



UNIVERSITA' DEGLI STUDI DI BERGAMO
DIPARTIMENTO DI INGEGNERIA GESTIONALE
QUADERNI DEL DIPARTIMENTO[†]

Department of Economics and Technology Management

Working Paper

n. 11 – 2009

The impact of aggregation level on lumpy demand management

by

Emilio Bartezzaghi, Matteo Kalchschmidt

[†] Il Dipartimento ottempera agli obblighi previsti dall'art. 1 del D.L.L. 31.8.1945, n. 660 e successive modificazioni.

COMITATO DI REDAZIONE[§]

Lucio Cassia, Gianmaria Martini, Stefano Paleari, Andrea Salanti

[§] L'accesso alla Collana dei Quaderni del Dipartimento di Ingegneria Gestionale è approvato dal Comitato di Redazione. I *Working Papers* della Collana costituiscono un servizio atto a fornire la tempestiva divulgazione dei risultati dell'attività di ricerca, siano essi in forma provvisoria o definitiva.

The impact of aggregation level on lumpy demand management

Emilio Bartezzaghi[♣] and Matteo Kalchschmidt^{*}

♣ Department of Management, Economics and Industrial Engineering - Politecnico di Milano

* Department of Economics and Technology Management - Università degli Studi di Bergamo.
Corresponding author: Viale Marconi 5, 24044 Dalmine (Italy), tel. +39.035.205.2360, fax +39.035.205.2077, email matteo.kalchschmidt@unibg.it

Abstract

Spare parts management is a rather complex issue. One of the reasons of its complexity is the lumpy pattern of the demand that spare parts frequently present. Several methods have been proposed to cope with this particular kind of problem and improvements have been proved compared to classical forecasting techniques. Literature has however devoted minor attention to the choice of aggregation level when demand is lumpy. This paper aims at studying whether aggregating data when demand is lumpy may be beneficial in terms of impact on inventory performances. An installation stock inventory model is considered and aggregation over time is taken into account; in particular for a single item different time buckets are considered and performances are evaluated in terms of service and inventory level. Based on simulation experiments on real demand data coming from the spare parts unit of a multinational white goods manufacturer, we identify that aggregation of data can significantly impact on inventory management performances. A contingency analysis based on demand characteristics allows us to draw some guidelines on when aggregation over time can be beneficial.

1. Lumpy demand management

In the last twenty years companies have always paid great attention on managing demand variability. Demand fluctuations are due to several reasons: quick changes in the final customer's preferences and taste are a common cause of demand variability (e.g., in the fashion industry demand for a given color can change dramatically from year to year). Marketing activities may also lead demand to suddenly change e.g., when promotional activities are conducted due to the high elasticity of demand to price. Competitors can also be a source of variability, since their behavior can influence how the demand distributes on each single company serving a specific market. The supply chain structure is also a significant cause of demand unsteadiness: the bullwhip effect (Lee et al., 1997) is a common phenomenon in different industrial contexts, leading to an increase in the variability of the demand over supply chain stages.

A vast amount of the literature has addressed the issue of designing managerial systems capable of coping with demand variability. This has been done by focusing on different leverages: from demand forecasting, aiming at increasing the capability of companies to understand variability, to production planning, trying to design efficient planning systems, capable of reacting towards sudden changes in the final demand, to inventory management, in order to manage the complex trade-off between inventory cost and service level, and so on.

This issue is common to all industrial contexts; however, a rather peculiar and complex situation is faced in the case of spare parts. The problem of managing spare parts demand is relevant for many reasons: first of all it influences the final product business since it affects post sale service quality. Moreover, it is a relevant business as the market is captive, thus very profitable and so firms have to pay relevant attention towards this issue. However, it is a very difficult business to cope with, since requirements are usually very dispersed over time and demand uncertainty is frequently very high.

Spare parts, in fact, often show very sporadic demand patterns for long periods of their life time. This is the case, for example, of service items that have to be stored for years as long as repair service has to be guaranteed even for products that have reached the end of their market life. Spare parts demand often tend to be highly variable and sporadic showing frequently a very peculiar pattern called *lumpy demand*.

Lumpy demand can be defined as (Bartezzaghi et al., 1999):

- *variable*, therefore characterized by relevant fluctuations (Wemmerlöv and Whybark, 1984; Wemmerlöv, 1986; Ho, 1995; Syntetos and Boylan, 2005);
- *sporadic*, as historical series are characterized by many days with no demand at all (Ward, 1978; Williams, 1982; Fildes and Beard, 1992; Vereecke and Verstraeten, 1994; Syntetos and Boylan, 2005);
- *nervous*, thus leading to show differences between successive demand observation, so implying that cross time correlation is low (Wemmerlöv and Whybark, 1984; Ho, 1995; Bartezzaghi and Verganti, 1995).

Managing inventories when demand is lumpy is thus a complex issue since companies have to cope with both a sporadic pattern, that usually induces high inventory investments, and highly variable order size, that make it difficult to estimate inventory levels and may affect service levels. For this reason, companies facing lumpy demand often experience both high inventory levels and unsatisfactory service levels at the same time.

