Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 22 (2017) 193–202

Transportation Research Procedia

www.elsevier.com/locate/procedia

19th EURO Working Group on Transportation Meeting, EWGT2016, 5-7 September 2016, Istanbul, Turkey

Modified top down approach for hierarchical forecasting in a beverage supply chain

Dejan Mirčetić [a,*], Svetlana Nikoličić[a], Đurđica Stojanović[a], Marinko Maslarić[a]

[a]*Faculty of Technical Sciences, Trg Dositeja Obradovica 6, Novi Sad 21000, Serbia*

Abstract

In this paper, we propose a new approach for hierarchical forecasting, which represents the modification of the commonly used "top down" approach. The proposed method is based on projecting the ratio of bottom and top level series into the future, rather than using average historical proportions and proportions of the historical averages, as in standard top down approaches. Forecasted projections are then used for determining how the "base" forecasts of the top series will be distributed to the revised forecasts for every series at the bottom level of the hierarchy. Revised forecasts for all series in the hierarchy are obtained in the same manner as in standard top down methodology. In order to estimate the accuracy of the proposed model, the simulation study is performed. Results demonstrate that the "modified" top down model significantly outperforms standard top down approaches. Beside comparison with the top down approach, the model is compared with the bottom up method, and two newly proposed approaches: top down forecasted proportions and optimal combination approach. Overall, the bottom up method shows the lowest forecasting error. Results also reveal that the modified top down model demonstrates good performance, since its forecasts are close to the forecasts of the top performing model. In the end, we demonstrate our method and others in forecasting the beverage distribution network.

Keywords: hierarchical forecasting; modified top down approach; time series; beverage supply chain.

* Corresponding author. Tel.: +381 21 485 2433;
Presenting author, *E-mail address:Dejan.Mircetic@uns.ac.rs*

10.1016/j.trpro.2017.03.026

1. Introduction

Forecasting is a critical business process for nearly every organization and often it is the very first step organizations must undertake when determining long-term capacity needs, annual business plans, and shorter-term operations and supply chain (Bozarth & Handfield, 2008). The application of forecasting methods in a supply chain (SC) started in the 1960s and it was related to the inventory management. In that time, researchers treated the two subjects as inter-related, driven by the practical requirements of designing and implementing inventory systems (Boylan & Syntetos, 2008). Today, these areas are separated in two disciplines, and developed without significant interaction. Boylan and Syntetos (2008) argue that more effort is needed to close the gap between inventory management and demand forecasting, and they state that "the interactions between forecasting and inventory models have been rather neglected". Forecasting is also important for production planning. Production managers need future demand forecast to plan and schedule a production and determine other related activities, like requirements planning and purchasing. Fildes and Beard (1992) dealt with the application of forecasting in production and inventory, and proposed the "ideal" design of a forecasting system for production and inventory-control. Beside production and inventory management, forecasting also impacts the physical distribution. Physical distribution is one of the key business processes in SC, which provides the delivery of the finished products to the market, and consequently obtaining the surplus of the observed SC. To do so, logistics must be consistent with the products it supports as customers tend not to place any difference between a product and the distribution system that supplies it (Hesse & Rodrigue, 2004). On the other hand, physical distribution is a significant generator of logistics costs, and managers are on the constant pressure to obtain as much full truck loads as possible in transportation. In order to achieve full truck loads, they need reliable information about the size of future good flows, type of goods, delivery directions, quantities, timing of demand, and other related information which significantly influence the decisions regarding the transportation scheduling and delivery. Therefore, while dealing with the organization and synchronization of SC processes, managers have special interest in acquiring information about future demand, disaggregated according to the information they need in order to plan the business processes in SCs (different demand attributes). This is the point where hierarchical forecasting (HF) is needed. Hierarchical time series represent multiple time series that are hierarchically organized and can be aggregated at several different levels in groups based on products, geography or some other features (R. J. Hyndman, Ahmed, Athanasopoulos, & Shang, 2011).

