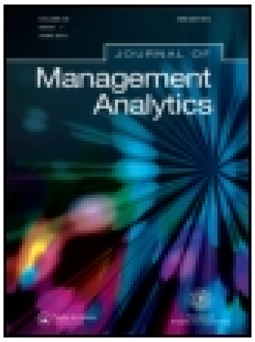

Journal of Management Analytics

ISSN: 2327-0012 (Print) 2327-0039 (Online) Journal homepage: https://www.tandfonline.com/loi/tjma20

Individual and combination approaches to forecasting hierarchical time series with correlated data: an empirical study

Hakeem-Ur Rehman, Guohua Wan, Azmat Ullah & Badiea Shaukat

Journal of Management Analytics, 2019
https://doi.org/10.1080/23270012.2019.1629342

Individual and combination approaches to forecasting hierarchical time series with correlated data: an empirical study

Hakeem-Ur Rehman, Guohua Wan*, Azmat Ullah and Badiea Shaukat

Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, People's Republic of China

(*Received 25 December 2018; revised 28 May 2019; accepted 5 June 2019*)

Hierarchical time series arise in manufacturing and service industries when the products or services have the hierarchical structure, and top-down and bottom-up methods are commonly used to forecast the hierarchical time series. One of the critical factors that affect the performance of the two methods is the correlation between the data series. This study attempts to resolve the problem and shows that the top-down method performs better when data have high positive correlation compared to high negative correlation and combination of forecasting methods may be the best solution when there is no evidence of the correlationship. We conduct the computational experiments using 240 monthly data series from the 'Industrial' category of the M3-Competition and test twelve combination methods for the hierarchical data series. The results show that the regression-based, VAR-COV and the Rank-based methods perform better compared to the other methods.

Keywords: hierarchical time series; individual forecasting methods; combination forecasting methods; correlation

1. Introduction

Hierarchical forecasting is an integral part of the operational decision making as managers often require information about demand from product family to stock keeping unit level. Considering a production planning and distribution problem, individual product level forecasts are needed to generate production plan (lot-sizing and scheduling), but an aggregate (multiple retail stores) forecast required for inventory decisions at the distribution facility. In the recent years, hierarchical time series forecasting has gained significant attention from the researchers (see, e.g. Chen & Boylan, 2009; Hyndman, Ahmed, Athanasopoulos, & Shang, 2011; Viswanathan, Widiarta, & Piplani, 2008; Widiarta, Viswanathan, & Piplani, 2008a, 2008b) and practitioners as it has wide applications in diverse fields such as in inventory control (Moon, Hicks, & Simpson, 2012; Moon, Simpson, & Hicks, 2013; Pennings & van Dalen, 2017), supply chain management (Huber, Gossmann, & Stuckenschmidt, 2017; Syntetos, Babai, Boylan, Kolassa, & Nikolopoulos, 2016), health care management (Athanasopoulos, Hyndman, Kourentzes, & Petropoulos, 2017) and in tourism management (Athanasopoulos, Ahmed, & Hyndman, 2009; Weatherford, Kimes, & Scott, 2001). Fliedner (2001) summarises the guidelines for the uses of the hierarchical forecasting approaches.

*Corresponding author. Email: ghwan@sjtu.edu.cn

'Top-down' and 'bottom-up' are most widely used methods in the literature of hierarchical forecasting to predict the aggregate or sub-aggregate level data. But it is not clear whether forecast should be made directly on the item level or indirectly by proportion down product family level forecast and vice versa. Some researchers support the use of a top-down method (Athanasopoulos et al., 2009; Fliedner, 1999; Grunfeld & Griliches, 1960; Park & Nassar, 2014) while others in favour of the bottom-up method (Schwarzkopf, Tersine, & Morris, 1988; Shlifer & Wolff, 1979; Widiarta, Viswanathan, & Piplani, 2007). This debate is still on that under what conditions one method provides better results compared to other. One of the important characteristics of the demand is a correlation that affects the performance of the hierarchical forecasting methods, but it is not clear whether positive or negative correlation. Schwarzkopf et al. (1988) examined the strength of bottom-up and top-down methods while considering the correlated demand between two products and concluded that estimation accuracy depends on the correlation between data series. Furthermore, the variance in the aggregated data also depends on the type of correlation (i.e. positive or negative) at the lowest (i.e. item) level between the series exist. If items have strong positive correlation then at the aggregate level, they may have high variance. While, if items have strong negative correlation then they may contain much low variance. They argue that strong negative correlation is having a positive impact on the performance top-down methods while Duncan, Gorr, and Szczypula (2001) claim that top-down methods perform better under positive correlation. Chen and Boylan (2009) used a simulation study to resolve this contradiction and argue that variability reduces due to a negative correlation which benefits the top-down method. Widiarta et al. (2007) compare the two classical hierarchical forecasting methods when the lowest level data of the hierarchy follows the AR(1) process. Their results reveal that bottom-up method is better when AR(1) of at least one lowest level data series from a concerned family is greater than 1/3, regardless the value of the correlation coefficient between error terms. Otherwise, both are almost the same. Nenova and May (2016) try to determine the optimal hierarchical forecasting method under the characteristic of the data set (i.e. correlation) using two-level oblique linear discriminant tree model and spoke in favour of bottom-up method compare to other considered techniques for inconsistent products. Widiarta et al. (2008b) concluded that top-down and bottom-up methods are equal, irrespective of the correlation between data when the lowest level data series follow the MA(1) process. From the judgmental perspective, Kremer, Siemsen, and Thomas (2016) studied the effect of correlation on forecasting that exists at the lower level of the hierarchy and found that bottom-up provided better results for the lowest level of the hierarchy. Whereas, the top-down method is more suitable when products are substitutes with each other at the lowest level in both long and short terms. The conclusions of the existing hierarchical forecasting literature on the comparative performance of two classical approaches still not clear to a certain extent.