Lumpiness may emerge as the consequence of different structural characteristics of the market. In particular, we may refer to the following main factors (Bartezzaghi et al., 1999):

- *low number of customers in the market*. Fewer customers usually induce sporadic requests for the product unit and, therefore, demand lumpiness increases;
- *high heterogeneity of customers*. Heterogeneous requests occur when the potential market consists of customers with considerably different sizes or buying behaviors (i.e. customers that order for very different lot sizes or with different frequencies); thus the higher the heterogeneity of customers, the higher the demand lumpiness;
- *low frequency of customers requests*. The higher the frequency of requests from a customer, the higher the number of different customers that ask for the unit in a given time bracket. Thus lumpiness increases as the frequency of each customer's purchase decreases;

- *high variety of customers requests.* Demand lumpiness increases also if each single customer has a variable reorder behavior over time. Customers may change significantly their buying behavior in specific periods of time due, for example, to promotional activities or to speculative buying;
- *high correlation between customers requests.* Lumpiness may occur also because customers' requests are strongly correlated with each other. Correlation, for example, may be due to imitation and managerial fads which induce similar behaviors in customers so that sudden peaks of demand may occur after periods of no requests.

Spare parts demand often shows this specific kind of variability. This is mainly due to the low frequency of requests for these items. In fact spare parts often tend to behave as “slow items” since requests for them are distributed over a long period of life. Besides, requests may change significantly between orders due to the fact that several different kinds of customers may be served by a single service unit. This is often the case when spare parts are ordered from independent units that provide the final customer with repair services. In this situation, reorder sizes are influenced by the specific reorder criteria adopted by each service provider. As a matter of fact, most of the contributions on lumpy demand management have specifically taken into consideration the spare parts case (see for example Croston, 1974; Petrovic et al., 1988; Cobbaert and Van Oudheusden, 1996; Shibuya et al., 1998; Liu and Shi, 1999; Willemain et al., 2004).

Due to its relevant impact on companies performances, lumpy demand management has received major attention in the current literature. Specifically the literature has provided several approaches (i.e. forecasting methods and inventory models) to cope with this kind of demand. Also the specific case of spare parts has been taken into account and specific methodologies have been provided to cope with demand variability in this particular case and proposing different models to improve inventory performance (some of the works in these field are: Petrovic et al., 1988; Cohen and Kleindorfer, 1989; Cobbaert and Van Oudheusden, 1996; David et al., 1997; Dekker et al., 1998; Shibuya et al., 1998; Liu and Shi, 1999; Teunter and Fortuin, 1999; Kalchschmidt et al., 2003; Syntetos and Boylan, 2005).

Literature on lumpy demand management has mainly focused on methods to better evaluate demand variability (i.e. forecasting methods; e.g., Syntetos and Boylan, 2005) and inventory models specific to this particular case (e.g., Teunter and Fortuin, 1999).

However these models usually don't address the problem of the aggregation level of data. This problem arises when the implementation of specific techniques takes place. In fact, when trying to implement forecasting and inventory models, practitioners often find out that this is much more complex than the simple design or selection of an appropriate algorithm and it involves the choice of the relevant pieces of information, the design of information systems, the control of data quality, and the definition of managerial processes. One critical issue concerning the implementation and adoption of forecasting and inventory management techniques is the choice of the appropriate level at which demand has to be managed. In particular, demand has to be defined over three dimensions:

1. one shall define the market he/she tries to forecast; e.g., one retailer might want to forecast demand at the single store level, while a manufacturer might be interested in the demand for the overall region or country; clearly the former forecasting problem is harder to tackle than the latter;
2. one shall define the product the demand refers to; e.g., for a given retailer it might be fairly difficult to predict the demand for a given product at the style-color-size-packaging level, whereas forecasting the total turnover for a given product category might not be that hard (Wacker and Lummus, 2002);
3. finally one needs to define the time frame of the forecasting problem, i.e., one shall define the time bucket and the forecasting horizon; indeed forecasting demand at the day level is much more complex than forecasting total yearly demand.

The choice of the aggregation level is important since according to the specific aggregation level chosen, demand variability may show specific peculiarities and thus different techniques may apply, thus affecting forecasting and inventory performances.

In the remainder of this work we will refer to the previous three dimensions as the *level of aggregation* of the forecasting problem. The smaller the market, the more detailed the definition of the product and the smaller the time bucket, the more the forecasting problem is detailed.

This work focuses on the impact of the aggregation level of data on inventory performances and we address in particular the specific case of lumpy demand. In fact limited contributions can be found regarding how the aggregation level may influence lumpy demand.

The remainder of this paper is thus structured as follows. In the next section the level of aggregation of demand will be described and literature contributions on this issue will be summarized. Then specific research objectives and methodology will be described. In the last two sections, empirical results will be described, a proper discussion of results will be provided and future research objectives will be highlighted.

2. The impact of data aggregation

As previously mentioned, when a forecasting problem has to be addressed, it is important to clearly state at which level of aggregation a forecast has to be provided. Typically this choice relates to the specific decision the forecast will be used for. For example, when yearly financial budget is under consideration, a company usually doesn't need a very detailed forecast: a forecast for some future months of sales at market level is going to be enough for this specific decision making process. On the contrary, if inventory decision is under consideration, probably a company will need to provide a forecast at the Stock Keeping Unit (SKU) level, for the next future days or weeks and regarding the part of the market that the specific warehouse is serving.