In literature, two general approaches are suggested for developing HF. First approach is labeled as a "top down" strategy, since a single forecast model is developed to forecast an aggregate - or family total which is then distributed to the individual items in the family based upon their historical proportion of the family total. The second is named as a "bottom up" strategy, since multiple forecast models based upon the individual item series are used to develop item forecasts (Fliedner, 1999). Recently, two new approaches are proposed: top down forecasted proportions and optimal combination approach. In this paper, we propose a new approach for HF, by modifying the top down methodology. Modification consists in proposing a new way for determining how the top level forecasts will be distributed to the bottom level series. Accordingly, we are modifying the top down procedure and producing the forecasting proportion ratios, which serve as disaggregating functions, in partitioning the top level forecast. Our motivation for developing a new approach based on the top down methodology is low performance of standard top down models in recent studies. Therefore, we have tried to develop the top down model which will be more accurate and thereby more competitive with other HF approaches.

The remainder of paper is organized as follows: next section provides theoretical background of the HF methodologies. Section 3 presents the core of the paper, where the simulation study, i.e. the evaluation of different HF models, also including the newly proposed modified top down approach (MTD), is performed. In Section 4, the forecasting of beverage distribution network is demonstrated. We conclude our paper in Section 5, where discussion of results and the final remarks are provided.

2. Methodology

In literature, there are five HF methodologies that are most commonly used. HF methodology is essentially the way of how the base forecasts are combined in order to produce revised forecasts. By base forecasts, we consider independent forecasts that are generated by some of the forecasting methods (for example: exponential smoothing, Holt's, etc.), while the revised forecasts represent the final forecasts for each node in the hierarchy. For better understanding of the basic principles of HF, the diagram of the hierarchical structure is provided in Fig. 1. The diagram considers a hierarchy with K = 2 levels and n = 13 series in total. The completely aggregated series at the top level (level 0) are disaggregated into n_1 = 4 component series at level 1. Each of these series is further subdivided into two series at level 2, i.e. n_2 = 8.

Nomenclature for HF

n	Total number of series in the hierarchy;
n_k	Number of the series in the bottom level of the hierarchy;
y_t	tth observation for "Total" series for $t = 1,\dots,T$;
$\boldsymbol{y}_t$	Vector of all observations at time t;
$y_{j,t}$	tth observation of the series which corresponds to the node j of the hierarchical tree;
$\boldsymbol{y}_{K,t}$	Vector of all observations in the bottom level of the hierarchy at time t;
S	Summing matrix of $n \times n_K$ dimensions, which dictates how the bottom level series are aggregated;
$\hat{y}_h$	h - step ahead "base" forecast generated for the "Total" series, having observed series until the T;
$\tilde{y}_h$	Final (i.e. revised) forecasts of the "Total" series, obtained by aggregating the "base" forecasts;
$\tilde{\boldsymbol{y}}_h$	Vector of all final forecasts in the hierarchy;
$\hat{\boldsymbol{y}}_{K,h}$	Vector of all h - step ahead "base" forecasts in the bottom level of the hierarchy;
$\tilde{\boldsymbol{y}}_{K,h}$	Vector of all final forecasts in the bottom level of the hierarchy;
p_j	Historical proportion of the bottom level series j and "Total" series;
$\boldsymbol{p}_j$	Vector of all historical proportions in the bottom level of the hierarchy;
$\hat{p}_{j,h}$	h - step ahead forecast of the proportion ratio at the node j, having observed time series until the moment T;
$\hat{\boldsymbol{p}}_{K,h}$	Vector of h - step ahead forecasts of the proportion ratios in the bottom level of the hierarchy;
$\hat{y}_{j,h}^{(l)}$	h - step ahead "base" forecast of the node which is l levels above j;
$\hat{S}_{j,h}^{(l)}$	Sum of the h - step ahead "base" forecasts below the node that is l levels above the node j.