DeLurgio (1998) argues that combining the two methods may improve the forecast accuracy when the performance of one method is not significantly better than the other one. Furthermore, various researchers claim that forecast combination produces, on average, better results compare to individual forecasting methods (Costantini & Pappalardo, 2010; Patton & Sheppard, 2009; Stock & Watson, 2004). Some authors also suggest that complex forecast combination methods do not always produce more accurate results compare to the simpler methods (Koning, Franses, Hibon, & Stekler, 2005; Stock & Watson, 2004). On the other hand, few studies concluded that best individual forecasting method might provide a smaller error, on average, compared to combination methods (Hibon & Evgeniou, 2005; Yang, 2004). In the light of above discussion, the contributions of this study are as follows:

- We tested that, is the positive or negative correlation between series affect the performance of individual hierarchical forecasting methods and determine under what conditions particular method perform better compare to other.
- We compare 12 major forecast combination methods to rank the top 5 most appropriate methods to improve the forecast accuracy for this hierarchical structure data series.
- One forecast combination method (i.e. Shift based) considered in this study is novel thus this work also contributed to the topic of forecast combination.

The remaining paper organises as follows. Section 2 briefly discusses and summarises the findings from the existing literature. Section 3 describes the individual and combination methods for forecasting hierarchical structure data series. Section 4 evaluates the performance of the alternative individual and combination forecasting methods using standard measures of accuracy. Section 5 concludes with a discussion of the results of the study.

2. Literature

The comparative performance of the hierarchical forecasting methods extensively studied in the literature using judgmental, empirical, simulation and analytical methods or their combinations (Hyndman et al., 2011; Kremer et al., 2016; Sbrana & Silvestrini, 2013). Using analytical methods, the comparative performance of forecasting methods are through the evaluation of forecast errors variability (Sbrana & Silvestrini, 2013; Schwarzkopf et al., 1988; Widiarta et al., 2008b) while the magnitude of forecast errors is used in the simulation and empirical studies (Hyndman et al., 2011; Moon et al., 2012; Viswanathan et al., 2008). Besides the forecasting methods themselves, the relative performance of the forecasting methods varies due to different features of the data (Moon et al., 2013; Nenova & May, 2016; Schwarzkopf et al., 1988; Widiarta et al., 2008a), which include, for example, correlation, the coefficient of variation, forecast horizon, the degree of substitutability etc. The literature of hierarchical forecasting didn’t give much attention to the combination methods. Only a few studies found in the literature those used these methods to improve the forecast accuracies for hierarchical time series (Hyndman et al., 2011; Moon et al., 2012).

In an early computational study, Shlifer and Wolff (1979) results reveal the superiority of the bottom-up method over the top-down method. Schwarzkopf et al. (1988) use the analytical study to investigate the strength of the two classical methods and spoke in favour of the bottom-up method, but also suggest that the top-down method is useful when the lowest level of the hierarchy contains missing data. Under AR(1) process, Widiarta et al. (2007) concluded that bottom-up method provides better results when at least one product demand having more than 0.3333 lag-1 autocorrelations. On the other hand, if lag-1 autocorrelation is less than 0.3333 then bottom-up and top-down methods are almost equal. Dangerfield and Morris (1992) studied the effect of correlation on the performance bottom-up and top-down methods using the M-competition data. The results of their study reveal that bottom-up method outperforms the top-down method in approximately 75% cases. Moreover, they argue that as the correlation between items increases the performance of bottom-up method become more significant. Gordon, Morris, and Dangerfield (1997) and Weatherford et al. (2001) also found similar results. Whereas, in the analytical study, Widiarta et al. (2008b) show that top-down and bottom-up methods are equally good at item-level, irrespective of the correlation between items when each item demand follows the MA(1) process. Sbrana and Silvestrini (2012, 2013) found similar results at the aggregate level when the demand for each item in the family follows the MA(1) process as well as in multivariate exponential smoothing framework, respectively. However, some studies concluded that top-down method performs better compared to the bottom-up with the proper selection of disaggregation scheme (Athanasopoulos et al., 2009; Gross & Sohl, 1990; Park & Nassar, 2014). Fliedner (1999), Grunfeld and Griliches (1960) also support the dominance of the top-down method. With extensive studies on the comparative performance of two classical (i.e. top-down and bottom-up) methods for a long time, the conclusion is yet unclear to a certain extent.

Kahn (1998) suggests combining the bottom-up and top-down methods to take the benefits of both forecastings methods, but he does not provide specific ideas. Hyndman et al. (2011) propose a regression-based an optimal combination method (OCM) where independent forecasts are generated for all the time-series at each level of the hierarchy, and then with the help of a regression model, optimal reconciled forecasts are obtained. With the help their empirical results, they concluded that OCM provides better results compare to bottom-up and top-down methods. Hyndman, Lee, and Wang (2016) extended OCM to a more general collection of time series by incorporating aggregation constraints and proposed an efficient algorithm to deal with the computation. They tested the performance of the hierarchical forecasting approaches using Australia labour market data and spoke in favour of OCM. Moon et al. (2012) use the hierarchical forecast methods for predicting the demand for spare parts in a case study of the South Korean Navy and conclude that the forecast combination methods perform better compared to both bottom-up and top-down methods. In the following paper, Moon et al. (2013) also advocate the superiority of the combination method for naval spare parts demand.

In the literature of hierarchical forecasting, only a few authors mentioned above discuss the forecast combinations methods. To fill this research gap, we tested twelve different forecast combination methods to rank up to top five best combination methods, to improve the forecast accuracy, for this hierarchical data set. Moreover, we also proposed a novel forecast combination method.

3. Hierarchical forecasting methods

We first describe the notations of the general hierarchical forecasting which are same as used by the Rehman, Wan, and Rafique (2017) followed by three individual and twelve combination methods to forecast the hierarchical time series. In order to illustrate the notations for general hierarchical forecasting, we consider the example of three-level hierarchy (see Figure 1), where level-0 represents the totally aggregated forecast series, level-1 represents the 1st level of disaggregation forecast series, level-2 represents the 2nd level of disaggregation and level-3 represents the lowest level of the forecast series (Table 1).

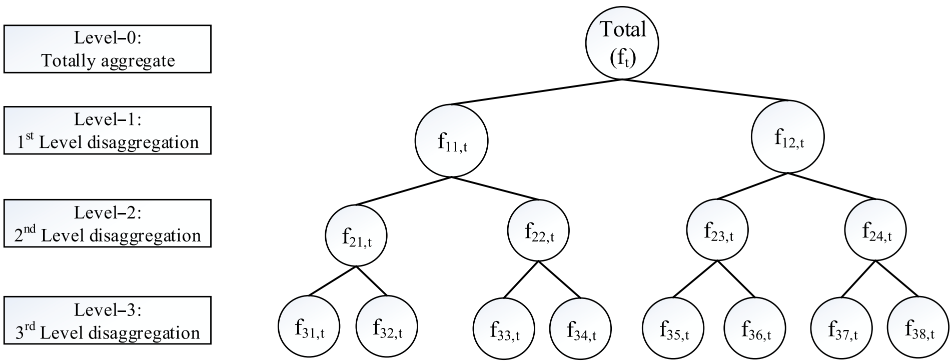

Figure 1. Three-level hierarchical tree diagram.