However, the level of aggregation at which the forecast has to be used (at thus provided) is not necessarily the same of the level of aggregation at which the forecast is evaluated. In particular, during the forecasting process, we face the problem of aggregation in two different situation. First, when information (i.e., past demand data) is collected, one has to choose at which level of aggregation these data should be used (we refer to this as the *data aggregation process*). Based on what companies choose here, different forecasting methods may then be selected, based on the characteristics of variability the demand shows at that level of aggregation (e.g., if monthly data is used then seasonality may be an important variability component to be considered; on the contrary if the same data is used at the yearly level seasonality becomes irrelevant, at least for the forecast evaluation process). Second, when the forecast has been evaluated it may need

to be aggregated or disaggregated in order to provide the final forecast at the required aggregation level that the decision making process needs (e.g., if a market forecast is needed for budget purposes, if we evaluate forecast at customer level we then need to aggregate all these forecasts by simply summing up them). We refer to this as the *forecast aggregation process*. Figure 1 summarizes these different situations. In this work we will focus on the data aggregation process.

As figure 1 exemplifies, given a certain aggregation level at which a forecast is needed, one can decide to obtain this forecast through a more aggregate forecast (thus a disaggregation process is needed to provide the final forecast), through a more disaggregated process (thus an aggregation is needed) or without any aggregation or disaggregation. Forecast evaluation can be performed at different data aggregation levels. Data in fact can be aggregated before evaluating the forecast or disaggregated (e.g., in the retail industry frequently companies estimate the demand in different ways according to whether data collected on demand refer to promotional periods or not).

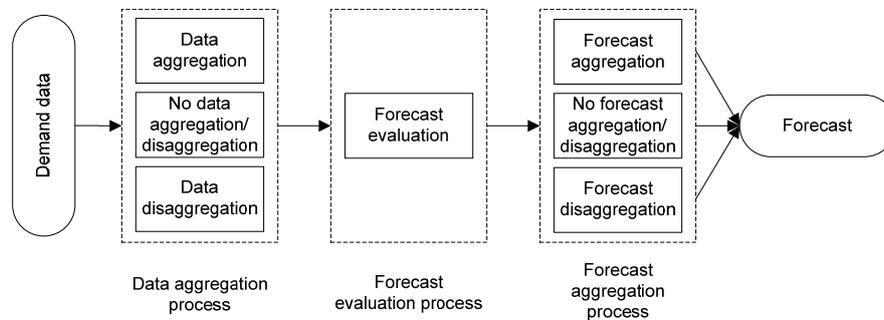


Fig. 1. Aggregation level options in the forecasting process

Based on these options companies may then structure their forecasting process differently. For example, one typical solution is when the forecast is evaluated at the same level of aggregation at which the forecast is used. In this situation, based on the aggregation level requested from the decision making process, the forecasting approach is selected according to data or information available, and no aggregation or disaggregation of the data is done. Another possibility is when the forecast is evaluated at a more detailed level compared to which the forecast is used. In this case, the forecast is evaluated

based on disaggregated data and then some aggregation of these is conducted before the final forecast is provided. This is often referred to as *Bottom-Up* approach (Orcutt et al., 1968; Zellner and Tobias, 2000; Weatherford et al., 2001). A typical example of this situation is when sales budget are developed: usually sales people are asked to provide their own forecast regarding their specific geographical area or regarding their specific customers. These forecasts are then aggregated all together to get to an overall market forecast. Another common situation is when the forecast is evaluated at a more aggregate level compared to which the forecast is used. In this situation, the data collection and the forecast evaluation is done at a more aggregate level compared to the one at which the forecast is then provided. This is often referred to as *Top-Down* approach (Theil, 1954; Grunfeld and Griliches, 1960; Lapide, 1998). For example, this is often used when a weekly forecast has to be divided at the daily level for production scheduling purposes, based on some estimation of the daily seasonality.

As previously mentioned, this choice has to be done on one or more of three different dimensions:

- *Product*: companies have to choose at which level of detail of the product structure they want to evaluate forecasts. The demand can be foreseen by leveraging on very detailed data (e.g., referring to SKU) or very aggregate ones (e.g., referring to product families).
- *Market*: the demand can be foreseen by taking into consideration very precise and detailed information (e.g., the demand of each single customer) or very aggregate one (e.g., the demand at market level).
- *Time bucket*: in the end, companies have to choose whether they want to rely on detailed or aggregate time buckets: demand can be foreseen at daily level, weekly level, monthly level, etc.

The literature on demand management and forecasting has devoted some attention to the problem of choosing the proper level of aggregation. Some contributions on this issue focus on the use of aggregation to estimate seasonality curves (Dalhart, 1974; Withycombe, 1989; Bunn and Vassilopoulos, 1993, 1999; Dekker et al., 2004). These works provided evidence that aggregating correlated time series can be helpful to better estimate seasonality since it can reduce random variability. Other works focus on the selection of the proper level of data aggregation (e.g., Chan, 1993; Gonzales, 1992; Weiss,

1984). Some authors argue that the top-down approach (i.e., evaluating forecast at aggregate level and then dividing it at detailed level) can be helpful as it is more efficient and more accurate in times of stable demand (Theil, 1954; Grunfeld and Griliches, 1960; Lapide, 1998). Other authors, however, reply that the bottom-up approach (i.e., building a forecast by evaluating the forecasts at detailed level and then aggregating them) is needed when there are differences across time series (Orcutt et al., 1968; Zellner and Tobias, 2000; Weatherford et al., 2001). Finally, other contributions (Miller et al., 1976; Barnea and Lakonishok, 1980; Fliedner, 1999) take a more contingent approach and show that the choice between the aggregate and detailed approach depends on the correlation among time series. Zotteri and Kalchschmidt (2007) analytically demonstrate that in fact aggregation is preferable only under certain circumstances (i.e., high demand variability, few periods of demand, etc.).