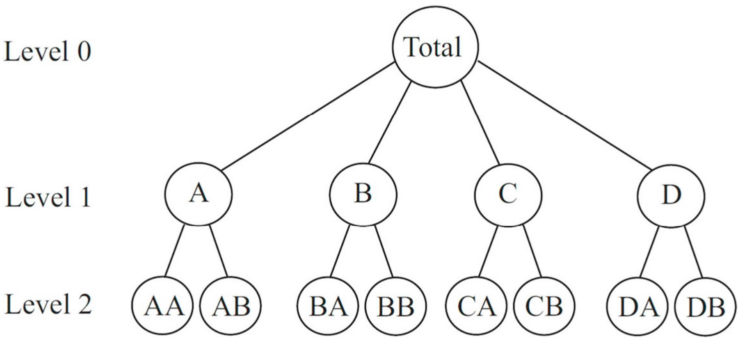

Fig. 1. Hierarchical structure of the simulated time series.

For any time t, the observations from the bottom level of series will aggregate to the observations of the series in the higher levels. With the introduction of the "summing matrix" (S) and using the matrix notation, this can be easily represented. The summing matrix determines how the higher levels in the hierarchy will be aggregated from the bottom level series. With the respect to the hierarchy from Fig. 1, previous can be expressed as:

$$\underbrace{\begin{bmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{C,t} \\ y_{D,t} \\ y_{AA,t} \\ y_{AB,t} \\ y_{BA,t} \\ y_{BB,t} \\ y_{CA,t} \\ y_{CB,t} \\ y_{DA,t} \\ y_{DB,t} \end{bmatrix}}_{\boldsymbol{y}_t} = \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{S} \cdot \underbrace{\begin{bmatrix} y_{AA,t} \\ y_{AB,t} \\ y_{BA,t} \\ y_{BB,t} \\ y_{CA,t} \\ y_{CB,t} \\ y_{DA,t} \\ y_{DB,t} \end{bmatrix}}_{\boldsymbol{y}_{K,t}}.$$

Or in more compact form:

$$\boldsymbol{y}_t = S \cdot \boldsymbol{y}_{K,t}. \tag{1}$$

2.1. Bottom up approach

Bottom up first involves generating the base forecasts for the bottom level series. To answer the question which models to use for base forecasts, we rely on past studies, especially on M-Competitions (Makridakis et al., 1982; Makridakis, Chatfield, Hibon, Lawrence, & Mills, 1993; Makridakis & Hibon, 2000), where simple models, such as the exponential smoothing, prove their superiority against much more statistically sophisticated methods. Therefore, we used the state space model (i.e. exponential tail smoothing models (ETS)) for forecasting the base time series. ETS models are derived from ES. They have the same point forecasts, but additionally, they provide the prediction intervals. For more details regarding the forecasting with ETS models refer to (R. Hyndman, Koehler, Ord, & Snyder, 2008). After generating the base forecasts, these forecasts are then aggregated to the upper levels, using the summing matrix S. In that manner, the final forecasts for each node are obtained, and for bottom level series, the initial base forecasts are regarded as the final forecasts, i.e. $\tilde{\boldsymbol{y}}_{K,h} = \hat{\boldsymbol{y}}_{K,h}$. Accordingly, the bottom up approach can be represented as follows:

$$\tilde{\boldsymbol{y}}_h = S \cdot \hat{\boldsymbol{y}}_{K,h}. \tag{2}$$

Main advantage of the bottom up approach is generating the initial base forecasts at the lowest disaggregated level of the hierarchy (i.e. bottom level). In that way there isn't any loss of information from the data, which may occur when dealing with the data from more aggregate levels of the hierarchy, as in the case of top down methodology. Hence, bottom up approach can better capture the dynamics of the individual series, but that also comes with the price since the data in the bottom level can be quite noisy and hard to model.

