University of Passau

SEMINAR PAPER

for the Master Seminar “Applied Statistics”

on the topic

“Point Predictions of Hierarchical Time-Series Using Hierarchical Forecasting Approaches”

Aleksandra Petrenko

2nd semester

M. Sc. Business Administration

Passau, 2022

Introduction

Hierarchical forecasting is not a new topic in statistics. The researchers have noticed the advantage of hierarchical time-series as they contain more information about a trend, dynamics, seasonality, etc., which is very useful for predictions. Data can be naturally classified by various attributes, therefore there is a demand for hierarchical forecasting methods.

Hierarchical forecasting is also a daily activity for supply chain planners in modern companies. The products are usually classified using product hierarchies or are differentiated by markets and geographical aspects. Each of these categories should be properly predicted as these forecasts are relevant for different management levels. In this paper, we will try to predict real retail data, that is why this research is of current interest.

The aim of this paper is to give an overview of the hierarchical point forecasting methods for time-series and see how these approaches perform at the real retail sales data.

The objectives of this paper are to:

- Conduct literature research on the topic hierarchical and grouped time-series forecasting,
- Determine main reconciliation approaches for calculating a forecast for hierarchical time-series,
- Find and prepare the data for testing the defined hierarchical forecasting techniques,
- Define the models for calculating base forecasts,
- Compare the prediction performance of different hierarchical forecasting methods for predicting baseline sales (sales without incentives like promotional events, holidays, etc.)
- Choose the best model for the data provided.

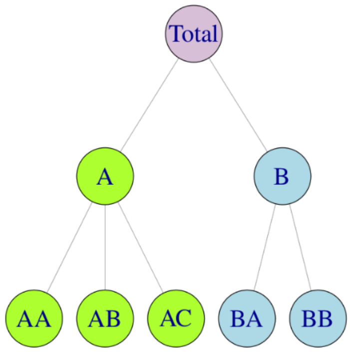
The tools used for conducting a statistical analysis for this paper are R, Gretl (mostly for investigation and plotting) and Excel (export and design).

Literature review

Hierarchical and grouped time-series definition

In practice, the time-series can usually be disaggregated by different attributes. For example, the sales data could be divided by product groups whereas these product groups could be split up into single products. The birth rate data could be classified by region or by sex.

Below is an example of a hierarchical structure of time series, where the number of levels is $k = 2$.



Picture 1. A two-level hierarchical tree diagram. (Hyndman & Athanasopoulos, 2018).

The top total series observation is denoted by y_t for each period $t = 1, \ldots, T$. The following bottom nodes can be denoted as $y_{j,t}$ where j is a corresponding node. For instance, the level $k = 1$ consists of time series observations $y_{A,t}$ and $y_{B,t}$, and each of them can be correspondingly spitted to $y_{AA,t}$, $y_{AB,t}$, $y_{AC,t}$ and $y_{BA,t}$, $y_{BB,t}$ at the level $k = 1$ for each period $t = 1, \ldots, T$.

The important constraint is that each node observation equals to sum of the corresponding observations at the following lower level, in which the node is split up, for each period t. In our example hierarchy the constraints can be written as follows

$$\begin{cases} y_t = y_{AA,t} + y_{AB,t} + y_{AC,t} + y_{BA,t} + y_{BB,t} \\ y_{A,t} = y_{AA,t} + y_{AB,t} + y_{AC,t} \\ y_{B,t} = y_{BA,t} + y_{BB,t} \end{cases}$$

or using the matrices

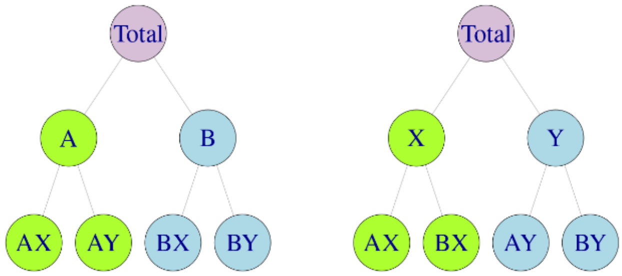
$$\begin{pmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{pmatrix}$$

Let n be a total number of series in the hierarchy and m – a number of series at the lowest level. Then the compact notation of hierarchical time series constraints is

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{y}_{k,t},$$

where $\boldsymbol{y}_t$ an n-dimensional vector of all the observations in the hierarchy at time t, $\boldsymbol{S}$ is a summing matrix $n \times m$, showing the way how the series are aggregated at different levels, and $\boldsymbol{y}_{k,t}$ is an m-dimensional vector of all the observations in the bottom-level of the hierarchy at time t. (Hyndman & Athanasopoulos, 2018; Abolghasemi et al., 2022).

However, in practice data classes do not necessarily have a determined hierarchical order. As in an example above, the birth rate can be classified firstly by region and then by sex or vice versa. The disaggregating factors in such series are both nested and crossed. Such data refer to grouped time-series, whose possible structure is depicted below.



Picture 1. A two-level grouped structure. (Hyndman & Athanasopoulos, 2018).

In this case, the time-series have additional constraints for the first level: not only for nodes A and B, but also for X and Y. This can be also written in the matrix notation explained above:

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{y}_{k,t},$$

$$\begin{pmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{X,t} \\ y_{Y,t} \\ y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{pmatrix}$$

Most of the reconciliation approaches support only hierarchical time-series. However, grouped time series can be interpreted as hierarchical time series that do not impose a unique hierarchical structure, so the order by which the series can be grouped is not unique. (Hyndman & Athanasopoulos, 2018).

Hierarchical point forecasting methods

The problem of hierarchical forecasting relates to the constraints discussed before and introduces additional complexity: namely, apart from selecting an appropriate forecasting model, the produced forecasts should be coherent. Coherence means that the forecasts of the lower levels of the hierarchy must sum up to the forecasts of the higher level.

Let $\widetilde{\boldsymbol{y}}_{n(h)}$ be a final h-step-ahead reconciled coherent forecast of all series and $\widehat{\boldsymbol{y}}_{n(h)}$ is a h-step-ahead independent base forecast of all series based on observations. Then all forecasting approaches for the hierarchical time series can be formulated as

$$\widetilde{\boldsymbol{y}}_{n(h)} = \boldsymbol{SG}\widehat{\boldsymbol{y}}_{n(h)},$$

where $\boldsymbol{G}$ is a $m \times n$ mapping matrix and $\boldsymbol{S}$ is a summing matrix. The $\boldsymbol{G}$ matrix depends on the hierarchical forecasting approach used. (Abolghasemi et al., 2022, Hyndman & Athanasopoulos, 2018).

1. *Bottom-up approach*

The bottom-up approach is the simplest method of hierarchical forecasting. This is the only approach which also works directly with the grouped time-series. The coherent forecast of the higher levels is produced just by summarizing the bottom level base forecasts. In this case, $\boldsymbol{G}$ can be constructed as $[\boldsymbol{0}_{m\times(n-m)}|\boldsymbol{I}_m]$ (a null-matrix and an identity matrix), so the shortened notation can be written as follows (Abolghasemi et al., 2022):

$$\widetilde{\boldsymbol{y}}_{n(h)} = \boldsymbol{S}\widehat{\boldsymbol{y}}_{n(h)}$$

2. *Top-down approaches*

The idea of this approach is to calculate forecast for the highest level – total time-series, and then split the forecast according to disaggregation proportions for the bottom level. As soon as the forecast for the lowest level is calculated, it is summed up for the middle levels to get coherent forecasts for the rest of the time-series. The matrix G can be constructed as $[\boldsymbol{p}_{m\times 1}\ |\boldsymbol{0}_{m\times(n-1)}]$, where the vector $\boldsymbol{p}$ is a vector of proportions which show how the forecast of the total level can be distributed on the lowest level. The short notation of this approach is:

$$\widetilde{\boldsymbol{y}}_{n(h)} = \boldsymbol{p}\widehat{\boldsymbol{y}}_{n(h)}$$

A disadvantage of all top-down approaches is that the coherent forecasts produced are always biased, so $E[\widehat{\boldsymbol{Y}}_{\boldsymbol{n}}(h)] \neq E[\boldsymbol{Y}_n(h)]$ (the evidence is provided in Hyndman et al, 2011). A part of information about the time series is lost due to aggregation. However, such approaches are quite simple, which is an advantage. (Abolghasemi et al., 2022, Hyndman & Athanasopoulos, 2018).

