The impact of the forecasting horizon when forecasting with group seasonal indices

The impact of the forecasting horizon when forecasting with group seasonal indices

Pim Ouwehand, Karel van Donselaar, Ton de Kok

WP-162

BETA publicatie	WP 162 (working paper)
ISBN	90-386-0676-1 978-90-386-0676-7
ISSN	1386-9213
NUR	804
Eindhoven	Maart 2006
Keywords	Forecasting / Exponential smoothing / Seasonality / Forecasting horizon

The impact of the forecasting horizon when forecasting with group seasonal indices

Pim Ouwehand*, Karel van Donselaar, Ton de Kok
Department of Technology Management
Technische Universiteit Eindhoven
P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Abstract

Due to increased product variety and shorter product life cycles, forecasting individual product demand has become increasingly difficult. Higher demand variation and less data make it harder to recognize seasonal patterns with standard exponential smoothing methods. In this study we try to improve individual item forecasts by using information about seasonal indices at a higher aggregation level, i.e., at the level of families of items with similar seasonal patterns. The method is an adaptation of the Holt-Winters procedure with the seasonal indices estimated at the product family level instead of at the item level. Empirical results (Dekker et al., 2004) for two Dutch wholesalers have shown significant improvement potential of this aggregation method over classical methods for one-period ahead forecasts. In this paper, we extend the previous study by looking at the impact of the forecasting horizon on the performance of the product-aggregation method. It is shown empirically that although the difference in forecast error decreases when forecasting further ahead, it still performs better than classical methods. Furthermore, it appears to be a more robust method.

Keywords: forecasting, exponential smoothing, seasonality, forecasting horizon.

1. Introduction

Being able to forecast customer demand is very important for companies that are part of a supply chain and face external demand, such as retailers, wholesalers and manufacturers. Better forecasts allow them to maintain lower inventory levels or higher service levels, to plan the use of labor more efficiently, and to make better-informed decisions when entering into contracts with suppliers. Forecasting is especially relevant for companies such as wholesalers, whose benefits lie in stocking a large variety of products.

Recent developments have made forecasting more difficult. Product ranges have increased, which causes for example substitution effects. Also, forecasts pertain to smaller amounts, which usually have a higher coefficient of variation. In addition, product life cycles have become shorter, resulting in less data to make reliable forecasts, and consumer behavior has changed, causing more irregularity in demand patterns. These four factors have made demand patterns become less predictable and as a result, the statistical quality of the forecasts has deteriorated. At the same time, forecasting has become more important, since

*Corresponding author. Tel.: +31-40-2472826; fax: +31-40-2464531.
E-mail address: p.ouwehand@tm.tue.nl (P. Ouwehand)

large numbers of stock-keeping units (SKU's) need to be kept in stock. Being able to make good forecasts and understand demand patterns can yield important competitive advantages and allows for large savings in terms of inventory investment, for better capacity planning, and for better decision-making.

A solution for dealing with these developments is the concept of *product-aggregation* or *group seasonal indices*. This approach tries to improve forecasts by simultaneously forecasting a group of products that exhibit a similar seasonal pattern. The aim of the method is to increases forecasting accuracy by using knowledge of the demand process at a higher aggregation level in forecasting demand at individual product level. Seasonal indices are estimated from the total demand for a group of products and used for forecasting individual items. For this method to work it is important that a group of products with similar seasonal patterns is constructed. The aggregation method is used as an adaptation to the Holt-Winters procedure, with the seasonal indices estimated at the product family level instead of at the item level. We discuss the method in more detail in Section 2.

The literature on aggregation approaches in forecasting can be divided into two lines of research. Both consider how to use knowledge at the aggregate level in order to improve forecasts at the individual item level. The first approach, called top-down forecasting, considers successive aggregation and disaggregation of the data. The data is first aggregated, a forecast is made at the aggregate level, and is then disaggregated back to the original aggregation level. The literature on this approach focuses mainly on the question when either top-down or direct disaggregate forecasts should be made (see e.g., Shlifer and Wolff, 1978, Schwarzkopf et al., 1988, Dangerfield and Morris, 1992). The second line of research focuses on products with similar seasonal patterns, and considers how to use knowledge of aggregate data in order to improve seasonal estimates at the individual item level. The latter is also the focus of the current article. There are only a few publications dealing with this aggregation approach. In these studies, the product-aggregation idea is first described and empirically tested (Dalhart, 1974; Withycombe, 1989; Bunn and Vassilopoulos, 1993, 1999; Dekker et al., 2004). The approaches described are similar to the aggregation concept studied in the current article. All studies show the improvement potential of product-aggregation over classical methods by applying basic forecasting and clustering techniques. Nevertheless, they only focus on making short-term forecasts, and do not investigate the impact of the forecasting horizon on the product-aggregation approach. Furthermore, they do not address sufficiently the issue of how to form product families. This article extends our earlier study (Dekker et al., 2004) by addressing these two issues. The primary focus is on the impact of the forecasting horizon, while the impact of the grouping method is explored briefly.

Firstly, we look at forecasting over a longer horizon. For the short term, the shortcomings of nonseasonal methods may not be so clear. Only in the longer term the benefits of seasonal methods, compared to nonseasonal methods, become apparent. In that case, a simple extrapolation of the level and trend component of demand is no longer sufficient to capture the development of the demand pattern, and obtaining a good estimate of the seasonal pattern becomes necessary for accurate forecasting. We investigate the influence of the forecasting horizon on the performance of the product-aggregation method and the impact on the forecasting errors when forecasting further ahead. We do this by looking at both h-step ahead point forecasts and forecasts for (cumulative) lead-time demand. Point forecasts are interesting from a capacity planning perspective, whereas cumulative forecasts are interesting for inventory management reasons (cf. Silver et al., 1998). We show empirically that although the difference in forecast error with classical methods decreases when forecasting further ahead, the product-aggregation method still performs better than classical methods. Besides, it turns out to be a more robust method.

