Newcastle University e-prints

Date deposited: 11 March 2010

Version of file: Author final, conference proceeding

Peer Review Status: Peer Reviewed

Citation for published item:

Moon S; Hicks C; Simpson A. [The development of a hierarchical forecasting method for predicting spare parts demand in the South Korean Navy](#). In: *Sixteenth Working Seminar on Production Economics*.2010,Innsbruck, Austria

Further information on publisher website:

Publishers copyright statement:

Use Policy:

The full-text may be used and/or reproduced and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not for profit purposes provided that:

- A full bibliographic reference is made to the original source
- A link is made to the metadata record in DRO
- The full text is not change in any way.

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

<p>Robinson Library, University of Newcastle upon Tyne, Newcastle upon Tyne. NE1 7RU. Tel. 0191 222 6000</p>
--

The Development of a Hierarchical Forecasting Method for Predicting Spare Parts Demand in the South Korean Navy

Seongmin Moon¹, Christian Hicks² and Andrew Simpson¹

¹ Newcastle University Business School, Citywall, Citygate, St James Boulevard, Newcastle upon Tyne, NE1 4JH, UK seongmin.moon@ncl.ac.uk, andrew.simpson@ncl.ac.uk

² Newcastle University Business School, Armstrong Building, Newcastle upon Tyne, NE1 7RU, UK chris.hicks@ncl.ac.uk

Abstract

In the South Korean Navy the demand for many spare parts is infrequent and the volume of items required is irregular. This pattern, known as non-normal demand, makes forecasting difficult. This paper uses data obtained from the South Korean Navy to compare the performance of various forecasting methods that use hierarchical and direct forecasting strategies for predicting the demand for spare parts. A simple combination of exponential smoothing models was found to minimise forecasting errors. A simulation experiment verified that this approach also minimised inventory costs.

Keywords: Hierarchical forecasting; Spare parts demand; Non-normal demand; Intermittent demand; Simulation

1. Introduction

The operational availability of weapon systems is highly dependent upon the timely supply of spare parts [1]. Operational availability is defined as “the ratio of time available when needed to the total time needed” [1, p. 260]. In order to sustain operational availability at a specified level, an adequate supply of spare parts to meet the requirements of repair and maintenance is necessary. In practice it is common for military forces to hold a large stock of spare parts. However, there is often little or no demand for a large proportion of the stock items. For example, the British Navy holds almost twice as many spare parts as are expected to be required [2]. The US Department of Defence (DoD) holds a 60% excess of spare parts, with 18% of the inventory (\$1.5 billion) having no demand [3]. Moreover, the ability to stock spare parts is constrained by limited budgets [4].

Hinton [3] claimed that the problems relating to spare parts inventory in the DoD arose because of inaccurate forecasting of inventory requirements. Although the procurement of spare parts may be initiated to meet specific requests, it is common for requirements to change after items have been ordered [3]. The South Korean military experience spare parts supply problems caused by inaccurate forecasts of spare parts demand [5].

A time series is defined as “a collection of observations made sequentially through time” [6, p. 1]. “A stochastic process is said to be a Gaussian (normal) distribution if the joint distribution of any set of Y_t 's is multivariate normal” [6, p. 36]. Demand that is characterised as having infrequent demand occurrences (intermittent demand), low average demand volumes (slow moving demand) or highly variable demand volumes (erratic demand) is an example of non-normal demand [7]. A large part of the time series for spare parts demand exhibits non-normal characteristics [8, 9]. The non-normal nature of the spare parts demand makes forecasting difficult [10].

A time series for individual items is known as an item level time series. An aggregated time series for more than two items is called a group level time series. A multi-level time series structure consists of item level time series and a group level time series in which the items are members. This is known as a hierarchical structure [11]. A forecasting strategy which ignores the hierarchical structure of time series and simply generates a forecast is variously known as a traditional forecasting, independent forecasting, or direct forecasting (DF) [12, 13]. A

forecasting strategy which derives forecasts at item level by prorating demand forecasts for the group in which the items are members is variously known as a family-based forecasting, pyramidal forecasting, dependent forecasting, derived forecasting, or hierarchical forecasting (HF) [12-14]. There are two sub-strategies for item level forecasts. Top-down forecasting (TDF) models a forecast at the top group level using the top group level time series of a hierarchical time series, and then creates lower level forecasts according to the item's percentage contribution within the group [15]. Combinatorial forecasting (CF) models forecasts at all levels of a hierarchical time series using all levels of the time series, and then creates lower level forecasts based on a combination of the forecasts at all levels [11, 16, 17]. When an item level time series is volatile and intermittent, a higher group level time series is usually less volatile and less intermittent. This is because the volatility and intermittency of an item level time series can be offset by other item level time series in the group [18]. This lower level of volatility and less intermittency of a group level time series can produce a more reliable item level time series forecast by using a hierarchical forecasting strategy [12].

An absolute measure of error evaluates a forecasting method in isolation; a relative measure of error evaluates one forecasting method relative to another method across a set of time series [10]. A limitation of absolute and relative measures is that they do not capture the monetary value or the service level for each item, so they do not measure the practical impact that a forecasting method has on the inventory system. It has been suggested that derivative measures are more practical [9, 19-22]. These use simulation to derive the impact of forecasting accuracy in terms of inventory and service levels achieved by the inventory system.

