European Journal of Operational Research 263 (2017) 412–418

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Production, Manufacturing and Logistics

Integrated hierarchical forecasting

Clint L. P. Pennings*, Jan van Dalen

Rotterdam School of Management, Erasmus University, T9-08, Burgemeester Oudlaan 50, Rotterdam 3062 PA, The Netherlands

ARTICLE INFO

Article history:
Received 8 December 2015
Accepted 18 April 2017
Available online 27 April 2017

Keywords:
Forecasting
Hierarchical
Top-down
Bottom-up
Decision-making

ABSTRACT

Forecasts are often made at various levels of aggregation of individual products, which combine into groups at higher hierarchical levels. We provide an alternative to the traditional discussion of bottom-up versus top-down forecasting by examining how the hierarchy of products can be exploited when forecasts are generated. Instead of selecting series from parts of the hierarchy for forecasting, we explore the possibility of using all the series. Moreover, instead of using the hierarchy after the initial forecasts are generated, we consider the hierarchical structure as a defining feature of the data-generating process and use it to instantaneously generate forecasts for all levels of the hierarchy. This integrated approach uses a state space model and the Kalman filter to explicitly incorporate product dependencies, such as complementarity of products and product substitution, which are otherwise ignored. An empirical study shows the substantial gain in forecast and inventory performance of generalizing the bottom-up and top-down forecast approaches to an integrated approach. The integrated approach is applicable to hierarchical forecasting in general, and extends beyond the current application of demand forecasting for manufacturers.

1. Introduction

For organizations, demand forecasting is essential as it drives production, inventory and planning decisions. Supply has to match demand as well as possible to avoid excess inventory and stockouts. Large manufacturers often have SKUs ranging in the thousands, spanning several product categories, each of which requires forecasts. Several decision makers are involved from operations, marketing, sales and finance, who require forecasts at various levels of aggregation. Forecasts are more easily discussed at an aggregated product level, but for production these forecasts have to be available at the SKU level.

SKUs naturally group together in hierarchies with individual sales per product at the base line, followed by several intermediary levels, denoting sales for groups of related products at increasingly general aggregation levels, such as product groups and categories, and with a top level that lists total sales. Two commonly used approaches in practice and research start from opposite ends of this hierarchy to generate forecasts for all series: bottom-up forecasting and top-down forecasting (Widiarta, Viswanathan, & Piplani, 2009). In bottom-up forecasting, base forecasts are generated for product demand at the lowest level in the hierarchy (Gordon, Morris, & Dangerfield, 1997). Subsequently, these are aggregated to determine forecasts at higher hierarchical levels. Bottom-up forecasting is commonly contrasted with top-down forecasting, in which aggregated demand forecasts are disaggregated downwards to determine forecasts at lower levels in the hierarchy (Kahn, 1998). Research stretches over three decades with mixed results as to a preference for either bottom-up or top-down forecast approaches (Syntetos, Babai, Boylan, Kolassa, & Nikolopoulos, 2016).

Both approaches generate forecasts for a selected part of the hierarchy, aggregated upwards or allocated downwards to obtain forecasts for the remaining series. This aggregation and allocation imply a potential loss of information, as the ignored series can only be recovered under stringent conditions. The loss of information is exacerbated as the selected series are forecasted separately. Product dependencies, such as complementarity of products and product substitution, are explicitly ignored. Yet product dependencies motivate combining similar products in groups and the existence of hierarchies.

Hyndman, Ahmed, Athanasopoulos, and Shang (2011) introduce a combination approach that uses forecasts of all series in the hierarchy. By taking a linear combination of the bottom-up and top-down forecasts at various hierarchical levels, their approach offers an ensemble of the bottom-up and top-down approaches. The combination entails a post-hoc revision of forecasts to ensure that forecasts add up consistently throughout the hierarchy. More forecasts are involved than in either the bottom-up and top-down approaches alone, but the initial forecasts are still generated independently.

* Corresponding author.
E-mail addresses: cpennings@rsm.nl (C.L.P. Pennings), jdalen@rsm.nl (J. van Dalen).

http://dx.doi.org/10.1016/j.ejor.2017.04.047
0377-2217/

The bottom-up, top-down and combination approaches use the hierarchy of products only after the initial forecasts are generated. By incorporating the hierarchical structure at an earlier stage, i.e., during the generation of forecasts, we introduce an integrated approach, superseding the traditional discussion of bottom-up versus top-down forecasting. This has at least two advantages. First, instead of selecting isolated series for forecasting, all the available data in the hierarchy can be used. Second, product dependencies can be explicitly incorporated, such as complementarity of products and product substitution, while these are otherwise ignored.

An empirical application evaluates the forecasting approaches for one of the largest manufacturers of consumer products, which has hundreds of brands spanning fourteen categories of food products, home and personal care. The empirical study shows a substantial gain in forecast and inventory performance of generalizing the bottom-up and top-down forecast approaches to an integrated approach.

The remainder of this paper is organized as follows. In Section 2, we present an overview of the relevant literature on hierarchical forecasting and the bottom-up, top-down, and combination approaches for forecasting. We especially focus on the use of the hierarchical structure, product dependencies and demand heteroscedasticity, and we critically evaluate several approaches. In Section 3, we introduce multiple state space models that are used as an integrated approach for hierarchical forecasting and outline the empirical study. For the empirical study, we compare approaches in terms of forecasting and inventory performance and use the company's own forecast as a benchmark. Section 4 lists the results and their implications, while Section 5 concludes and gives suggestions for future research.