In Figure 1, there are four, two and two families at level-3, level-2 and level-1, respectively, according to the hierarchical structure. For example, the members of the first family at level-3, level-2 and level-1 are '$f_{31,t}, f_{32,t}$', '$f_{21,t}, f_{22,t}$' and '$f_{11,t}, f_{12,t}$', respectively, whereas the members of the second family at level-3 and level-2 are '$f_{33,t}, f_{34,t}$'and '$f_{23,t}, f_{24,t}$' and so on.

Table 1. Notations.

Notation	Definition
Indices	
$\boldsymbol{l}, \boldsymbol{k}$	Level of the hierarchy (= 0, 1, 2, …, $K-1$)
$\boldsymbol{i}, \boldsymbol{j}$	Forecasted values at time 't' (= 1, 2, …, N)
$\boldsymbol{q}$	Family of forecasted values (= 1, 2, …, Q)
$\boldsymbol{m}, \boldsymbol{n}$	Hierarchical forecast approaches (= B, T, and C)
Index sets	
$\boldsymbol{S_{qlt}}$	Set of forecasted values that belong to $\boldsymbol{q^{th}}$family of $\boldsymbol{l^{th}}$ hierarchy level at time '$\boldsymbol{t}$'
Parameters and data	
$\boldsymbol{p_{jl}}$	Proportion for $\boldsymbol{j^{th}}$ series at the $\boldsymbol{l^{th}}$ hierarchy level
$\boldsymbol{f_{kit}}$	kth hierarchy level forecast of series 'i' at time 't' (i.e. independent / base forecast)
$\boldsymbol{F_{0t}}$	Base forecasts at the totally aggregate level
$\boldsymbol{f^{m}_{lit}}$	mth hierarchical forecast method at lth hierarchy level forecast of series 'i' at time 't'
$\boldsymbol{\sigma^2_{ljm}}$	Forecast error variance of the jth series at lth hierarchy level produced by mth forecast method
$\boldsymbol{\sigma_{lj,mn}}$	Forecast error covariance of the jth series at lth hierarchy level produced by mth and nth forecast methods
$\boldsymbol{MSE_{ljm}}$	Mean square error (MSE) of the jth series at lth hierarchy level produced by mth forecast method
$\boldsymbol{R_{ljm}}$	Performance rank of the mth forecast method based on MSE for the jth series at lth hierarchy level
$\boldsymbol{d_{jt}}$	jth Bottom level data series at time 't'
$\boldsymbol{D_t}$	Sum of all the bottom level data series at time 't'
Variables	
$\boldsymbol{w_{ljm}}$	Weight for the jth forecast series at lth hierarchy level produced by mth forecast method

3.1. *Individual hierarchical forecasting methods*

The details of three most popular and commonly used individual hierarchical forecasting methods are as follows.

3.1.1. *Bottom-up method (BU):*

Bottom-Up method firstly obtained 'f_{Kjt}' that is forecasts of the each 'j^{th}' series at the bottom level (i.e. lowest level) and then summing upwards with respect to the structure of hierarchy. The key advantage of using bottom-up method is that no information is lost due to an aggregation of the data (Dangerfield & Morris, 1992). On the other hand, it is more challenging to predict the bottom level data because of higher variability and intermittency (Schwarzkopf et al., 1988; Shlifer & Wolff, 1979).

Under the bottom-up method, the forecast $f_{ljt}^{B}$of the series in period t can be mathematically represented as follows:

$$f_{ljt}^{B} = \sum_{j \in S_{qlt}} f_{(l+1),jt} \quad \forall l,\ t \tag{1}$$

3.1.2. *Top-down static method (TD-S)*

In this method firstly 'F_{0t}' generated that is the base forecast for the totally aggregated data and then disaggregation performs with the help of appropriate proportion method to obtain the lower level forecast according to the hierarchical structure. Usually, item level data series are noisy, in such situation top-down method is more accurate than bottom-up (Fliedner, 1999; Grunfeld & Griliches, 1960).

Under the top-down method, the forecast of the series period $'t'$ can be mathematically represented as follows:

$$f_{ljt}^{T} = p_{jl} F_{0t} \quad l = \ 1, 2, \ \ldots, K \tag{2}$$

Gross and Sohl (1990) provide various disaggregation schemes which depend on the historical data due to having static nature to compute the propositions, we call this 'Top-Down Static Method (TD-S)'. Out of those disaggregation methods, they claim that method 'A' produces reasonably good forecasts. This method is based on equal weight average which can be mathematically represented as follows:

$$p_{jl} = \frac{\sum_{t=1}^{n} \frac{d_{jt}}{D_t}}{n} \ \ \forall\, l$$

3.1.3. *Top-down dynamic method (TD-D)*

In order to improve top-down method having static nature proportions to disaggregate the aggregate data, Athanasopoulos et al. (2009) produce dynamic proportions for disaggregation which are based on forecasted proportions of lower level data series. We call this 'top-down dynamic method (TD-D)'.

These dynamic proportions can be computed as follows.

$$p_{jl} = \prod_{l=0}^{K-1} \frac{f_{ljt}}{\sum_{j \in S_{q,(l+1),t}} f_{(l+1),jt}} \tag{3}$$

3.2. Forecast combination methods

In order to develop the combination methods, we consider two best individual hierarchical forecasting approaches namely bottom-up and top-down dynamic approach (see Table 3). Let the bottom-up and top-down dynamic methods forecasts be 'f_{ljt}^{B}' and 'f_{ljt}^{T}', respectively. Using combination weights, we can combine these two forecasts as follows:

$$f_{ljt}^{C} = w_{ljB} f_{ljt}^{B} + w_{ljT} f_{ljt}^{T} \tag{4}$$

where, 'f_{ljt}^{C}' represents the combined forecast value.

We tested the following '3' average, '2' variance, '3' regression, '1' inverse-mean square error, and rank as well as two shift based forecast combination methods.