The aim of this paper is to establish an appropriate forecasting strategy for predicting the demand for consumable spare parts in the South Korean Navy. The objectives are to: i) clarify the nature of the spare parts demand in the Navy; ii) compare the performance of alternative forecasting strategies for predicting spare parts demand at item level; and iii) evaluate the performance of the forecasting methods with absolute, relative and derivative measures.

The remainder of this paper is organised as follows. Section 2 and 3 reviews the theoretical framework for forecasting strategies and accuracy measures respectively. Section 4 describes a case study that employed data obtained from the Navy. Section 5 and 6 present the forecasting methods and the accuracy measures used in this paper respectively. This is followed by results and analysis in Section 7. Finally, Section 8 presents the conclusions of this work.

2. Forecasting Strategies

This section summarises the results of research that has compared the forecasting performance of various item level forecasting strategies. Table 1 compares the results of eleven major studies that have compared the performance of different forecasting strategies that have used top-down forecasting, direct forecasting and combinatorial forecasting. In the literature the methods have been compared using analytical models, simulation, empirical studies or some combination of these approaches. The researchers cited obtained data from a range of contexts and sources. The number of items varied from 2 to 477 with 2-4 hierarchical levels. Makridakis, et al. [23] proposed 1,001 time series, known as M-competition, that used to compare forecasting methods. Autoregressive moving average (ARMA) models, sometimes called Box-Jenkins models are often applied to time series data as tools for understanding and predicting future values. The model is usually referred to as an ARMA (p , q) model (3) where p is the order of the autoregressive (AR) part (1) and q is the order of the

moving average (MA) part (2). The autoregressive integrated moving average (ARIMA) model (4) is a generalisation of an ARMA model. The model is normally referred to as an ARIMA (p, d, q) model where p , d , and q refer to the order of the autoregressive, integrated, and moving average parts of the model respectively.

Reference	Strategy	M	Context	Data		Group		Forecasting performance (superior strategy)
				Source	Demand characteristics	No. of items	No. of Levels	
Shlifer and Wolff [24]	TDF & DF	A	Sales			Several	2	DF; TDF for long forecasting horizon
Schwarzkopf et al. [15]	TDF & DF	A	Production line			2	2	No superiority
Gross and Sohl [25]	TDF & DF	E	production line	Steel sales data	Low correlations	3-7	2	DF in 98.4% of cases
Dangerfield and Morris [26]	TDF & DF	E	Production line	M-competition data	Various seasonal & trend patterns	2	2	DF in 73% of cases
Fliedner and Lawrence [12]	TDF & DF	E	Spare parts distribution	Demand for automotive spare parts	Erratic	95 - 477	3	DF 1.71% smaller MPE
Kahn [17]	TDF & CF	E	Sales	Sales data	Various seasonal patterns	14	3	CF
Dekker et al. [27]	TDF, DF & CF	E	Wholesale (drinks & tubes)	Sales data	2 different seasonal patterns	13-29	2	CF presents 3.5-5.0% & 7.3-12.0% smaller MAD than DF & TDF
Widiarta et al. [28]	TDF & DF	A S	Production line	AR(1)		2	2	Performance depends on lag-1 autocorrelation
Hyndman et al. [11]	TDF, DF & CF	S E	Tourist arrivals	ARIMA Tourism data	Various seasonal & trend patterns	56	4	CF presents 0.07% & 8.59% smaller MAPE than DF & TDF
Widiarta et al. [29]	TDF & DF	A S	Production line	MA (1)		2	2	No superiority
Widiarta et al. [30]	TDF & DF	S	Production line	AR(1), MA(1), & ARMA(1,1)		2	2	Performance depends on the degree of product substitutability

CF = combinatorial forecasting; DF = direct forecasting; TDF = top-down forecasting; M = method of comparison (A = analytic study; E = empirical study; S = simulation); MAD = mean absolute deviation; MPE = mean percentage error; MAPE = mean absolute percentage error.

Table 1. A review of forecasting strategies and performance

$$\text{AR}(p): y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + z_t, \quad (1) \quad \text{MA}(q): y_t = z_t + \theta_1 z_{t-1} + \dots + \theta_q z_{t-q}. \quad (2)$$

$$\text{ARMA}(p, q): y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + z_t + \theta_1 z_{t-1} + \dots + \theta_q z_{t-q}. \quad (3)$$

$$\text{ARIMA}(p, d, q): w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + z_t + \theta_1 z_{t-1} + \dots + \theta_q z_{t-q}. \quad (4)$$

where: y_t = a value at time t ; ϕ 's, θ 's, p , q , and d = constants; z_t = a random variable which are mutually independent and identically distributed at time t ; B^d = the backward shift operator such that $B^d y_t = y_{t-d}$; ∇^d = the difference operator such that $\nabla^d y_t = y_t - y_{t-d}$; $w_t = \nabla^d y_t = (1 - B)^d y_t$.

Analytical studies have compared forecasting performance in terms of the variance of forecasting errors, whereas simulation and empirical studies have compared forecasting performance in terms of the magnitude of forecasting errors. The relative performance of

hierarchical forecasting methods varied, although combinatorial forecasting was found to be superior to top-down and direct forecasting in three studies and direct forecasting was found to be superior to top-down forecasting in four studies. The varying performance of the forecasting methods could be attributed to different statistical features of the data, for example, variations in the number of items in a group or sources of data used [17, 31]. Some authors noted that certain demand features, such as a long forecasting horizon and/or a high degree of substitutability make hierarchical forecasting better than direct forecasting. Hence, the characteristics of demand can be significant in selecting appropriate forecasting strategies. Hierarchical forecasting has been used in many contexts such as marketing, manufacturing, and travelling. The characteristics of spare parts demand is more intermittent and more variable [32]. Direct forecasting methods such as exponential smoothing [33], Croston's method [34], modified Croston's method [35], and the weighted moving average method [36] have been recognised as appropriate forecasting methods for non-normal demand.