Secondly, we explore the issue of how to form product families such that they can benefit from the aggregation method. All articles on product-aggregation recognize that finding groups of similar items is important for the aggregation method to work properly. Only Bunn and Vassilopoulos (1993, 1999) apply a statistical clustering method in order to form the product families but do not regard this to be the main research issue. In this article, we show that the way the group of products is composed has a large influence on the results, and that a standard statistical clustering approach is not necessarily appropriate. We conclude that it is important to find an optimal way of grouping products when applying product-aggregation but that this is not a straightforward task and thus needs further research.

2. Group seasonal indices

Standard exponential smoothing methods are univariate and only use the history of the demand for the particular product for which they make a forecast. More accurate estimates of seasonality can be obtained if products with similar seasonal patterns are first combined into a product group. Seasonal indices can then be estimated from the total demand of a group of products. These indices in turn are used as the estimate of the seasonal component for each of the products. In situations where the random component is relatively large compared to the seasonal pattern, this aggregation approach can help to reduce noise and thus aid seasonal methods in determining more accurately the seasonal pattern. It is known that in general aggregating data reduces variability. The coefficient of variation goes down because random fluctuations cancel out. However, only five publications were found that apply this to forecasting in order to improve seasonal estimates.

Dalhart (1974) first proposed combining products into product classes to obtain a better estimate of the seasonal component. He first estimated the seasonal indices for all products in a product group individually and then averaged them to obtain the estimates for the group seasonal indices. These indices were then used to make forecasts. The concept was tested for simulated data for 100 time series and showed substantial improvements compared to using the individual seasonal indices, although no true out-of-sample forecasts were made.

Withycombe (1993) proposed a different method for calculating the seasonal indices, by first aggregating the demands for all component series and calculating the seasonal indices at the aggregate level. The improved seasonal estimates were used to deseasonalize each series in the group, before extrapolation by Double Exponential Smoothing, and to reseasonalize afterwards. Seasonal indices were not updated, like in the Holt-Winters method. He demonstrated the concept for 29 products (6 product classes with 4 or 5 products) with monthly data from a computer peripherals supplier, and reported an average decrease in MSE of 11% per product class compared to forecasting for individual products. The product classes were determined by the marketing department, so no grouping procedure was applied.

Bunn and Vassilopoulos (1993, 1999) extended the studies of Dalhart and Withycombe by providing a broader comparison of methods and addressing the issue of forming the groups according to statistical criteria. Besides Dalhart's and Withycombe's method, they suggested several combinations of these methods for deseasonalizing and reseasonalizing the data. Deseasonalized demand was forecasted by Holt's Two Parameter Exponential Smoothing. They also introduced alternative ways of calculating initial estimates of the seasonal indices. But again, no updating of seasonal indices was done. They used monthly data of 54 highly seasonal series from 5 product classes from a large UK department store chain. They found that Withycombe's method led to an average decrease in MSE of 6% (compared to forecasting for individual products) and on average outperformed Dalhart's method. Although they made 1-, 2- and 3-period ahead forecasts, not much attention was paid to the influence of the forecast horizon.

Dekker et al. (2004) used a different approach towards aggregation. Instead of deseasonalizing and reseasonalizing with fixed seasonal indices, they suggested an approach which stays closer to the classical Holt-Winters method. In fact, it is equivalent to multiplicative Holt-Winters for individual items with the seasonal indices determined at the product family level. The advantage of this approach is that it allows for updating of seasonal indices each period, when additional information becomes available. Another difference with the previous articles is that they provided a broader comparison of methods by comparing the aggregation method with Holt-Winters, Simple Exponential Smoothing and combined forecasting. The combined forecasting was performed by taking a weighted average of the Naïve 1 method from the well-known M-Competitions (e.g., Makridakis and Hibon, 2000) and the Holt-Winters' method. Furthermore, they considered demand data that is more difficult to forecast. Firstly, they looked at weekly demand data instead of monthly data, which in general has a higher coefficient of variation. Secondly, they looked at different seasonal patterns, namely stochastic seasons that change slightly from year to year, both in timing and in magnitude. Finally, they considered more products and larger product families.

The current article article builds on Dekker et al. (2004) and therefore we use the same aggregation method as used there. The aggregation method is adapted from Winters' method and works as follows. First, the demand at the family level is determined by adding up all individual product demands. Next, the seasonal indices are estimated through a multiplicative decomposition approach using a ratio-to-moving average procedure (see Silver, Pyke and Peterson, 1998 or Makridakis, Wheelwright and Hyndman, 1998 for details). These seasonal indices are then used to forecast at individual product level, using Winters' seasonal method. Each period when additional data becomes available the seasonal indices are updated on the family level using exponentially smoothing updating formulae. Level and trend estimates are updated on the individual item level (cf. Dekker et al., 2004).

The product-aggregation concept assumes that the seasonal effect is the same for all products in the family, and therefore that the seasonal patterns at the family level is the 'true' seasonal pattern. It is therefore important that the products in the family all have a similar seasonal pattern. We study this by applying two alternative ways of grouping products. The first approach is based on a visual inspection of the seasonal patterns. The second approach is a statistical clustering approach. Both are described in more detail in the next section.

3. Description of the data and experiments

3.1 Data

The data we use represents weekly sales figures of four product groups from two large Dutch wholesalers. The products comprise both slow moving and fast moving items and exhibit clear seasonal patterns. The product classes cover beers, soft drinks in regular bottles, and soft drinks in small packages from a supermarket chain, and plastic tubes from an electro-technical wholesaler. For all products five years of weekly data are available.

The selected products come from large datasets of each of the product classes. We first select all products with 5 years of complete demand history. From these groups of about 20-40 products, we pick subsets of products with similar seasonal patterns. This is done in two ways: by visual inspection and by statistical clustering.

The first approach, visual inspection, results in groups that are described in Dekker et al. (2004). Here, we use the same groups, which are obtained as follows. A ratio-to-moving average procedure is performed on the full five years of data for each product to obtain 52 seasonal indices per product. These are graphed and products with similar patterns are selected. After obtaining the first results, we eliminate some of the products that influence the results of the product-aggregation method negatively. This shows, as will be elaborated later on, that finding the appropriate product families is very important for this method. Figure 1 shows the seasonal patterns for all products in the product families of beers. Figure 2 shows the seasonal indices of the aggregate demand per product family. As can be seen, the product families of soft drinks (regular bottles) and beers exhibit similar patterns. The group of soft drinks in small packages show slightly different seasonal patterns.

