ORDER QUANTITY DISTRIBUTIONS: ESTIMATING AN ADEQUATE AGGREGATION HORIZON

Poul Svante Eriksen[1], Peter Nielsen[2]

[1] *Aalborg University, Department Mathematical Sciences, Aalborg, Denmark*
[2] *Aalborg University, Department Mechanical and Manufacturing Engineering, Aalborg, Denmark*

Corresponding author:
Peter Nielsen
Aalborg University
Department Mechanical and Manufacturing Engineering
Aalborg, Denmark
phone: +45 9940 8932
e-mail: peter@m-tech.aau.dk

ABSTRACT
In this paper an investigation into the demand, faced by a company in the form of customer orders, is performed both from an explorative numerical and analytical perspective. The aim of the research is to establish the behavior of customer orders in first-come-first-serve (FCFS) systems and the impact of order quantity variation on the planning environment. A discussion of assumptions regarding demand from various planning and control perspectives underlines that most planning methods are based on the assumption that demand in the form of customer orders are independently identically distributed and stem from symmetrical distributions. To investigate and illustrate the need to aggregate demand to live up to these assumptions, a simple methodological framework to investigate the validity of the assumptions and for analyzing the behavior of orders is developed. The paper also presents an analytical approach to identify the aggregation horizon needed to achieve a stable demand. Furthermore, a case study application of the presented framework is presented and concluded on.

KEYWORDS
Demand distribution, Poisson processes, order quantities, Coefficient of Variation, Planning&Control.

Introduction

To achieve good customer service, companies aim at delivering products on time at competitive prices. One aspect in achieving competitive performances is companies' abilities to match customer expectations. In many situations resources are dedicated prior to knowing actual customer demand. This makes forecasting customers' ordering behavior of critical importance. Most research on this topic has been dedicated to either forecasting demand [1] or how to react in a quick robust manner to customer demand through e.g. better planning [2]. The research presented in this paper focuses on another aspect that has received limit research attention, namely the fact that demand is composed of individual customer orders. The fact that the demand a company faces is actually composed of a large number of presumed i.i.d. customer orders that can be aggregated is challenged in this research through an investigation of actual behavior and modelling of what to do when customer orders are in fact not i.i.d. This paper expands on the work reported in [3] through adding a framework, extending the complexity of the model, elaborating on the current state and adding more numerical examples. This paper focuses on customer order quantity distributions, analyses their behavior and investigates the consequence of transforming these into demand rates that are used for planning purposes. It also presents an analytical model to calculate the aggregation horizon necessary to ensure a stable distribution of demand for operational planning purposes. In this research main aim is to establish the specific need for aggregation in

Management and Production Engineering Review

the form of the number of customer orders needed to aggregate to achieve a system-wide desired stability. In this context a stable distribution of demand is defined as a demand pattern approximately symmetrically distributed with a given (low) Coefficient of Variation (CV) (estimated variance of the distribution divided by its estimated mean) [4] as this is a good proxy for stability. Here it is important to note that the assumption of stationary, symmetrical distributed demand is found in many areas of application and is a pre-requisite for e.g. achieving cyclical steady state behavior in manufacturing environments [5] This assumption is commonly used, despite it being widely recognized that it is an over-simplification of the actual demand situation [6–8]. The first step towards more adequately modeling the actual customer behavior, rather than just demand rates, must be to build an understanding of the behavior. A logical second step is to investigate if it is possible to aggregate customer ordering behavior in a manner that ensures that we live up to these assumptions. If this aggregation is not possible the research community must develop methods that are able to overcome the short comings of the current models or at the very least investigate the consequences of violating the assumptions. In this work the first two steps are addressed. An analytical model that can be used to estimate the aggregation horizon (in the form of number of customer orders) necessary to ensure a stable demand rate in a given manufacturing environment is presented. So when customer orders are not directly corresponding to our assumptions, one can estimate how much information must be aggregated to achieve something that matches our requirements. The presented method is particularly suited for manufacturing environments with high order frequencies as the illustrative example and case show.

The remainder of the paper is structured as follows. First, we present a discussion of the standard assumptions associated with demand behavior and the link between demand and actual individual customer orders from various perspectives within manufacturing planning and control. Following this a framework for analyzing the behavior of orders is developed and presented. Third, an analytical approach able to identify the aggregation horizon needed to achieve a stable demand is developed. The analytical approach illustrates the flexibility needs or desired lead time in practical manufacturing environments. Fourth, a case study application of the framework and analytical approach is included to illustrate the applicability. Finally conclusions and potential avenues of further research are presented.