In practice there can be some hidden features in the pattern of demand for spare parts, such as seasonality, or some other trend in the time series. An advantage of hierarchical forecasting is that it can bring out these hidden demand features so as to decrease forecasting errors [16]. However, the literature has paid little attention to the use of hierarchical forecasting for the intermittent demand at item level. This is the pattern of non-normal demand associated with spare parts demand. This paper attempts to fill this research gap.

3. Measures of Forecasting Accuracy

The non-normal property of spare parts demand requires a careful choice of accuracy measures. Absolute, relative, and derivative measures are possible alternatives.

3.1. Absolute Measures of Accuracy

There are several absolute measures that can be used to calculate forecasting errors. The mean percentage error (MPE) and mean absolute percentage error (MAPE) that were employed in some of the research referenced in Table 1 are inappropriate for intermittent demand because it is difficult to define the terms for periods with zero demand [37]. The mean squared error (MSE) and root mean square error (RMSE) can be used for intermittent demand. These place heavier weight to larger errors than other methods [25, 38]. Hence, the MSE and RMSE can be useful when larger errors cause greater costs in proportion to small errors [39]. The mean absolute deviation (MAD) is less sensitive to outliers [38]. Hence, the MAD is also useful in order to avoid heavier weight on larger errors.

3.2. Relative Measures of Accuracy

For the purpose of comparing two alternative forecasting strategies, Widiarta et al. [30] employed the ratio of RMSE obtained with hierarchical and direct forecasting methods (expressed as " $RMSE_{(HF)}/RMSE_{(DF)}$ "). Dangerfield and Morris [26] argued that this measure can be biased and proposed the natural log of the ratio of the mean absolute deviation obtained with hierarchical and direct forecasting as an unbiased measure.

$$\text{Log relative error (MAD)} = \ln(MAD_{HF}/MAD_{DF}). \quad (5)$$

A positive value indicates that a DF is superior to a HF, whereas a negative value indicates that a HF is superior to DF.

3.3. Derivative Measures of Accuracy

A derivative measure assesses the performance of a forecasting method with respect to its impact on the performance of an inventory system. This can be measured in terms of total inventory costs (expressed as the sum of inventory carrying costs and inventory stock-out

costs) or the inventory fill rate [40]. The inventory fill rate can be expressed as (6) [22].

$$\text{Inventory fill rate} = 1 - \frac{\text{mean shortage}}{\text{mean demand}} . \quad (6)$$

where: shortage = demand quantity – (stock on hand + delivery quantity).

It is reasonable to quantify the inventory carrying and inventory stock-out costs as a proportion of an item's unit variable cost [21]. In many circumstances the stock-out costs can dramatically outweigh the unit variable cost. For example, the absence of a £10,000 spare part might cause a £100 million warship to be non-operational. This could even lead to a military defeat that could cause casualties and deaths.

4. Case Study

A large stock of spare parts is held for 184 warships in the Navy [41]. However, there is no demand for a large proportion of the stock items. While 45,557 warship spare parts were held, the demand in 2008 was 26,415 [5]. Table 2 presents the annual demand volume of spare parts in the Navy in 2008. 52.3% of spare parts had a demand of one or zero.

Annual demand	No.	%
0 ~ 1	23,838	52.3
2 ~ 5	11,159	24.5
6 ~	10,560	23.2

Table 2. Annual demand volume of spare parts [5]

In order to analyse spare parts demand patterns, historical spare parts consumption records (Jan 2002 – Nov 2007) for three types of warships were collected. This is because these three types of warships have consumed a large volume of spare parts and many of these warships use the same pieces of equipment. A large proportion of the spare parts demand time series are comprised of zero demand periods. Such time series are extremely difficult to analyse. Hence, spare parts which had no demand during five years out of six were screened out in this research. Then, 300 items were randomly chosen using a random sampling procedure with a random number generator.

Figure 1. Grouping structure

In order to use a HF method, a form of a group needs to be defined. An exemplified grouping structure for this research is presented as in Figure 1. The 300 items were classified by the 8 equipment groups which the items are used for. Then, the items within each equipment group were sub-classified into 36 groups using the NATO Supply Classification Group (NSCG) coding system which classifies spare parts by their functions. Research [42] has claimed that grouping based on the dollar-volume (defined as demand per year $\times$ item price) increases the accuracy of HF significantly. Hence, the 300 items were ranked in terms of the dollar-volume within the same equipment and NSCG. Much of the research shown in Table 1 limited the group size to two items. Hence, the two nearest (i.e. the most homogeneous) items in terms of

the dollar-volume form a group. The 300 items can be considered as 150 groups for HF, each containing two items.

Although the 300 items were classified by the 8 equipment groups, it was difficult to find a significantly superior forecasting method for spare parts for an equipment group consisted of a small number of items. Hence, the 8 equipment groups were combined into 3 groups (i.e. Gun & Radar, Main engine, and Generator & Air compressor) for analysis. This is based on the links of the pieces of equipment. For example, Guns and Radar are combined into an equipment group, because some of Guns are linked with Radar in combat data systems [41].