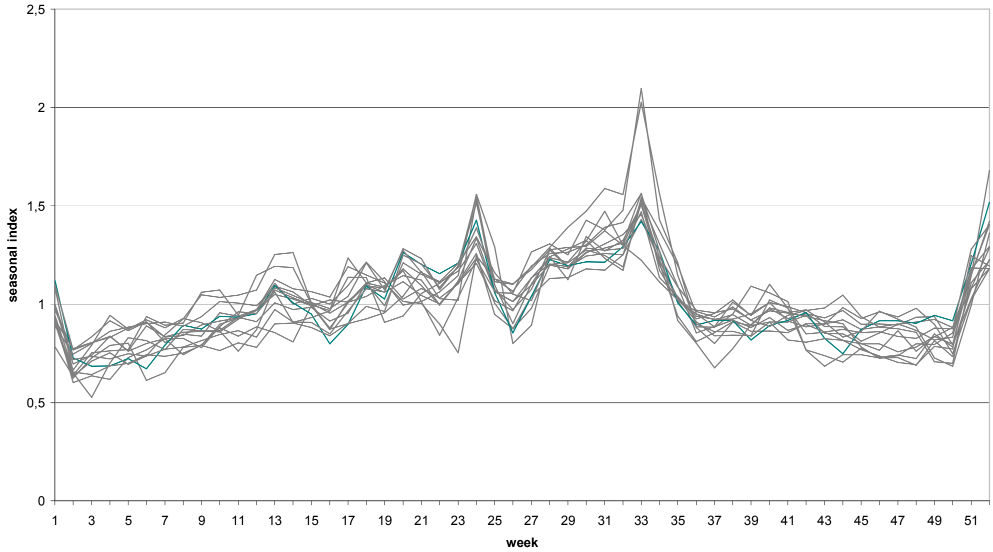

Figure 1: Seasonal patterns for product family 'beers' (composed by visual inspection)

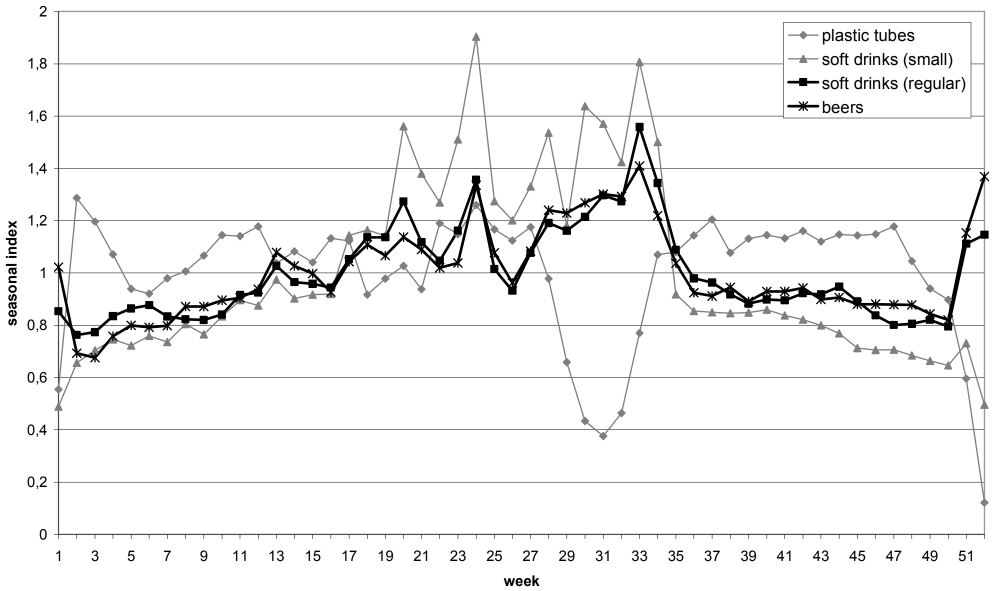

Figure 2: Seasonal patterns for aggregate demand per product family (visual inspection)

The second approach is a cluster analysis approach. We group similar products according to their seasonal indices, which are estimated by the ratio-to-moving average decomposition method. For the cluster analysis, each time series is represented by a variable with 52 observations, one for each weekly seasonal index. More specifically, we apply hierarchical clustering (see e.g., Johnson and Wichern, 1998). Hierarchical clustering is an agglomerative method in which all products initially are in distinct clusters, and are grouped together sequentially. A cluster is built up by joining objects that are close to each other. This is done according to a linkage criterion. In our case, we use the commonly used *average linkage*, which adds the object to a cluster that has smallest average distance to all objects already in the cluster (the average of the individual similarities is considered). As a distance measure, we use Pearson's correlation, since this measure is able to capture a pattern of higher and lower values of the seasonal indices (the relative magnitudes within variables are important here). In order to ensure the products in each cluster were similar, we select the largest cluster with a correlation coefficient of 0.8 or higher. Figure 3 shows the seasonal patterns for all products in the statistical cluster of plastic tubes. Table 1 shows the number of products in the original product classes and in the selected subsets.