2. Theoretical background

Hierarchical forecasting has different forms pertaining to temporal and contemporaneous aspects. Here, we exclusively focus on contemporaneous hierarchies, specifically on products aggregated in groups and categories. This section summarizes the relevant theoretical background on hierarchical forecasting and the approaches of bottom-up, top-down, and the combination approach of Hyndman et al. (2011) for forecasting. We especially focus on the use of the hierarchical structure, product dependencies and heteroscedasticity in product demand, and critically evaluate approaches.

Over three decades of forecasting literature show mixed results as to a preference for either top-down or bottom-up forecasting (Syntetos et al., 2016). This is not surprising as the performance of the approaches depends on the underlying demand process of products (Lütkepohl, 1984). Due to the additive nature of the hierarchy, in which sums of product sales determine group sales, which, in turn, add up to determine category sales, the underlying demand process is transformed at various levels of the hierarchy. Aggregation can lead to substantial information loss, which makes bottom-up forecasting seem favorable (e.g., Edwards & Orcutt, 1969; Orcutt, Watts, & Edwards, 1968; Zellner, 1969). However, if no important information is lost, benefits can be gained if random noise cancels out (Fliedner, 1999), which makes top-down forecasting seem more favorable. A wide variety of performance is seen as the nature and extent of differences between top-down and bottom-up are dependent upon context and the assumed demand processes (Wei & Abraham, 1981). Examples show that conclusions may revolve around differences in demand processes or parameter settings (Widiarta, Viswanathan, & Piplani, 2007; 2009).

Dependencies between the demand for different products are a key characteristic of the demand process, and hence a main driver of differences in performance between top-down and bottom-up approaches (Chen & Boylan, 2009; Kohn, 1982; Schwarzkopf, Tersine, & Morris, 1988; Tiao & Guttman, 1980). A particular type of demand dependency does not unequivocally make either bottom-up or top-down more favorable (Fliedner & Mabert, 1992; Fliedner, 2001; Sohn & Lim, 2007). Stronger negative cross-correlations between individual demand series lead to less variation at an aggregate level, but imply differences between individual product sales. In contrast, stronger positive correlations between individual demand series lead to more variable aggregate sales, but imply that differences at the individual product level are smaller.

This explains why empirical studies are unable to consistently show one approach outperforming the other. Dangerfield and Morris (1992) compare bottom-up and top-down approaches on empirical data and conclude that bottom-up forecasting is more accurate, especially when products are highly correlated. By contrast, Fliedner (1999) concludes that stronger positive and negative correlations improve the forecast at the aggregate level to such an extent that the top-down approach is more accurate.

An important difference between the bottom-up and top-down approaches is that the latter requires additional measures to allocate an aggregate forecast downwards to lower levels in the hierarchy. Gross and Sohl (1990) compare various ways of determining allocation proportions. A common allocation is based on averaging historical sales proportions, where the unweighted proportion p_j for each product j is determined as its sales y_j relative to the total sales in the product category y over time period T.

$$p_j = \frac{1}{T}\sum_{t=1}^{T}\frac{y_{j,t}}{y_t} \tag{1}$$

A common alternative is based on a single, total proportion observed over all time periods, leading to a weighted allocation:

$$p_j = \sum_{t=1}^{T} y_{j,t} \Big/ \sum_{t=1}^{T} y_t \tag{2}$$

Both allocations perform well in practice (Gross & Sohl, 1990).

The two approaches of top-down and bottom-up can also be combined at intermediary levels in the hierarchy, known as the middle-out approach. Forecasts are generated at a particular level and then aggregated upwards using the bottom-up approach, and allocated downwards using a top-down approach.

Recently, Athanasopoulos, Ahmed, and Hyndman (2009) and Hyndman et al. (2011) introduced a different approach, labeled the combination approach, which uses the hierarchical structure to create revised forecasts. This forecasting approach follows two steps: (1) generate independent forecasts for each series in the hierarchy, (2) weight these forecasts according to the hierarchical structure to determine the final forecasts. These final forecasts adhere to the hierarchical structure in the sense that aggregates of the forecasts at the bottom level exactly match forecasts at higher levels in the hierarchy.

The combination approach proposed by Hyndman et al. (2011) is a continuation of earlier work on revising measurements of macro-economic indicators (e.g., Byron, 1978; Solomou & Weale, 1991; 1993; 1996; Stone, Champernowne, & Meade, 1942; Weale, 1985; 1988). A salient difference is that Hyndman et al. (2011) have underlying time series of sales available for each forecast. We introduce notation for hierarchical series to discuss the combination approach, focusing on sales without loss of generality. We have a large vector $\boldsymbol{y}_t$ which contains the n sales series at all levels of the hierarchy. Sales at higher levels are determined by aggregating sales of m products at the lowest level $\boldsymbol{b}_t$. $\boldsymbol{y}_t$ is an $n \times 1$ matrix determined by linear combinations of the $m \times 1$ vector $\boldsymbol{b}_t$ containing sales at the base product level, using an $n \times m$ design matrix $\boldsymbol{S}$ to link sales at each level of the hierarchy with the base level sales:

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t \tag{3}$$

For forecasting, we are interested in the expected $\boldsymbol{y}_t$ and $\boldsymbol{b}_t$. Hyndman et al. (2011) determine the unknown forecasts of product sales, $\hat{\boldsymbol{b}}_t$, as a function of initial forecasts $\hat{\boldsymbol{y}}_t$ generated for each series in the hierarchy by regression, supposing:

$$\hat{\boldsymbol{y}}_t = \boldsymbol{S}\hat{\boldsymbol{b}}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}) \tag{4}$$