Background and motivation

Manufacturing Planning & Control addresses a large number of decisions that require coordination. Many of these decisions are directly related to customer orders and the characteristics of these. It must be accepted that the demand faced by companies arrives in the form of customer orders and that customer orders can consist of multiple order lines and that each order (line) will contain the following parameters relevant for operational planning purposes: desired due date (typically the same for the whole order), product identification and order quantity. For operational purposes these parameters are transformed into various forms of models for demand (also predictive models) depending on how they are used in the planning processes and manufacturing system design (e.g. line balancing). Typically, demand is modelled in terms of total demand of an independent product for a planning period. Inside this planning period the total demand is translated into a demand rate by uniformly distributing the demand over the whole planning period [7, 9], see Fig. 1.
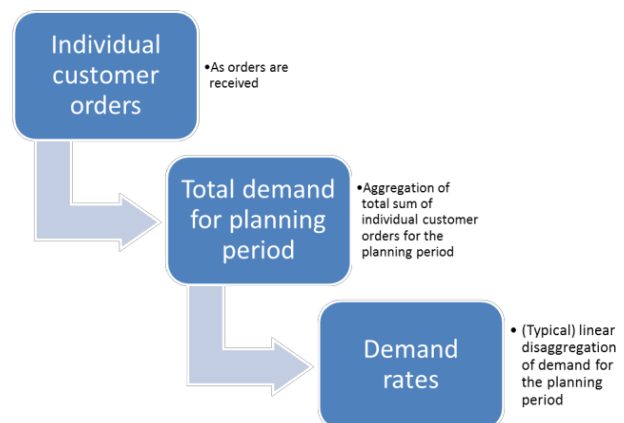

Fig. 1. Flow of translating customer orders into demand rates used in planning.

The demand rate has traditionally been used in inventory management so individual customer orders are translated into time dependent demand rates (typically from a known distribution) through an aggregation/disaggregation process as stated above. This demand rate is then used for calculating re-order points, lot-sizes, timing, safety stock etc. [10, 11]. The same is found within line balancing where customer orders are likewise transformed into demand rates (also dubbed release or launch rates) [12–14] for individual products and product mix [15] for a given line. Similarly, lot-sizing rules rely on strict assumptions such as demand rates: are

deterministic or stem from a stationary stochastic process [16]. It is interesting to note that these assumptions are maintained despite the fact that demand rates are composed of a number of individual customer orders. These individual customer orders should be by their very nature behave stochastically, but also potentially arrive in a coordinated manner. While the stochastic aspect is clear, there are a number of well-established reasons to expect that customer orders do not arrive in a random uncoordinated manner to the system. The main reasons are batch ordering (customer orders more than one item due to e.g. transportation costs) and multiple customer orders of similar quantities due to e.g. sales discounts, so that multiple orders are received from different customers in a coordinated manner.

Regardless of which purpose the demand rate models are used for, a low variation of demand is desirable in all situations. A typical measure of demand stability used within manufacturing planning and control is the Coefficient of Variation (CV). Bobko and Whybark [4] deem CV a robust measure of demand volatility and it is used in e.g. Tsubone and Furuta [17] and Pujawan and Kingsman [18] to characterize demand behavior. To lower variation and thereby reduce uncertainty of the demand the general approach is to aggregate individual customer orders into demand rates for both individual products and product families [9, 19, 20] over a sufficient time horizon or postponing allocations [21]. However, as recognized in [21] postponement is not cost free as it involves investing in flexibility. The aggregation can be conducted along a number of dimensions [19, 22] (typically time and products [23]) until a stabilized customer ordering behavior is identified. Aggregating in time has the disadvantage of reducing responsiveness as this implies buffering in customer ordering lead time by increasing manufacturing lead time. Buffering in abstraction level (i.e. planning on product family level rather than individual product level) has the disadvantage of significant information loss. So regardless of aggregation dimension, the transformation comes with a loss of detail level or responsiveness. So in practice the less need to aggregate information, the better planning can be performed. To facilitate the aggregation and disaggregation of information a number of assumptions must be made [24, 25]. These assumptions are simple in nature and can basically be reduced to assuming a one-to-one relationship exists when aggregating and disaggregating [19, 26] and complete independence of demand. As a result, order quantities and the number of orders per period are assumed to be constant in many of the methods applied [11], even

if this is obviously an assumption that has limited likelihood of being true. Especially if one considers the fact that it is highly unlikely that the individual customer orders are in fact completely independent of each other. It is also interesting to note that this assumption is central and if violated, the disaggregation of aggregate plans may result in suboptimal or even infeasible disaggregate plans through e.g. too little / too much time allocated for setup time. This assumption can be relaxed if customer orders are identically independently distributed, stem from a symmetrical distribution with a low CV and a sufficiently large time period can be used for aggregation. In this case demand would be considered to be i.i.d., but when we then disaggregate into demand rates further complications can arise. Some research indicates that there can be significant costs associated with assuming e.g. Gaussian distributed demand if the demand is in fact non-Gaussian distributed [6]. Tadikamalla [8] concludes that the high values of CV associated with asymmetrical distributions of demand lead to poor performance of inventory management techniques and that the symmetry/asymmetry of demand is critical for e.g. inventory costs [27]. Similar conclusions along with practical considerations can be found in [28], where it is confirmed that the CV of demand has a significant performance impact. Depending on the manufacturing system's external context, the assumption of independent distributed observations may likewise not hold true. A chronological list of order lines for delivery can expect to exhibit non-stationary behavior. So systems addressing orders in a First-Come-First-Serve manner will potentially be highly sensitive to this assumption. A typical manifestation of this will be if an order contains order lines for more than one product from the same product family leading to systematic short term amplification of demand rates [7, 29]. This is critical if the order lines are for products within the same product family or related in a manner that means that their demand will be aggregated within the planning environment on one of the traditional dimensions [23]. On the individual product level it seems reasonable in some situations to assume independently distributed order quantities. This is especially so if many customers purchase the same products independently of each other and no significant order quantity discounts are in place.