Limited contributions can, however, be found regarding aggregation in the case of lumpy demand. Specifically, several of the mentioned contributions considered frequently the case of stationary and continuous demand. Unfortunately this is not always the case: spare parts usually show a lumpy pattern and it is not completely clear whether literature findings still hold here.

Similarly, contributions on the aggregation level selection have mainly focused on forecasting, i.e., the impact of aggregation on forecasting accuracy. Limited contributions have considered simultaneously the impact on forecasting and inventory management systems. In fact, literature on lumpy demand management has argued and, sometimes, proved that designing an integrated forecasting and inventory management system may be much more beneficial than focusing on just one of the two (Kalchschmidt et al., 2003). In this situation the forecasting method applied has to focus on estimating demand characteristics that the chosen inventory system needs to define reorder policies.

This work aims at providing a better understanding of how aggregation may influence inventory performance when demand is lumpy. In particular here attention is devoted to the case of aggregation over time (temporal aggregation). This choice is due to the fact that limited contributions can be found on this specific issue (also in the stationary and stable demand case). As a matter of fact, the vast amount of contributions on this topic

usually refer to aggregation over product and over market dimensions (see previously mentioned contributions), while only limited contributions can be found regarding temporal aggregation for non-lumpy demand (some contributions can be found in Johnston and Harrison, 1986; Snyder et al., 1999; Dekker et al., 2004) and very few specific to the lumpy demand case (Nikolopoulos et al., 2009). For all these reasons, in the remainder of the paper only temporal aggregation will be considered.

3. Objectives and methodology

The goal of this work is to study whether temporal aggregation of lumpy demand may be beneficial in terms of impact on inventory performances. Specifically, the objectives are:

1. Analyze the impact of temporal aggregation level in a specific situation, namely spare parts demand.
2. Evaluate the impact of demand characteristics (e.g., lumpiness) on the choice of the proper level of aggregation.

In order to achieve these goals a simulation analysis based on real demand data has been considered. Demand data has been collected from the Spare Parts Management Division of a major multinational white goods manufacturer. The company provided us with daily level demand data for all its spare parts SKUs over almost one year period (specifically 209 working days). The company manages more than 68.000 SKUs; among these, almost 52.000 have less than two orders per year. We decided to focus only on those items that presented at least two orders over the available data set. Then we selected from 16.875 SKUs that guaranteed this requirement a sample of 1.000 SKUs chosen at random. Among these, 926 SKUs at the end were considered (some SKUs were omitted due to item specific problems, such as item recoding or termination).

This data set was divided in two samples: the first one, based on the first 105 days was allocated for fitting purposes of the selected model, while the second one (based on the remaining 104 days) was used for testing performance. Table 1 synthesizes some descriptive statistics on the overall sample.

Demand			Demand Interarrival		Demand Size		Order frequency
Average (units)	CV		Average (days)	CV	Average (units per order)	CV	(n. of days with non zero demand)
Min	0,01	0,86	1,00	0,00	1,00	0,00	2,00
25%ile	0,03	3,89	7,21	0,66	1,01	0,00	4,00
Median	0,10	6,19	18,34	0,84	1,81	0,64	9,00
75%ile	0,44	8,46	40,20	1,01	4,00	1,20	28,00
Max	53,44	13,58	104,50	4,09	269,67	4,84	209,00

Table 1. Descriptive statistics for the considered sample

As it can be noted the demand variability is quite high (median CV is above 6). This is due to both variability in the demand size (median CV of demand size is 0.64) but also to the demand intermittency (median n. of days with non zero demand is 9 out of 209 days of demand). Thus, coherently with our definition of lumpy demand and with previous contributions (Syntetos and Boylan, 2005), we can conclude that considered data has in fact a lumpy pattern (both size is variable and demand is intermittent).

To achieve our research goals, we based our analyses on a simulation model with the following characteristics:

1. In order to estimate inventory levels, we adopted Syntetos and Boylan (S&B) unbiased variation of Croston's method (Syntetos and Boylan, 2001). We selected this forecasting approach since it is actually considered a reliable method for the case of lumpy demand compared to other known solutions (Syntetos and Boylan, 2001 and 2005; Altay et al., 2008). This approach reduces the bias in the estimation of the average demand in case of lumpy pattern. We refer to Syntetos and Boylan (2001) for a detailed description of the approach and to Kalchschmidt et al. (2003) and Syntetos and Boylan (2005) for comparisons with other methodologies in the case of lumpy demand. The smoothing parameters were set at 0.2 and the C parameter was set to 200 (see cited papers for details on the model). Some tests were also run with other values for these parameters. Even if differences

arise when parameters change, these do not affect significantly the results of our analyses. For this reason and for brevity sake we omitted these analyses here.

2. The safety stock is defined according to the actual demand variability and the desired service level. Specifically we simulated different scenarios according to different average service levels i.e., 94% (the average service level the company was achieving) and 99.1% (the desired service level the company was aiming to have). We considered these two levels of performance since they represent what the company considered as reference. For brevity sake we will show the results only for the former case.
3. The reorder model considered is an order-up-to system with daily revisions of inventory levels; backlog is allowed.
4. Deliveries from suppliers are assumed constant and equal to 20 days for all items. The company based its own reorder politics on this specific value. We argue that according to the specific lead time suppliers provide, the impact of the aggregation level on inventory performances may change. However, we claim that the considerations we draw from our analyses are not affected explicitly from this specific assumption. We discuss this issue in deeper detail in the conclusions.