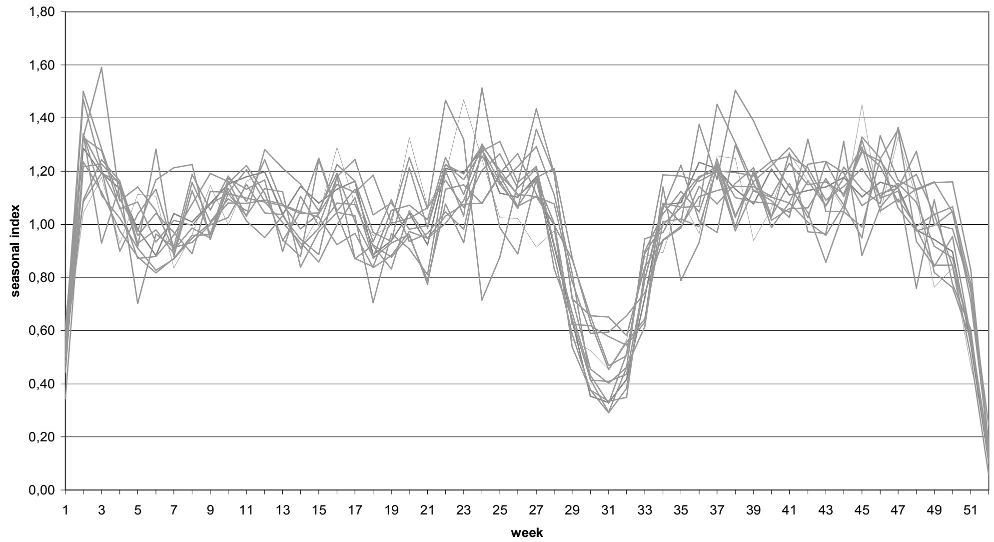

Figure 3: Seasonal patterns for product family 'plastic tubes' (statistical cluster)

Table 1: Number of products in each of the groups

Product class	Number of products		
	Product class	Visual Inspection*	Statistical clusters
Beers	21	14	15
Soft drinks regular	41	29	16
Soft drinks small	19	13	19
Plastic tubes	29	11	10

*The groups formed by visual inspection are the same as used in Dekker et al. (2004).

3.2 Experiments

In the experiments, several forecasting methods are compared, which are listed in Table 2. Two of these are standard exponential smoothing methods. Single Exponential Smoothing (SES) makes a forecast based only on an estimate of the level of demand. Multiplicative Holt-Winters' (HW) also incorporates a seasonal pattern. The two classical methods are compared with methods based on product-aggregation and combining forecasts with the Naïve 1 method. The Naïve 1 (Random Walk) method is a very simple method that just takes 'this week's demand as a forecast for next week's demand', or $\hat{x}_{t,t+1} = x_t$. Method HWP is the product-aggregation method as described above. Methods HWC and HWPC take a weighted average of methods HW and HWP respectively with the Naïve 1 method (with C = combined). Since the data do not show a clear trend, no methods based on a trend are considered.

Table 2: Overview of forecasting methods

Method	Description
SES	Single Exponential Smoothing
HW	Holt-Winters' Exponential Smoothing
HWC	Holt-Winters' Exponential Smoothing combined with Naïve 1
HWP	Product-aggregation
HWPC	Product-aggregation combined with Naïve 1

For each series, the data is divided into three parts. The first two years are used for clustering the products and finding initial estimates of the smoothing parameters and the seasonal indices. The following two years are used to fit the model to the data. The remaining year (the so-called hold-out period) is used for making out-of-sample forecasts and computing the accuracy measures. Forecasts are made over several forecasting horizons, namely 4, 8, and 12 weeks. Cumulative forecasts are also made over a horizon of 24 weeks, since the average age of the forecasts then equals 12 weeks.

Fitting of the model to the data, using the second portion of the data, in this case means finding the optimal settings for the smoothing parameters and the weights for averaging (for the combined forecast). This is done by minimizing the within-sample h-step ahead 'forecast' errors, or residuals (rather than one-step ahead). The parameters are optimized on a grid search basis, which means that a discrete set of parameter values for each parameter is tested. The combination of parameter values giving the smallest forecast error is chosen. When reporting the results over the hold-out period, the same accuracy measure is used as is used for optimizing the parameters over the test period. Table 3 summarizes the parameter settings for the experiments.

Table 3: Parameter settings for experiments

Parameter	Value
Demand history	260 weeks
Periods used for clustering and initialization	104 weeks
Fitting	104 weeks
Hold-out period	52 weeks
Forecast horizon	4, 8, 12, 24 weeks
α, γ	0.01; 0.05; 0.1; 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9
w	0.1; 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9

For comparing the forecasting methods, we consider several well-known accuracy measures. Firstly, we look at two measures that are predominantly used in the M3-competition, namely the symmetric Mean Absolute Percentage Error (sMAPE) and average ranking. Secondly, we look at two accuracy measures that are used in the articles on aggregation discussed above, namely the Mean Absolute Deviation (MAD) and the Mean Squared Error (MSE). The comparisons of the methods are based on the total sMAPE, MAD or MSE per product family, summed over all products in the family. Note that, if we consider the total sMAPE per family, the relative magnitudes of the products (in terms of demand per week) are not taken into account. When we look at the total MAD or MSE, these magnitudes have a stronger effect on the total accuracy measure. Average rankings are computed by sorting, for each product and for each forecasting horizon, the relevant accuracy measure from the smallest (taking rank 1) to the largest. Once the ranks for all series in a product family are determined, the mean rank is calculated.

4. Results

In Tables 8-11 in the appendix, the results are presented for groups that are based on visual inspection and for groups composed by statistical clustering. For the former kind of groups results are shown for both point forecasts and cumulative forecasts, for the latter only results for point forecasts are presented.

The groups that are composed by statistical clustering contain different products than the groups found by visual inspection and thus it would be difficult to compare the two grouping approaches. Therefore, both groups are complemented with products from the other group so that comparisons are based on the same set of products. In each of the four cases, the groups based on visual inspection and statistical clustering consisted of largely the same products. To give an impression of the sizes of the original and complemented groups, Table 4 lists the average total demand per week of all products in each group.

Table 4: Average total demand per week per product group

Product class	Average demand		
	Visual Inspection	Statistical clusters	Complemented groups
Beers	82566	110075	134393
Soft drinks regular	250039	125689	250039
Soft drinks small	103290	182819	182819
Plastic tubes	9410	14219	14836

Tables 8-11 show the results for the complemented product groups. For either group, the added products are forecasted individually, and thus are not incorporated in aggregation procedures. A forecast is made with HW (in cases where HWP is used for the original group) or HWC (in case of HWPC). In Ouwehand et al. (2004), the results of the original non-complemented groups are discussed as well. This provides a better comparison with Dekker et al. (2004), since there the same non-complemented groups are studied.