3.2.1. *Average-based*

i. Arithmetic mean (AM)

In this method, we combined the two individual forecasts by assigning equal weights to each approach, that is $w_{ljB} = w_{ljT} = 1/2$.

ii. Geometric mean (GM)

The arithmetic mean is a simple and most commonly used average method to combine the forecasts. However, other averages such as geometric mean could also be valuable for combing forecasts. The geometric mean is a nonlinear method which has an advantage over the arithmetic mean is that it always gives a lower value. Moreover, one of the desirable property needed from the forecast combination method is shrinkage which geometric mean provides in a way comparable to the arithmetic mean. Faria and Mubwandarikwa (2008) argue that geometric combination has an advantage over linear combination methods because these combinations are externally Bayesian. Only a few studies used geometric mean due to having special situations, such as volatility forecasting (Patton & Sheppard, 2009) and grey forecasting (Chen, Ding, & Zhang, 2007; Sun, 2012), to combine the forecasts. It provides combined forecast as follows:

$$f_{ljt}^{C} = \sqrt{(f_{ljt}^{B})(f_{ljt}^{T})} \tag{5}$$

iii. Harmonic mean (HM)

The third average method we consider is HM which is also a nonlinear method to combine the forecasts. It is well known that the relationship between three averages is

'AM $\geq$ GM $\geq$ HM'; thus, HM provides some shrinkage. This method also gets the little attention from the researchers (Chen et al., 2007). It produces forecast combination using the following formula.

$$f_{ljt}^{C} = \frac{2(f_{ljt}^{B})(f_{ljt}^{T})}{f_{ljt}^{B} + f_{ljt}^{T}} \tag{6}$$

3.2.2. *Variance-based*

i. Variance-covariance method (VAR-COV)

In this method, we combined the two individual hierarchical forecasts by assigning the weights to the forecast methods according to error variance produced by each method. The key idea underlying this approach is to combine the two individual forecast approaches by giving more weight to the approach that produces less error of variance and less weight to the other one. Rehman et al. (2017) derived the optimum weights using nonlinear programming which is as follows:

$$w_{ljB} = \begin{cases} 0, & \sigma_{lj,BT} \geq \sigma_{ljT}^2 \\ 1, & \sigma_{lj,BT} \geq \sigma_{ljB}^2 \\ \dfrac{\sigma_{ljT}^2 - \sigma_{lj,BT}}{\sigma_{ljB}^2 + \sigma_{ljT}^2 - 2\sigma_{lj,BT}}, & \text{else} \end{cases}, \tag{7}$$

$$w_{ljT} = \begin{cases} 0, & \sigma_{lj,BT} \geq \sigma_{ljB}^2 \\ 1, & \sigma_{lj,BT} \geq \sigma_{ljT}^2 \\ \dfrac{\sigma_{ljB}^2 - \sigma_{lj,BT}}{\sigma_{ljB}^2 + \sigma_{ljT}^2 - 2\sigma_{lj,BT}}, & \text{else} \end{cases} \tag{8}$$

ii. Variance-no-covariance method (VAR-NO-COV)

In the VAR-NO-COV method, we assume that the forecast errors obtained from two individual forecasting approaches are independent. Therefore, considering the independence of these forecast errors, we can rewrite the weights obtained from the VAR-COV method as follows:

$$w_{ljB} = \frac{\sigma_{ljT}^2}{\sigma_{ljB}^2 + \sigma_{ljT}^2} \tag{9}$$

$$w_{ljT} = \frac{\sigma_{ljB}^2}{\sigma_{ljB}^2 + \sigma_{ljT}^2} \tag{10}$$

3.2.3. *Regression-based*

i. Optimal combination method (OCM)

In OCM, we first produce an independent base forecast for each series in the hierarchy with the help of appropriate forecasting methods. Then the optimal combination method is used to combine these base forecasts and to generate a set of reconciled forecasts for each level according to the structure of the hierarchy (Hyndman et al., 2011). The main idea of this method can be derived using a linear regression model as follows

$$\hat{f}_t = S\beta_t + \varepsilon$$

where $\hat{f}_t$ is a vector of base forecast of the entire hierarchy, S is a summing matrix that captures the structure of the hierarchy, 'β_t' is the unknown expected value of the bottom level 'K', and 'ε' has zero mean and covariance matrix $Var(\varepsilon) = \Sigma$.

In the above regression model, Hyndman et al. (2011) show that $\hat{\beta}_t = (S^{/}S)^{-1}S^{/}\hat{f}_t$ is the best linear unbiased estimator for 'β_t'. This unbiased estimator provides set of reconciled forecasts given by

$$f_t^C = (S^{/}S)^{-1}S^{/}\hat{f}_t, \tag{11}$$

ii. Least squares estimation (LSE)

To estimate the 'w_{ljB}' and 'w_{ljT}', we presented the two variations of the linear regression. These variations allowed according to amount of flexibility in the weights. Granger and Ramanathan (1984) provided the formulation of these linear regression formulations as follows:

$$f_{ljt}^C = w_{ljo} + w_{ljB}f_{ljt}^B + w_{ljT}f_t^T \tag{12}$$

$$f_{ljt}^C = w_{ljB}f_{ljt}^B + w_{ljT}f_{ljt}^T \tag{13}$$

Equation (12), we call it LSE-1 model, contains an intercept to remove the biases from the 'w_{ljB}', 'w_{ljT}' and predications which is helpful if there is any belief that individual hierarchical forecasting approach could be biased. By assuming both the individual forecasts are unbiased, Equation (13), we call it LSE-2 model, is presented.

3.2.4. *The inverse of mean square error method (INV-MSE)*

This method is similar to the VAR-NO-COV method in which the weights are determined using the proportion of inverse of the mean square error (MSE) produced by each forecasting method (Stock & Watson, 1998). These weights can be computed as follows:

$$w_{ljB} = \frac{\mathrm{MSE}_{ljT}}{\mathrm{MSE}_{ljB} + \mathrm{MSE}_{ljT}} \tag{14}$$

$$w_{ljT} = \frac{\text{MSE}_{ljB}}{\text{MSE}_{ljB} + \text{MSE}_{ljT}} \tag{15}$$

3.2.5. *Rank-based method (rank)*

Aiolfi and Timmermann (2006) introduced a combination method in which weights are computed relative to the inverse of their performance rank. Compare to the INV-MSE method, this method is less sensitive to outliers but having discrete nature. Mean square error use for measuring the performance for ranking purposes. These weights can be determined using the following formulas:

$$w_{ljB} = \frac{R_{ljT}}{R_{ljB} + R_{ljT}} \tag{16}$$

$$w_{ljT} = \frac{R_{ljB}}{R_{ljB} + R_{ljT}} \tag{17}$$

3.2.6. *Shift-based methods (SHIFT)*

As discussed above, few authors concluded that, on average, best individual forecasting method might provide a smaller error compare to combination methods (Hibon & Evgeniou, 2005; Yang, 2004). However, the strategy of forecast combination may not be useful in this situation. On the other hand, when the performance of individual forecasting approaches is closer to each other, or one is not outperforming the other than forecast combination would be a better strategy. Moreover, various studies concluded that forecast combination strategy is, on average, better compare to individual forecasting methods (Costantini & Pappalardo, 2010; Patton & Sheppard, 2009; Stock & Watson, 2004).