Each day of the simulation we update model parameters and evaluate inventory performances in terms of inventory levels and service level (i.e., served quantity compared to actual demand). If inventory is not enough to fulfill daily demand a backlog is accounted and demand is served as soon as inventory is available.

Simulations were run according to five different aggregation levels of demand data. Specifically we considered aggregation on a one day level (1d, data as it is), a two days level (2d, demand is aggregated between two subsequent days), three days level (3d), ten days level (10d) and thirty days level (30d). Other intermediate aggregation levels were run but here they are omitted for sake of brevity. Since the performances of the systems under investigation are based on two objectives (service level and inventory level), in order to compare the different scenarios we run all simulations so to guarantee a 94% service level on the average of the test period. We then can directly compare inventory levels to identify the impact of the data aggregation process.

4. Simulation results and discussion

Figure 2 shows the average inventory level of the considered items on the testing period for different aggregation levels.

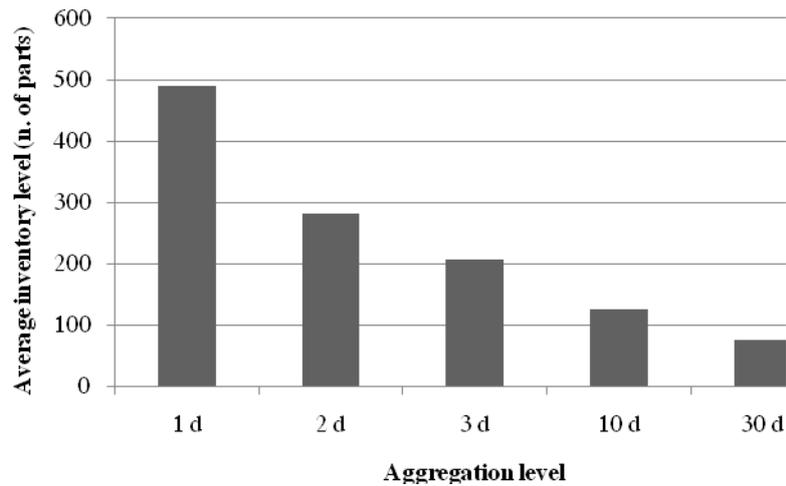


Fig. 2. Average inventory levels for the different temporal aggregation levels considered (average service level is 94.1%)

As it can be noted, the average inventory level required to guarantee a 94% average service level reduces as we aggregate demand data. The extent to which inventories benefit from aggregation is impressive (in particular when comparing the more detailed levels with the more aggregate ones) and the benefit of further aggregation tends to reduce on higher horizons. In order to verify that these average results were consistent at SKUs level (and to avoid eventual bias due to few peculiar cases, e.g., high volume skus) we ran nonparametric tests on the equality of average inventories between the different simulation runs (we based our analyses on Friedman's test¹). All tests were significant at 0.99 level, thus we can claim that on a relevant portion of our SKUs, aggregating demand improves inventory performances.

¹ The Friedman test is the nonparametric equivalent of a one-sample repeated measures design or a two-way analysis of variance with one observation per cell. Friedman tests the null hypothesis that k related variables come from the same population. For each case, the k variables are ranked from 1 to k . The test statistic is based on these ranks.

Even if on average the temporal aggregation seems to pay off, a more detailed analysis showed that this is not true for all items. Table 2, in fact, highlights that some items don't benefit from aggregation but, on the contrary, face a worsening of the inventory level. As we can see, among the considered SKUs, on average almost 22% show worse performance when demand is aggregated, while almost 9% on average are not affected by the aggregation level.

	From 1 day to 2 days	From 2 days to 3 days	From 3 days to 10 days	From 10 days to 30 days
Improvement	817	529	711	517
Indifference	59	80	76	105
Worsening	50	317	139	304
Total	926	926	926	926

Table 2. Distribution of SKUs for different aggregation levels, classified according to whether they improve performance by aggregating demand, stay the same, or worsen.

The identified phenomenon seems to apply differently on the items considered, thus we take a contingent approach to identify what are the key drivers that influence the optimal aggregation level. In order to identify discriminant contingent factors we ran multiple comparisons among three groups of items (namely those for which aggregation improves performance, those where aggregation is indifferent and those where aggregation lead to worse performance) for all the considered aggregation levels. T-tests on the equality of means were run among the different groups on the following variables²:

- Average demand;
- Standard deviation of demand;
- Coefficient of variation of demand;
- Asymmetry of demand;

² For space sake we omit all statistical analyses. All contingencies have been evaluated at daily level since this was the most detailed level available.

- Lumpiness of demand: lumpiness has been measured according to the following expression:

$$Lumpiness = \frac{CV^2}{\mu \cdot LT}$$

where CV is the coefficient of variation of demand, μ is the average demand and LT is the replenishment lead time;

- Number of days with non zero demand;
- Average size of demand;
- Standard deviation of demand size;
- Coefficient of variation of demand size;
- Average interarrival;
- Standard deviation of interarrival;
- Coefficient of variation of interarrival;

Among all, two variables seem to be constantly changing between the considered groups at the different aggregation levels: coefficient of variation of demand size (CV_s) and average interarrival between successive orders. These two measures are negatively correlated among themselves (Pearson correlation index is -0.51 with $p < 0.001$) due to the fact that both measures are affected by the number of days of non zero demand. In order to get rid of the effect of the days of actual demand, we define a standard coefficient of variation of demand size as follows:

$$CV_s^2 = \frac{CV^2}{\sqrt{n}}$$

where CV_s is the coefficient of variation of demand size and n is the number of days of non zero demand. This indicator reduces the bias that CV_s has due to the number of days of actual observations.