There are multiple ways how to find the vector of proportions:

2.1 Average historical proportions

This method was described in Gross & Sohl, 1990 as a "top-down method A". Each proportion is an averaged historical relation between bottom and total time-series for all periods t.

$$p_i = \frac{1}{T}\sum_{t=1}^{T}\frac{y_{i,t}}{y_t}, \qquad i = 1, \dots, m$$

2.2 Proportions of the historical averages

One more method from Gross & Sohl, 1990 – "method F". The proportion is derived as a relation between an average level of the bottom time-series over all periods of time and of total time-series over all periods of time.

$$p_i = \sum_{t=1}^{T}\frac{y_{i,t}}{T}\Bigg/\sum_{t=1}^{T}\frac{y_t}{T}, \qquad i = 1, \dots, m$$

2.3 Forecast proportions

This approach is more advanced because it captures the dynamics and trends of each time series, which can obviously change over time. First, base forecasts for each time-series are calculated. As soon as these forecasts are not coherent, we use them just to obtain the proportions. If we consider one full branch of the hierarchical tree, the proportion for the end node of this branch is calculated as follows: we obtain the relations of each node of this branch to the whole corresponding level and multiply these relations. To keep it simple, below is an example for the hierarchy from the Picture 1, namely for the $y_{AA,t}$.

$$\tilde{y}_{AA,t} = \left(\frac{\hat{y}_{AA,t}}{\hat{y}_{AA,t} + \hat{y}_{AB,t} + \hat{y}_{AC,t}}\right)\left(\frac{\hat{y}_{A,t}}{\hat{y}_{A,t} + \hat{y}_{B,t}}\right)\hat{y}_t$$

The mathematical notation for this approach is:

$$p_i = \prod_{l=0}^{K-1}\frac{\hat{y}_{i,h}^{(l)}}{\hat{S}_{i,h}^{(l+1)}}, j = 1, \dots, m,$$

Where $\hat{y}_{i,h}^{(l)}$ is a h-step-ahead initial forecast of the series that corresponds to the node which is l levels above i and $\hat{S}_{i,h}^{(l+1)}$ is the sum of the h-step-ahead initial forecasts below the node that is l levels above node i and are directly connected to that node. (Hyndman & Athanasopoulos, 2018).

3. *Middle-out approach (combined approach)*

Middle-out approach is a combination of methods above. First, we choose one of the middle levels to begin with and generate the coherent forecasts for the levels lower using a top-down approach (in our case – with forecast proportions) and a bottom-up approach for the levels higher than the chosen middle level. The $\boldsymbol{G}$ matrix will be a combination of the above two. Villegas & Pedregal, 2018 argue that a middle-out approach is just a subset of combined approaches, where the methods you combine depend on the selecting appropriate values for the diagonal elements of the summing matrix and the data affecting it. (Hyndman & Athanasopoulos, 2018).

4. *Optimal reconciliation approach*

This approach deals with the biasedness of the coherent forecasts obtained with a top-down approach. Hyndman et al. 2011 provide evidence that for unbiased forecasts the below relation must be true. Moreover, they show that this statement is never true for the top-down approaches, as mentioned above. The statement is a constraint for the mapping matrix $\boldsymbol{G}$.

$$\boldsymbol{SGS} = \boldsymbol{G}$$

Wickramasuriya et al. (2019) prove that the variance-covariance matrix of the h-step-ahead coherent forecast errors is constructed as follows:

$$\boldsymbol{V}_h = Var[\boldsymbol{y}_{T+h} - \widetilde{\boldsymbol{y}}_h] = \boldsymbol{SGW}_h\boldsymbol{G}'\boldsymbol{S}'$$

The matrix $\boldsymbol{W}_h = Var[(\boldsymbol{y}_{T+h} - \widehat{\boldsymbol{y}}_h)]$ is a variance-covariance matrix of the base forecast errors. Now we must find a matrix $\boldsymbol{G}$ minimizing the error variances of coherent forecasts, namely, minimizing the sum of all error variances on the diagonal of $\boldsymbol{V}_h$ – the trace $tr(\boldsymbol{V}_h)$. Moreover, the condition $\boldsymbol{SGS} = \boldsymbol{G}$ should be fulfilled. Wickramasuriya et al. (2019) proves that such a matrix is constructed as follows:

$$\boldsymbol{G} = (\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}$$

The optimal reconciled forecast is calculated as:

$$\widetilde{\boldsymbol{y}}_h = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}\widehat{\boldsymbol{y}}_h$$

This is called a minimum trace estimator “MinT”.

The next challenge is to estimate the matrix $\boldsymbol{W}_h$. There are several approaches to simplify the estimation. Some of them offer the estimators which do not depend on the data. (Hyndman & Athanasopoulos, 2018).

4.1 OLS estimator

We simply ignore the $\boldsymbol{W}_h$ and set it equal to identity matrix $\boldsymbol{I}$. The matrix $\boldsymbol{G}$ is now calculated as an OLS-estimator.

$$\boldsymbol{G} = (\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$$

4.2 WLS estimator using variance scaling

We ignore everything apart from a diagonal and set $\boldsymbol{W}_h = \boldsymbol{\Lambda}_v = \ diag(\boldsymbol{W}_1)^{-1}$.

$$\widehat{\boldsymbol{W}}_1 = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{e}_t\, \boldsymbol{e}'_t$$

The vector $\boldsymbol{e}_t$ is an n-dimensional vector of residuals of the models for base forecasts stacked in the same order as data. The approach scales the base forecasts using the variance of residuals and therefore is referred to as WLS estimator using variance scaling. The matrix $\boldsymbol{G}$ is specified as follows:

$$\boldsymbol{G} = (\boldsymbol{S}'\boldsymbol{\Lambda}_v\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{\Lambda}_v$$

4.3 WLS estimator using structural scaling

Almost as in the previous approach, we set $\boldsymbol{W}_h = \boldsymbol{\Lambda}_s = \ diag(\boldsymbol{S1})^{-1}$. This approach assumes that the bottom-level base forecast errors are not correlated between nodes. Each element of the diagonal $\boldsymbol{\Lambda}_s$ matrix contains the number of forecast error variances contributing to each node. This estimator only depends on the structure of the aggregations. The matrix $\boldsymbol{G}$ is now calculated as:

$$\boldsymbol{G} = (\boldsymbol{S}'\boldsymbol{\Lambda}_s\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{\Lambda}_s$$

4.4 Minimal trace (sample estimate of the residual covariance matrix)

We set $\boldsymbol{W}_h = \widehat{\boldsymbol{W}}_{sam}$. Here we assume that the error covariance matrices are proportional to each other, and we directly estimate the full one-step covariance matrix. The most obvious and simple way would be to use the sample covariance.

$$\boldsymbol{G} = (\boldsymbol{S}'\widehat{\boldsymbol{W}}_{sam}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\widehat{\boldsymbol{W}}_{sam}^{-1}$$