Aggregation of demand data is for all applicable domains assumed to lead to stabilized behavior [19, 30]. A number of methods have been proposed to address the issue of interrelated demand from the perspective of improving the performance of demand forecasts [31, 32]. These methods are focused on giv-

ing an accurate estimate of demand per planning period for e.g. individual SKUs or product families. However, they fail to address the significant issue of how much aggregation is required to achieve a given stability in the short term for operational planning purposes.

The general approach in literature is to assume that order quantities stem from a symmetrical distribution preferably with a low Coefficient of Variation (CV) and are independently distributed. However, for demand within a product family the orders may in actuality stem from a dependent distribution, since orders consist of multiple order lines, typically for more than one product within the same product family. So if linear relationships in aggregation are an oversimplification of the issue, what should be done instead? In this paper the assumptions used to aggregate and disaggregate demand information are investigated and the impact on various planning approaches is evaluated. The aim is to design a framework for robust estimation of order quantities and their behavior. Specifically if the order quantities can be assumed to be independently distributed and stem from a symmetrical distribution. The next step is the development of an analytical model describing the CV for the order quantities. In this step it is assumed that the order quantities can stem from an arbitrary distribution and are in fact dependently distributed in time due to the nature of how orders in a production family are received. Through this we aim to facilitate a practical method for calculating the operational aggregation horizon (in the form of number of orders) required to attain a desired CV of demand distribution. The aim is not to predict demand and volume, but rather to identify the short term needs for aggregation of orders in manufacturing environments to allow to batch orders and achieve a level utilization / load in the manufacturing system. This method can be used for such diverse activities as setting realistic delivery lead times or identifying flexibility requirements. It can also be used to estimate if it is reasonable to translate a demand model into demand rates and the potential consequence of this translation.

The aggregation horizon is a critical term in determining the need for setup time and thus in determining the actual available manufacturing time. This becomes especially critical in the situations where each new order arriving at the manufacturing system incurs a fixed setup cost.

Figure 2 shows a simple example of this this dilemma, where there is a fixed volume of 1000 units to be produced with a fixed processing time of 10 time units. The x and y axis are respectively the set-

up time per order (varied from 0 to 10 time units per order) and the number of orders used to produce the 1000 units (varied from 1 to 1000). The z-axis illustrates the total required manufacturing time (setup + processing time) to produce the 1000 units with the time illustrated on the right hand scale. As Fig. 2 illustrates, the combination of a low number of orders to achieve a given production volume and low setup times compared to processing times lead to a situation where the variation in order quantities would have limited impact. However, it also underlines that manufacturing systems faced with large setup times compared to processing times are very sensitive to the order quantity distributions. In practice one typically assumes a given allocation of time needed for setup, but underlying this is an assumption that a given volume is always sold in a particular number of orders leading to a constant need for setup time.

Manufacturing times dependence on order quantities and setup times
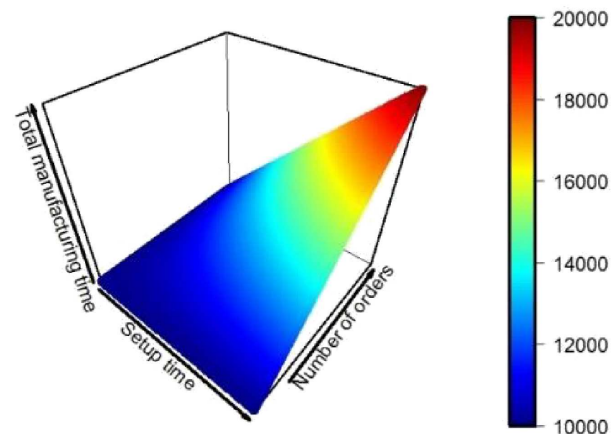

Fig. 2. Illustration of the link between the number of orders used to achieve a given volume, setup time and total manufacturing time.

From literature [27] it is known that the symmetry of the demand distributions is critical for the subsequent performance of the manufacturing system. This implies that prior to evaluating the order quantity distributions one should test for the presence of symmetry in the distributions. If these tests are negative, the order quantities are considered to stem from a skewed distribution and are therefore also assumed to be non-constant. In this paper the following three different statistical tests will be applied to evaluate the symmetry of the distributions:
• The MGG-test [33].
• The CM-test [34].
• The Mira-test [35].

From the literature review it is also known that the individual orders for each product can likely be considered to be independently distributed in the

case of multiple customers purchasing the same product. However, the aggregate behavior for a product family cannot necessarily be assumed to abide by the same behavior. To test for independence of the order quantities in the FCFS system the autocorrelation function is used [36].

Based on this analysis of literature the simple framework shown in Fig. 3 seems like an appropriate method for establishing the behavior of order quantity distribution. The framework consists of four steps (see Fig. 3). First, the data must be cleaned and prepared for analysis. Second, the order quantity distributions are investigated for symmetry. Third, the distributions are checked for independence using autocorrelation before finally it is concluded if aggregation is needed to achieve a stable planning environment.
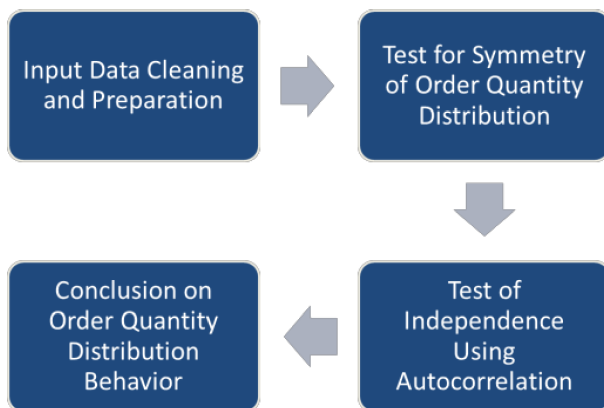

Fig. 3. A simple framework for establishing some of the critical behaviors of order quantity distributions.