2.2. Top down approach

Opposite to the bottom up, the top down approaches first involve generating the base forecasts for the "Total" series ($\hat{\boldsymbol{y}}_{\boldsymbol{h}}$), and then disaggregating these downwards, in order to obtain the revised forecasts of the series in the bottom level hierarchy ($\tilde{\boldsymbol{y}}_{\boldsymbol{K},\boldsymbol{h}}$). Disaggregating the top level forecasts is performed using the disaggregating proportions ($p_1, \ldots, p_{nk}$), which dictate how the top level base forecasts will be distributed to the bottom level series of the hierarchy. When the bottom level series are determined, all other series can be easily calculated using the summing matrix S. The summing matrix is created according to the structure of how series are disaggregated in an observed case. For the top down approach, the top level revised forecasts are equal to the top level base forecasts, i.e. $\tilde{\boldsymbol{y}}_{\boldsymbol{h}} = \hat{\boldsymbol{y}}_{\boldsymbol{h}}$. The two most commonly used top down approaches are average historical proportions (TD1) and proportions of the historical averages (TD2). Their common feature is that they both use historical proportions of the data in order to specify disaggregating proportions. In the case of TD1 proportions are determined as follows:

$$p_j = \frac{1}{T}\sum_{t=1}^{T}\frac{y_{j,t}}{y_t}; \tag{3}$$

for $j = 1,\ldots, n_k$. Each proportion p_j reflects the average of the historical proportions of the bottom level series ($y_{j,t}$) relative to the total aggregate series (y_t), for the period $t = 1,..., T$. In the TD2 approach, proportions are determined in the following manner:

$$p_j = \frac{\sum_{t=1}^{T}\frac{y_{j,t}}{T}}{\sum_{t=1}^{T}\frac{y_t}{T}}. \tag{4}$$

In this case, p_j reflects the relationship of the average historical value of the bottom level series ($y_{j,t}$), relative to the average value of the total aggregate series (y_t), for the observed period $t = 1,..., T$. With respect to the given hierarchy, the general top down approach could be described using the matrix notation, as follows:

$$\underbrace{\begin{bmatrix} \tilde{y}_h \\ \tilde{y}_{A,h} \\ \tilde{y}_{B,h} \\ \tilde{y}_{C,h} \\ \tilde{y}_{D,h} \\ \tilde{y}_{AA,h} \\ \tilde{y}_{AB,h} \\ \tilde{y}_{BA,h} \\ \tilde{y}_{BB,h} \\ \tilde{y}_{CA,h} \\ \tilde{y}_{CB,h} \\ \tilde{y}_{DA,h} \\ \tilde{y}_{DB,h} \end{bmatrix}}_{\tilde{\boldsymbol{y}}_h} = \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{S} \cdot \underbrace{\begin{bmatrix} \tilde{y}_{AA,h} \\ \tilde{y}_{AB,h} \\ \tilde{y}_{BA,h} \\ \tilde{y}_{BB,h} \\ \tilde{y}_{CA,h} \\ \tilde{y}_{CB,h} \\ \tilde{y}_{DA,h} \\ \tilde{y}_{DB,h} \end{bmatrix}}_{\tilde{\boldsymbol{y}}_{K,h}}$$

Using the matrix and vector multiplication, former could be written in a more compact form:

$$\tilde{\boldsymbol{y}}_h = S \cdot \tilde{\boldsymbol{y}}_{K,h}; \tag{5}$$

where $\tilde{\boldsymbol{y}}_{k,h}$ is obtained by multiplying the top level forecast ($\hat{y}_h$), by $\boldsymbol{p}_j=[p_{AA},...,p_{DB}]^T$ which consists of historical proportions of bottom series relative to the top level, i.e. $\tilde{\boldsymbol{y}}_{K,h} = \hat{\boldsymbol{y}}_h \cdot \boldsymbol{p}_j$. Accordingly, the final equation for general top down approaches (TD1 and TD2), could be expressed as follows:

$$\tilde{\boldsymbol{y}}_h = S \cdot \hat{y}_h \cdot \boldsymbol{p}_j. \tag{6}$$

2.3. Top down approach based on forecasted proportions

Top down forecasted proportions (top down FP) represent the alternative approach for obtaining the disaggregating proportions (p_j). The logic of the presented method consists of using all individual forecasts of all series in the hierarchy in order to obtain proportions, in a manner that is consistent with the observed hierarchy. Instead of using the historical proportions in the static manner, as seen in the case of TD1 and TD2 models, this approach is using the forecasted proportions to obtain p_j.