4.5 Minimal trace (shrinkage estimate of the residual covariance matrix)

The previous approach works well when the number of time-series at the bottom level is not large in comparison to the number of time periods in the time series. To improve it, the shrinkage estimator $\boldsymbol{W}_h = \widehat{\boldsymbol{W}}_{shr} = \lambda diag(\widehat{\boldsymbol{W}}_{sam}) + (1-\lambda)\widehat{\boldsymbol{W}}_{sam}$ was introduced, where λ is an intensity parameter. It shrinks the off-diagonal elements towards zero. By parameterizing the shrinkage in terms of variances and correlations, rather than variances and covariances, and assuming that the variances are constant, Schäfer & Strimmer, 2005 proposed the following shrinkage intensity parameter:

$$\lambda = \frac{\sum_{i \neq j} \widehat{Var(\hat{r}_{ij})}}{\sum_{i \neq j} \hat{r}_{ij}^2}$$

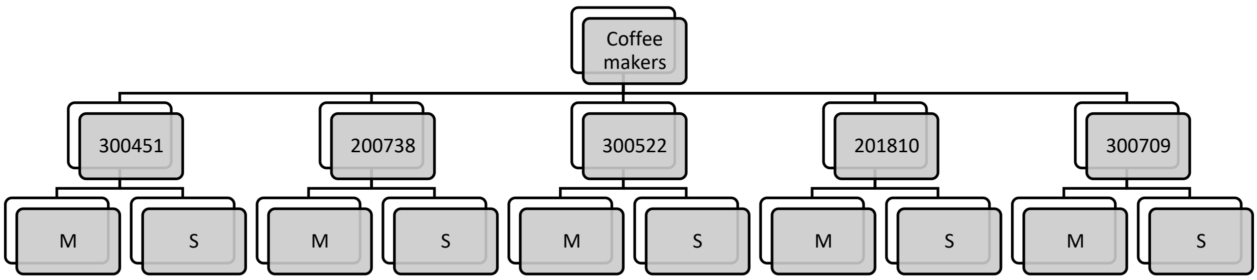
$\hat{r}_{ij}$ is the ij-element of the sample correlation matrix of the in-sample 1-step ahead base forecast errors. As an advantage over previous methods, this shrinkage estimator also accounts for the relationships between the time-series while the shrinkage parameter regulates the complexity of the matrix $\boldsymbol{W}_h$. (Hyndman & Athanasopoulos, 2018; Oliveira & Ramos, 2019).

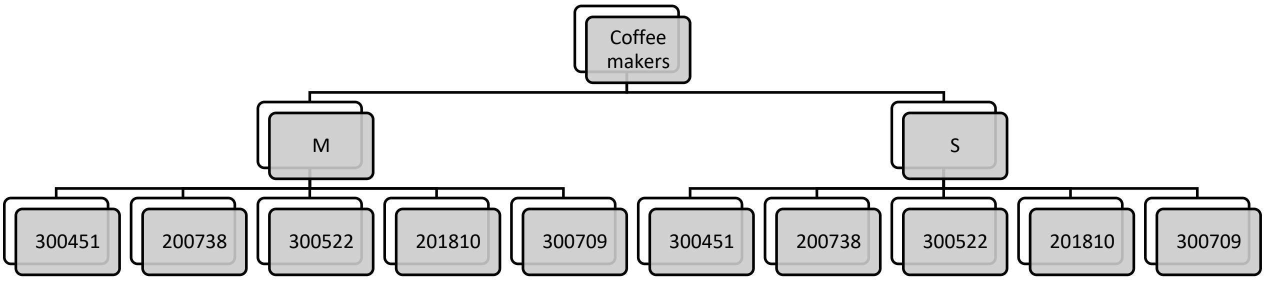
Empirical study

Data description

The sales data of two large German electronics stores M and S was chosen for the analysis. The grouped time-series show the sales of the category "coffee makers" in pieces, the product category consists of 5 SKUs (stock-keeping-units – the smallest product hierarchy category, product items with unique material numbers). The data is aggregated on a week level because supply planning for stores is done on weekly basis. The period of the data is from 41-2020 week till 24-2022 week.

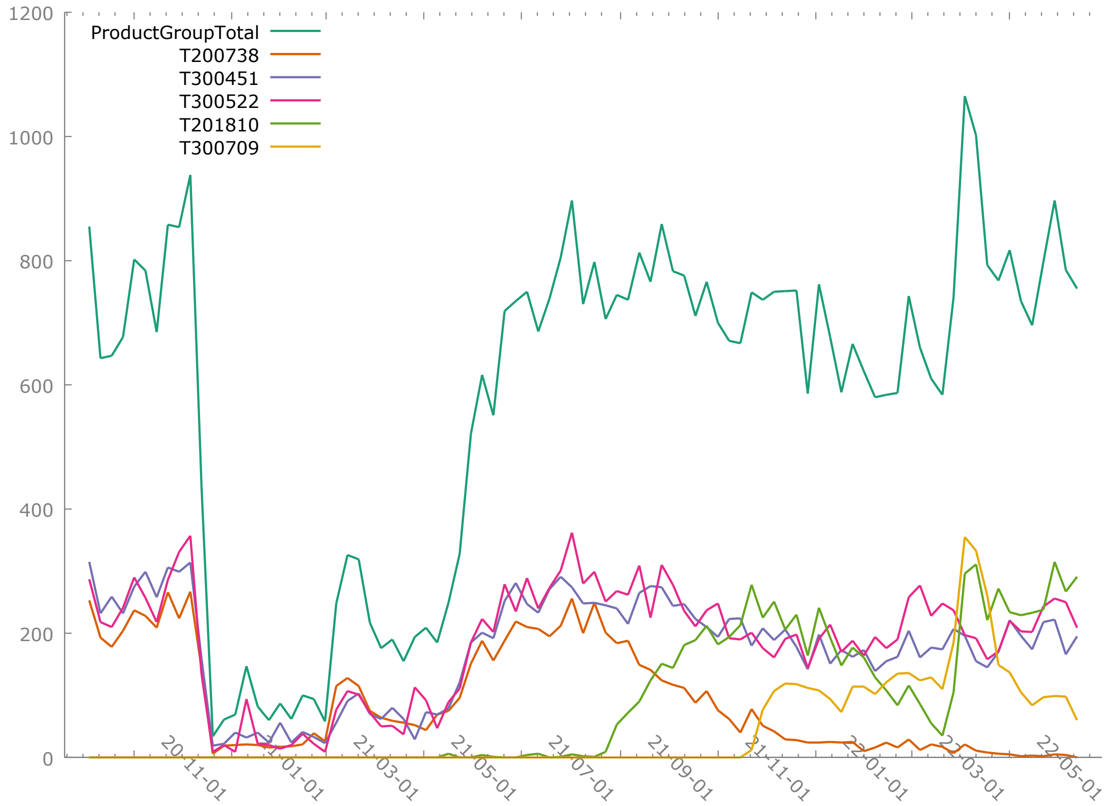
The data is grouped because there is no pre-defined hierarchical order. The products are classified by SKU and retailer. The hierarchical schemes of the data are depicted below.