They assume, as a simplifying approximation, that $\boldsymbol{\varepsilon}_t$ can be derived from a linear combination, using design matrix $\boldsymbol{S}$, of only the errors at the base level $\boldsymbol{\epsilon}$: $\boldsymbol{\varepsilon}_{n\times 1} = \boldsymbol{S}_{n\times m}\boldsymbol{\epsilon}_{m\times 1}$. They use a generalized inverse of $\boldsymbol{\Sigma}$ to estimate $\hat{\boldsymbol{b}}_t$, which can then be obtained using simple ordinary least squares involving only $\boldsymbol{S}$ and $\hat{\boldsymbol{y}}$ (for details see Hyndman et al., 2011, p.2583). Note that $\hat{\boldsymbol{y}}_t$ contains forecasts for all series, including the product demand forecasts at the lowest level. The revised forecasts are then generated as:

$$\tilde{\boldsymbol{y}}_t = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'\hat{\boldsymbol{y}}_t \tag{5}$$

They conclude that their method is "optimal" and has "minimum variance amongst all combination forecasts under some simple assumptions" (Hyndman et al., 2011, p.2579). For the multivariate analogue of the Gauss–Markov theorem to apply we require homoscedasticity and absence of cross-correlation in the error terms, requiring the variance–covariance matrix $\boldsymbol{\Sigma}$ to be $\sigma^2\boldsymbol{I}$. However, instead of equal variance of sales over products, it is likely that some series have lower variation and are, as a result, easier to forecast than other series. Series with lower variation give more insight into the underlying demand and so we should give more weight to the observations of these series. A recent extension can use the inverse of the variances of the base forecasts to estimate the revised forecasts, which partly addresses this concern (Hyndman, Lee, & Wang, 2016). Yet, products are still treated independently. For a manufacturer with multiple products, many dependencies may exist among product sales due to complementarity of products and product substitution. The restrictions on the variance-covariance matrix $\boldsymbol{\Sigma}$ are violated in practice when there are heteroscedasticity and product dependencies, which can make its proposed estimator highly inefficient, resulting in large mean square forecast errors.

The bottom-up and top-down approaches are based on subsets of all demand series. The bottom-up approach only uses the series at the bottom of the hierarchy as input, while the top-down approach takes the series at the top of the hierarchy as input. These approaches exclusively focus on different parts of the hierarchy, in effect ignoring information. A bottom-up approach cannot benefit from possible noise canceling out at higher hierarchical levels, while the top-down approach suffers from information loss due the use of aggregated series (Fliedner, 1999; Gordon et al., 1997; Kahn, 1998). Both approaches create initial forecasts for the series in the selected parts of the hierarchy only. In the bottom-up approach, forecasts for series at higher hierarchical levels are derived by aggregating the forecasts. In the top-down approach, forecasts for lower hierarchical levels are determined by allocating forecasts downwards in the hierarchy. Moreover, forecasts are generated for each series independently. In contrast, the combination approach can use information from different series more flexibly, as it can use a selection of forecasts generated by both of the other approaches. For example, it can use forecasts at higher levels from a top-down approach, and forecasts at lower levels derived from a bottom-up approach. However, it only uses the hierarchy after forecasts are generated to reconcile forecasts. As the forecasts are generated separately, the hierarchy is not applied to consider the underlying time series of sales for initial forecast generation.

All three approaches ignore the hierarchy when generating the forecasts, and, as a consequence, ignore product dependencies and possible heteroscedasticity in the demand. Exploiting the hierarchy that characterizes the original sales series circumvents the discussion between bottom-up and top-down approaches by directly tackling the underlying demand process.

3. Methodology

We introduce state space models as an integrated hierarchical forecasting approach. As empirical evidence is needed (Moon, Hicks, & Simpson, 2012; Rostami-Tabar, Babai, Ducq, & Syntetos, 2015; Syntetos et al., 2016), we apply the approach to real world sales data from a global supplier in fast-moving consumer goods. We then introduce the company and explain our empirical study in which we compare the performance of the integrated hierarchical approach to the bottom-up and top-down approaches and the combination approach.

3.1. Integrated approach

Multivariate state space models allow an alternative approach for hierarchical series, because they can be efficiently estimated using the Kalman filter to decompose the prediction error (Durbin & Koopman, 2012; Harvey, 1989). Hence, they have many applications in forecasting (e.g., Fildes, Nikolopoulos, Crone, and Syntetos, 2008; De Gooijer and Hyndman, 2006; Hyndman, Koehler, Ord, and Snyder, 2008; Snyder et al., 2012; Syntetos, Boylan, & Disney, 2009). The local level model is a special case:

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{S}\boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_\varepsilon) \\ \boldsymbol{\mu}_t &= \boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_\eta) \end{aligned} \tag{6}$$

It links observed quantities $\boldsymbol{y}_t$ to unobserved time-varying components $\boldsymbol{\mu}_t$ using a design matrix $\boldsymbol{S}$.

The state space model in Eq. (6) can use the same design matrix $\boldsymbol{S}$ as in Eq. (3), reflecting the hierarchy. This means that $\boldsymbol{\mu}_t$ refers to demand at the product level. Sales are not directly observed by manufacturers, but various, possibly conflicting, estimates of sales are available for the different levels in the hierarchy as $\boldsymbol{y}_t$, necessitating the inclusion of measurement noise $\boldsymbol{\varepsilon}_t$.