Figure 2 shows the distribution of the 926 SKUs considered according to these two variables.

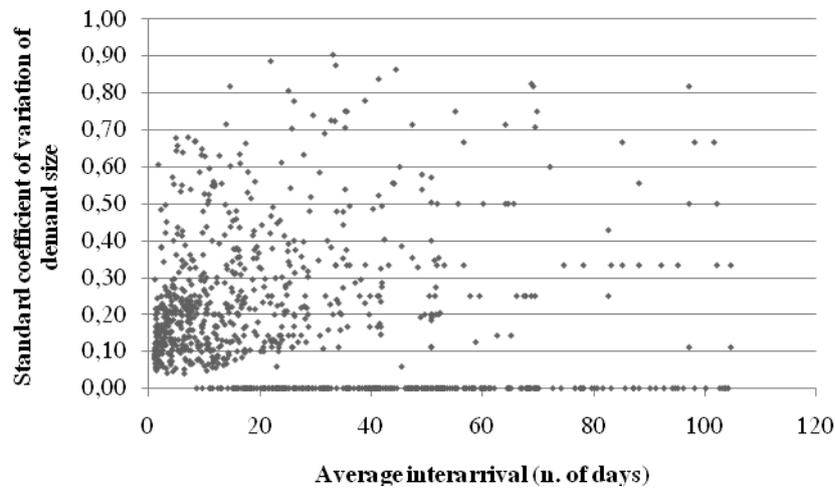


Fig. 2. SKUs distribution according to the standard coefficient of variation of demand size ($\frac{\sigma}{\mu}$) and the average interarrival

Based on these two variables, we divided the different SKUs in clusters. In particular, we ran first a hierarchical cluster analysis, in order to identify the proper number of groups. Specifically, we applied a hierarchical cluster analysis with between-groups linkage based on Squared Euclidean distance. Through the analysis of the dendrogram, we identified four as a proper number of clusters. A k-means cluster analysis with four as number of clusters led us to identify the following groups of items (Table 3 provides information regarding the final cluster centroids):

- High variability and highly sporadic (HVHS): these items are characterized by both high variability of demand size and sporadic pattern (on average less than 4 orders per year).
- Low variability and highly sporadic (LVHS): these items show a sporadic pattern (on average less than 4 orders per year) but demand size tends to be quite stable.
- High variability and sporadic (HVS): these items are ordered more frequently (on average more than 13 orders per year), however with highly variable quantities
- Low variability and sporadic (LVS): these items are ordered frequently (on average more than 13 orders per year) and with very stable quantities.

	Std. CV of demand size			Interarrival	
	N	Mean	Std. Dev.	Mean	Std. Dev.
HVHS	99	0,450	0,198	56,262	19,603
LVHS	237	0,011	0,038	50,798	21,020
HVS	136	0,472	0,129	15,856	8,828
LWS	454	0,154	0,076	10,364	8,281
Combined	926	0,196	0,194	26,426	24,267

Table 3. Descriptive statistics of final clusters centroids

The clusters obtained are also coherent to previous contributions that group SKUS according to similar variables (e.g., Syntetos et al., 2005).

In order to compare the different aggregation levels, we defined the Average Inventory Reduction (AIR) as the average percentage reduction of inventories between two different aggregation levels. In particular, AIR is defined as:

$$AIR[i - j] = \frac{1}{k} \sum_k \frac{Inv. Lev. [i]_k - Inv. Lev. [j]_k}{Inv. Lev. [j]_k}$$

Where $Inv. Lev. [i]_k$ is the inventory level for item k at aggregation level i.

Based on these clusters we evaluated the average inventory reduction (AIR) between the different aggregation levels within each cluster. Table 4 summarizes this comparison.

These results show again that improvement in inventory performances are widespread in the sample, and thus they confirm our previous evidence. Quite interestingly, however the improvements obtained through a more aggregated forecast are significant when the demand is not highly sporadic. In fact, when sporadic behavior is limited all reductions are significant (based on pair comparisons at SKU level). Significant benefits may occur also when the demand is highly sporadic but only if the variability is limited; in fact, in most of the comparisons there is a significant reduction even if in one case a

significant increase can also be seen. When the demand is highly sporadic and the demand size is highly variable, no significant improvements can be found; quite interestingly even if on average some reductions can still be found here, they are not statistically significant.

		Interarrival			
		Sporadic	Highly Sporadic		
Demand size variability	High	AIR[2-1]:	-51.1%	AIR[2-1]:	-20.0%
		AIR[3-2]:	*	AIR[3-2]:	-10.9%
		AIR[10-3]:	-33.6%	AIR[10-3]:	+4.1%
		AIR[30-10]:	*	AIR[30-10]:	+9.5% *
			-47.1%		
			*		
	Low	AIR[2-1]:	-39.5%	AIR[2-1]:	-44.3%
		AIR[3-2]:	*	AIR[3-2]:	*
		AIR[10-3]:	-23.8%	AIR[10-3]:	-12.5%
		AIR[30-10]:	*	AIR[30-10]:	-19.8%
			-37.5%		*
			*		+18.5%
	-44.3%		*		
	*				

Table 4. Average inventory reduction for each cluster between the different aggregation levels (AIR[i-j]: average inventory reduction with *i* days aggregation level compared to *j* days aggregation level; * $p < 0.05$, based on pair comparison of each SKU)

These results suggest that aggregating demand seems to be a reliable approach when demand is lumpy. However when demand sporadic behavior and variability are extremely high (i.e. HVS skus), this approach is not helpful and can in some cases lead to worse performances. This result eventually suggests that in this last situation, demand forecasting can be highly inefficient and one should design inventory management solutions based on other approaches. In our case it should also be noted that this situation affects only a limited part of the inventory problem we are addressing. In fact these items

account for no more than 10% of the considered SKUs that are responsible for less than 2% of demand volumes and less than 1% of the average inventory level.