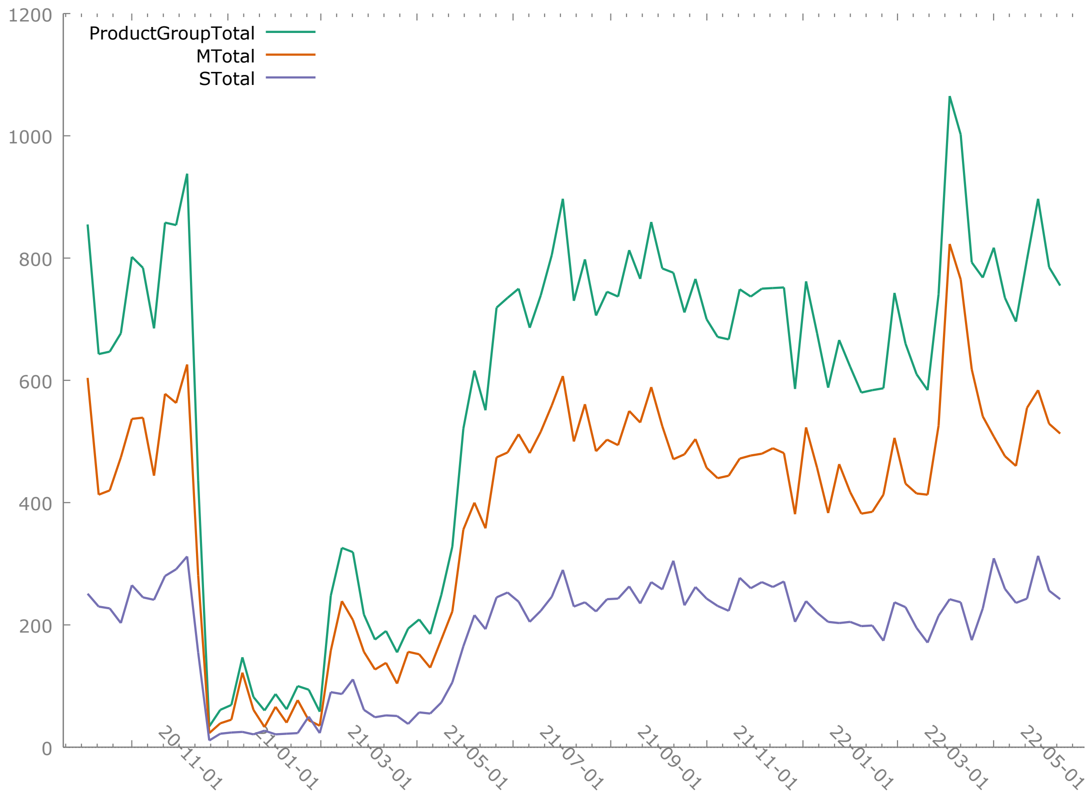


Picture 1. Two possible hierarchies of coffee makers: classification by SKU and by retailer.

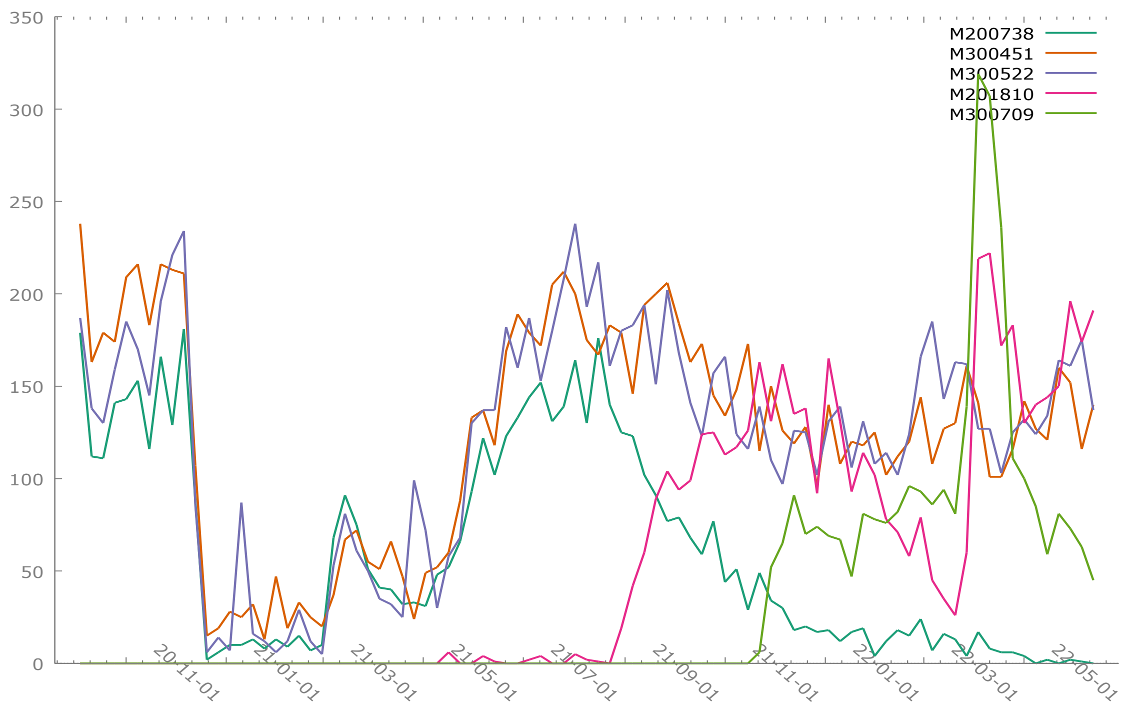
The time series for different levels look like as follows:



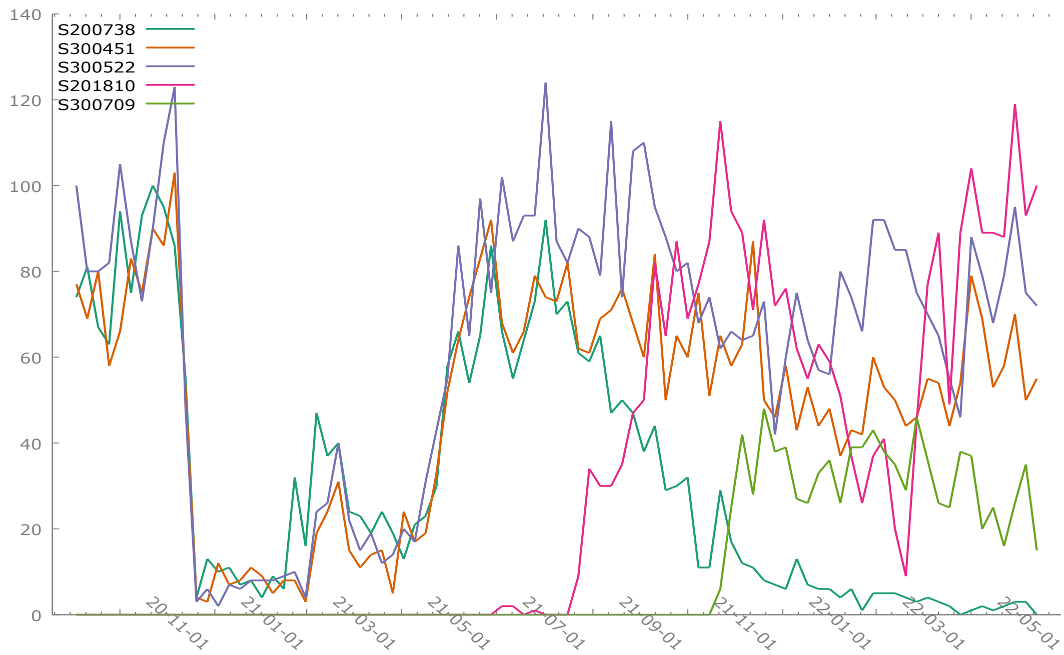
Picture 1. Total sales of coffee makers and sales of each SKU for both retailers.



Picture 1. Total sales of coffee makers and total sales for both retailers.