Based on these situations, we propose two novel approaches namely SHIFT-1 and SHIFT-2. The idea of the Shift-1 approach is whether to employ forecast combinations or not whereas, in shift-2 approach, deciding between two forecast combinations methods (i.e. if two individual forecasts are equally good then assign equal weights to each approach otherwise use the idea of the VAR-COV method). In order to employ these methods, we first need to apply the significant statistical test to test the hypothesis:

> The two individual hierarchical forecasting approaches are equally good.

To test the hypothesis, we use the Wilcoxon's signed rank test at the 5% significance level.

i. SHIFT-1 method

In this method, we apply the equal weight method (i.e. AM) if the hypothesis is accepted otherwise we use the best forecasting method (i.e. deciding either shift to forecast combination strategy or not).

ii. SHIFT-2 method

In the SHIFT-2 method, we apply the equal weight method (i.e. AM) if the hypothesis is accepted otherwise use the VAR-COV idea to assign the weights (i.e. deciding either shift to AM or VAR-COV method).

4. Performance evaluation of forecasting approaches

4.1. *The data*

We now test the performance of the individual and combination approaches for hierarchical forecasting under correlated data. We consider a three-level hierarchy shown in Figure 2.

Where, level 0 titled as 'Total' represents the completely aggregated demand (i.e. multi-family demand), level 1 shows family level-1 demand which categorises with respect to positive and negative correlation type series whereas level 2 shows family level-2 demand which is artificially constructed based on the correlation between the two-series using 'Industrial' category data from M3-Computation study. In this industrial category, it composes of 334 monthly data series. These data series are selected and grouped according to the correlation between the two series, for example, High positive ($r \geq 0.7$), Moderate positive ($0.4 \leq r < 0.7$), Low positive ($0 < r < 0.4$), No correlation ($r = 0$), Low negative ($-0.4 < r < 0$), Moderate negative ($-0.7 < r \leq -0.4$), and High negative ($r \leq -0.7$). From each of these groups, the pairwise 20 data series (i.e. $6 \times 2 \times 20 = 240$) are selected randomly and represented in a hierarchical structure shown in Figure 2 (e.g. high positive correlation (H^+), Moderate Positive Correlation (M^+), etc.), and the individual series from the respective family represent the item level demand at level 3 in the hierarchy. The detail about the total number of series at each level of the hierarchy presented in Table 2.

Here, we are interested in generating forecasts at 1st, 2nd and 3rd level of the hierarchy. To generate the forecast at all levels of the hierarchy, firstly direct forecasts are generated for each series at Level 1, 2 and 3 using appropriate forecasting method. Exponential smoothing (Hyndman, Koehler, Snyder, & Grose, 2002) methods are used to forecast the individual time series as these methods are relatively simple but robust, and they perform quite well in forecasting competition against more

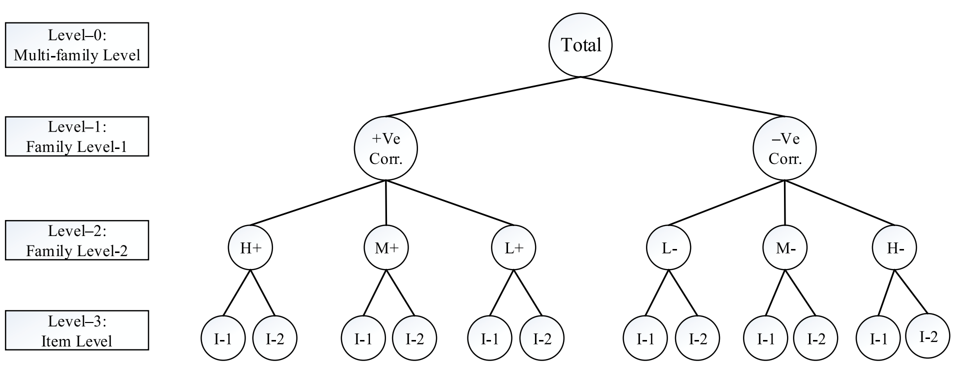

Figure 2. Three-level hierarchical tree diagram based on correlation between data.

Table 2. The details of the hierarchy.

Level	Number of series	Samples	Total series per level
0 – Total	1	–	1
1 – Categories	2	–	2
2 – Correlation	6	20	120
3 – Independent series	12	–	240

sophisticated methods (Makridakis & Hibon, 2000). The 'forecast' package (Hyndman et al., 2015) of the statistical software R, 3.2.3, is used to implement these methods. We refer to these as 'base' forecasts; these base forecasts are then combined to produce the desired hierarchical forecasts in a manner that is consistent with the structure of the hierarchy.

The individual hierarchical forecasting approaches presented in Section 3 are evaluated using out-of-sample forecasts for level 1, 2 and 3 of the hierarchy. To evaluate and compare the performance of the alternative hierarchical forecasting approaches at aggregate levels and sub-aggregate levels, we present the two standard measures of forecast accuracies, namely, root mean square error (RMSE) and mean absolute percentage error (MAPE), as well as the percentage of time particular approach, outperform the others (i.e. performance percentage) in Table 3.

The results in Table 3 show that the performance of the top-down approaches varies with respect to high positive or negative correlations. Similar to Duncan et al. (2001), our study results also reveal that the performance of top-down methods is quite good in high positive correlation, compare to high negative correlation. At Level-2 and 3 of the hierarchy when lowest level data series have a high negative correlation, the performance of top-down dynamic approach (TD-D) decreases 15% to 25% and 15% to 17.5%, respectively, with respect to the measures of accuracy. The results also suggest that the performance of the top-down approach also not as good as when lowest level series have other types of positive correlation (i.e. Low and moderate) compare to a negative correlation. The overall results show the dominance of the bottom-up method compares to the top-down method in all scenarios.