5. Conclusions

This work provides evidence that the temporal aggregation of data may be beneficial in spare parts inventory management and forecasting. The presented results show a clear effect of the aggregation of data over inventory performance, thus they emphasize the importance of paying proper attention in defining the aggregation level at which demand is managed. This consideration is coherent with previous results on this topic in the case of non-lumpy demand (see literature review for details) and provides evidence that also when demand is sporadic or lumpy, this issue has to be taken in high consideration.

A second contribution relates to the contingent analysis of the impacts of aggregation. The analyses show that even if the impact is usually significant, the characteristics of the demand significantly influence the possibility of improving inventory performances by leveraging on temporal aggregation. In particular, results provide evidence that when the demand is sporadic, impressive inventory reduction can be gained by leveraging on data aggregation. However when the demand occurrence is extremely low (in our case less than four orders in one year), leveraging on data aggregation may be effective if variability in demand size is not extremely high. On the contrary, if both sporadic nature of demand and variability of demand size are extremely high, impacts can be limited. This last situation, however, is limited to few cases in our sample (almost 10% of considered SKUs). This result is coherent with other contributions in the field, claiming that when demand lumpiness is too high, companies should not invest too much in forecasting those patterns due to the extreme uncertainty of the situation.

This work also highlights some interesting issues that future studies should devote attention to. First of all, it would be important to define criteria that can provide companies with a clear a-priori determination of the aggregation level they should adopt. In fact this work, provides some guidelines for companies willing to understand whether

they should aggregate data or not. Such a contribution is important for managers since it can provide them with some guidelines to better manage their spare parts inventories.

We would also like to draw attention to some limitations of this work. First of all, we considered a specific situation in terms of data (available from one single company), thus one can doubt about the possibility to generalize these results. We argue, however, that these results are at some extent of general validity since even if the data come from a single company they represent a typical situation faced in the spare parts business. Indeed future studies should consider other data sets from other companies to verify these results. A second issue relates to the specific forecasting technique that we adopted to manage demand. This work focuses on one specific forecasting method (i.e. Syntetos and Boylan's method). It would be important to evaluate to which extent these results are method-specific and thus how the selection of the aggregation level should take the forecasting method adopted into account. We consider that some specificities of the applied method may have an impact since different methods rely on the estimation of different parameters that can be influenced heterogeneously by the aggregation level. We argue, however, that our results still constitute a relevant contribution for this topic, also due to the fact that the adopted method is considered to provide superior performances compared to others for the specific case of lumpy demand.

In the end, we would like to draw attention on the assumptions we made on suppliers lead time. As we mentioned, we assumed lead times constantly equal to 20 days for all items. Our results are for sure influenced by this assumption that we made in order to simplify analyses. We argue, however, that the overall conclusions of our work is not affected by this supposition. We claim that relaxing this assumption would be important for providing more reliable guidelines for companies and thus future works should address this topic.

References

- Altay, N., Rudisill F., Litteral, L.A., 2008, Adapting Wright's modification of Holt's method to forecasting intermittent demand, *International Journal of Production Economics*, 111 (2):389-408
- Barnea, A., Lakonishok, J., 1980. An analysis of the usefulness of disaggregated accounting data for forecasts of corporate performance. *Decision Sciences* 11, 17–26.
- Bartezzaghi E., R. Verganti, 1995, Managing demand uncertainty through order overplanning, *International Journal of Production Economics*, 40, 107-120
- Bartezzaghi E., Verganti V., Zotteri G., 1999, A simulation framework for forecasting uncertain lumpy demand, *Int. J. Production Economics*, 59: 499-510
- Bunn, D.W., Vassilopoulos, A.I., 1993. Using group seasonal indices in multi-item short-term forecasting. *International Journal of Forecasting* 9, 517–526.
- Bunn, D.W., Vassilopoulos, A.I., 1999. Comparison of seasonal estimation methods in multi-item short-term forecasting. *International Journal of Forecasting* 15, 431–443.
- Chan, W., 1993. Disaggregation of annual time-series with common seasonal patterns. *Journal of Econometrics* 55, 173–200.
- Cobbaert K., D. Van Oudheusden, 1996, Inventory models for fast moving spare parts subject to sudden death obsolescence, *International Journal of Production Economics*, 44, 239 – 248
- Cohen M.A., P.R. Kleindorfer, 1989, Near – optimal service constrained stocking policies for spare parts, *Operations Research*, 37 (1)
- Croston J.D., 1974, Stock level for slow-moving items, *Operational Research Quarterly*, 25 (1):123-130
- Dalhart, G., 1974. Class seasonality—a new approach. In: *Proceedings of 1974 Conference of American Production and Inventory Control Society*, Reprinted in *Forecasting*, 2nd ed. APICS, Washington, DC, pp. 11–16.
- David I., E. Greenshtein, A. Mehrez, 1997, A dynamic programming approach to continuous review obsolescent inventory problem, *Naval Research Logistics*, 44 (8)
- Dekker R., M.J. Kleijn, P.J. De Rooij, 1998, A spare parts stocking policy based on equipment criticality, *International Journal of Production Economics*, 56 – 57, 69 – 77
- Dekker, M., Van Donselaar, K., Ouwehand, P., 2004. How to use aggregation and combed forecasting to improve seasonal demand forecast. *International Journal of Production Economics* 90, 151–167.
- Fildes R., C. Beard, 1992, Forecasting system for production and inventory control, *International Journal of Operation and Production Management*, 12 (5), 4-27