The next step is to investigate how to stabilize the demand behaviour experienced in the form of customer orders through proper aggregation. In practices this means determining the number of orders to aggregate to achieve a sufficiently low variance of the aggregated distribution. The proposed model is able to determine the number of consecutive orders to aggregate to achieve a distribution with a satisfactory behaviour. In this case a symmetrical distribution, using the central limit theorem [37] with a low CV is considered to be satisfactory for planning purposes. The more aggregate the planning level, the lower CV is desired.

An analytical model of the aggregation needs

To facilitate the formulation of the model a few assumptions are necessary. We limit these to make the model generally applicable. First, the orders for the whole product family are assumed to come from an unbroken stationary time series. This assumption seems to be reasonable as the data used will be a set of customer orders that can, within reasonable limits, be assumed to be an unbroken time series. This is of course especially true if one investigates relatively short time periods where no changes have occurred to the product family composition. Second, we assume that it makes sense to aggregate the information in this manner (this is a standard assumption in Manufacturing Planning & Control). This assumption is of course context dependent and thus the method should be applied only in contexts where this is reasonable. Third, we assume that the orders are handled in a FCFS manner so that they are processed for manufacturing in the sequence they are received. This means that the sequence we receive orders in is also the sequence in which they released for production. This assumption is reasonable in most manufacturing systems that serve customers to some extend from inventory, it may be less reasonable in manufacturing systems that are e.g. manufacturing to order. So, again situational application of the model is advised. Fourth, we assume that orders arrive to the system following a Poisson process. This is a reasonable assumption in most manufacturing systems and while it is not easy to document it also has limited impact if the assumption is somewhat relaxed.

We consider a time period of length T, where the number $N(T)$ of orders has mean λT and standard deviation $\tau\sqrt{T}$. In case of a Poisson process we have $\tau^2 = \lambda$. To describe the volatility of the order quantities the coefficient of variation (CV) is used. The size Q of an order is assumed to have mean μ and standard deviation σ, i.e. Q has coefficient of variation $CV(Q) = \sigma/\mu$. Furthermore, to simplify the model in its first version the autocorrelation function of successive order quantities is assumed to be zero except for lag 1. This means that the dependency exists only between each pair of orders. The autocorrelation of lag one is denoted ρ. Let $Q_1 \ldots Q_{N(T)}$ be the successive orders in the period. The aggregated order quantity is $= \sum_{i=1}^{N(T)} Q_i$, which has mean conditional on $N(T)$ given by $E\left(AQ|N(T)\right) = \mu N(T)$ and variance given by:

$$
\begin{aligned}
Var(AQ|N(T)) &= \sigma^2(N(T) + 2\rho\left(N(T) - 1\right)) \\
&\approx \sigma^2(1 + 2\rho)N(T),
\end{aligned}
\tag{1}
$$

under the assumption that $N(T)$ is sufficiently large so that we can assume that $N(T) \approx N(T) - 1$. This will be the case for most manufacturing environments where it is sensible to aggregate orders in this manner due to the number of orders that will be re-

ceived. For manufacturing environments with a low number of orders, other methods than aggregating orders are used to plan the production, see e.g. [38]. The mean of AQ is then given by $E(AQ) = \mu\lambda T$, whereas the variance is (approximated by):

$$
\begin{aligned}
Var(AQ) &= E(Var(AQ|N(T))) \\
&\quad + Var(E(AQ|N(T)) \\
&= \sigma^2(1+2\rho)\lambda T + \mu^2\tau^2 T.
\end{aligned}
\tag{2}
$$

This leaves us with a coefficient of variation given by

$$
\mathrm{CV}(AQ) = \frac{\sqrt{\sigma^2(1+2\rho)\lambda + \mu^2\tau^2}}{\mu\lambda\sqrt{T}}.
\tag{3}
$$

Aiming at a time horizon yielding a coefficient of variation given by CV_0, this is obtained when

$$
T = \frac{\sigma^2(1+2\rho)\lambda + \mu^2\tau^2}{(\mu\lambda\mathrm{CV}_0)^2} = \frac{\mathrm{CV}(Q)^2(1+2\rho)\lambda + \tau^2}{(\lambda\mathrm{CV}_0)^2}.
\tag{4}
$$