For all product families, we report the relative difference in sMAPE (Table 8), MAD (Table 9), MSE (Table 10) of all methods compared to standard Holt-Winters (HW). We also present the average ranks of the methods, based on the MSE's (Table 11). As mentioned, the comparisons of the methods are based on the total sMAPE, MAD or MSE per product family.

5. Influence of forecasting horizon

For a forecasting horizon of one period ahead we have seen the following (Dekker et al., 2004). The classical Holt-Winters method (HW) performed poorly. Compared to HW, both methods HWC and HWP showed a large decrease in sMAPE/MAD/MSE. The decrease was larger for method HWC than for method HWP. The combination of these two methods, method HWPC, usually performed best, although the simple exponential smoothing (SES) could show similar or even slightly better results. Therefore, when forecasting one period ahead, one should always consider whether a seasonal method should be used in the first place. The average rankings roughly subscribed to these conclusions.

For a forecasting horizon greater than one period ahead, the results are not always so clear. Firstly, it is not always the case anymore that a particular method consistently performs better or worse than another method. There are some differences across product families, across horizons, and across accuracy measures. Secondly, the forecast errors (in terms of the different accuracy measures) need not increase monotonically with increasing lead-time. For some forecast horizons the error may even decrease. This is in line with Chatfield and Yar (1991), where it was shown that the forecast error variance does not necessarily increase monotonically with h. This behavior is typical of nonlinear models (Chatfield, 2000), which the multiplicative Holt-Winters is. Nevertheless, it makes the results more difficult to interpret and draw generic conclusions. Below, we give a general discussion of the results, meaning that there may be some exceptions to these global conclusions. The three main conclusion that can be drawn from the results are:

- The added value of the aggregation approach decreases as the horizon increases, but improved accuracy remains
- The aggregation approaches appear to be more robust than Holt-Winters
- Aggregation shows potential for both point forecasts and cumulative forecasts

Below these three conclusions are illustrated by highlighting some results from Tables 8-11 in the appendix.

Decreasing but remaining added value

For a forecasting horizon greater than one period ahead (*h>1*), we can draw the following general conclusions for point forecasts:

- Single Exponential Smoothing (SES) deteriorates rapidly after *h=1*.
- Holt-Winters' (HW) only slightly decreases in accuracy for increasing *h*.
- Methods HWPC and HWC lose their value compared to HWP and HW respectively.
- Product-aggregation (HWP) continues to give an improvement over Holt-Winters (HW), albeit small, but is more robust.

Table 4: difference with method HW when sMAPE is minimized

	method	soft drinks small packages visual inspection group, point forecasts			
Horizon		1	4	8	12
	HW				
	SES	-25%	0%	31%	37%
	HWC	-23%	-3%	-6%	2%
	HWP	-9%	-7%	-7%	-8%
	HWPC	-24%	-3%	-7%	-3%

Table 4 illustrates these conclusions for a particular product group. For a forecasting horizon of more than one period ahead, all methods show a clear increase in forecast error from *h=1* to *h=4* (the absolute values of the accuracy measures are not shown, see Ouwehand et al, 2004). For all product families, the increase is particularly large for method SES. This method performed well compared with other methods for *h=1*, but for *h>1* the forecast errors (in terms of the different accuracy measures) go up strongly, and more than for other methods. The largest increase mostly seems to take place from *h=1* to *h=4*. This is not surprising for stationary series with most seasonal indices relatively close to 1 and a smooth seasonal pattern (products from the supermarket chain) or where abrupt deviations from the smooth pattern only account for a limited number of periods (products from the electro-technical supplier) (see Figure 2).

Method HW performed worst of all methods for *h=1*. For *h>1* this method hardly shows an increase in forecast error. It sometimes even performs better than for smaller *h*. At the same time the accuracy of SES decreases a lot when going from $h=1$ to $h=4$. Apparently, choosing a seasonal method over a nonseasonal method becomes necessary for accurately forecasting the longer term, as opposed to the short term, where a seasonal method was not necessarily better than a non-seasonal one.

The added value of the alternative methods (HWC, HWP and HWPC) compared to method HW decreases for larger *h* (see Tables 8-11). Combining forecasts (HWC and HWPC) tends to lose its improvement potential compared to HW and HWP. The reason for this lies in the averaging with the Naïve 1 method. This causes the most recent actual demand to have a large influence on the forecast for *h* periods ahead. In the short term these two periods may be correlated, but for the longer term this is less likely; this week's demand does not say very much about the demand *h* weeks ahead when *h* gets larger. In some cases, however, combining forecasts continues to improve. This can be explained as a random effect since the demand patterns are quite stationary. For trending demand patterns this is not expected to be the case.

We conclude that for point forecasts the added value of the aggregation approach decreases – the improvement over Holt-Winters becomes less as the forecast horizon increases – but in almost all cases still an improvement can be seen (in case of sMAPE for 3 out of 4 product groups, in case of MAD and MSE for 4 out of 4 product groups).

Robustness

Although the positive effect of method HWP decreases for larger *h*, it often remains the best method for point forecasts. It always (with one exception) shows an improvement over HW, although this may become small. In Table 7, the mean average ranks for MSE over all product families are computed, together with the average standard deviations. This serves as an indication of the robustness of all methods. In the short run especially the combined methods (HWC and HWPC) show a low rank, while for longer horizons HWP shows a decreasing rank. Together, these three methods have the lowest ranks, meaning that on average they outperform other methods. As explained, in the short run, combining methods yields good results. For the longer term, its value diminishes. The fact that combined methods in the long term still show a slightly better rank than HWP is maybe due to randomness. However, HWP always has a lower rank and a lower standard deviation than Holt-Winters, showing it is a more robust method. Methods HW and SES often show a high rank.

Point forecasts vs. Cumulative forecasts

For the cumulative forecasts, the sMAPE's mostly decrease when *h* increases and the total MAD and MSE grow less than proportionally with *h*, while the number of periods contained in each forecast (*h*) increases. In other words, forecasts become relatively more accurate. This can be explained by the fact that we make forecasts at an increasingly higher (temporal) aggregation level, which in general is less difficult and thus gives a relatively higher accuracy.