For a practical application, it is necessary to change model (6) to more flexibly capture the market for consumer goods, which is characterized by possible trends and seasonality. Product sales at all levels of the hierarchy $\boldsymbol{y}_t$ are determined by using the design matrix $\boldsymbol{S}$ on the base product sales $\boldsymbol{\mu}_t$ and seasonal sales γ_t, where seasonality differs per week of the year:

$$\boldsymbol{y}_t = \boldsymbol{S}(\boldsymbol{\mu}_t + \boldsymbol{A}\gamma_t) + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_\varepsilon) \tag{7}$$

This formulation for the seasonality is based on a single overall seasonal pattern within a product group that affects products differently. The seasonal effect can be scaled differently for each product, using a diagonal loading matrix $\boldsymbol{A}$ (Durbin & Koopman, 2012).

$$\boldsymbol{A} = \mathrm{diag}(\psi_1, \ldots, \psi_n) \tag{8}$$

We include weekly seasonality through a trigonometric specification to limit the number of parameters in γ_t to estimate (Durbin & Koopman, 2012). Moreover, rather than having a constant seasonality effect throughout the three year period, we allow seasonality to change over time to cope with market changes. Thus, seasonality is stochastic and its influence can change over the 52 weeks per year. The formulation allows for seasonal effects that are smoothed over time and ensures that the contributions of the seasonal errors ω_{jt} and ω^*_{jt} are not amplified by the trigonometric functions (Proietti, 2000):

$$\begin{aligned}
\gamma_t &= \sum_{j=1}^{26} \gamma_{jt}, \quad s = 52, \quad \lambda_j = \frac{2\pi j}{s} \\
\gamma_{j,t} &= \gamma_{j,t-1} \cos\lambda_j + \gamma^*_{j,t-1} \sin\lambda_j + \omega_{jt}, \qquad \omega_{jt} \sim N(0, \sigma^2_\omega) \\
\gamma^*_{j,t} &= -\gamma_{j,t-1} \sin\lambda_j + \gamma^*_{j,t-1} \cos\lambda_j + \omega^*_{jt}, \qquad \omega^*_{jt} \sim N(0, \sigma^2_\omega)
\end{aligned} \tag{9}$$

The product sales $\boldsymbol{y}_t$ have independent measurement noise with diagonal matrix $\boldsymbol{\Sigma}_\varepsilon$, as the product dependencies within product groups are contained in the underlying states $\boldsymbol{\mu}_t$. The sales per product $\boldsymbol{\mu}_t$ follow an autoregressive process with diagonal coefficient matrix $\boldsymbol{\Gamma}$ and disturbances $\boldsymbol{\eta}_t$. Possible cross-correlations are represented by the variance–covariance matrix $\boldsymbol{\Sigma}_\eta$. Sales for products are likely to change over time due to market developments, so that sales can have short-term positive or negative trends. An additive trend $\boldsymbol{\beta}_t$ is included to incorporate market developments (Durbin & Koopman, 2012):

$$\begin{aligned}
\boldsymbol{\mu}_t &= \boldsymbol{\beta}_t + \boldsymbol{\Gamma}\boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\eta) \\
\boldsymbol{\Gamma} &= \mathrm{diag}(\gamma_1, \ldots, \gamma_n) \\
\boldsymbol{\beta}_t &= \boldsymbol{\beta}_{t-1} + \boldsymbol{\zeta}_t, \qquad \boldsymbol{\zeta}_t \sim N(\mathbf{0}, \sigma^2_\zeta \boldsymbol{I})
\end{aligned} \tag{10}$$

Using this approach, forecasts for aggregate levels are not generated independently, but derived as the sums of forecasts of base product sales. The Kalman filter traces the forecast errors for each series at each level back to the underlying states, so that the forecasts use information from all series. Many alternative formulations are possible for the model, but two variations are most appealing. If the matrix $\boldsymbol{\Gamma}$ is restricted to the identity matrix, the model without seasonality is a local linear trend model (Durbin & Koopman, 2012; Harvey, 1989):

$$\boldsymbol{\mu}_t = \boldsymbol{\beta}_t + \boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\eta) \tag{11}$$

If the matrix $\boldsymbol{\Gamma}$ is used for $\boldsymbol{\beta}_t$ instead, the model is an additive damped trend model (Gardner & McKenzie, 1985):

$$\begin{aligned}
\boldsymbol{\mu}_t &= \boldsymbol{\beta}_t + \boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\eta) \\
\boldsymbol{\Gamma} &= \mathrm{diag}(\gamma_1, \ldots, \gamma_n) \\
\boldsymbol{\beta}_t &= \boldsymbol{\Gamma}\boldsymbol{\beta}_{t-1} + \boldsymbol{\zeta}_t, \qquad \boldsymbol{\zeta}_t \sim N(\mathbf{0}, \sigma^2_\zeta \boldsymbol{I})
\end{aligned} \tag{12}$$

The first formulation of an integrated approach, see Eqs. (7)–(10), constitutes our main model. The local linear trend model, see Eq. (11), and additive damped trend model, see Eq. (12), are used for comparison. Non-integrated versions can be created by applying the models to each series in the hierarchy separately and independently, ignoring the design matrix $\boldsymbol{S}$, and restricting all variance-covariance matrices to be diagonal. Seasonal patterns are then also estimated per series, rather than as an overall pattern. If the integrated models outperform alternatives that include their non-integrated counterparts, we derive support that this is because of the integration, rather than the specific formulation of levels and trends. Any model can be further refined after estimation by inspecting estimated values. If needed, particular parameters can be dropped or their values restricted. However, this makes the practical application of the model more time-consuming. Moreover, tweaking the model to particular cases undermines its robustness for a wide range of product groups within the company, which is why this is not considered for this study. The focus here is on whether a robust performance gain can be realised, and whether the integrated models – the main model, the local linear trend model, and the additive damped trend model – can consistently outperform alternatives.