Expressed in terms of the mean number of orders we obtain

$$
E(N(T)) = \frac{\mathrm{CV}(Q)^2(1+2\rho)\lambda + \tau^2}{\lambda(\mathrm{CV}_0)^2}
\tag{5}
$$

and in case of a Poisson process this simplifies to:

$$
E(N(T)) = \frac{\mathrm{CV}(Q)^2(1+2\rho) + 1}{(\mathrm{CV}_0)^2}.
\tag{6}
$$

So for any given distribution of Q, for any given company and period it is (given sufficient observations) possible to estimate $\mathrm{CV}(Q)$ and ρ. In the case where there is no significant autocorrelation present the term simplifies further to:

$$
E(N(T)) = \frac{\mathrm{CV}(Q)^2 + 1}{(\mathrm{CV}_0)^2}.
\tag{7}
$$

Using this formulation it is possible to decide upon a desired stability of order quantities for the system in the form of CV_0 and calculate the number of orders ($E(N(T))$) necessary to aggregate over to achieve this stability. We can also extend the problem to the more general case of ρ_l, where the autocorrelation of order quantity sizes depends on more than one lag, and l denotes this lag. In this case through the same derivation process as previously we arrive at the following maintaining the assumption of a poison arrival process:

$$
E(N(T)) = \frac{\mathrm{CV}(Q)^2\left(1 + 2\sum_{k=1}^{L}\rho_k\right) + 1}{(\mathrm{CV}_0)^2},
\tag{8}
$$

where L is the maximum autocorrelation lag one wishes to include and k is the specific lag. This is under the assumption that $N(T)$ is so sufficiently large

that it is reasonable to assume that $N(T) \approx N(T) - L$ i.e. $N(T) \gg L$. Again for most practical cases one would not find the need to include large lags in the model, as it is unreasonable to assume that very old order quantities in the sequence will influence the current observation. Especially if one recalls that the reasoning behind this assumption is that if there is a structure it is due to a customer placing multiple orders at once. Extending the models beyond a few lags is thus unlikely as it results in a significantly more complex model and this will significantly limit practical possibilities of application and also mean that one runs the risk of overfitting.

Case

To illustrate the simple framework and analytical approach formulated in the previous section an analysis is conducted. The test is based on a set of orders for a product family. The data set contains information about product ID (i.e. the specific product in the product family), order quantity (the quantity ordered of each product at each time) and delivery date (the desired delivery date requested by the customer). The data covers a 6 year period and the analysis was conducted in the statistical analysis tool R [39]. Over the six years 54,243 orders where received and the product family contains 2,160 unique products that are physically similar, but still distinctly differ from each other. The data was arranged as received by the manufacturing company based on delivery due date. The data was also sub-organized, i.e. sequenced, based on system creation time. This sequencing reflects the manner in which they are processed and handled by the company. This means that the customer orders follow something resembling a FCFS systematic and order lines for products from the same family from the same order are sequenced after each other as they are simultaneously received. From this description alone it should be clear that while orders may arrive in a stochastic manner they are at the very least not completely randomly distributed, but somehow structured i.e. some form of dependence can exist.

Firstly, we investigate the distribution of order quantity for the whole product family regardless of the product ID. Here we establish that 99% of the order quantities had a value below 61. Meaning that 99% of the customer order lines were for quantities of less than 61 units. The histogram of these order quantities is shown in Fig. 4. It is worth noting that the order quantity tail is rather long, with a maximum value of 273. The overall picture shows an expo-

nential decay with isolated peaks at multiples of 10 and to some extend multiples of 5 and 12. This illustrates the nature of demand, where some order quantities for reasons such as batching, palletizing, transportation, discounts are preferred. The three test for symmetry were conducted and all had highly significant p-values below 10^{-15}. The estimated coefficient of variation was $\mathrm{CV}(Q) = 1.2$ for the data set. This equally indicates a very skewed order quantity distribution.
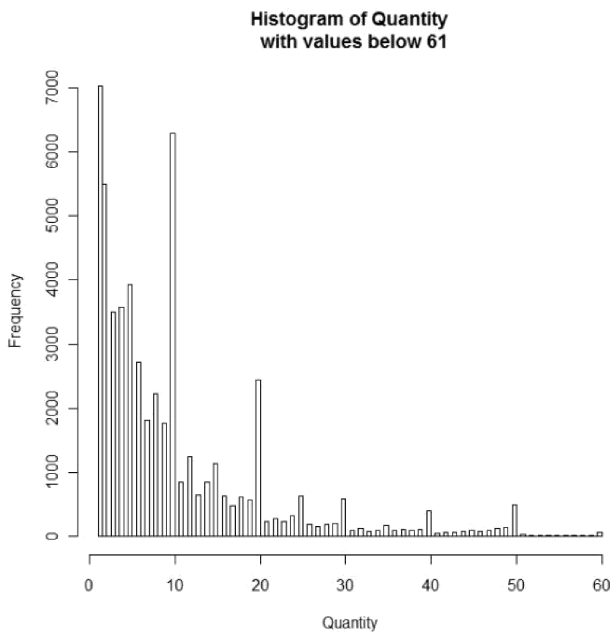

Fig. 4. Truncated histogram of order quantities containing the whole data set.

To further investigate the data and the behaviour of the order quantity distributions an analysis was conducted for the 151 products that had more than 100 orders in the data set. This number of observations is the minimum requirement for having sufficient observations for applying the symmetry tests. These products make up only 6% of the total number of products, but they comprise 66% of the total order quantity. It thus seems reasonable to assume that if sufficient observations had been present for the remaining 94% products the conclusions would be similar to those reached for the smaller sample. The results of the symmetry tests when adopting a 5% significance level can be seen in Table 1.

Table 1
Results of symmetry tests.