The results are comparable to those for point forecasts. However, the results show more variation. In most cases the advantage of the alternative methods (HWC, HWP, HWPC) compared to method HW decreases when *h* increases (See Table 5), and in some cases they perform worse than HW. Mostly the relative improvements over HW are in between those for point forecasts for *h=1* and either *h=4, h=8* or *h=12*. This is expected since the cumulative forecasts are sums of point forecasts at different horizons. Method SES performs considerably worse than for point forecasts. This can be explained by the fact that if we only base our forecast on the level of demand, we miss out on the seasonal pattern. When we sum several forecasts at different horizons, always some of these forecasts are in error.

Table 5: difference with method HW when sMAPE is minimized

	method	soft drinks small packages visual inspection group, cumulative forecasts			
Horizon		4	8	12	24
	HW				
	SES	-15%	23%	81%	95%
	HWC	-21%	-17%	0%	-6%
	HWP	-8%	-3%	-16%	-10%
	HWPC	-20%	-15%	-8%	-12%

6. Statistical clusters

In order to explore the issue of how to group products with similar seasonal patterns, we examine a standard statistical clustering method (see Section 3). The aim is to see whether high correlation among seasonal patterns would guarantee successful application of method HWP. Tables 8-11 show the results for two different grouping methods. Since both groups were complemented with products from the other group, the results for methods SES, HW and HWC are the same for both grouping methods. For aggregation-based methods (HWP and HWPC) the results are different. In each of the four cases, the two kinds of groups consisted of largely the same products. Despite this large overlap between the groups, the results on some occasions differed substantially. For example, compare the results for the two grouping methods for the same product group in Table 6.

Table 6: difference with method HW when sMAPE is minimized

	method	soft drinks small packages visual inspection group, point forecasts				soft drinks small packages statistical clusters, point forecasts			
Horizon		1	4	8	12	1	4	8	12
	HW								
	SES	-25%	0%	31%	37%	-25%	0%	31%	37%
	HWC	-23%	-3%	-6%	2%	-23%	-3%	-6%	2%
	HWP	-9%	-7%	-7%	-8%	-11%	4%	23%	9%
	HWPC	-24%	-3%	-7%	-3%	-23%	1%	6%	15%

Also, for the same type of products, there can be differences between the two grouping methods for different accuracy measures. Under sMAPE visual inspection may be preferred, while under MSE statistical clustering is preferred. And for the same accuracy measures there can be differences between performance of the grouping methods as well. For the 'soft drinks regular' method HWP (under MSE) performs better for the visual inspection groups than for cluster groups, while for 'soft drinks small packages' the opposite is the case.

Since the two kinds of groups both consist of products with similar seasonal patterns, we conclude that results for HWP depend on the exact combination of products that are in a group. This shows that the way the products are grouped together can have a large influence on the results, and that a standard statistical clustering method is not necessarily appropriate. A good statistical cluster (with products that are very similar) does not necessarily imply that it is optimal for product-aggregation. Therefore, it is important to further study how groups should be formed and find a refined clustering method.

Table 7: mean average ranks and standard deviations (in brackets) for MSE over all product groups

method	visual inspection point forecasts				statistical clusters point forecasts			
horizon	1	4	8	12	1	4	8	12
HW	4,4 (1,1)	3,8 (1,5)	3,7 (1,5)	3,3 (1,5)	4,2 (1,1)	3,6 (1,5)	3,5 (1,6)	3,1 (1,5)
SES	2,5 (0,8)	3,1 (1,1)	3,5 (0,9)	3,7 (1,1)	2,3 (0,9)	2,9 (1,2)	3,3 (1,1)	3,4 (1,2)
HWC	2,3 (1,0)	2,3 (1,1)	2,3 (1,1)	2,4 (1,2)	2,2 (0,9)	2,2 (0,9)	2,1 (1,0)	2,2 (1,1)
HWP	3,5 (1,1)	3,3 (1,2)	2,9 (1,3)	2,6 (1,3)	3,9 (0,9)	3,6 (1,0)	3,1 (1,3)	2,9 (1,3)
HWPC	2,0 (1,0)	2,2 (1,0)	2,4 (1,1)	2,7 (1,1)	2,1 (0,9)	2,4 (1,1)	2,6 (0,9)	3,0 (0,9)

Despite the differences between the two grouping approaches, the aggregation-based forecasts of both groups consistently improved on the results of Holt-Winters. In Table 7, as a measure of robustness the mean of the average ranks per product group are shown, together with the average standard deviation of the four groups. The low ranks and standard deviations of the new methods (HWC, HWP, HWPC) show the good performance and robustness of these methods.

7. Conclusions

In this article, we have investigated the influence of the forecasting horizon when forecasting with group seasonal indices methods. Furthermore, we have looked at how to group products to benefit from these product-aggregation methods. Both Makridakis and Hibon (2000) and Chatfield (2000) report that no forecasting method performs well both across forecasting horizons and across accuracy measures. Furthermore, for nonlinear models, forecast errors usually do not increase monotonically when the forecast horizon increases (Chatfield, 2000). Although we have found this to be true for the product-aggregation method as well, in general we can conclude that product-aggregation performs relatively well across horizons and across accuracy measures, although the advantage over other methods when using this approach decreases for longer horizons. For the short term it is a method that can improve on forecasts made by Holt-Winters significantly. For the longer horizon its main advantage lies in its higher robustness compared to Holt-Winters. However, the way products are clustered is important. In this article we have seen that a standard clustering algorithm need not be optimal and thus that forming product families is not a straightforward task. Therefore, we suggest that better ways of grouping products should be developed.

References

Bunn, D.W. and A.I. Vassilopoulos (1993). Using group seasonal indices in multi-item short-term forecasting, *International Journal of Forecasting* 9, 517-526.

Bunn, D.W. and A.I. Vassilopoulos (1999). Comparison of seasonal estimation methods in multi-item short-term forecasting, *International Journal of Forecasting* 15, 431-443.

Chatfield, C. (2000). Time-series forecasting, Chapman & Hall, London.

Chatfield, C. and M. Yar (1991). Prediction intervals for multiplicative Holt-Winters, *International Journal of Forecasting* 7, 31-37.