Data aggregation	Statistical feature	300 item time series				150 group time series			
		Gun/RD	ME	GE/AC	Total	Gun/RD	ME	GE/AC	Total
Yearly	Cv(vol)	0.68	0.70	0.93	0.75	0.56	0.63	0.90	0.68
	Correlation	0.46	0.37	0.56	0.43	-	-	-	-
	Skewness	0.86	0.91	1.42	1.02	0.70	0.98	1.63	1.08
	Pr zero	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Quarterly	Cv(vol)	1.25	1.28	1.44	1.31	1.01	1.13	1.32	1.15
	Correlation	0.43	0.36	0.47	0.39	-	-	-	-
	Skewness	1.71	1.94	2.42	2.02	1.43	1.98	2.64	2.05
	Pr zero	0.28	0.17	0.20	0.20	0.11	0.05	0.07	0.06
Monthly	Cv(vol)	2.12	2.13	2.37	2.18	1.69	1.86	2.14	1.90
	Correlation	0.41	0.33	0.41	0.36	-	-	-	-
	Skewness	3.16	3.81	4.48	3.87	2.76	3.88	4.89	3.94
	Pr zero	0.61	0.46	0.50	0.49	0.44	0.26	0.30	0.29

Key: Cv(vol) = coefficient of variation in demand volume; Pr zero = proportion of zero demand periods; Gun/RD = Gun & Radar; ME = Main engine; GE/AC = Generator & Air compressor.

Table 3. Statistical features of the time series obtained from the South Korean Navy

The Navy generates forecasts based on yearly aggregated data sets [5]. However, yearly aggregated data cannot reflect seasonality. Research has found that combinatorial forecasting combined with a model that considers seasonality produced superior forecast [17, 27]. In order to develop a combinatorial forecasting method reflecting seasonality, quarterly or monthly aggregated data sets are required. In this research the time series for 300 items and their 150 groups was aggregated into yearly, quarterly, or monthly aggregation to compare the performance of forecasts produced using these different aggregation approaches. Table 3 presents the average statistical features of the time series for 300 items as well as 150 pairs. The statistics considered were: the coefficient of variation in demand volume (7) [43], correlation between item level time series in a group [29], skewness [8], and the proportion of periods with zero demand [7].

$$\text{Coefficient of variation in demand volume} = s/\bar{y}. \quad (7)$$

where: s = standard deviation of demand volume; $\bar{y}$ = mean demand volume.

There was high Cv(vol) and significant correlation (0.33 ~ 0.56) in all the data, and skewness greater than 1.0 and high Pr zeros in quarterly and month data. This identified that the time series were variable, correlated with each other, were significantly skewed toward the left and were highly intermittent. The group time series had lower variability and intermittency than the item time series (as Cv(vol) and Pr zero were lower). The non-normal demand features with the group time series were reduced. This suggests that HF was superior to DF [12, 31]. Gun/RD had higher intermittency than the other groups as indicated by higher Pr zero value. ME was characterised as lower correlation and lower intermittency owing to lower correlation and lower Pr zero. GE/AC was characterised as higher variability and more deviating from a

normal distribution owing to higher Cv(vol) and higher skewness.

Additional information for the 300 items was obtained: i) the long forecasting horizon consisted of procurement lead time ranging from 3 to 18 months with a 12 months review cycle; ii) a large number of spare parts were substitutable. This is because the Navy purchased a series of equipment from the same manufacturers to ensure stability of supply and continued technical support. Previous research (Table 1) has indicated that the long forecasting horizon and substitutability are features which make HF better than DF.

5. The Development of Hierarchical Forecasting Methods for Naval Spare Parts

This section develops a range of direct and hierarchical forecasting methods for forecasting demand for 300 items. Each forecast for an item was produced and measured once a year: 01/05, 01/06, and 01/07 (in accordance with the review cycle). Thus, 900 forecasts (300 items $\times$ 3 years) were generated using each forecasting method. Each forecast was based upon all the available previous periods. For example, the forecast for 2006 (or 2005) used data for the periods between 01/02 and 12/05 (or 12/04).

5.1 Direct Forecasting Methods

Simple exponential smoothing (SES) can be represented by (8). SES has been shown to be superior to moving average or other complex models when used with hierarchical forecasting [12, 27, 31].

$$\hat{y}_t(1) = \alpha y_t + (1 - \alpha) \hat{y}_{t-1}(1). \quad (8)$$

where: $\hat{y}_t(1)$ = 1 period ahead forecast made at time t ; y_t = demand for an item at time t ;

α = smoothing parameter ($0 < \alpha < 1$).

In the assessment of forecasting strategies for spare parts in the Korean Navy this paper focuses upon SES because the objective was to compare the alternative forecasting strategies not the individual forecasting methods. The optimal smoothing parameter of SES was identified using a forecasting software package, 'R 2.6.2 – forecast'[44].