3.2. Empirical study

The data have been made available by one of the world's largest manufacturers of consumer products, which has hundreds of brands spanning fourteen categories of food products, home and personal care. For forecasting, several decision makers are involved from operations, marketing, sales and finance, who require forecasts at various aggregation levels. Forecasts are often discussed at an aggregated product level, but for production these forecasts are transformed to the SKU level.

The forecasting methods offered by the company's IT systems are univariate. As a consequence, the company's forecasting matches the top-down and bottom-up approaches described previously. Forecasts are generated for products at a level called 'forecast unit,' which represents an SKU. A forecast unit can consist of several 'distribution units' to account for small changes in products, such as in artwork or ingredients. Distribution units are ignored in the forecasting process, as well as in this present study. Similar forecast units group together in so called 'forecast groups.'

Data pertaining to the regular sales of SKUs and product groups, excluding promotions, was obtained from the company. The data includes the statistical forecast generated by the company and the final forecast after judgmental adjustment. This data was collected for the years 2010, 2011 and 2012, yielding 156 weekly time series, which allows us to examine trends and seasonality. The first two years of data constitute the training sample, and the final year serves as the holdout sample. Because of changes to product lines and promotions, historical data of three years is not consistently available for all products. For these reasons, two particular product categories have been selected: foods and personal care.

For foods, we examine two product groups of mayonnaise and ice cream. For personal care, we consider one product group of hair products. Product groups consist of several forecast groups and forecast units. Mayonnaise sauce include three different forecast groups: mayonnaise pots, mayonnaise bottles and mayonnaise with a screw cap. Mayonnaise pots consist of twelve forecast units. Mayonnaise bottles consist of thirteen forecast units. Mayonnaise with a screw cap is excluded, because historical data is incomplete. Ice cream has two forecast groups, labeled jars and cones. Jars consist of 37 forecast units. Cones consist of seventeen forecast units. The hair products consist of sixteen forecast groups of which one forecast group is analyzed: shampoo, which consists of 27 forecast units. See Table 1 for an overview.

Table 1
Product groups at the company. Overview of forecast groups and forecast units. Total number of forecast units analyzed is 106.

Product category	Product groups	Forecast groups	Forecast units
Foods	Mayonnaise	Mayonnaise pots	12
		Mayonnaise bottles	13
	Ice cream	Jars	37
		Cones	17
Personal care	Hair products	Shampoo	27

We apply the integrated hierarchical models outlined in Section 3.1–the main model of Eqs. (7)–(10), the local linear trend model of Eq. (11), and the additive damped trend model of Eq. (12)–to each forecast group. We use the Kalman filter for forecast error decomposition to efficiently apply maximum likelihood to determine all unknown quantities, including the initial states using the BFGS algorithm, a quasi-Newton method of numerical optimization, with 500 random starts. For n forecast units, we have to estimate $\frac{n^2}{2} + \frac{13n}{2} + 6$ parameters. See Table 2 for an overview of the number of estimates needed per forecast group.

The training period is also used to estimate the bottom-up approach, the top-down approach, and the combination approach, all of which are described next. The training period is further split into an initial 18 months, used to estimate parameters, and a subsequent 6 months, used to select the forecasting methods that are evaluated over the holdout period.

Table 2
Parameters and observations per forecast group. Overview per forecast group of the number of forecast units in the group, the number of parameters that have to be estimated, and the number of observations available in the training period.

Forecast group	Units	Parameters	Observations (training)
Mayonnaise pots	12	156	1248
Mayonnaise bottles	13	175	1352
Ice cream jars	37	931	3848
Ice cream cones	17	261	1768
Hair shampoo	27	546	2808

The bottom-up approach generates forecasts at the product level by considering each product series in isolation. Various methods are considered as inputs. Simple exponential smoothing, Holt's, Holt-Winters, various ARIMA formulations, and the univariate versions of the integrated models are applied to each product series. Parameters are optimized to minimize mean square forecast error.

The top-down approach generates forecasts at the top level of the hierarchy, and allocates these downwards to create forecasts for the base products. Simple exponential smoothing, Holt's, Holt-Winters, various ARIMA formulations, and the univariate versions of the integrated models are applied to the top series, and parameters are optimized to minimize mean square forecast error over the initial 18 months of the training sample. In addition to generating forecasts, the forecasts have to be allocated downwards to the lower levels in the hierarchy, for which we use the two most common ways specified in (1) and (2), determined using the first 18 months (Gross & Sohl, 1990).

The combination approach of Hyndman et al. (2011) requires forecasts for each series at each level, which can subsequently be revised into final forecasts using (5). The required forecasts in $\hat{\boldsymbol{y}}$ are the best performing forecasts from the bottom-up and top-down approaches. The best performing forecasts are selected based on the performance over the last 6 months of the training set. For each series, the method is selected that minimizes mean squared forecast error. The forecasts of the selected methods are then revised to obtain the final forecasts of the combination approach, which is then applied to the hold out set.