Dalhart, G. (1974). Class seasonality - A new approach, *American Production and Inventory Control Society 1974 Conference Proceedings*, reprinted in Forecasting, 2nd ed. (APICS, Washington, DC) 11-16.

Dangerfield, B.J. and J.S. Morris (1992). Top-down or bottom-up: Aggregate versus disaggregate extrapolations, *International Journal of Forecasting* 8, 233-241.

Dekker, M., K.H. van Donselaar and P. Ouwehand (2004). How to use aggregation and combined forecasting to improve seasonal demand forecasts, *International Journal of Production Economics* 90, 2, 151-167.

Holt, C.C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages, Office of Naval Research, Memorandum No. 52, Carnegie Institute of Technology, Pittsburgh, Pennsylvania.

Johnson, R.A. and D.W. Wichern (1998). *Applied multivariate statistical analysis*, 4th ed., Prentice Hall, Upper Saddle River.

Makridakis, S. and M. Hibon (2000). The M3-Competition: results, conclusions and implications, *International Journal of Forecasting* 16, 451-476.

Makridakis, S., S. Wheelwright and J. Hyndman (1998). *Forecasting Methods and Applications*, John Wiley, New York.

Ouwehand, P., K.H. van Donselaar and A.G. de Kok (2004). The impact of the forecasting horizon when forecasting with group seasonal indices, working paper, Beta Research School, Technische Universiteit Eindhoven.

Schwarzkopf, A.B., R.J. Tersine and J.S. Morris (1988). Top-down versus bottom-up forecasting strategies, *International Journal of Production Research* 26, 11, 1833-1843.

Shlifer, E. and R.W. Wolff (1978). Aggregation and proration in forecasting, *Management Science* 25, 6, 594.
Silver, E.A., D.F. Pyke, and R. Peterson (1998). *Inventory Management and Production Planning and Scheduling*, 3rd. ed., John Wiley & Sons, New York.
Winters, P.R. (1960). Forecasting Sales by exponentially weighted moving averages, *Management Science* 6, 324-342.
Withycombe, R. (1989). Forecasting with combined seasonal indices, *International Journal of Forecasting* 5, 547-552.

This research was supported by the Technology Foundation STW, applied science division of NWO and the technology programme of the Dutch Ministry of Economic Affairs.

Appendix

Table 8: difference with method HW when sMAPE is minimized

	method	visual inspection point forecasts				visual inspection cumulative forecasts				statistical clusters point forecasts			
Horizon		1	4	8	12	4	8	12	24	1	4	8	12
beers	HW												
	SES	-12%	4%	12%	13%	12%	37%	55%	70%	-12%	4%	12%	13%
	HWC	-14%	-5%	-4%	-2%	-6%	-5%	-1%	6%	-14%	-5%	-4%	-2%
	HWP	-6%	-1%	1%	2%	-2%	6%	12%	20%	1%	3%	6%	4%
	HWPC	-15%	-6%	-2%	0%	-8%	-5%	-1%	20%	-12%	-3%	2%	5%
soft drinks regular													
	HW												
	SES	-17%	-20%	-14%	-12%	-22%	-10%	3%	2%	-17%	-20%	-14%	-12%
	HWC	-18%	-16%	-16%	-8%	-21%	-19%	-15%	-10%	-18%	-16%	-16%	-8%
	HWP	-11%	-20%	-19%	-18%	-24%	-27%	-26%	-14%	-4%	-2%	-2%	-3%
	HWPC	-20%	-23%	-18%	-18%	-30%	-28%	-23%	-16%	-19%	-17%	-5%	-6%
soft drinks small packages													
	HW												
	SES	-25%	0%	31%	37%	-15%	23%	81%	95%	-25%	0%	31%	37%
	HWC	-23%	-3%	-6%	2%	-21%	-17%	0%	-6%	-23%	-3%	-6%	2%
	HWP	-9%	-7%	-7%	-8%	-8%	-3%	-16%	-10%	-11%	4%	23%	9%
	HWPC	-24%	-3%	-7%	-3%	-20%	-15%	-8%	-12%	-23%	1%	6%	15%
plastic tubes													
	HW												
	SES	1%	7%	8%	8%	17%	16%	8%	14%	1%	7%	8%	8%
	HWC	-7%	-3%	-2%	-3%	0%	-4%	-6%	-2%	-7%	-3%	-2%	-3%
	HWP	-4%	-3%	-3%	-4%	-2%	-3%	-5%	2%	-3%	-2%	-2%	-1%
	HWPC	-11%	-4%	-7%	-8%	-3%	-4%	-6%	2%	-9%	-5%	-4%	-3%

Table 9: difference with method HW when MAD is minimized

	method	visual inspection point forecasts				visual inspection cumulative forecasts				statistical clusters point forecasts			
horizon		1	4	8	12	4	8	12	24	1	4	8	12
beers	HW												
	SES	-16%	-1%	1%	1%	5%	30%	39%	70%	-16%	-1%	1%	1%
	HWC	-17%	-4%	-2%	-2%	-11%	1%	-2%	4%	-17%	-4%	-2%	-2%
	HWP	-7%	-4%	-3%	-1%	-10%	-1%	-2%	5%	-3%	-1%	0%	-1%
	HWPC	-17%	-6%	-4%	-2%	-13%	1%	-2%	5%	-17%	-6%	-4%	-3%
soft drinks regular													
	HW												
	SES	-24%	-17%	-12%	-9%	-12%	-9%	6%	13%	-24%	-17%	-12%	-9%
	HWC	-25%	-14%	-14%	-5%	-17%	-17%	-15%	-5%	-25%	-14%	-14%	-5%
	HWP	-16%	-12%	-13%	-9%	-18%	-25%	-19%	-6%	-7%	-3%	-2%	-3%
	HWPC	-26%	-18%	-19%	-10%	-22%	-27%	-20%	-7%	-25%	-14%	-14%	-5%
soft drinks small packages													
	HW												
	SES	-33%	-10%	15%	14%	-22%	22%	58%	63%	-33%	-10%	15%	14%
	HWC	-30%	-13%	-10%	-4%	-27%	-11%	-8%	-5%	-30%	-13%	-10%	-4%
	HWP	-7%	0%	-6%	-6%	-5%	3%	-4%	-15%	-14%	0%	11%	-3%
	HWPC	-32%	-11%	-7%	0%	-24%	-5%	1%	-3%	-31%	-9%	1%	11%
plastic tubes													
	HW												
	SES	0%	13%	12%	11%	19%	23%	16%	8%	0%	13%	12%	11%
	HWC	-10%	-2%	-4%	-5%	-3%	-2%	-5%	-2%	-10%	-2%	-4%	-5%
	HWP	-4%	-1%	-4%	-3%	-2%	-3%	-3%	-5%	-6%	-6%	-7%	-3%
	HWPC	-11%	-4%	-9%	-7%	-4%	-3%	-5%	-4%	-12%	-5%	-10%	-7%