Data aggregation method		10 direct forecasting methods	
<i>m</i>	monthly aggregated data	<i>um</i>	a forecast with monthly aggregated unadjusted data
<i>q</i>	quarterly aggregated data	<i>tm</i>	a forecast with monthly aggregated data adjusted for linear trend
<i>y</i>	yearly aggregated data	<i>sm</i>	a forecast with monthly aggregated data adjusted additive seasonality
		<i>tsm</i>	a forecast with monthly aggregated data adjusted for linear trend and additive seasonality
		<i>uq</i>	a forecast with quarterly aggregated unadjusted data
SES		<i>tq</i>	a forecast with quarterly aggregated data adjusted for linear trend
		<i>sq</i>	a forecast with quarterly aggregated data adjusted additive seasonality
Data adjustment method		<i>tsq</i>	a forecast with quarterly aggregated data adjusted for linear trend and additive seasonality
<i>u</i>	unadjusted data	<i>uy</i>	a forecast with yearly aggregated unadjusted data
<i>t</i>	data adjusted for linear trend	<i>ty</i>	a forecast with yearly aggregated data adjusted for linear trend
<i>s</i>	data adjusted for additive seasonality		
<i>ts</i>	data adjusted for linear trend and additive seasonality		

Figure 2. 10 direct forecasting methods

In order to use SES, trend and seasonal components from time series have to be measured or

removed [6]. This is because SES should generally be used for non-seasonal time series showing no trend [6]. Research has found that SES coupled with additive seasonality minimised forecasting errors for intermittent time series [45]. Thus, the time series were adjusted for linear trend and additive seasonality. Three data adjustment methods were conducted: i) linear trend values were removed from data; ii) additive seasonal deviations were removed from data; or iii) both linear trend values and additive seasonal deviations were removed from data. These data adjustment methods were combined with the three data aggregation methods so as to produce ten direct forecasting methods using SES at both group and item levels as shown in Figure 2.

5.2 Hierarchical Forecasting Methods

Various proration methods for top-down forecasting [25] and proration methods for combinatorial forecasting [16] were investigated. Four proration methods capable of producing a long horizontal forecast were employed for this research, because the Navy required a long horizontal forecast as stated. Two proration methods for TDF [25] are defined by (9) and (10); two proration methods for CF [16] are defined by (11) and (12). It is noteworthy that previous researchers [11, 17, 27] argued that combinatorial forecasting is superior to top-down forecasting, and DeLurgio [16] argued that a simple combination is as good as a more complex combination method (e.g. weighted combination).

$$\text{Top-down1 (TD1):} \quad f_{i,t+p} = F_{t+p} \times \sum_{i=1}^n \frac{y_{i,t}}{Y_t} / n \quad (9)$$

$$\text{Top-down2 (TD2):} \quad f_{i,t+p} = F_{t+p} \times \frac{\sum_{i=1}^n y_{i,t}}{n} / \frac{\sum_{i=1}^n Y_t}{n} \quad (10)$$

$$\text{Simple Combination (SC):} \quad f_{i,t+p} = \frac{1}{2} (F_{t+p} + \sum_{i=1}^N f_{i,t+p}) \times \frac{f_{i,t+p}}{\sum_{i=1}^N f_{i,t+p}} \quad (11)$$

$$\text{Weighted Combination (WC):} \quad f_{i,t+p} = (w_1 F_{t+p} + w_2 \sum_{i=1}^N f_{i,t+p}) \times \frac{f_{i,t+p}}{\sum_{i=1}^N f_{i,t+p}} \quad (12)$$

where: $y_{i,t}$ = demand of item i at time t ($i = 1, 2, \dots, N$; $t = 1, 2, \dots, n$); Y_t = aggregate demand for a group of N at time t ; $f_{i,t+p}$ = item level forecast of demand i , p periods ahead made at time t ; F_{t+p} = group level forecast, p periods ahead made at time t ; SSE_i = sum of squared errors for demand i ; $w_i = 1/SSE_i / \sum (1/SSE_i)$ = weights of individual forecasts ($\sum w_i = 1.0$).

10 TD1 and 10 TD2 methods were generated with the 10 group level direct forecasts (F_{t+p}). 100 SC and 100 WC methods were produced with the combinations of the 10 group level direct forecasts (F_{t+p}) and the 10 item level direct forecasts ($f_{i,t+p}$). In total 220 hierarchical forecasting methods were generated. In order to represent a hierarchical forecasting method, an abbreviation scheme was used as shown in Figure 3. For example, a forecasting method generated by SC between tq at group level and um at item level is abbreviated to $SCtqum$ as shown.

Figure 3. Abbreviation for a hierarchical forecasting method

6. Formulation of Measurements

This section formulates measurements for assessing the 10 direct and 220 hierarchical forecasting methods for the 300 items. Three groups of accuracy measures were used for this

research: i) MAD and RMSE were used as absolute measures; ii) the log relative errors for MAD and RMSE were used as relative measures; and iii) the total inventory costs and the inventory fill rate were used as derivative measures. The total inventory costs were calculated using (13). The stock-out costs were assumed to be twice the inventory carrying costs (to reflect the high costs of stock outs).

$$\text{Total inventory costs} = \text{unit variable cost} \times (\text{mean inventory per month} \times 0.2 + \text{mean stock-out per month} \times 0.4). \quad (13)$$

Figure 4. Simulation process

In order to obtain the derivative performance of a forecasting method, a simulation of each forecasting method was conducted. The simulated inventory system was a periodic review, order-up-to-level system [46] which is similar to the Navy's system. Figure 4 illustrates the simulated inventory system. The deterministic discrete event simulation used real data (historical consumption and lead time data from 01/02 to 11/07) to generate forecasts. Each demand was assumed to occur on the first day of each month from 01/05 to 11/07. Performance was measured over the same period. Procurement decisions based upon the forecasts was generated 3 times for each item.