Test	MGG	CM	Mira
Ratio of non-symmetrical order quantity distributions	77.5%	76.9%	67.5%

The main conclusion is that the individual products' customer ordering behaviour seems to reflect the whole product family's order quantity distribution. I.e. the majority of the individual products' order quantity distributions are skewed and it thus seems very reasonable to assume that some form of aggregation is necessary to live up to stand assumptions needed to aggregate and disaggregate the customer order information into demand and demand rates. CV is also a good indicator for symmetry. Figure 5 shows the histogram of the estimated CV's for the 151 products. As can be seen the majority of the CV values are above 0.5. This also underlines that there is significant skewness of the order quantity distributions for the individual products.
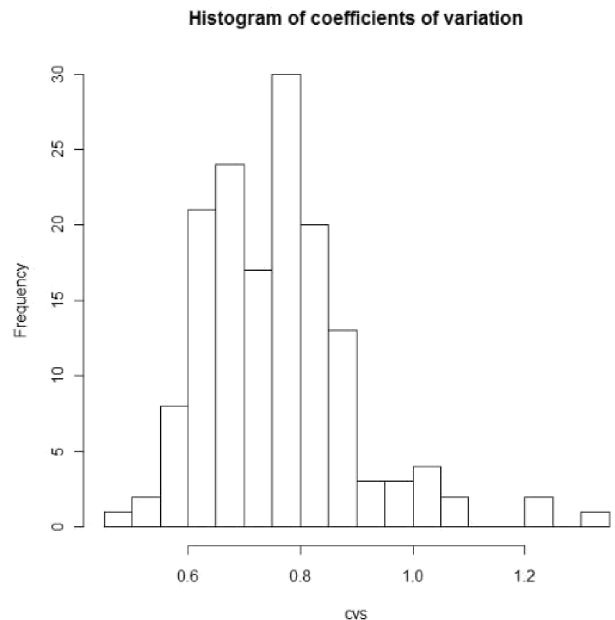

Fig. 5. Histogram of order quantity CV's for 151 products.

For the order quantity time series when using the data from the product family as a whole, there is an estimated mean of $\mu = 10.75$ and an estimated lag one correlation $\rho = 0.1105$ while the subsequent lags are in fact non-significantly correlated. As there is no data on arrival times of orders to the system, but rather delivery dates, it is fair to assume a Poisson arrival process. This lack of information regarding customer order arrivals is a typical problem encountered as this information is simply not logged by most companies' ERP-systems. For us it means that the aggregation level is determined by:

$$E(N(T)) = \frac{\mathrm{CV}(Q)^2(1+2\rho)+1}{(\mathrm{CV}_0)^2}. \qquad (9)$$

Based on the data estimates, this yields the equation for determining aggregation level in terms of orders given by

$$E(N(T)) = \frac{2.761}{(\mathrm{CV}_0)^2}. \quad (10)$$

If one uses a $\mathrm{CV}_0 = 0.3$ (achieving an approximate Gaussian distribution [6] or standard values for assuming symmetry), we obtain the integer round up $E(N(T)) = 31$. In Fig. 6, the aggregated distribution of order quantities using an aggregation over 31 orders is displayed. The distribution is visibly closer to symmetry, but it should be noted that it still has a statistically significant deviation from symmetry. The skewness is clearly much lower as reflected by the estimated coefficient of variation, which is 0.283 for this distribution. This is very close to the target of 0.3. The lack of complete symmetry can potentially be explained by having excluded correlation lags higher than 1. These may in fact have slightly influenced the result and if they had been included the need to aggregate to achieve a given CV would have increased.
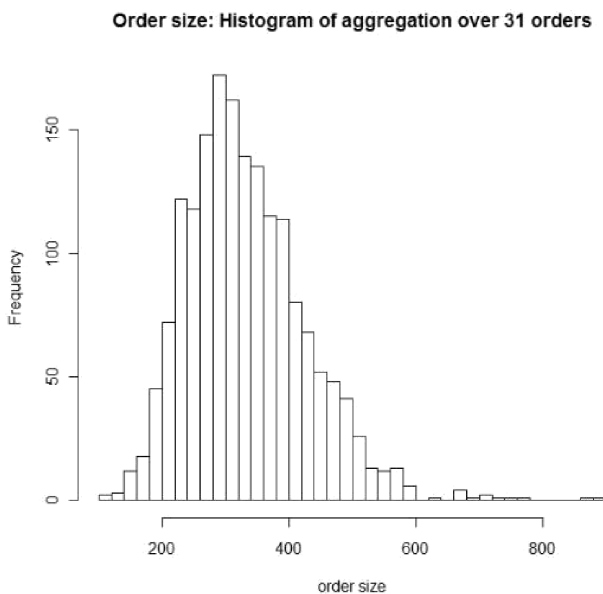

Fig. 6. Truncated histogram of order quantities containing the whole data set.

The aggregated order quantity distribution shown in Fig. 6 clearly shows the benefit of aggregation to achieve symmetry.