Table 10: difference with method HW when MSE is minimized

	method	visual inspection point forecasts				visual inspection cumulative forecasts				statistical clusters point forecasts			
horizon		1	4	8	12	4	8	12	24	1	4	8	12
beers	HW												
	SES	-34%	-5%	-3%	-5%	27%	71%	102%	169%	-34%	-5%	-3%	-5%
	HWC	-36%	-11%	-1%	-3%	-15%	-10%	-3%	3%	-36%	-11%	-1%	-3%
	HWP	-6%	-1%	-3%	-1%	-3%	4%	-5%	16%	-6%	-6%	-5%	-5%
	HWPC	-36%	-10%	-5%	-5%	-13%	-10%	-3%	21%	-36%	-14%	-9%	-6%
soft drinks regular													
	HW												
	SES	-54%	-22%	-21%	-10%	-23%	-15%	-10%	25%	-54%	-22%	-21%	-10%
	HWC	-55%	-16%	-21%	-10%	-30%	-26%	-26%	-17%	-55%	-16%	-21%	-10%
	HWP	-47%	-24%	-27%	-19%	-27%	-35%	-37%	-7%	-8%	-4%	-5%	-6%
	HWPC	-59%	-27%	-33%	-18%	-38%	-41%	-40%	-11%	-56%	-18%	-24%	-14%
soft drinks small packages													
	HW												
	SES	-79%	-53%	-41%	-26%	-47%	54%	125%	332%	-79%	-53%	-41%	-26%
	HWC	-72%	-42%	-40%	-11%	-40%	-27%	-11%	483%	-72%	-42%	-40%	-11%
	HWP	-4%	-2%	-4%	-5%	4%	14%	20%	20%	-63%	-42%	-43%	-35%
	HWPC	-73%	-41%	-35%	-2%	-36%	-3%	33%	639%	-78%	-56%	-35%	8%
plastic tubes													
	HW												
	SES	4%	33%	28%	26%	72%	54%	37%	45%	4%	33%	28%	26%
	HWC	-11%	-4%	-11%	-11%	-4%	-4%	-7%	5%	-11%	-4%	-11%	-11%
	HWP	-3%	-4%	-6%	-3%	-13%	-15%	-14%	2%	-1%	-10%	-7%	0%
	HWPC	-10%	-7%	-13%	-11%	-12%	-14%	-15%	9%	-11%	-10%	-14%	-8%

Table 11: average ranks when MSE is minimized

	method	visual inspection point forecasts				visual inspection cumulative forecasts				statistical clusters point forecasts			
horizon		1	4	8	12	4	8	12	24	1	4	8	12
beers	HW	4,2	3,1	3,0	2,9	2,6	2,3	2,5	2,0	3,8	2,8	2,6	2,4
	SES	2,3	3,5	3,6	4,0	3,9	4,3	4,5	4,6	1,9	3,3	3,2	3,4
	HWC	2,0	1,9	2,4	2,4	1,9	1,8	1,9	2,2	1,8	1,6	2,1	2,1
	HWP	3,8	3,9	3,3	3,1	3,6	3,7	3,1	2,9	4,6	4,3	3,8	3,9
	HWPC	2,4	2,4	2,6	2,4	2,7	1,8	1,9	3,0	2,9	3,1	3,2	3,1
soft drinks regular													
	HW	4,6	4,3	4,4	4,0	4,1	3,3	3,1	2,7	4,1	3,7	3,9	3,6
	SES	2,7	3,0	3,4	3,5	3,3	4,0	4,1	3,9	2,3	2,3	2,8	3,0
	HWC	2,5	3,1	2,7	2,9	2,8	2,7	2,9	2,7	2,0	2,3	2,0	2,4
	HWP	3,3	3,0	2,9	2,2	3,0	2,8	2,6	2,9	4,0	3,8	3,4	3,0
	HWPC	1,6	1,6	1,6	2,3	1,9	2,1	2,3	2,8	1,7	2,1	1,9	2,1
soft drinks small packages													
	HW	4,6	4,2	3,7	2,9	3,1	2,7	2,3	1,8	4,8	4,1	3,8	2,9
	SES	1,6	2,2	3,4	3,4	2,1	3,5	3,9	4,2	1,6	2,3	3,3	3,3
	HWC	2,4	1,8	1,8	2,3	2,1	1,7	1,9	2,8	2,6	2,0	1,9	2,3
	HWP	3,9	3,6	2,2	1,8	4,2	3,2	2,8	1,7	3,9	4,1	2,1	1,7
	HWPC	1,9	2,7	3,3	4,1	3,0	3,4	3,6	4,1	2,1	2,5	3,8	4,7
plastic tubes													
	HW	4,1	3,6	3,5	3,2	3,5	3,6	3,5	2,8	4,0	3,8	3,5	3,3
	SES	3,2	3,7	3,7	3,8	4,2	3,4	3,5	3,5	3,3	3,6	4,0	3,9
	HWC	2,4	2,5	2,3	2,2	2,5	3,0	2,9	3,0	2,5	2,8	2,4	2,1
	HWP	3,0	2,9	3,2	3,5	2,5	2,5	2,5	2,4	2,9	2,5	3,0	3,1
	HWPC	2,0	2,0	2,0	2,0	2,0	2,2	2,3	2,9	1,8	1,8	1,6	2,2