7. Forecasting Results and Analysis

This section analyses forecasting results using the 10 DF methods and 220 HF methods for the 300 spare parts for the 3 years with the 3 groups of measurements.

7.1 Direct Forecasting

Table 4 presents mean ranks for the performance and the total inventory costs for the 10 DF methods for the 300 spare parts. As recommended by Sani and Kingsman [21], Friedman's nonparametric test was used due to the non-normal nature of spare parts demand. The number of treatments was 10 (i.e. No. of DF methods) and the number of blocks was 300 (i.e. No. of items). The performance of *tsm* dominates followed by that of *um* (*p*-value of mean ranks = 0.000), with the exception of the total inventory costs as shown.

		<i>um</i>	<i>tm</i>	<i>sm</i>	<i>tsm</i>	<i>uq</i>	<i>tq</i>	<i>sq</i>	<i>tsq</i>	<i>uy</i>	<i>ty</i>
MAD	Mean rank	4.84	6.16	5.22	4.50	5.53	6.28	5.76	5.42	5.24	6.02
RMSE	Mean rank	4.87	6.15	5.13	4.43	5.59	6.29	5.91	5.38	5.07	6.15
Inventory	Mean rank	5.14	5.24	5.82	4.54	6.01	5.19	6.36	5.05	5.35	5.59
cost	Costs (£1,000)	398	411	454	404	417	411	512	472	700	606

Table 4. Forecasting performance of direct forecasting methods

Table 5 provides the best DF methods for each equipment group in terms of Friedman's test for the total inventory costs and MAD. Each cell presents the mean rank of the forecasting method (p -value) under each measure in each equipment group.

		Inventory costs (p -value)	MAD (p -value)
Gun/RD	<i>um</i>	3.85 (0.001)	3.73 (0.002)
ME	<i>tsm</i>	4.46 (0.000)	4.52 (0.000)
GE/AC	<i>tsm</i>	4.25 (0.000)	3.87 (0.000)

Table 5. Direct forecasting methods for equipment groups

7.2 Hierarchical Forecasting

The 220 HF methods were compared with the best DF method (*tsm*). 36 (16.4%) and 39 (17.7%) HF methods were superior to *tsm* in terms of the log relative error for MAD and RMSE respectively. The common top 11 HF methods in both the log relative error for MAD and RMSE are presented in Table 6. No TD1, only 1 TD2, 8 SC and 2 WC methods are included.

	MAD		RMSE			MAD		RMSE	
	LN	Rank	LN	Rank		LN	Rank	LN	Rank
<i>SCtqum</i>	-32.10	1	-36.38	1	<i>WCtqum</i>	-22.34	7	-24.34	7
<i>SCtsqum</i>	-27.00	2	-28.66	4	<i>SCtmsm</i>	-21.87	8	-25.35	6
<i>SCtsmum</i>	-26.02	3	-27.68	5	<i>SCtsmsm</i>	-20.63	9	-22.71	9
<i>SCtqsm</i>	-25.25	4	-28.81	2	<i>SCtsqsm</i>	-18.54	10	-20.30	11
<i>SCtmum</i>	-24.75	5	-28.76	3	<i>WCtsqum</i>	-17.85	11	-20.80	10
<i>TD2tsm</i>	-22.98	6	-23.81	8					

LN = the sum of log relative errors over the 300 items.

Table 6. Top 11 hierarchical forecasting methods in terms of relative measures

The simulation results from the 220 HF methods were compared with those of the best DF (*um*). 35 (15.9%) HF methods were superior to *um* in terms of the total inventory costs. The top 10 HF methods in terms of the total inventory costs are presented in Table 7. No TD1 and TD2 methods are included, whereas 9 SC and 1 WC methods are included. *SCtqum* was found to be the best forecasting method as shown in Table 6 and Table 7.

	Inventory costs (£1,000)			Fill rate		Inventory costs (£1,000)			Fill rate
	Costs	HF- <i>um</i>	Rank			Costs	HF- <i>um</i>	Rank	
<i>SCtqum</i>	354	-43	1	0.79	<i>SCtyum</i>	366	-32	6	0.83
<i>SCtyum</i>	379	-39	2	0.83	<i>SCtmum</i>	371	-27	7	0.79
<i>SCtqsm</i>	359	-39	3	0.79	<i>WCtqum</i>	372	-26	8	0.83
<i>SCtquq</i>	361	-36	4	0.78	<i>SCtsmum</i>	373	-25	9	0.83
<i>SCtyuq</i>	365	-32	5	0.82	<i>SCtmsm</i>	374	-25	10	0.79

HF-*um* = the inventory costs of each HF method deducted by the inventory costs of *um*.

Table 7. Top 10 hierarchical forecasting methods in terms of total inventory costs

	Best DF	Best HF	MAD		RMSE		Simulation		
			LN	Rank	LN	Rank	HF-DF (£1,00)	Rank	Fill rate
Gun/RD	<i>um</i>	<i>SCtyum</i>	-2.3	8	-1.6	6	-8.2	2	0.82
ME	<i>tsm</i>	<i>SCtquy</i>	-12.6	1	-13.1	1	-398.1	1	0.79
GE/AC	<i>tsm</i>	<i>SCtquq</i>	-2.7	9	-3.1	10	-26.1	13	0.82

HF-DF = the inventory costs of each HF method deducted by the inventory costs of the best DF method.