Tables 4–6 show the out-of-sample results of the combination methods at level '1', '2' and '3' of the hierarchy, respectively. The following observations about these twelve forecast combination methods extract out of 20 samples:

- At level-1 of the hierarchy on positive correlation node, we presented up to top five combination methods with respect to their performance (i.e. the percentage of time particular approach outperform the others), according to MAPE, are LSE-1, then OCM, then LSE-2, and then VAR-COV. Whereas on negative correlation node, methods are LSE-1, then LSE-2, then RANK, then SHIFT-2 and then OCM. But, according to RMSE, at both node LSE-1 and then OCM outperformed the others.
- At level-2 of the hierarchy on high positive correlation (H^+) node, methods with respect to their performance, according to MAPE, are LSE-1, then VAR-COV, then OCM, then HM and then LSE-2. On moderate positive

Table 3. Standard measures of accuracy for out-of-sample individual forecasting methods at each level of the hierarchy with the performance percentages in brackets.

Hierarchy level		BU		TD-S		TD-D	
		RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
Level-1	Positive correlation	1156.10 (70%)	3.20 (65%)	2171.49 (5%)	6.38 (5%)	1348.18 (25%)	3.68 (30%)
	Negative correlation	1118.75 (80%)	2.79 (75%)	2006.74 (0%)	5.34 (0%)	1253.11 (20%)	3.16 (25%)
Level-2	High positive correlation (H^+)	819.34 (60%)	5.95 (60%)	1472.57 (0%)	11.00 (0%)	852.24 (40%)	6.24 (40%)
	Moderate positive correlation (M^+)	538.53 (75%)	5.34 (80%)	1403.30 (0%)	14.21 (0%)	572.22 (25%)	5.63 (20%)
	Low positive correlation (L^+)	569.52 (60%)	4.31 (55%)	741.60 (5%)	5.72 (5%)	629.25 (35%)	4.81 (40%)
	Low negative correlation (L^-)	693.79 (65%)	6.44 (65%)	790.62 (15%)	7.25 (5%)	715.98 (20%)	6.74 (30%)
	Moderate negative correlation (M^-)	601.83 (80%)	4.42 (75%)	1115.25 (5%)	8.67 (5%)	660.86 (15%)	4.80 (20%)
	High negative correlation (H^-)	479.43 (85%)	3.09 (75%)	680.47 (0%)	4.51 (0%)	483.16 (15%)	3.12 (25%)
Level-3	High positive correlation (H^+)	528.32 (65%)	7.11 (62.5%)	995.46 (0%)	13.94 (2.5%)	541.43 (35%)	7.28 (35%)
	Moderate positive correlation (M^+)	379.83 (75%)	12.95 (72.5%)	465.11 (0%)	17.08 (2.5%)	378.18 (25%)	12.80 (25%)
	Low positive correlation (L^+)	234.49 (60%)	3.90 (65%)	458.68 (5%)	7.91 (2.5%)	256.48 (35%)	4.37 (32.5%)
	Low negative correlation (L^-)	176.80 (67.5%)	3.77 (72.5%)	256.72 (5%)	5.20 (2.5%)	191.31 (27.5%)	4.00 (25%)
	Moderate negative correlation (M^-)	296.85 (80%)	4.66 (80%)	1157.73 (0%)	21.97 (0%)	317.47 (20%)	5.10 (20%)
	High negative correlation (H^-)	381.98 (80%)	6.20 (82.5%)	1368.73 (0%)	24.88 (0%)	385.85 (20%)	6.27 (17.5%)

Table 4. Standard measures of accuracy for out-of-sample combination forecasting methods at level-1 of the hierarchy.

	Positive correlation		Negative correlation	
Level-1	RMSE	MAPE	RMSE	MAPE
AM	1235.70	3.37	1153.19	2.92
GM	1235.95	3.37	1153.52	2.92
HM	1236.20	3.37	1153.84	2.92
OCM	1236.89	3.38	1149.50	2.93
LSE-1	1107.45	3.14	1093.49	2.72
LSE-2	1108.04	3.16	1118.40	2.79
VAR-COV	1112.97	3.12	1118.74	2.79
VAR-NO-COV	1221.57	3.34	1145.80	2.90
INV-MSE	1221.24	3.34	1145.82	2.90
RANK	1204.56	3.30	1133.60	2.86
SHIFT-1	1235.70	3.37	1153.19	2.92
SHIFT-2	1235.70	3.37	1153.19	2.92

correlation (M^+) node, LSE-1, then VAR-COV, then LSE-2, then OCM and then RANK. On low positive correlation (L^+) node, LSE-1, then VAR-COV, then OCM, then LSE-2 and then RANK. On low negative correlation (L^-) node, LSE-1, then LSE-2, then VAR-COV, then RANK, and then GM. On moderate negative correlation (M^-) node, LSE-1, then LSE-2, then SHIFT-1, then VAR-COV, and the OCM. Whereas on the high negative correlation (H^-) node, LSE-1, then VAR-COV, and then LSE-2. But, according to RMSE, at all nodes LSE-1, OCM, VAR-COV, RANK and HM outperformed the others.

- At level-3 of the hierarchy data series from the high positive correlation (H^+) node, combination methods with respect to their performance, according to MAPE, are LSE-1, then OCM, then VAR-COV, then LSE-2, and then SHIFT-1. For data series from the moderate positive correlation (M^+) node, methods are LSE-1, then LSE-2, then VAR-COV, then OCM and then HM. For data set from the low positive correlation (L^+) node, methods are LSE-1, then LSE-2, then HM, then VAR-COV and then RANK. For data series from the low negative correlation (L^-) node, LSE-1, then LSE-2, then VAR-COV, then OCM, and then HM. For data series from the moderate negative correlation (M^-) node, LSE-1, then LSE-2, then VAR-COV, then OCM and then SHIFT-1 Whereas for data series from the high negative correlation (H^-) node, LSE-1, then LSE-2, then OCM, then VAR-COV and then SHIFT-1. But, according to RMSE, at all nodes LSE-1, LSE-2, and OCM outperformed the others.
- So, the top five combination methods those outperform all the others at all levels of the hierarchy are LSE-1, LSE-2, OCM, VAR-COV and RANK. Moreover, the dominance of all the regression-based and VAR-COV combination methods observe at all levels of the hierarchy in all the cases. Whereas the worst methods for all the levels of the hierarchy are turned out to be

Table 5. Standard measures of accuracy for out-of-sample combination forecasting methods at level-2 of the hierarchy.