Picture 1. Sales of each SKU for retailer M.



Picture 1. Sales of each SKU for retailer S.

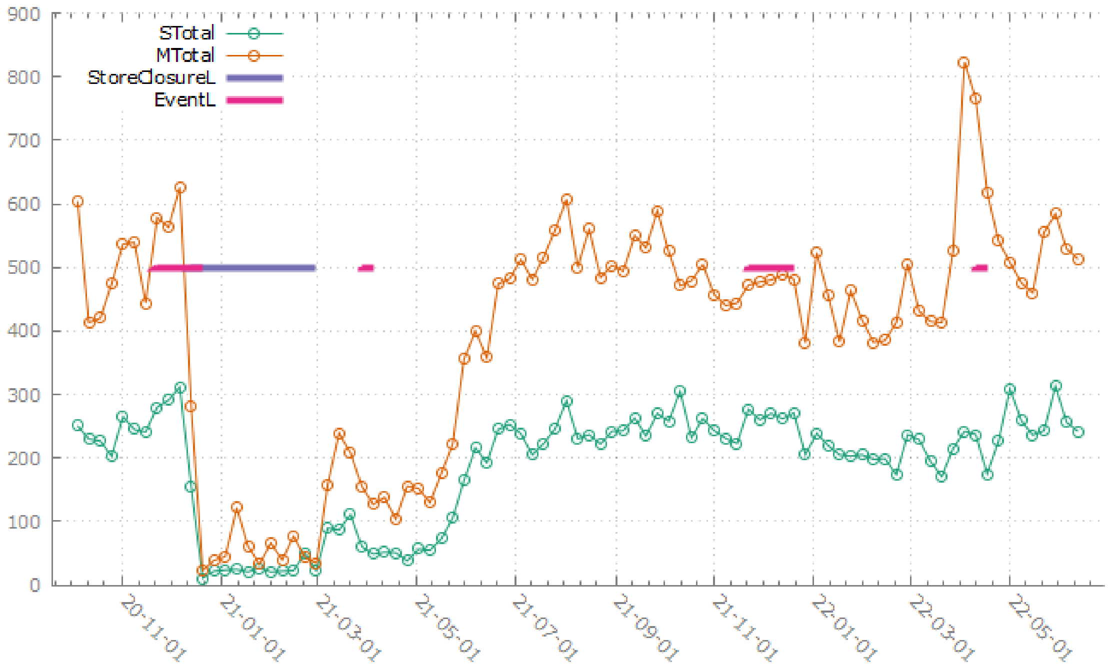
In addition to sales data, there is information in the dataset about the promotional events in the stores (Black Friday, Christmas and Easter) and temporary store closures during the coronavirus pandemic in Germany in 2020-2021.

Data pre-processing

If we look at the graphs in the previous section, we may see that the retailer M has a larger share of sales as the retailer S. The trend of total sales in both electronics markets is quite similar. Moreover, if we compare the sales of each SKU in both stores, the trends and patterns of time-series are similar, too. If we take the total sales of each SKU, we may observe cannibalization (when one new similar product “cannibalizes” the sales of the other substitute product) or product replacements. These dependences and “reallocation” of sales tell us that the usage of hierarchical forecasting could be useful as we can predict the points considering other time-series.

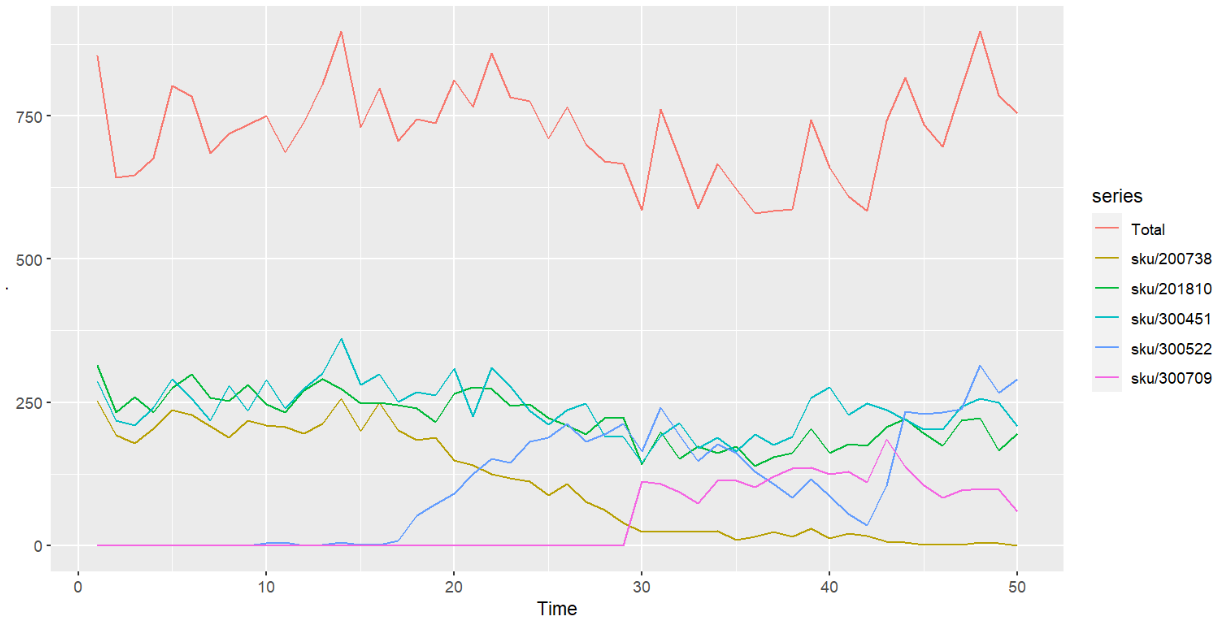
The aggregated sales data for the retailers M and S as well as the time periods of promotional events in the stores and temporary store closures are depicted on the graph below. There is an obvious impact of these events: the store closures have dramatically dropped the sales and caused the structural break in the time series whereas some of the successful promotional campaigns have significantly raised the sales. However, this impact is seen not strictly within the weeks of special events: the sales raised not only during the campaign, but also a week before and a week after, which could be connected not only with promotions, but also with holidays themselves. The same is with store closures: it took a little time to recover the previous sales level.

After trying to figure out some seasonal or cyclic patterns, it can be concluded that none of them are presented in the data. The analysis was conducted using not only weekly, but also monthly data (on a higher aggregation level). For the detection, the correlograms and periodograms of each time series was used. An automated function in R was also used to prove the presence of seasonality.