Figure 7 shows the need for aggregation as a function of the target CV (CV_0). It also indicates that in this particular case it could be problematic to achieve a low CV through aggregation of orders as this would require aggregating over a very large number of orders. This is a consequence of the high degree of asymmetry of the distribution of the orders. The aggregation has in this case been done on the product family level. It is worth to note that out of the 151 products with more than 100 orders in the studied period few have values of CV below 0.6. Assuming a similar degree of dependence as on the product family level and a target CV_0 of 0.3, and a current CV of 0.6, this implies the need to aggregate 5 orders, or in the case of a CV_0 of 0.1 a need to aggregate 43 orders. This could indicate that the particular company would find it very difficult to aggregate in time on a product level to achieve a sufficient stable distribution of demand. This also implies that it will be unreasonable to disaggregate the demand into demand rates using a simply linear disaggregation model. This research thus supports the findings of [7]. Nielsen et al. [7] underlines that demand rates are in fact non-constant over a planning period, but rather behave in a structured non-constant manner.
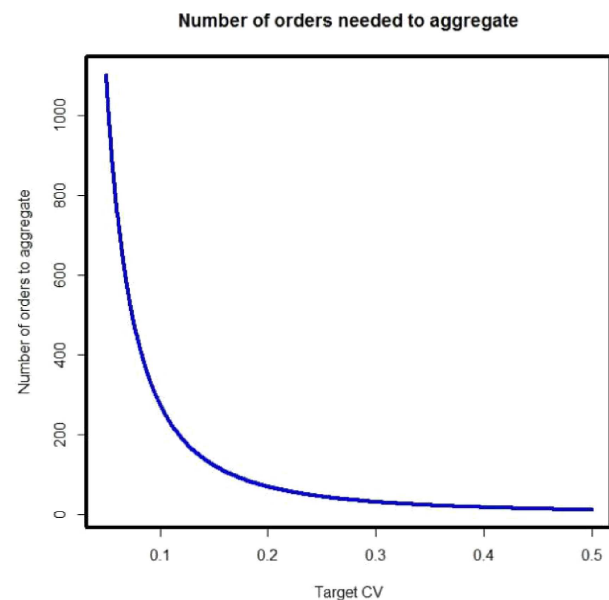

Fig. 7. An overview of the number of orders required to aggregate to achieve a given target CV_0.

Conclusions and further research

This paper addressed the nature order quantity distributions and a potential method to mitigate this through aggregating the orders in a manner ensuring low variation. To get to this point we have taken a number of steps and reach a number of significant conclusions. First, using the method it is possible to establish that the tested case data behaves as expected rather than what is assume in literature. The order quantities are, when the data for the product family is considered as a whole, neither symmetrical-

ly nor independently distributed. Second, the same behaviour as seen for the whole product family is also seen for the majority of the individual products. Although the product level analysis this analysis is limited to products where sufficient data is present to allow one to conduct the analysis. Third, the developed analytical model allows one to calculate the number of orders needed to aggregate to achieve a target CV. This means that it is now possible to estimate the needed aggregation horizon to achieve a target stability in the form of CV. In the particular case aggregating 31 orders means one achieves a CV of approximately 0.3 on the product family set. This aggregation gives an order quantity distribution that is close to symmetrical. The conclusion is that although the individual customer orders are not behaving as expected, it is possible to aggregate over the order series so that demand can be considered to stem from a symmetrical distribution. The reverse implication is of source that order quantities can only be considered to be symmetrically distributed with a sufficiently low CV if orders are in fact aggregated. This underlines that fact that the assumptions used in literature are (at least in this case) far from correct. On a product family level this may not be a problem (in this example a desired CV of 0.1 on aggregate level would require aggregating 277 orders, 0.5% of the total number of orders, c. 2 weeks of orders in the particular case). The trouble, however occurs when the individual products are considered. Here the high CV's indicate a need for aggregating rather much information, and even aggregating over 10 orders would require use data (i.e. orders) gathered over a very long time period. This means that the translation from total demand to demand rates is highly problematic as this is done under the assumption that demand is received completely randomly during a planning period. This assumption of independence of demand is apparently untrue, at least in this case. This also seems to lend empirical support to the assumptions in [19, 20] regarding the stability of aggregate demand and underlines why it can be challenging to model the demand behaviour of individual products, even over a long time horizon.

The method that is presented in this paper has a limited application due to the assumption of a stationary distribution of order sizes. However, if the method is applied in environments with high order frequency (i.e. many orders per planning period) and is used for short term operational purposes a stationary distribution of order sizes seems to be a reasonable assumption.

Future work will focus on the impact on line balancing implied through the very large variation of

order quantities and the challenges this will present in achieving well balanced production lines.

References

[1] Chand S., Hsu V.N., Sethi S., *Forecast, solution, and rolling horizons in operations management problems: A classified bibliography*, Manufacturing and Service Operations Management, 4, 1, 25–43, 2002.

[2] Sitek P., Wikarek J., *A declarative framework for constrained search problems*, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 5027 LNAI, pp. 728–737, 2008.

[3] Eriksen P.S., Nielsen P., *Order quantity distributions in make-to-order manufacturing: At what level of aggregation do they respect standard assumptions?*, IFIP Advances in Information and Communication Technology, 384 AICT, pp. 82–90, 2012.

[4] Bobko P.B., Whybark D.C., *The coefficient of variation as a factor in MRP research*, Decision Sciences, 16, 420–27, 1985.

[5] Bocewicz G., Nielsen I., Banaszak Z., *Iterative multimodal processes scheduling*, Annual Reviews in Control, 38, 1, 113–22, 2014.

[6] Lau H.-S., Lau A.H.-L., *Nonrobustness of the Normal Approximation of Lead-Time Demand in a (Q, R) System*, Naval Research Logistics, 50, 149–66, 2003.

[7] Nielsen P., Nielsen I., Steger-Jensen K., *Analyzing and evaluating product demand interdependencies*, Computers in Industry, 61, 869–76, 2010.