General formula for calculating the forecasted proportions is the following:

$$p_j = \prod_{l=0}^{K-1} \frac{\hat{y}_{j,h}^{(l)}}{\hat{S}_{j,h}^{(l+1)}}. \tag{7}$$

2.4. Optimal combination approach

The approach first considers generating the independent base forecasts for all series in the hierarchy. The problem with independent base forecasts is that they will not provide “aggregate consistency”, since they will not sum up according to the hierarchy. The optimal combination approach solves this problem by combining the base forecasts to produce a set of revised forecasts that are as close as possible to the independent forecasts, but also meet the requirement that forecasts at upper levels in the hierarchy are the sum of the associated lower level forecasts. As a result of this kind of forecast combining, this approach uses all the available information within the hierarchy. The general idea is derived from the representation of the h - step ahead base forecasts for the whole of the hierarchy by the linear regression model, while the final forecasts are determined in the following manner:

$$\tilde{\boldsymbol{y}}_h = S(S'S)^{-1}S'\hat{\boldsymbol{y}}_h. \tag{8}$$

For more details regarding the optimal combination approach, refer to (R. J. Hyndman et al., 2011).

2.5. Modified top down approach

In this paper, we propose a new method for obtaining the disaggregating proportions, which consists in forecasting the ratio between the bottom level ($y_{j,t}$), and the top level series (y_t), into h - step ahead period, and in using those forecasts as final values of disaggregated proportions ($\hat{p}_{j,h}$). Previous could be expressed as follows:

$$\hat{p}_{j,h} = \widehat{\frac{y_{j,h}}{y_{t,h}}}. \tag{9}$$

In this manner, the relationship between bottom level series ($y_{j,t}$) and top level series (y_t) is observed as a function in time, rather than a single average value, as in the case of TD1 and TD2. Hence, the underlying relationship between top and bottom levels is “preserved” and it is not deteriorated by averaging, since the average can mislead statistics in cases when the fluctuations of series values around its mean are not normally distributed.

Since in MTD proportions are observed as functions, which are projected h – steps ahead (Eq. 9), certain modifications to the (Eq. 6) are needed in order to perform the forecasting of final “revised” forecasts. With respect to the Eq. 9 and Eq. 6, revised forecasts for MTD could be expressed in the matrix form as follows:

$$\underbrace{\begin{bmatrix} \tilde{y}_h \\ \tilde{y}_{A,h} \\ \tilde{y}_{B,h} \\ \tilde{y}_{C,h} \\ \tilde{y}_{D,h} \\ \tilde{y}_{AA,h} \\ \tilde{y}_{AB,h} \\ \tilde{y}_{BA,h} \\ \tilde{y}_{BB,h} \\ \tilde{y}_{CA,h} \\ \tilde{y}_{CB,h} \\ \tilde{y}_{DA,h} \\ \tilde{y}_{DB,h} \end{bmatrix}}_{\tilde{\boldsymbol{y}}_h} = \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{S} \cdot \underbrace{\underbrace{\begin{bmatrix} \hat{p}_{AA,h} \\ \hat{p}_{AB,h} \\ \hat{p}_{BA,h} \\ \hat{p}_{BB,h} \\ \hat{p}_{CA,h} \\ \hat{p}_{CB,h} \\ \hat{p}_{DA,h} \\ \hat{p}_{DB,h} \end{bmatrix}}_{\hat{\boldsymbol{p}}_{K,h}} \cdot [\hat{y}_h]}_{\tilde{\boldsymbol{y}}_{K,h}}$$

In a more compact form, former could be written in a following manner:

$$\tilde{\boldsymbol{y}}_h = S \cdot \hat{y}_h \cdot \hat{\boldsymbol{p}}_{K,h}. \tag{10}$$