To measure forecast accuracy, we use the mean average percentage error (MAPE). This particular measure is selected to obscure the scale of the original series, required in view of the confidentiality of the data:

$$\mathrm{MAPE}_i^j = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{\hat{y}_{i,t}^j - y_{i,t}}{y_{i,t}}\right| \tag{13}$$

where $y_{i,t}$ denotes actual observations and $\hat{y}_{i,t}^j$ the forecast obtained from method j. We take averages of MAPE for each level in the hierarchy.

In addition to evaluating forecast accuracy, we assess the impact of various methods on inventory performance in terms of stock investment and service levels using an order-up-to (T, S) policy, where T is a constant review time, and S is the order-up-to level. The service level is measured as the sales fill rate, defined as the proportion of demand, d_t, that can be fulfilled immediately from stock, where at time t the amount of stock on hand is given by I_t, and sales by O_t:

$$\text{Service level} = \frac{\sum_{t=1}^{T} O_t}{\sum_{t=1}^{T} d_t} \tag{14}$$

$$O_t = \min\{d_t, I_t\}$$

Each product has an associated leadtime, which ranges between 4 and 8 weeks, and the service level target is 95% for all products. The state space models give a forecast distribution, but all other methods only provide point forecasts. By splitting the holdout sample in two, the forecast errors over the first half can be used to approximate the forecast distribution. The mean square error during this period can be used as a measure of variance, which characterizes a fitted distribution, such as the normal distribution. However, this can be restrictive in a practical setting, such as assuming symmetry of forecast errors. An alternative, more flexible forecast distribution is given by bootstrapping the forecast errors, using a thousand draws with replacement, to derive an empirical one-period ahead forecast distribution. For each forecast approach, we simulate demand 10,000 times over the leadtime using the one-period ahead forecast distribution, where forecasts in each time period are conditional on the simulated demand in the preceding time period. Finally, we determine the quantity that satisfies the service level target for this leadtime distribution, deriving the order-up-to-level S for each approach. We then simulate inventory levels over time, and determine inventory performance over the remaining holdout sample by calculating the actual service level attained.

Average inventory is used as a proxy for the required stock investment. We include the company's own forecast in this evaluation, and scale the average inventory for each method using the average inventory needed when we use the company's own forecast.

4. Results

The empirical application of an integrated approach to forecasting hierarchical series demonstrates that forecast accuracy and inventory performance can be substantially improved with respect to the bottom-up, top-down and the combination approaches.

Table 3 shows the forecast accuracy of the integrated approach, combination approach, and the best performing bottom-up and top-down approaches for the five forecasting groups. The results show that the forecast performance of the integrated approach is substantially and consistently higher than that of the other approaches. The main integrated model dominates the other approaches in terms of forecast accuracy over all forecast groups.

Table 3
Forecast accuracy (MAPE) for empirical data. This table shows the forecast accuracy, in terms of MAPE, of the integrated approach, the combination approach, and the best performing bottom-up and top-down approach for each of the five forecast groups.

	Foods				Personal care	
	Mayonnaise		Ice cream		Hair	
	Pots (%)	Bottles (%)	Jars (%)	Cones (%)	Shampoo (%)	Average (%)
Integrated: main	39.54	38.20	34.93	41.20	29.11	36.60
Integrated: local linear trend	47.37	53.05	38.72	68.95	35.33	48.58
Integrated: damped trend	43.82	46.51	37.62	62.43	34.31	44.94
Combination approach	48.08	61.79	41.95	91.10	38.57	56.30
Bottom-up/top-down	61.04	64.66	46.92	83.37	49.63	61.12

Compared to the best performing bottom-up and top-down approaches, the main integrated model leads to an improvement in forecast accuracy of between 26%, for ice cream jars, and 51%, for ice cream cones. Though the main integrated model has its worst performance for ice cream cones, its MAPE is less than half the MAPE of the bottom-up and top-down approaches. Similar to the main integrated model, the other integrated models outperform the bottom-up, top-down and combination approaches, which suggests that using the design matrix $\boldsymbol{S}$ and allowing for multivariate distributions are most important in improving forecast performance.

The performance of the combination approach appears to be unstable, as it does not persistently outperform the best performing bottom-up and top-down approach in all forecast groups. In the case of ice cream cones, its MAPE is worse than the MAPE of the bottom-up/top-down approach. In all other cases, its improvement in forecast accuracy over the bottom-up and top-down approaches ranges between 4%, for mayonnaise bottles, and 22% for hair shampoo. Overall, the combination approach constitutes an improvement over the best performing bottom-up and top-down approach, but is in turn outperformed by the integrated approach.

The superior forecast performance of the integrated approach translates into substantial financial savings for inventory management. Table 4 summarizes inventory performance, as measured by achieved service level (see Eq. (14)) and required stock investment relative to the company, of the three approaches and the company's own forecast. For all approaches, the realized service levels are lower than the target of 95%, which means that all approaches underestimate the variation in sales. In the case of ice cream, not a single approach is able to achieve a service level higher than 85%. The main integrated model is closer to the target of 95% than all other methods, except in the case of ice cream jars, but the differences in service levels between the three approaches is small. The largest difference is seen for hair shampoo, where the main integrated model has a service level of 89%, whereas the other two approaches even perform worse than the company's own forecast with a service level of 75%. Overall, the more accurate point forecasts of the integrated approach allow for a better approximation of the forecast distribution, which results in higher service levels for the forecast groups.