- Fliedner, G., 1999. An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation. *Computers & Operations Research* 26, 1133–1149.
- Gonzales, P., 1992. Temporal aggregation and systemic sampling in structural time-series models. *Journal of Forecasting* 11, 271–281.
- Grunfeld, Y., Griliches, Z., 1960. Is aggregation necessarily bad? *Review of Economics and Statistics* 42, 1–13.
- Ho C.J., 1995, Examining the impact of demand lumpiness on the lot-sizing performance in MRP systems, *International Journal of Production Research*, 33 (9), 2579-99.
- Johnston, F.R., Harrison, P.J., 1986. The variance of lead-time demand. *Journal of the Operational Research Society* 37, 303–308
- Kalchschmidt M., Zotteri G., Verganti R., 2003, Inventory management in a multi-echelon spare parts supply chain, *Int. J. Production Economics* 81–82: 397–413
- Lapide, L., 1998. New developments in business forecasting. *The Journal of Business Forecasting* Summer, 28–29
- Lee, H.L., Padmanabhan V., Whang S.J., 1997, Information Distortion in a Supply Chain: The Bullwhip Effect. *Management Science*, Vol. 43, No. 4), pp. 546 – 558.
- Liu L., D.H. Shi, 1999, An (s, S) model for inventory with exponential lifetimes and renewal demands, *Naval Research Logistics*, 46 (1)
- Miller, J.G., Berry, W., Lai, C-Y.F., 1976. A comparison of alternative forecasting strategies for multi-stage production inventory systems. *Decision Sciences* 7, 714–724.
- Nikolopoulos K., Syntetos A.A., Boylan J.E., Petropoulos F., Assimakopoulos V., 2009, ADIDA: An aggregate – disaggregate intermittent demand approach to forecasting, Working paper: 330/09, University of Salford, ISSN 1751-2700
- Orcutt, G., Watts, H.W., Edwards, J.B., 1968. Data aggregation and information loss. *American Economic Review* 58, 773–787.
- Petrović R., A. Šenborn, M Vujošević, 1988, A new adaptive algorithm for determination of stocks in spare parts inventory systems, *Engineering Costs and Production Economics*, 15, 405 – 410
- Shibuya T., T. Dohi, S. Osaki, 1998, Optimal continuous review policies for spare parts provisioning with random lead times, *International Journal of Production Economics*, 55, 257 – 271
- Snyder, R.D., Koehler, A.B., Ord, J.K., 1999. Lead-time demand for simple exponential smoothing. *Journal of the Operational Research Society* 50, 1079–1082.
- Syntetos A.A., Boylan J.E., 2001, On the bias of intermittent demand estimates, *International Journal of Production Economics*, 71: 457-466
- Syntetos A.A., Boylan J.E., 2005, The accuracy of intermittent demand estimates, *International Journal of Forecasting*, 21: 303-314

- Teunter R.H., L. Fortuin, 1999, End – of - life service, *International Journal of Production Economics*, 59, 487 – 497.
- Theil, H., 1954. *Linear Aggregation of Economic Relations*. North-Holland, Amsterdam.
- Vereecke A., P. Verstraeten, 1994, An inventory model for an inventory consisting of lumpy items, slow movers and fast movers, *International Journal of Production Economics* 35, 379-389
- Wacker, J.G., Lummus, R., 2002. Sales forecasting for strategic resource planning: practical implications and research directions. *International Journal of Production and Operations Management* 22 (9), 1014–1031
- Ward J.B., 1978, Determining reorder points when demand is lumpy, *Management Science*, 24 (6), 623-632.
- Weatherford, L.R., Kimes, S.E., Scott, D.A., 2001. Forecasting for hotel revenue management—testing aggregation against disaggregation. *Cornell Hotel and Restaurant Administration Quarterly* August, 53–64.
- Weiss, A.A., 1984. Systematic sampling and temporal aggregation in time-series models. *Journal of Econometrics* 12, 547–552
- Wemmerlöv U., D.C. Whybark, 1984, Lot-sizing under uncertainty in a rolling schedule environment, *International Journal of Production Research*, 22 (3), 467-484.
- Wemmerlöv U., 1986, A time phased order-point system in environments with and without demand uncertainty: A comparative analysis of no-monetary performance variables, *International Journal of Production Research*, 24 (2), 343-358.
- 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.M., 1982, Reorder levels for lumpy demand, *Journal of the Operational Research Society*, 33, 185-189.
- Withycombe, R., 1989. Forecasting with combined seasonal indices. *International Journal of Forecasting* 5, 547–552.
- Zellner, A., Tobias, J., 2000. A note on aggregation, disaggregation and forecasting performance. *Journal of Forecasting* 19, 457–469
- Zotteri G., Kalchschmidt M., 2007. A model for selecting the appropriate level of aggregation in forecasting processes, *International Journal of Production Economics*, 108, 74-83