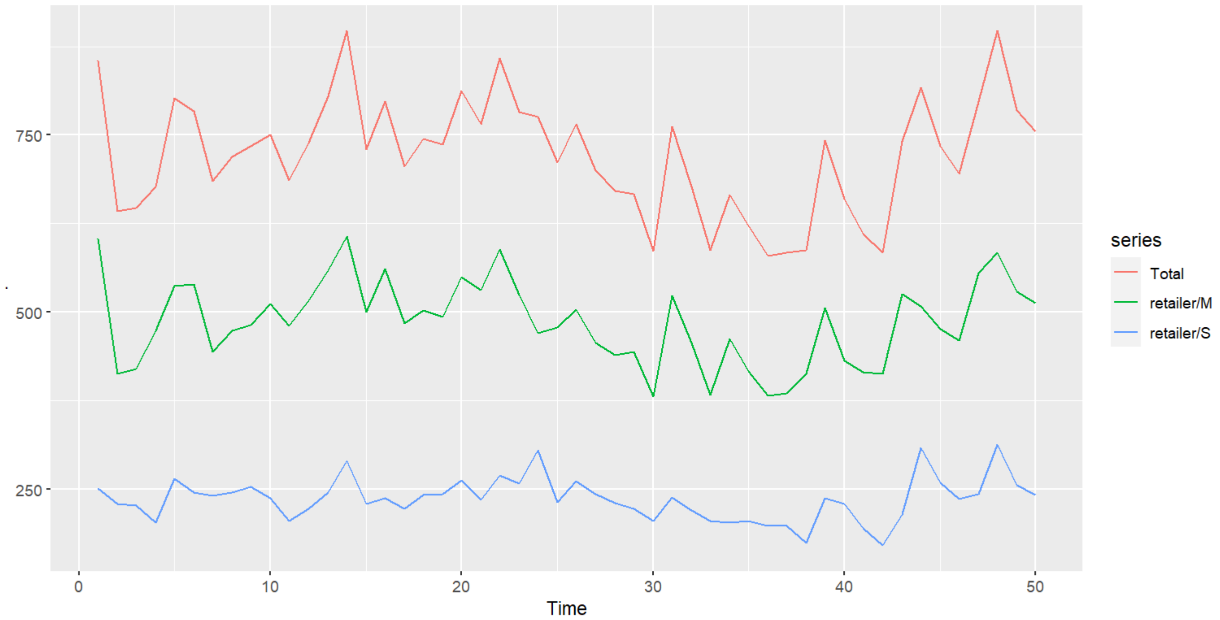


Picture 1. Aggregated sales of coffee makers in Germany for the retailers M and S and the approximate timelines of promotional events and store closures.

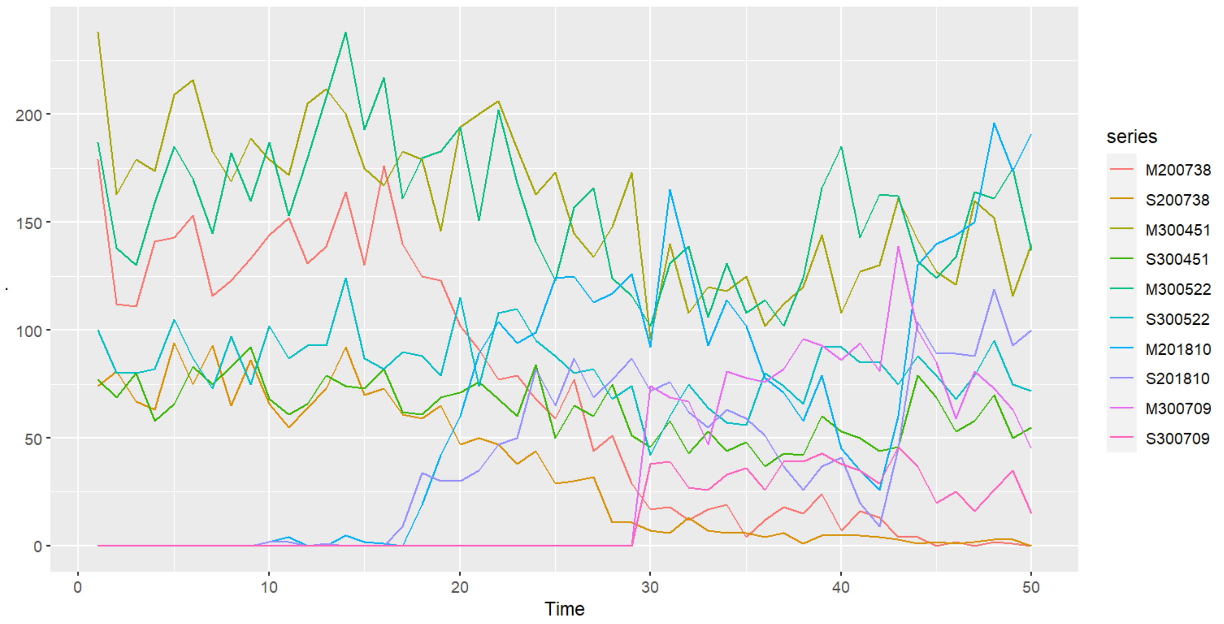
The aim of this paper is to predict base sales filtering out the external events, that is why we simply exclude the periods of sales which were influenced by the events named above. As we do not see any cyclic/seasonal patterns, we simply replace the numbers of weeks with the periods 1 – 50. The eliminated weeks are 48-2020–24-2021, 47-2021–51-2021 and 14-2022–17-2022. This is how the data looks like after cleaning, the trend has become more visible.



Picture 1. Total sales of coffee makers and sales of each SKU for both retailers.



Picture 1. Total sales of coffee makers and total sales for both retailers.



Picture 1. Sales of each SKU for both retailers.

Forecast calculation

As soon as hierarchical forecasts will be calculated with the package “hts” for hierarchical forecasting, there is a huge restriction of the package: namely, it is impossible to derive model parameters from the functions in this package. The forecast functions show only the forecasts, however, the model is chosen automatically using “ets()” or “auto.arima()” functions. Therefore, it was decided to use the automated function to derive model parameters of base forecasts to check the model, which was fitted. Then the automatic hierarchical forecasting was done. The time-series were split into a training sample and a test sample, where the test sample was equal to chosen horizon.

Base forecasts

First, the time-series data was slightly transformed: a constant 0.00001 was added and a Box-Cox transformation with the parameter $\lambda = 0$ was applied (Shumway & Stoffer, 2010).

$$y_t = \begin{cases} (x_t^{\lambda} - 1)/\lambda, & \lambda \neq 0 \\ log x_t, & \lambda \neq 0 \end{cases}$$

In our case, this transformation ensures non-negative forecast values. The forecast values also must be integer, but we remove this constraint for the sake of simplicity by just rounding the forecast afterwards. The method chosen for deriving base forecasts and implementing it is ARIMA without drift. This decision was made for several reasons: the residuals were proved with the graphs and Ljung-Box test to check whether the residuals are autocorrelated or not and whether they look like a white noise or not. All the p-values were large enough not to reject the null-hypothesis: so there was no autocorrelation in the residuals. Moreover, the overall behavior of the forecasting function was estimated with graphs: the ARIMA models without drift caught the main patterns and trends better than, for example, ETS models. To cut it short, he details on coefficients of each base model can be seen in the code.

Hierarchical forecasts

After that, the hierarchical forecasting was performed on the training set. The evaluation included the methods reviewed above:

1. Bottom-up approach (bu)
2. Top-down approach
    a. Average historical proportions (tdgsa)
    b. Proportions of the historical averages (tdgsf)
    c. Forecast proportions (tdfp)
3. Middle-out (combined) approach (mo)
4. Optimal reconciliation approaches
    a. OLS estimator (comb-ols)
    b. WLS estimator (comb-wls)
    c. WLS estimator using structural scaling (comb-nseries)
    d. Minimal trace: sample estimate of the residual covariance matrix (comb-mint-sam)
    e. Minimal trace: shrinkage estimate of the residual covariance matrix (comb-mint-shr)

Prediction performance assessment

As soon as we have grouped data, both possible variants of hierarchies were evaluated. We refer to the hierarchies Total-Retailer-SKU as RS and Total-SKU-Retailer as SR. The Box-Cox parameter was set to 0. The calculation was performed for a short and long horizon: 3 and 8 weeks. Then the errors RMSE and MAPE were calculated.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \times 100\%$$

Discussion of the results

The results of the calculated errors are presented in the tables below. The mean of the errors are deliberately not given because the errors on different hierarchy levels as well as overall dynamics of different error measures are analyzed to choose an appropriate model.