	Correlation					
Level-2	H^{+}	M^{+}	L^{+}	L^{-}	M^{-}	H^{-}
Root mean square error						
AM	833.22	548.73	591.10	697.75	615.19	470.23
GM	833.27	548.89	591.13	697.77	615.02	470.39
HM	833.33	549.05	591.16	697.80	614.86	470.55
OCM	834.86	551.89	586.73	706.26	615.36	464.12
LSE-1	782.71	531.60	563.67	686.76	586.96	455.25
LSE-2	806.73	531.72	564.23	692.55	601.61	470.15
VAR-COV	808.01	538.21	565.34	693.44	601.61	470.15
VAR-NO-COV	832.60	547.77	588.18	697.39	612.53	470.12
INV-MSE	832.57	547.73	588.18	697.41	612.51	470.22
RANK	828.01	543.80	581.95	694.82	606.94	470.84
SHIFT-1	819.34	548.73	591.10	697.75	615.19	479.43
SHIFT-2	808.01	548.73	591.10	697.75	615.19	470.15
Mean absolute percentage error						
AM	6.08	5.41	4.49	6.51	4.52	3.01
GM	6.08	5.41	4.49	6.51	4.52	3.02
HM	6.08	5.41	4.49	6.51	4.51	3.02
OCM	6.11	5.46	4.45	6.60	4.54	2.97
LSE-1	5.68	5.36	4.31	6.41	4.33	2.95
LSE-2	5.86	5.38	4.31	6.41	4.41	3.02
VAR-COV	5.87	5.34	4.29	6.44	4.41	3.02
VAR-NO-COV	6.08	5.41	4.47	6.51	4.50	3.01
INV-MSE	6.08	5.41	4.47	6.51	4.49	3.01
RANK	6.04	5.37	4.42	6.47	4.44	3.02
SHIFT-1	5.95	5.41	4.49	6.51	4.52	3.09
SHIFT-2	5.87	5.41	4.49	6.51	4.52	3.02

INV-MSE, VAR-NO-COV and Average-based methods. The performance of shift based methods found to be moderate.

We applied Wilcoxon's signed rank test to check the statistically significant difference between individual and combination forecast methods. It is a non-parametric test for testing the statistically significant difference between two models with respect to their performance. Due to space consideration, we only apply this test to the MAPE values (not for RMSE values). The results show of the performance of top-5 ranked combination forecast methods significantly perform better (at the 5% significance level) compare to the individual methods at all levels of the hierarchy. As each pair of seven methods has smaller p-values than the level of significance, with a Bonferroni correction, as measured by MAPE. Moreover, we also found, in some cases, that the performance of two forecast combination methods (i.e. OCM and RANK) is as good as bottom-up (BU) method.

Table 6. Standard measures of accuracy for out-of-sample combination forecasting methods at level-3 of the hierarchy.

	Correlation					
Level-3	H^+	M^+	L^+	L^-	M^-	H^-
Root mean square error						
AM	533.69	378.01	241.20	178.91	299.76	381.17
GM	533.71	378.03	241.17	178.94	299.68	381.21
HM	533.73	378.05	241.14	178.98	299.60	381.24
OCM	535.31	379.87	237.75	176.19	300.67	378.81
LSE-1	511.25	372.57	227.51	174.61	294.75	378.07
LSE-2	524.52	377.58	234.06	174.65	295.43	380.39
VAR-COV	524.85	377.84	234.25	176.33	296.11	380.84
VAR-NO-COV	533.54	378.00	240.23	178.30	299.14	381.15
INV-MSE	533.53	378.00	240.23	178.35	299.09	381.15
RANK	531.63	378.39	237.95	177.01	297.08	380.83
SHIFT-1	528.32	378.01	241.20	178.91	299.76	381.98
SHIFT-2	524.85	378.01	241.20	178.91	299.76	380.84
Mean absolute percentage error						
AM	7.16	12.81	4.05	3.76	4.77	6.15
GM	7.16	12.81	4.05	3.76	4.77	6.15
HM	7.16	12.81	4.05	3.76	4.77	6.15
OCM	7.23	12.81	4.01	3.73	4.77	6.16
LSE-1	6.76	12.70	3.90	3.71	4.65	6.10
LSE-2	7.08	12.78	3.87	3.71	4.65	6.13
VAR-COV	7.07	12.78	3.90	3.75	4.68	6.16
VAR-NO-COV	7.15	12.81	4.03	3.74	4.76	6.15
INV-MSE	7.15	12.81	4.03	3.74	4.76	6.15
RANK	7.14	12.85	3.99	3.73	4.72	6.16
SHIFT-1	7.11	12.81	4.05	3.76	4.77	6.20
SHIFT-2	7.07	12.81	4.05	3.76	4.77	6.16

5. Conclusions

In this paper, we first try to clarify some of the contradiction in the literature about the performance of the top-down method when data contains positive or negative correlation. The results of our study reveal that the performance of top-down method under high positive correlation is much better, compare to high negative correlation. But, these results also indicated that bottom-up approach outperformed the top-down approach, either the lowest level data series have a positive or negative correlation, at each level of the hierarchy.

Secondly, we investigated the idea of forecast combination methods using hierarchical structure time series. We tested twelve different forecast combination methods and provided the best five combination methods. The results of this empirical study indicated the improvements in the accuracy with the help of forecast combinations relative to the individual methods. Moreover, results also showed the extent to which these combination methods performed well compare to individual forecasts. Data used in this study is the real-time series taken from the industrial category of M3-Competition (Makridakis & Hibon, 2000).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Program of Shanghai Subject Chief Scientist [16XD1401700] and National Natural Science Foundation of China [71421002].

References

Aiolfi, M., & Timmermann, A. (2006). Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics*, *135*(1–2), 31–53.

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

Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Petropoulos, F. (2017). Forecasting with temporal hierarchies. *European Journal of Operational Research*, *262*(1), 60–74.

Chen, H., & Boylan, J. E. (2009). The effect of correlation between demands on hierarchical forecasting. In K. D. Lawrence & R. K. Klimberg (Eds.), *Advances in business and management forecasting* (Vol. 6, pp. 173–188). Bingley: Emerald Group Publishing.