Table 8. The best hierarchical forecasting methods for 3 equipment groups

The performance the 220 HF methods for the 3 equipment groups was investigated. The 220 HF methods were compared with the best DF for each group. The best HF methods for each equipment group that rank within the top 10 in terms of the log relative errors for MAD and RMSE and are superior to the best DF in terms of the total inventory costs are presented in Table 8. Employing these forecasting methods for the equipment groups, the total inventory costs, which result from using only the generally best forecasting method (i.e. *SCtqum*), can be reduced. If these are used for Gun/RD, ME, and GE/AC respectively, the total inventory costs for all the 300 items were calculated as £333,502. These costs are 5.4% smaller than the total inventory costs from *SCtqum*, that is, £352,619.

8. Conclusions

The nature of the spare parts demand, which is non-normal (however, less non-normal at group level time series) and substitutable in many spare parts, and has long forecasting horizons, was identified by the data obtained from the South Korean Navy. A reduction of non-normality at group level time series [12, 31], substitutability [30], long forecasting horizons [24] are features that could make hierarchical forecasting better than direct forecasting.

This paper has demonstrated that for Naval spare parts: i) combinatorial forecasting is superior to top-down forecasting and direct forecasting; ii) the simple combination between the forecast with quarterly aggregated data adjusted for linear trend at group level and the forecast with monthly aggregated unadjusted data at item level (*SCtqum*) is generally the most superior forecasting method among the forecasting methods tested; and iii) the three simple combinations such as *SCtyum*, *SCtquy*, and *SCtquq* are recommendable for the three equipment groups (Gun & Radar, Main engine, and Generator & Air compressor respectively), which have relatively different demand features. The forecasting performance was evaluated with the three groups of measurement. With these three-fold measurements, reliability and internal validity of the results are claimed to be established. The superiority of combinatorial forecasting to top-down forecasting was consistent with the literature [11, 17, 27]. Especially, simple combination was the most superior proration method among the proration methods tested. This superiority of simple combination corroborates DeLurgio [16].

This paper has proposed a forecasting strategy for the Navy as: i) *SCtqum* should be considered preferentially as a forecasting method for spare parts; ii) in order to reduce the risk of a wrong decision and guarantee the best practical decision verification of a forecasting method using simulation before implementing the forecast should be conducted; and iii) for the 3 equipment groups the different forecasting methods identified above should be used.

References

1. The US Naval Institute, 1977. *Naval Operations Analysis*. MD, USA: US Naval Institute.
2. Fowler, A., 2003. The Royal Navy: JIT based ship spares provisioning In: Johnston, R., Chambers, S., Harland, C., Harrison, A., Slack, N. (Eds.). *Cases in Operations Management*. 3rd edn. London: Prentice Hall, 300 - 307.
3. Hinton Jr., H. L., 1999. *Defense inventory, continuing challenges in managing inventories and avoiding adverse operational effects*. Washington, D.C.: US General Accounting Office.
4. Rustenburg, W. D., Houtum, G. J. v., Zijm, W. H. M., 2001. Spare parts management at complex technology-based organizations: an agenda for research. *International Journal of Production Economics*, 71, 177 - 193.
5. Seon, M.-S., U, J.-U., 2009. A study on forecasting of repair part demands of Korean

- Military: focused on Navy. *The Korean Journal of Defense Analysis*, 85, 201 - 234.
6. Chatfield, C., 2004. *The Analysis of Time Series: an Introduction*, 6th edn. London: Chapman & Hall/CRC.
 7. Boylan, J. E., Syntetos, A. A., Karakostas, G. C., 2008. Classification for forecasting and stock control: a case study. *Journal of Operational Research Society*, 59, 473 - 481.
 8. Businger, M. P., Read, R. R., 1999. Identification of demand patterns for selective processing: a case study. *International Journal of Management Science*, 27, 189 - 200.
 9. Eaves, A. H. C., Kingsman, B. G., 2004. Forecasting for the ordering and stock-holding of spare parts. *Journal of the Operational Research Society*, 55, 431 - 247.
 10. Syntetos, A. A., Boylan, J. E., 2005. The accuracy of intermittent demand estimates. *International Journal of Forecasting*, 21 (2), 303 - 314.
 11. Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., 2007. Optimal combination forecasts for hierarchical time series *Monash University*, 1 - 21.
 12. Flidner, E. B., Lawrence, B., 1995. Forecasting system parent group formation: An empirical application of cluster analysis. *Journal of Operations Management*, 12 (2), 119 - 130.
 13. Miller, J. G., Berry, W. L., Lai, C.-Y. F., 2007. A comparison of alternative forecasting strategies for multi-stage production-inventory systems. *Decision Sciences*, 7 (4), 714 - 724.
 14. Zotteri, G., Kalchschmidt, M., Caniato, F., 2005. The impact of aggregation level on forecasting performance. *International Journal of Production Economics*, 93 - 94 (8), 479 - 491.
 15. Schwarzkopf, A. B., Tersine, R. J., Morris, J. S., 1988. Top-down versus Bottom-up Forecasting Strategies *International Journal of Production Research*, 26 (11), 1833 - 1843.
 16. DeLurgio, S. A., 1998. *Forecasting Principles and Applications*. NY: Irwin McGraw-Hill.
 17. Kahn, K. B., 1998. Revisiting top-down versus bottom-up forecasting. *The Journal of Business Forecasting*, Summer 14 - 19.
 18. Widiarta, H., Viswanathan, S., Piplani, R., 2009. Forecasting aggregate demand: an analytical evaluation of top-down versus bottom-up forecasting in a production planning framework. *International Journal of Production Economics*, 118 (1), 87 - 94.
 19. Wemmerlöv, U., 1989. The behavior of lot-sizing procedures in the presence of forecast errors *Journal of Operations Management*, 8 (1), 37 - 47
 20. Gardner Jr., E. S., 1990. Evaluating forecast performance in an inventory control system. *Management Science*, 36 (4), 490 - 499.
 21. Sani, B., Kingsman, B. G., 1997. Selecting the best periodic inventory control and demand forecasting methods for low demand items. *Journal of the Operational Research Society*, 48, 700 - 713.
 22. Heuts, R. M. J., Strijbosch, L. W. G., Schoot, E. H. M. v. d., 1999. A combined forecast-inventory control procedure for spare parts. *Flexible Automation & Intelligent Manufacturing*, 1 - 15.
 23. Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., Winkler, R., 1982. The accuracy of extrapolation (time series) methods: results of a forecasting competition. *Journal of Forecasting*, 1 (2), 111 - 153.
 24. Shlifer, E., Wolff, R. W., 1979. Aggregation and proration in forecasting. *Management Science*, 25 (6), 594 - 603.
 25. Gross, C. W., Sohl, J. E., 1990. Disaggregation methods to expedite product line forecasting. *Journal of Forecasting*, 9, 233 - 254.
 26. Dangerfield, B. J., Morris, J. S., 1992. Top-down or bottom-up: aggregate versus disaggregate extrapolations. *International Journal of Forecasting*, 8 (2), 233 - 241.