3. Design of the simulation

In order to evaluate the performance of the proposed method, we first performed a simulation study. The simulation study is performed following the guidance provided in (R. J. Hyndman et al., 2011), with some minor adjustments. The data for each series were generated from the *Autoregressive Integrated Moving Average* (ARIMA) process. Orders (p, d, q) of the ARIMA process were chosen randomly, from a uniform distribution. For the degree of differencing (d), number of autoregression (p) and moving average (q) terms, possible values were restricted to 0, 1 and 2. The parameter values for p and q terms were also generated from a uniform distribution. The simulation process simulated the series at the bottom level, while higher levels were obtained by summing the simulated bottom series, according to the hierarchy in Fig. 1. In order to allow relationships between series in the hierarchy, the simulated bottom level series were constructed to have correlated errors. The covariance matrix is designed to allow the positive error correlation between bottom level series, which are derived from the same series at *level 1* (A, B, C, or D). Also, the covariance matrix allows the error correlation of bottom series which are derived from A and B, or C and D series (i.e. within the correlation between AA, BB, BA, BB and CA, CB, DA, DB). The simulation was repeated for 100 times, generating 100 different scenarios of bottom level series. Each series in the bottom level had 56 simulated observations. For the sake of convenience and simpler presentation, we regarded them as a quarterly observation, with the time scale ranging from 2002 to 2015. Since we were dealing with simulated time series, any other label could be assigned to the observations (e.g. monthly, weekly, daily, etc.). The same applies to the assigned time scale for series. The only reason why the authors have chosen this attributes for simulated series lies in the fact that we are used to dealing with quarterly time series when dealing with HF.

3.1. Evaluation of forecasting performances

In order to evaluate the forecasting performance of each HF approach, we calculated the test error for each simulated series in the hierarchy from Fig. 1. Each simulated series was divided into training and test observations. Training data were initially set with 12 observations (2002:Q1 - 2005:Q4), and test set with 8 observations (2006:Q1 - 2008:Q4). After setting the training and test data, 1- to 8 - step - ahead forecasts were produced and the test error was calculated (for each series in the hierarchy). The process continued by increasing the training set by one observation, re-estimating the models, and again producing 1- to 8 - step - ahead forecasts. Training and test sets were iteratively updated until training data reached 48 observations (2002:Q1 - 2013:Q4) and the test captured the period of (2014:Q1 - 2015:Q4), which at the end produced 32 different test and train data sets. Created forecasts of all 32 forecasts for all series in the hierarchy were used to evaluate the out-of-sample forecast performance of each of the methods considered. As the error measure we chose the root mean square error (RMSE). The final test error was obtained by averaging all individual RMSE errors of the simulated scenarios.

3.2. Simulation results

In Table 1, RMSE for each series in the hierarchy from Fig. 1, is presented. Results reveal that TD1 and TD2 performed much poorer than other methods. Possible reasons could be found in the fact that the majority of the simulated series had strong and volatile trend, which TD1 and TD2 were unable to adequately incorporate in the future estimates of historical proportions p_j. As a consequence, these models produce forecasts with large residuals. Conversely, in those simulations where there was the lack of trend, or the trend was week, TD1 and TD2 performed very competitive, and in some cases they were the top performing models. Table also reveals small differences between the bottom up, optimal combination and MTD models. Overall, the bottom up shows best results (based on the criteria of minimum average RMSE), slightly beating the optimal combination, although the subsequent tests show that this difference is not statistically significant. Bottom up and optimal combination are followed by MTD, which also demonstrate good performance since its RMSE's are close to the RMSE's of the bottom up and optimal combination models. Overall score of the top down FP is almost three times worse than the bottom up approach! We believe that the reasons are negative values in series. The top down FP failed to produce accurate forecasts in those simulation scenarios where negative values in series were also present, and, as a consequence, unattainable RMSE deficit has been made. Unlike other top down approaches, MTD effectively dealt with the trend and negative values in series, and showed robustness producing stable forecasts regardless of the structure of the bottom level series.