Stock investment for each approach is relative to the stock investment required by the company. The company is consistently outperformed in all product groups by all approaches. The best performing bottom-up and top-down approach gives a better performance than the company's forecast in all product groups, with the exception of hair shampoo. The outcome for hair shampoo cannot be directly compared to the performance of the company forecast, because though the bottom-up and top-down approach entails a 3 percentage point decrease in stock investments, the service level also drops by 3 percentage points. For all other groups, the bottom-up and top-down approaches can substantially improve company performance, by achieving higher service levels with lower stock investments.

The inventory performance of the combination and the bottom-up and top-down approaches can often not be directly compared, because one of the two achieves a higher service level at the expense of a larger stock investment. The overall performance between the two is similar. In the cases of mayonnaise bottles and ice cream cones, the combination approach performs better, but no definite conclusions can be drawn in the other cases.

The main integrated model results in a dramatic drop in the required stock investment, reducing the stock investment needed, based on the company's current forecast, by 15% in all product groups. The biggest reduction is in ice cream jars, where the main integrated approach reduces the required stock investment by 20%. The largest reduction offered by another approach is given by the

Table 4
Inventory performance. This table shows the inventory performance of the three approaches and the company's own forecast based on a service level target of 95%. Stock investment for each approach is relative to the stock investment needed based on the company's own forecasts.

	Foods								Personal care			
	Mayonnaise				Ice cream				Hair			
	Pots		Bottles		Jars		Cones		Shampoo		Average	
	Stock (%)	Service (%)	Stock (%)	Service (%)	Stock (%)	Service (%)	Stock (%)	Service (%)	Stock (%)	Service (%)	Stock (%)	Service (%)
Integrated: main	87.51	94.86	84.39	89.22	79.99	80.05	81.46	84.03	91.76	89.24	85.02	87.48
Integrated: local linear trend	91.95	93.75	87.48	88.42	84.86	81.07	85.09	83.81	93.17	86.11	88.51	86.63
Integrated: damped trend	92.06	93.70	86.75	88.03	82.15	79.62	84.23	82.39	92.44	84.90	87.53	87.53
Combination approach	96.23	92.00	91.13	87.27	92.41	74.06	91.06	81.33	95.79	70.08	93.32	80.95
Bottom-up/top-down	94.28	89.89	92.19	84.07	94.49	82.47	93.40	81.31	96.81	71.47	94.23	81.84
Company's own	100	87.81	100	78.85	100	74.06	100	79.66	100	74.50	100	78.98

combination approach, also for ice cream cones, which entails a reduction of 9%. The main integrated model gives its smallest stock investment reduction for hair shampoo, which is still equal to 8%. The other integrated models also outperform the other approaches. The integrated approach offers substantial gains.

5. Discussion and conclusion

We introduce an integrated hierarchical forecasting approach to forecast the demand of products at different, but hierarchically-related aggregation levels. The approach supersedes the traditional comparison of bottom-up and top-down approaches (Fliedner, 1999; Kahn, 1998), by generating forecasts at all hierarchical levels and incorporating all available information, rather than only using selected parts of available data. The integrated approach avoids ex-post revising of forecasts, as is done in the combination approach (Hyndman et al., 2011), since generated forecasts are already reconciled and respect the additive restrictions placed on the series by the hierarchy.

Our empirical study shows the substantial gain, in terms of forecast performance as well as inventory performance, of generalizing the bottom-up and top-down forecast approaches to an integrated approach. All available information is used and complementarity of products and product substitution, which are otherwise ignored, are explicitly taken into account. Other features of the series, such as seasonality, are easily incorporated as well. Though the gain in performance is dependent upon the precise formulation of the model, the different integrated models applied here all outperform the alternatives.

The integrated approach is applicable to hierarchical forecasting in general, and extends beyond the current application of forecasting for manufacturers. Even overlapping groups of products can be easily accommodated. The large reductions in stock investments, show that the forecast performance directly translates to large financial gains, and is highly relevant for forecasting processes at companies. The advantages of formulating the integrated approach as a state space model are that outliers, missing values, and extra information, such as pertaining to promotions, can be easily, and flexibly, included (Durbin & Koopman, 2012; Harvey, 1989). The results of the empirical study show that future research has to broaden its scope beyond the bottom-up and top-down approaches, as these approaches are too restrictive, by ignoring dependencies and only using parts of the available data, which comes at a financial cost.

Acknowledgment

This research was financially supported by DINALOG as part of the 4C4More project. We also thank two anonymous reviewers for their valuable comments.