27. Dekker, M., Donselaar, K. v., Ouwehand, a. P., 2004. How to use aggregation and combined forecasting to improve seasonal demand forecasts. *International Journal of Production Economics*, 90 (2), 151 - 167.
28. Widiarta, H., Viswanathan, S., Piplani, R., 2006. On the effectiveness of top-down strategy for forecasting autoregressive demands. *Naval Research Logistics*, 54 (2), 176 - 188.
29. Widiarta, H., Viswanathan, S., Piplani, R., 2008. Forecasting item-level demands: an analytical evaluation of top-down versus bottom-up forecasting in a production-planning framework *IMA Journal of Management Mathematics*, 19, 207 - 218.
30. Widiarta, H., Viswanathan, S., Piplani, R., 2008. Evaluation of hierarchical forecasting for substitutable products. *International Journal of Services and Operations Management*, 4 (3), 277 - 295.
31. Fliedner, G., 1999. An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation. *Computers & Operations Research*, 26 (10 -11), 1133 - 1149.
32. Ghobbar, A. A., Friend, C. H., 2002. Sources of intermittent demand for aircraft spare parts within airline operations. *Journal of Air Transport Management*, 8, 221 - 231.
33. Markland, R. E., 1970. A comparative study of demand forecasting techniques for military helicopter spare parts. *Naval Research Logistics Quarterly*, 17 (1), 103 - 119.
34. Croston, J. D., 1972. Forecasting and stock control for intermittent demands. *Operational Research Quarterly* 23 (3), 289-303.
35. Syntetos, A. A., Boylan, J. E., 2001. On the bias of intermittent demand estimates. *International Journal of Production Economics*, 71 (1 - 3), 457 - 466.
36. Ghobbar, A. A., Friend, C. H., 2003. Evaluation of forecasting methods for intermittent parts demand in the field of aviation: a predictive model. *Computers & Operations Research*, 30 (14), 2097 - 2114.
37. Hyndman, R. J., Makridakis, S., Wheelwright, S. C., 1998. *Forecasting: Methods and Applications*, 3rd edn. NY: John Wiley & Sons.
38. Eaves, A. H. C., 2002. *Forecasting for the Ordering and Stock-holding of Consumable Spare Parts*. Thesis, University of Lancaster, Lancaster, UK.
39. Kling, J. L., Bessler, D. A., 1985. A comparison of multivariate forecasting procedures for economic time series. *International Journal of Forecasting*, 1 (1), 5 - 24.
40. Petrovica, D., Royb, R., Petrovic, R., 1998. Modelling and simulation of a supply chain in an uncertain environment *European Journal of Operational Research*, 109 (2), 299 - 309.
41. Saunders, S., 2009. *Jane's Fighting Ships 2009 - 2010*. Coulsdon, Surrey, UK: Jane's Information Group.
42. Fliedner, E. B., Mabert, V. A., 1992. Constrained forecasting: some implementation guidelines. *Decision Sciences*, 23 (5), 1143 - 1161.
43. Williams, T. M., 1984. Stock control with sporadic and slow-moving demand. *Journal of Operational Research Society*, 35 (10), 939 - 948.
44. The R project for statistical computing, 2009. *The R project for statistical computing*, Available at:<<http://www.r-project.org/>> (Accesed on 24 October).
45. Gardner Jr., E. S., Diaz-Saiz, J., 2002. Seasonal adjustment of inventory demand series: a case study. *International Journal of Forecasting*, 18 (1), 117 - 223.
46. Silver, E. A., Pyke, D. F., Peterson, R., 1998. *Inventory Management and Production Planning and Scheduling*, 3rd edn. NY: John Wiley & Son, Inc.