Table 1. RMSE for each series in the hierarchy (average across 100 simulation scenarios)

	RMSE					
	Bottom up	TD1	TD2	Optimal	Top down FP	MTD
Total	55.33	58.69	58.69	**54.98**	58.69	59.63
A	21.53	631.815	160.75	**21.52**	39.19	27.36
B	20.83	389.311	276.77	**20.80**	48.05	24.56
C	15.71	433.86	314.77	**15.69**	54.31	20.52
D	17.30	712.78	460.16	**17.29**	50.31	21.79
AA	**9.37**	346.33	133.14	9.61	25.99	12.19
AB	**14.50**	462.27	110.19	14.75	32.01	20.00
BA	**11.46**	254.28	182.36	11.73	35.33	15.57
BB	**12.72**	209.61	123.06	12.93	44.88	15.86
CA	**8.97**	300.04	171.51	9.47	28.34	11.38
CB	**9.27**	190.46	150.84	9.56	30.10	13.92
DA	**10.37**	229.47	215.90	10.60	25.64	13.45
DB	**9.50**	523.21	249.89	9.86	29.04	13.80
Average	**16.68**	364.78	200.62	16.83	38.60	20.77

Fig. 2 confirms the conclusions drawn from Table 1. It is notable that the bottom up, optimal combination, MTD and top down FP have a narrow interquartile range, which indicates that these models have forecasting errors which are consistent across all the series in the hierarchy. On the other hand, the dispersion of TD1 and TD2 is much larger than in the case of previous models, and clearly TD1 and TD2 underperform compared to competing models. Empirically determined differences in Table 1 and Fig. 2 are tested for their significance. The p-value corresponding to the χ^2 statistic of the Kruskal–Wallis test is lower than 0.05, suggesting that one or more forecasts are significantly different. For identifying the pairs of forecasts which are significantly different from each other, the Dunn's post-hoc test is applied. Dunn's test shows that the forecasts of TD1 and TD2 models are significantly different from the forecasts of all other models, but not mutually different. Also, the forecasts of the top down FP model are statistically different from all other forecasts. Further, the test reveals that there is no statistically significant difference between the forecasts of the bottom up, optimal combination and MTD models.

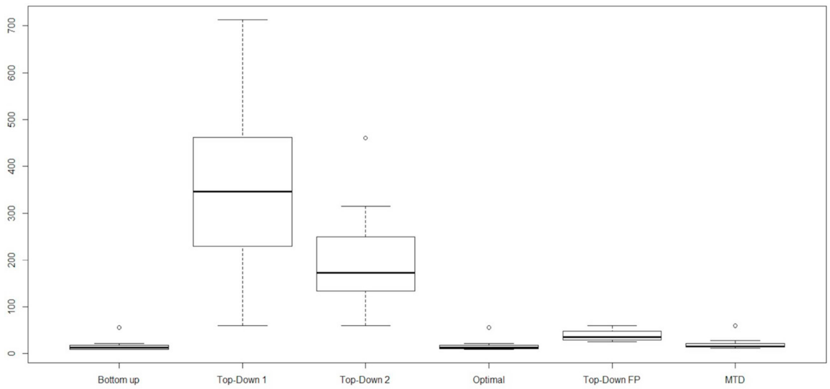

Fig. 2. RMSE's of competing models tested on simulated series.

4. Hierarchical forecasting of the beverage distribution chain

The observed beverage distribution chain consists from a central warehouse (CW) and several distribution centers (DC). Produced beverages are distributed to final consumers, either through direct delivery (DD) or regional DC's. There are four DC's which supply regions for which they are responsible.

Table 2. RMSE for each series in the beverage distribution chain

	RMSE					
	Bottom up	TD1	TD2	Optimal	Top down FP	MTD
Total (CW)	3807.77	3604.31	3604.31	3634.24	3604.31	**3503.07**
DD	1793.19	1435.77	**1434.42**	1777.47	1775.86	1700.26
DC1	1738.33	1501.33	1528.11	1709.76	1693.59	**1501.32**
DC2	1171.17	1058.71	1036.35	1147.76	1138.21	**1023.55**
DC3	543.32	650.77	645.58	**524.83**	532.78	597.20
DC4	1871.95	1560.93	**1533.57**	1864.83	1860.27	1545.55
Average	1820.95	1635.30	**1630.39**	1776.48	1767.50	1645.15

In forecasting the beverage distribution chain, TD2 model showed best performance, followed by TD1 and MTD (Table 2), but its performance did not prove to be statistically significant as the Kruskal–Wallis test failed to find any significant difference between models. We argue that the good performance of TD2 could be attributed to the lack of trend and the consistent pattern of bottom level series with pronounced seasonality. For more details regarding the structure of time series refer to Fig. 1 in (Mircetic, Ziramov, Nikolicic, Maslaric, & Ralevic, 2015).