As for the horizon 3 weeks, the forecasting using classification SR and RS performed both well, but SR achieved on average a little bit better results for all levels of the hierarchy. Top-down models using historical proportions have not captured the changes in the demand and cannibalization: they predicted the ending products too high and the new products too low. The bottom-up approach performing good in some scientific papers has not the best results in our case. The optimal reconciliation approaches, a middle-out approach and a top-down approach with forecast proportions overperform traditional methods. Minimum trace methods produce robust superior results. The middle-out approach has an average performance. The decision for the best method for the horizon 3 weeks is the minimal trace: sample estimate of the residual covariance matrix (comb-mint-shr).

Let us look at the prediction performance of the models at the horizon of 8 weeks. Here the hierarchy RS has better results in terms of forecast. The bottom-up approach has the worst performance for the 0 and 1 level of the hierarchy RS. Again, optimal reconciliation approaches overperform other methods, both minimal trace methods show the best results. Top-down models using historical proportions improved the prediction performance on larger horizons. The middle-out approach has also an average performance on larger horizons. The shrinkage method as well as sample covariance method have also on average the best and quite robust results. The decision for the best method for the horizon 3 weeks is again the minimal trace: sample estimate of the residual covariance matrix (comb-mint-shr). As soon as hierarchy RS is better for forecasting at the horizon 8 and similar to SR for the shorter horizon, we choose the hierarchy RS.

	Hierarchy	Method	Total	M	S	M200738	M201810	M300451	M300522	M300709	S200738	S201810	S300451	S300522	S300709
1	RS	bu	119,85	75,85	45,00	1,01	71,11	15,46	18,82	23,69	1,49	21,88	8,53	10,76	12,41
2	RS	tdgsa	94,96	53,59	41,93	73,60	123,80	31,97	15,90	27,66	36,20	66,18	10,71	11,46	13,86
3	RS	tdgsf	94,96	53,38	42,11	76,16	125,02	32,71	16,02	29,99	37,53	66,92	10,85	11,50	14,80
4	RS	tdfp	94,96	51,76	43,33	0,99	67,45	18,74	11,58	23,79	1,48	21,31	8,27	10,31	12,33
5	RS	mo	103,86	58,09	46,02	1,00	69,24	17,85	13,27	22,45	1,49	22,50	8,28	10,62	12,53
6	RS	comb-ols	99,10	56,86	42,49	4,34	68,31	16,80	14,83	22,82	1,01	21,47	8,23	10,30	11,99
7	RS	comb-wls	101,98	57,13	45,03	0,60	61,49	15,58	17,87	23,54	1,44	22,05	8,43	10,62	12,44
8	RS	comb-nseries	105,67	63,20	42,90	2,64	69,34	16,13	16,19	22,87	0,91	21,62	8,22	10,29	12,04
9	RS	comb-mint-sam	104,05	65,93	40,08	9,72	61,69	14,94	22,32	27,10	2,97	13,54	9,22	10,98	10,58
10	RS	comb-mint-shr	99,79	55,19	44,88	0,69	60,72	15,48	18,33	23,70	1,14	22,09	8,41	10,27	12,47

Picture 1. RMSE for the hierarchy RS, horizon 3 weeks.

	Hierarchy	Method	Total	200738	201810	300451	300522	300709	M200738	S200738	M300451	S300451	M300522	S300522	M201810	S201810	M300709	S300709
1	SR	bu	119,85	2,42	23,09	25,04	91,78	21,40	1,01	1,49	15,46	8,53	18,82	10,76	71,11	21,88	23,69	12,41
2	SR	tdgsa	94,96	109,79	41,61	22,29	189,72	40,59	73,60	36,20	31,97	10,71	15,90	11,46	123,80	66,18	27,66	13,86
3	SR	tdgsf	94,96	113,67	42,50	22,56	191,69	43,93	76,16	37,53	32,71	10,85	16,02	11,50	125,02	66,92	29,99	14,80
4	SR	tdfp	94,96	2,31	22,97	28,21	65,28	20,21	1,02	1,36	15,22	8,61	20,86	11,47	55,94	14,67	22,02	12,68
5	SR	mo	82,32	2,28	23,55	25,09	60,02	21,40	1,02	1,35	15,99	8,50	18,85	10,77	52,93	13,57	23,69	12,41
6	SR	comb-ols	94,71	2,42	23,05	24,54	70,71	19,46	1,34	1,20	15,58	8,24	18,64	10,51	59,86	15,53	22,80	12,49
7	SR	comb-wls	90,07	2,45	23,17	25,30	66,02	21,12	1,15	1,47	15,61	8,51	19,04	10,79	45,50	21,27	23,40	12,44
8	SR	comb-nseries	98,62	1,73	23,16	24,37	75,04	20,47	0,85	1,04	15,63	8,32	18,54	10,49	62,29	16,08	23,38	12,23
9	SR	comb-mint-sam	79,48	3,40	22,73	23,01	45,94	17,91	3,29	0,90	15,11	8,69	18,29	9,80	31,06	16,25	18,03	15,33
10	SR	comb-mint-shr	85,06	1,54	23,28	23,88	64,36	20,97	0,80	1,09	15,91	8,43	18,53	10,17	43,46	21,72	23,08	12,70

Picture 1. RMSE for the hierarchy SR, horizon 3 weeks.

	Hierarchy	Method	Total	M	S	M200738	M201810	M300451	M300522	M300709	S200738	S201810	S300451	S300522	S300709
1	RS	bu	10,85	11,29	11,06	703316,92	34,74	9,55	10,62	39,84	5125983,14	17,32	12,58	9,66	33,14
2	RS	tdgsa	8,45	7,86	9,47	248665587,05	65,92	22,33	8,62	39,20	127246081,08	62,35	18,32	14,46	36,97
3	RS	tdgsf	8,45	7,81	9,57	257179438,84	66,58	22,95	8,52	43,60	131662448,07	63,08	18,50	14,50	42,07
4	RS	tdfp	8,45	7,30	10,62	665420,08	33,94	11,28	4,93	41,66	5132005,61	16,90	12,43	9,50	32,89
5	RS	mo	9,99	8,85	12,13	653012,14	35,06	11,60	5,40	39,26	5036306,31	18,30	11,50	8,62	32,78
6	RS	comb-ols	9,01	8,18	10,55	18963338,95	34,11	10,86	7,33	39,98	4229392,34	17,04	12,29	9,42	31,11
7	RS	comb-wls	9,36	8,26	11,44	2006106,33	32,20	9,79	9,88	39,84	4370752,82	17,72	12,28	9,28	32,86
8	RS	comb-nseries	9,38	8,58	10,91	12749913,24	34,39	10,43	8,46	39,71	2934795,34	17,22	12,05	9,25	30,43
9	RS	comb-mint-sam	10,28	10,82	9,01	35820395,68	32,42	9,94	11,10	46,38	5887471,58	8,53	10,65	7,35	33,99
10	RS	comb-mint-shr	9,41	8,31	11,48	3632683,54	31,89	9,87	10,13	39,93	2843392,76	17,68	12,15	8,84	33,46

Picture 1. MAPE for the hierarchy RS, horizon 3 weeks.