Chen, H., Ding, K., & Zhang, J. (2007, November 18–20). *Properties of weighted geometric means combination forecasting model based on degree of logarithm grey incidence*. IEEE international conference on Grey Systems & Intelligent Services, Nanjing, China.

Costantini, M., & Pappalardo, C. (2010). A hierarchical procedure for the combination of forecasts. *International Journal of Forecasting*, *26*(4), 725–743.

Dangerfield, B. J., & Morris, J. S. (1992). Top-down or bottom-up: Aggregate versus disaggregate extrapolations. *International Journal of Forecasting*, *8*(2), 233–241.

DeLurgio, S. A. (1998). *Forecasting principles and applications.* New York: McGraw-Hill.

Duncan, G. T., Gorr, W. L., & Szczypula, J. (2001). Forecasting analogous time series. In J. S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners* (pp. 195–213). Boston, MA: Kluwer Academic.

Faria, A. E., & Mubwandarikwa, E. (2008). The geometric combination of Bayesian forecasting models. *Journal of Forecasting*, *27*(6), 519–535.

Fliedner, G. (1999). An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation. *Computers and Operations Research*, *26*(10–11), 1133–1149.

Fliedner, G. (2001). Hierarchical forecasting: Issues and use guidelines. *Industrial Management & Data Systems*, *101*(1), 5–12.

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*, *Fall 1997*, 13–16.

Granger, C. W. J., & Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, *3*, 197–204.

Gross, C. W., & Sohl, J. E. (1990). Disaggregation methods to expedite product line forecasting. *Journal of Forecasting*, *9*(3), 233–254.

Grunfeld, Y., & Griliches, Z. (1960). Is aggregation necessarily bad? *The Review of Economics and Statistics*, *42*(1), 1–13.

Hibon, M., & Evgeniou, T. (2005). To combine or not to combine: Selecting among forecasts and their combinations. *International Journal of Forecasting*, *21*(1), 15–24.

Huber, J., Gossmann, A., & Stuckenschmidt, H. (2017). Cluster-based hierarchical demand forecasting for perishable goods. *Expert Systems with Applications*, *76*, 140–151.

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.

Hyndman, R. J., Athanasopoulos, G., Bergmeir, C., Cinelli, C., Khan, Y., Mayer, Z., … Zhou, Z. (2015). Package "forecast": Forecasting functions for time series and linear models. Retrieved from https://github.com/robjhyndman/forecast.github.com/robjhyndman/forecast

Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, *18*(3), 439–454.

Hyndman, R. J., Lee, A. J., & Wang, E. (2016). Fast computation of reconciled forecasts for hierarchical and grouped time series. *Computational Statistics and 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.

Koning, A. J., Franses, P. H., Hibon, M., & Stekler, H. O. (2005). The M3 competition: Statistical tests of the results. *International Journal of Forecasting*, *21*(3), 397–409.

Kremer, M., Siemsen, E., & Thomas, D. J. (2016). The sum and its parts: Judgmental hierarchical forecasting. *Management Science*, *62*, mnsc.2015.2259.

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

Moon, S., Hicks, C., & Simpson, A. (2012). The development of a hierarchical forecasting method for predicting spare parts demand in the South Korean Navy – a case study. *International Journal of Production Economics*, *140*(2), 794–802.

Moon, S., Simpson, A., & Hicks, C. (2013). The development of a classification model for predicting the performance of forecasting methods for naval spare parts demand. *International Journal of Production Economics*, *143*(2), 449–454.

Nenova, Z. D., & May, J. H. (2016). Determining an optimal hierarchical forecasting model based on the characteristics of the data set: Technical note. *Journal of Operations Management*, *44*, 62–68.

Park, M., & Nassar, M. (2014). *Variational Bayesian inference for forecasting hierarchical time series*. In International conference on machine learning (ICML), workshop on divergence methods for probabilistic inference, Beijing.

Patton, A. J., & Sheppard, K. (2009). Optimal combinations of realised volatility estimators. *International Journal of Forecasting*, *25*(2), 218–238.

Pennings, C. L. P., & van Dalen, J. (2017). Integrated hierarchical forecasting. *European Journal of Operational Research*, *263*(2), 412–418.

Rehman, H. U., Wan, G., & Rafique, R. (2017). *A hybrid approach to forecasting hierarchical time series*. Shanghai: Shanghai Jiao Tong University.

Sbrana, G., & Silvestrini, A. (2012). Comparing aggregate and disaggregate forecasts of first order moving average models. *Statistical Papers*, *53*(2), 255–263.

Sbrana, G., & Silvestrini, A. (2013). Forecasting aggregate demand: Analytical comparison of top-down and bottom-up approaches in a multivariate exponential smoothing framework. *International Journal of Production Economics*, *146*(1), 185–198.

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

Shlifer, E., & Wolff, R. W. (1979). Aggregation and proration in forecasting. *Management Science*, *25*(6), 594–603.

Stock, J. H., & Watson, M. W. (1998). *A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series* (No. w6607). National Bureau of Economic Research.

Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, *23*, 405–430.

Sun, L. (2012). Properties of weighted geometric means combination forecasting model based on absolute of Grey incidence. *Advanced Materials Research*, *490–495*, 442–446.

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.

Viswanathan, S., Widiarta, H., & Piplani, R. (2008). Forecasting aggregate time series with intermittent subaggregate components: Top-down versus bottom-up forecasting. *IMA Journal of Management Mathematics*, *19*(3), 275–287.

Weatherford, L. R., Kimes, S. E., & Scott, D. A. (2001). Forecasting for hotel revenue management: Testing aggregation against disaggregation. *The Cornell Hotel and Restaurant Administration Quarterly*, *42*(4), 53–64.

Widiarta, H., Viswanathan, S., & Piplani, R. (2007). On the effectiveness of top-down strategy for forecasting autoregressive demands. *Naval Research Logistics*, *54*, 176–188.

Widiarta, H., Viswanathan, S., & Piplani, R. (2008a). Evaluation of hierarchical forecasting for substitutable products. *International Journal of Services and Operations Management*, *4*(3), 277–295.

Widiarta, H., Viswanathan, S., & Piplani, R. (2008b). Forecasting item-level demands: An analytical evaluation of top-down versus bottom-up forecasting in a production-planning framework. *IMA Journal of Management Mathematics*, *19*(2), 207–218.

Yang, Y. (2004). Combining forecasting procedures: Some theoretical results. *Econometric Theory*, *20*(1), 176–222.