiding, C. H., Teller, C., & Thorstenson, A. (2008). An inventory control project in a major Danish company using compound renewal demand models. *IMA Journal of Management Mathematics, 19*, 145-162.
Mak, K., & Hung, C. (1993). A Simple Model for Computing (s, S) Inventory Policies when Demand is Lumpy. *International Journal of Operations & Production Management, 6*, 62-68.
Matheus, P., & Gelders, L. (2000). The (R, Q) inventory policy subject to a compound Poisson demand pattern. *International Journal of Production Economics, 68*, 307-317.
Mitchell, C., Rappold, R., & Faulkner, W. (1983). An analysis of Air Force EOQ data with an application to reorder point calculation. *Management Science, 29*, 440-446.
Naddor, E. (1978). Sensitivity to distributions in inventory systems. *Management Science, 24*, 1769-1772.
Porras, E., & Dekker, R. (2008). An inventory control system for spare parts at a refinery: An empirical comparison of different re-order point methods. European Journal of Operational Research, 184, 101-132.
Roscoe, J. T., & Byars, J. A. (1971). An investigation of the restraints with respect to sample size commonly imposed on the use of the chi-square statistic. *Journal of the American Statistical Association, 66*, 755-759.
Sato, K. I. (1999). *Lévy processes and infinitely divisible distributions*. Cambridge: CambridgeUniversity Press.
Satterthwaite, F. (1942). Generalized Poisson distribution. *The Annals of Mathematical Statistics*, 410-417.
Shale, E., Boylan, J., & Johnston, F. (2005). Forecasting for intermittent demand: the estimation of an unbiased average. *Journal of the Operational Research Society, 57*, 588-592.

Shale, E., Boylan, J. E., & Johnston, F. R. (2008). Characterizing the frequency of orders received by a stockist. *IMA Journal of Management Mathematics, 19*, 137-143.
Shenton, L. R., & Bowman, K. O. (1977). *Maximum likelihood estimation in small samples*. London: C. Grifiin & Co.
Strijbosch, L., & Moors, J. (2005). The impact of unknown demand parameters on (R, S)-inventory control performance. *European Journal of Operational Research, 162*, 805-815.
Syntetos, A. A., Babai, M. Z., & Altay, N. (2012). On the demand distribution of spare parts. *International Journal of Production Research*, *50*, 2101-2117.
Syntetos, A.A., Lengu, D., Babai, M.Z., (2013). A note on the demand distributions of spare parts, *International Journal of Production Research*, in press (doi:10.1080/00207543.2013.798050).
Syntetos, A. A., Boylan, J. E., & Croston, J. D. (2005). On the categorization of demand patterns. *Journal of the Operational Research Society, 56*, 495-503.
Syntetos, A. A., Keyes, M., & Babai, M. (2009). Demand categorisation in a European spare parts logistics network. *International Journal of Operations & Production Management, 29*, 292-316.
Teunter, R. H., Babai, M. Z., & Syntetos, A. A. (2010a). ABC classification: service levels and inventory costs. *Production and Operations Management, 19*, 343-352.
Teunter, R. H., Syntetos, A. A., & Babai, M. Z. (2010b). Determining order-up-to levels under periodic review for compound binomial (intermittent) demand. *European Journal of Operational Research, 203*, 619-624.
Upton, G. and Cook, I. (1996), *Understanding statistics*. Oxford: Oxford University Press.
Van Kampen, T. J., Akkerman, R., & van Donk, D. P. (2012). SKU classification: a literature review and conceptual framework, *International Journal of Operations & Production Management*, *32*, 850 – 876.
Ward, J. (1978). Determining reorder points when demand is lumpy. *Management Science, 24*, 623-632.
Wackerly, D.D., Mendenhall, W. & Scheaffer R. L. (2002). *Mathematical Statistics with Applications*, Boston, Massachussets: Duxbury Press.
Watson, R. B. (1987). The effects of demand-forecast fluctuations on customer service and inventory cost when demand is lumpy. *The Journal of the Operational Research Society, 38*, 75-82.
Willemain, T. R., Smart, C. N., & Schwarz, H. F. (2004). A new approach to forecasting intermittent demand for service parts inventories. *International Journal of Forecasting, 20*, 375-387.
Williams, T. (1984). Stock control with sporadic and slow-moving demand. *The Journal of the Operational Research Society, 35*, 939-948.
Zhao, Y. (2009). Analysis and evaluation of an Assemble-to-Order system with batch ordering policy and compound Poisson demand. *European Journal of Operational Research, 198*, 800-809.