References

Athanasopoulos, G., Ahmed, R. A., & Hyndman, R. J. (2009). Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting, 25*(1), 146–166.
Byron, R. P. (1978). The estimation of large social account matrices. *Journal of the Royal Statistical Society. Series A (General), 141*(3), 359.
Chen, H., & Boylan, J. E. (2009). The effect of correlation between demands on hierarchical forecasting. In *Advances in business and management forecasting: 6* (pp. 173–188). Emerald Group Publishing Limited.
Dangerfield, B. J., & Morris, J. S. (1992). Top-down or bottom-up: aggregate versus disaggregate extrapolations. *International Journal of Forecasting, 8*(2), 233–241.
De Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of time series forecasting. *International Journal of Forecasting, 22*(3), 443–473.
Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford University Press.
Edwards, J. B., & Orcutt, G. H. (1969). Should aggregation prior to estimation be the rule? *The Review of Economics and Statistics, 51*(4), 409–420.
Fildes, R., Nikolopoulos, K., Crone, S. F., & Syntetos, A. A. (2008). Forecasting and operational research: a review. *Journal of the Operational Research Society, 59*(9), 1150–1172.
Fliedner, E. B., & Mabert, V. A. (1992). Constrained forecasting: some implementation guidelines. *Decision Sciences, 23*(5), 1143.
Fliedner, G. (1999). An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation. *Computers & Operations Research, 26*(10), 1133–1149.
Fliedner, G. (2001). Hierarchical forecasting: issues and use guidelines. *Industrial Management & Data Systems, 101*(1), 5–12.
Gardner, E. S., & McKenzie, E. (1985). Forecasting trends in time series. *Management Science, 31*(10), 1237–1246.
Gordon, T. P., Morris, J. S., & Dangerfield, B. J. (1997). Top-down or bottom-up: which is the best approach to forecasting? *The Journal of Business Forecasting Methods & Systems, 16*(3), 13–16.
Gross, C. W., & Sohl, J. E. (1990). Disaggregation methods to expedite product line forecasting. *Journal of Forecasting, 9*(3), 233–254.
Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press.
Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis, 55*(9), 2579–2589.
Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach*. Springer Science & Business Media.
Hyndman, R. J., Lee, A. J., & Wang, E. (2016). Fast computation of reconciled forecasts for hierarchical and grouped time series. *Computational Statistics & Data Analysis, 97*, 16–32.
Kahn, K. B. (1998). Revisiting top-down versus bottom-up forecasting. *The Journal of Business Forecasting Methods & Systems, 17*(2), 14–19.
Kohn, R. (1982). When is an aggregate of a time series efficiently forecast by its past? *Journal of Econometrics, 18*(3), 337–349.
Lütkepohl, H. (1984). Linear transformations of vector ARMA processes. *Journal of Econometrics, 26*(3), 283–293.
Moon, S., Hicks, C., & Simpson, A. (2012). The development of a hierarchical forecasting method for predicting spare parts demand in the South Korean navy. *International Journal of Production Economics, 140*(2), 794–802.
Orcutt, G. H., Watts, H. W., & Edwards, J. B. (1968). Data aggregation and information loss. *The American Economic Review*, 773–787.
Proietti, T. (2000). Comparing seasonal components for structural time series models. *International Journal of Forecasting, 16*(2), 247–260.
Rostami-Tabar, B., Babai, M. Z., Ducq, Y., & Syntetos, A. A. (2015). Non-stationary demand forecasting by cross-sectional aggregation. *International Journal of Production Economics, 170*, 297–309.
Schwarzkopf, A. B., Tersine, R. J., & Morris, J. S. (1988). Top-down versus bottom-up forecasting strategies. *The International Journal of Production Research, 26*(11), 1833–1843.
Snyder, R. D., Ord, J. K., & Beaumont, A. (2012). Forecasting the intermittent demand for slow-moving inventories: a modelling approach. *International Journal of Forecasting, 28*(2), 485–496.
Sohn, S. Y., & Lim, M. (2007). Hierarchical forecasting based on AR-GARCH model in a coherent structure. *European Journal of Operational Research, 176*(2), 1033–1040.
Solomou, S., & Weale, M. (1991). Balanced estimates of UK GDP 1870–1913. *Explorations in Economic History, 28*(1), 54–63.
Solomou, S., & Weale, M. (1993). Balanced estimates of national accounts when measurement errors are autocorrelated: the UK, 1920-38. *Journal of the Royal Statistical Society. Series A (Statistics in Society), 156*(1), 89.
Solomou, S., & Weale, M. (1996). UK national income, 1920–1938: the implications of balanced estimates. *The Economic History Review, 49*(1), 101–115.
Stone, R., Champernowne, D. G., & Meade, J. E. (1942). The precision of national income estimates. *The Review of Economic Studies, 9*(2), 111.
Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S., & Nikolopoulos, K. (2016). Supply chain forecasting: theory, practice, their gap and the future. *European Journal of Operational Research, 252*(1), 1–26.
Syntetos, A. A., Boylan, J. E., & Disney, S. M. (2009). Forecasting for inventory planning: a 50-year review. *Journal of the Operational Research Society, 60*(1), S149–S160.
Tiao, G. C., & Guttman, I. (1980). Forecasting contemporal aggregates of multiple time series. *Journal of Econometrics, 12*(2), 219–230.
Weale, M. (1985). Testing linear hypothesis on national account data. *The Review of Economics and Statistics, 67*(4), 685.
Weale, M. (1988). The reconciliation of values, volumes and prices in the national accounts. *Journal of the Royal Statistical Society. Series A (Statistics in Society), 151*(1), 211.
Wei, W. W. S., & Abraham, B. (1981). Forecasting contemporal time series aggregates. *Communications in Statistics – Theory and Methods, 10*(13), 1335–1344.
Widiarta, H., Viswanathan, S., & Piplani, R. (2007). On the effectiveness of top-down strategy for forecasting autoregressive demands. *Naval Research Logistics, 54*(2), 176–188.
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.
Zellner, A. (1969). On the aggregation problem: a new approach to a troublesome problem. In K. A. Fox, J. K. Sengupta, & G. V. L. Narasimham (Eds.), *Economic Models, Estimation and Risk Programming* (pp. 365–374). Springer.