5. Conclusions and discussion

In this paper, we present an alternative approach for HF. The proposed approach is compared with several others in a simulation study, and the results are shown in Table 1. Overall, the bottom up and optimal combination models have performed best, producing almost identical forecasts. Good performance of the bottom up model was a quite surprising result, especially in the higher hierarchy levels. These findings could be linked to previous researches of (Athanasopoulos, Ahmed, & Hyndman, 2009; R. J. Hyndman et al., 2011), in which the bottom up also performed very well, but it was slightly beaten by the optimal combination and top down FP. Reasons for this should be searched in the patterns of the bottom level series and in the forecasting scenario settings (length of forecasting horizon, number of disaggregating hierarchy levels, etc.). The majority of simulated series had patterns with little noise. Therefore, individual ETS models were able to detect those patterns and incorporate them in the future forecasts. In that manner, the bottom up forecasts were able to stay competitive in the upper levels of the hierarchy. We argue that in the case of some other simulation scenario with more disaggregated nodes in the bottom level, as well as with noise in series and deteriorated patterns, it could be possible for the bottom up to show poorer performance. In the mentioned researches of (Athanasopoulos et al., 2009; R. J. Hyndman et al., 2011), top down models were only competitive in the top level. As we moved down the hierarchy, they performed poorly and were clearly outperformed by competing models. The same could be seen in the present research (Table 1). Results also demonstrate that the MTD improved the forecasting accuracy of the top down methodology and significantly outperformed TD1 and TD2 models. MTD also performed better than the top down FP, and its forecasts could not be statistically distinguished from the forecasts of the bottom up and optimal combination models. In forecasting the beverage distribution hierarchy, TD2, TD1 and MTD showed best performance, but the difference from other models was not statistically significant (Table 2). Therefore, all models had statistically indistinguishable forecasts. Bearing in mind the demonstrated results, calculating the proportions as in Eq. (9) proved to be a useful upgrade to the top down methodology. Therefore, we recommend the substitute of standard top down approaches with MTD.

Acknowledgements

This research was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, through the Project TR 36030.

References

Athanasopoulos, G., Ahmed, R. A., & Hyndman, R. J. (2009). Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting, 25*(1), 146–166.

Boylan, J. E., & Syntetos, A. A. (2008). Editorial: Special issue on demand forecasting for inventory management. *IMA Journal of Management Mathematics, 19*(2), 99-100. doi: doi:10.1093/imaman/dpn003

Bozarth, C. C., & Handfield, R. B. (2008). *Introduction to operations and supply chain management* (2nd ed.). New Jersey: Pearson Prentice Hall.

Fildes, R., & Beard, C. (1992). Forecasting Systems for Production and Inventory Control. *International Journal of Operations & Production Management, 12*(5), 4-27.

Fliedner, G. (1999). An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation. *Computers & Operations Research, 26*(10), 1133-1149.

Hesse, M., & Rodrigue, J.-P. (2004). The transport geography of logistics and freight distribution. *Journal of Transport Geography, 12*(3), 171–184. doi: 10.1016/j.jtrangeo.2003.12.004

Hyndman, R., 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., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics and Data Analysis, 55*(9), 2579–2589.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., & Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting, 1*(2), 111-153.

Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., & Mills, T. (1993). The M2-competition: A real-time judgmentally based forecasting study. *International Journal of Forecasting, 9*(1), 5-22. doi: doi:10.1016/0169-2070(93)90044-N

Makridakis, S., & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting, 16*(4), 451-476.

Mircetic, D., Ziramov, N., Nikolicic, S., Maslaric, M., & Ralevic, N. (2015). *Demand Forecasting in Beverage Supply Chain*. Paper presented at the 3rd Olympus International Conference In Supply Chains (3rd Olympus ICSC), Greece, Athens.