[8] Tadikamalla P.R., *A Comparison of Several Approximations to the Lead Time Demand Distribution*, Omega, 12, 575–81, 1984.

[9] Vollmann W., Berry D., Whybark T.E., Jacobs F., *Manufacturing Planning and Control for Supply Chain Management*, McGraw-Hill, Singapore, 2005.

[10] Silver E.A., *Operations Research in Inventory Management: A Review and Critique*, Operations Research, 29, 628–645, 1981.

[11] Silver E.A., *Inventory management: An overview, Canadian publications, practical applications and suggestions for future research*, INFOR: Information Systems and Operational Research, 46, 1, 15–28, 2008.

[12] Becker C., Scholl A., *A survey on problems and methods in generalized assembly line balancing*, European Journal of Operational Research, 168, 694–715, 2006.

[13] Nilakantan J.M., Ponnambalam S.G., Jawahar N., Kanagaraj G., *Bio-inspired search algorithms to*

solve robotic assembly line balancing problems, Neural Computing and Applications, 26, 6, 1379–1393, 2015.

[14] Nilakantan M.J., Ponnambalam S.G., Jawahar N., *Design of energy efficient RAL system using evolutionary algorithms*, Engineering Computations, 33, 2, 580–602, 2016.

[15] Nicholas J., *Competitive Manufacturing Management: Continuous Improvement, lean production and customer-focused quality*, New York: McGraw-Hill, 1998.

[16] Brander P., Levén E., Segerstedt A., *Lot sizes in a capacity constrained facility-a simulation study of stationary stochastic demand*, International Journal of Production Economics, 93–94, 375–386, 2005.

[17] Tsubone H., Furuta H., *Replanning timing in hierarchical production planning*, International Journal of Production Economics, 44, 53–61, 1996.

[18] Pujawan I.N., Kingsman B.G., *Properties of lot-sizing rules under lumpy demand*, International Journal of Production Economics, 81–82, 295–307, 2003.

[19] Fliedner G., *Hierarchical forecasting: issues and use guidelines*, Industrial Management & Data Systems, 101, 1, 5–12, 2001.

[20] Silver E.A., Pyke D.F., Peterson R., *Inventory Management and Production Planning and Scheduling*, third ed., John Wiley & Sons, 1998.

[21] Graman G.A., Sanders N.R., *Modelling the tradeoff between postponement capacity and forecast accuracy*, Production Planning & Control, 20, 206–215, 2009.

[22] Zotteri G., Kalchschmidt M., Caniato F., *The impact of aggregation level on forecasting performance*, International Journal of Production Economics, 93–94, 479–491, 2005.

[23] Wijngaard J., *On Aggregation in Production Planning*, Engineering Costs and Production Economics, 6, 259–66, 1982.

[24] Axäster S., *Aggregation of product data for hierarchical production planning*, Operations Research, 29, 44–756, 1981.

[25] Axäster S., *On the feasibility of aggregate production plans*, Operations Research, 34, 796–800, 1986.

[26] Hax A.C., Meal H.C., *Hierarchical Integration of Production Planning and Scheduling*, Logistics, M.A. Geisler [Ed.], vol. 1 of Studies in the Management Sciences, North-Holland/American Elsevier, 1973.

[27] Zotteri G., *The impact of distributions of uncertain lumpy demand on inventories*, Production Planning & Control, 11, 32–43, 2000.

[28] Feng K., Rao U.S., Raturi A., *Setting planned orders in master production scheduling under demand uncertainty*, International Journal of Production Research, 49, 13, 4007–4025, 2011.

[29] Nielsen P., Do N.A.D., Eriksen T., Nielsen I., *Towards an analysis methodology for identifying root causes of poor delivery performance*, Foundations of Management, 6, 2, 31–42, 2014.

[30] Theil H., *Linear aggregation of economic relations*, first ed., North-Holland Publishing Company, 1955.

[31] Chen A., Blue J., *Performance analysis of demand planning approaches for aggregating, forecasting and disaggregating interrelated demands*, International Journal of Production Economics, 128, 586–602, 2010.

[32] Fliedner G., *An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation*, Computers & Operations Research, 26, 1133–1149, 1999.

[33] Miao W.E., Gel Y.R., Gastwirth J.L., *A new test of symmetry about an unknown median*, Random Walk, Sequential Analysis and Related Topics – A Festschrift in Honor of Yuan-Shih Chow. World Scientific Publisher, Singapore, pp. 199–214, 2006.

[34] Cabilio P., Masaro J., *A simple test of symmetry about an unknown median*, The Canadian Journal of Statistics, 24, 349–361, 1996.

[35] Mira A., *Distribution-free test for symmetry based on Bonferroni's measure*, Journal of Applied Statistics, 26, 959–972, 1999.

[36] Box G.E.P., Jenkins G., *Time series analysis: forecasting and control*, Holden-Day, 1976.

[37] Walpole R.E., Myers R.H., Myers S.L., Ye K., *Probability & Statistics for Engineers & Scientists*, Prentice Hall, New York, 2002.

[38] Nielsen I., Wójcik R., Bocewicz G., Banaszak Z., *Towards constraint programming driven methodology for online multi-project planning and control*, Information Systems Architecture and Technology, Copyright by Wroclaw University of Technology, ISBN 978-83-7493-857-0, pp. 65–74, 2014.

[39] R-project (webpage) www.r-project.org, 2016.