	Hierarchy	Method	Total	200738	201810	300451	300522	300709	M200738	S200738	M300451	S300451	M300522	S300522	M201810	S201810	M300709	S300709
1	SR	bu	10,85	2914644,76	10,43	9,97	28,46	21,57	703316,92	5125983,14	9,55	12,58	10,62	9,66	34,74	17,32	39,84	33,14
2	SR	tdgsa	8,45	187955236,54	19,56	7,72	64,69	39,17	248665587,05	127246081,08	22,33	18,32	8,62	14,46	65,92	62,35	39,20	36,97
3	SR	tdgsf	8,45	194420325,54	20,12	7,65	65,37	43,72	257179438,84	131662448,07	22,95	18,50	8,52	14,50	66,58	63,08	43,60	42,07
4	SR	tdfp	8,45	2621985,93	9,94	10,44	21,08	21,73	632700,31	4611284,08	9,90	12,06	11,08	8,62	28,72	13,17	36,53	33,08
5	SR	mo	6,55	2688662,03	10,98	9,98	19,18	21,57	648789,00	4728547,90	10,02	13,15	10,63	9,64	27,00	12,61	39,84	33,14
6	SR	comb-ols	8,21	2870412,64	10,13	9,48	22,70	19,79	5081776,44	659059,87	9,75	12,07	10,24	8,77	30,18	13,43	38,51	32,57
7	SR	comb-wls	7,40	1900602,74	10,52	9,95	21,25	21,47	932035,80	4733127,33	9,59	12,64	10,60	9,51	23,64	16,80	39,31	33,05
8	SR	comb-nseries	8,17	184288,97	10,41	9,60	23,89	20,37	2027130,43	2395629,52	9,61	12,55	10,34	9,25	31,13	12,85	39,49	30,83
9	SR	comb-mint-sam	6,22	4292555,14	9,96	9,15	13,94	22,49	12545705,67	3960575,27	10,64	14,77	10,17	9,47	15,71	10,49	28,21	45,11
10	SR	comb-mint-shr	6,98	975694,55	10,52	9,44	20,72	21,84	2256360,40	4207681,34	9,62	12,79	10,31	9,39	22,56	17,20	38,61	33,54

Picture 1. MAPE for the hierarchy SR, horizon 3 weeks.

	Hierarchy	Method	Total	M	S	M200738	M201810	M300451	M300522	M300709	S200738	S201810	S300451	S300522	S300709
1	RS	bu	203,75	122,21	85,86	10,89	127,51	22,14	19,24	27,18	2,65	78,85	16,18	9,63	10,37
2	RS	tdgsa	167,88	107,57	65,42	68,21	108,43	16,19	21,48	62,79	33,79	65,69	11,84	10,08	19,89
3	RS	tdgsf	167,88	107,28	65,69	71,09	109,83	16,24	21,05	65,30	35,28	66,64	11,74	10,01	20,90
4	RS	tdfp	167,88	101,10	71,52	11,49	126,07	18,84	23,97	28,92	3,00	77,62	13,32	14,85	11,03
5	RS	mo	168,42	101,46	71,68	11,48	126,10	18,90	23,86	28,91	3,00	77,63	13,35	14,78	11,02
6	RS	comb-ols	170,26	102,87	72,14	14,64	123,40	19,93	21,09	28,65	5,55	75,93	14,07	11,50	11,27
7	RS	comb-wls	174,40	101,13	77,75	11,23	107,60	21,71	19,60	27,25	3,98	76,43	15,20	11,43	10,48
8	RS	comb-nseries	179,90	108,70	75,86	13,47	124,67	20,53	20,45	28,10	4,75	76,72	14,63	10,94	11,00
9	RS	comb-mint-sam	161,80	100,16	66,67	4,56	111,56	18,04	21,51	27,55	2,57	66,74	17,31	14,51	11,32
10	RS	comb-mint-shr	169,73	98,05	76,25	10,80	105,30	21,05	19,68	27,21	4,04	75,76	15,36	12,06	10,54

Picture 1. RMSE for the hierarchy RS, horizon 8 weeks.

	Hierarchy	Method	Total	200738	201810	300451	300522	300709	M200738	S200738	M300451	S300451	M300522	S300522	M201810	S201810	M300709	S300709
1	SR	bu	203,75	13,28	33,75	23,93	205,31	35,34	10,89	2,65	22,14	16,18	19,24	9,63	127,51	78,85	27,18	10,37
2	SR	tdgsa	167,88	101,99	20,44	27,18	173,43	81,67	68,21	33,79	16,19	11,84	21,48	10,08	108,43	65,69	62,79	19,89
3	SR	tdgsf	167,88	106,35	20,21	26,59	175,79	85,23	71,09	35,28	16,24	11,74	21,05	10,01	109,83	66,64	65,30	20,90
4	SR	tdfp	167,88	15,61	27,56	32,46	196,22	36,36	12,05	3,90	18,79	14,47	23,73	12,54	121,66	75,66	27,77	10,74
5	SR	mo	194,61	14,74	33,31	24,34	198,50	35,34	11,43	3,64	21,87	16,07	19,44	9,77	123,11	76,46	27,18	10,37
6	SR	comb-ols	174,68	18,92	29,93	27,20	196,02	36,45	13,50	5,84	20,57	14,40	20,46	11,31	122,80	74,38	27,57	11,21
7	SR	comb-wls	187,62	14,71	32,38	25,58	195,04	35,42	11,83	3,15	21,20	15,97	20,21	10,01	117,20	78,73	27,22	10,40
8	SR	comb-nseries	188,65	16,00	31,88	25,40	199,73	35,81	12,12	4,32	21,36	15,31	19,79	10,38	124,66	76,19	27,35	10,71
9	SR	comb-mint-sam	177,33	14,29	34,69	23,48	188,63	38,66	13,87	1,14	22,37	16,91	18,45	10,39	116,89	72,56	28,73	12,12
10	SR	comb-mint-shr	185,64	14,92	32,49	25,83	194,50	35,51	12,16	3,02	21,18	16,11	20,18	10,29	116,86	78,51	27,24	10,48

Picture 1. RMSE for the hierarchy SR, horizon 8 weeks.

	Hierarchy	Method	Total	M	S	M200738	M201810	M300451	M300522	M300709	S200738	S201810	S300451	S300522	S300709
1	RS	bu	24,68	21,87	29,63	39451310,28	75,22	12,24	12,12	27,78	5379153,91	82,72	18,77	11,55	40,36
2	RS	tdgsa	19,84	19,02	20,75	261821114,38	63,70	10,21	10,69	66,96	##########	66,48	13,84	8,67	56,21
3	RS	tdgsf	19,84	18,96	20,87	272606735,94	64,85	10,27	10,64	70,78	##########	67,66	13,87	8,60	61,08
4	RS	tdfp	19,84	17,68	23,46	41986414,86	74,16	11,49	14,09	32,56	5851094,35	81,21	14,85	17,87	46,09
5	RS	mo	19,92	17,76	23,54	41947747,76	74,18	11,50	14,03	32,48	5845705,75	81,23	14,89	17,78	46,02
6	RS	comb-ols	20,17	18,03	23,73	57085337,19	72,50	11,61	12,51	31,89	9252091,61	79,17	15,60	13,78	47,42
7	RS	comb-wls	20,74	17,68	26,17	41113399,87	61,77	12,05	12,18	28,02	7147335,07	79,76	17,18	13,70	41,82
8	RS	comb-nserie	21,48	19,20	25,34	51624455,55	73,31	11,66	12,29	30,52	8282133,84	80,15	16,12	13,20	45,72
9	RS	comb-mint-s	19,01	17,50	21,31	9558448,23	65,73	11,08	12,50	30,42	5438946,72	67,79	20,45	17,34	47,73
10	RS	comb-mint-s	20,10	17,05	25,52	39482748,37	60,35	11,75	12,20	28,19	7323117,67	78,93	17,42	14,38	42,44

Picture 1. MAPE for the hierarchy RS, horizon 8 weeks.

	Hierarchy	Method	Total	200738	201810	300451	300522	300709	M200738	S200738	M300451	S300451	M300522	S300522	M201810	S201810	M300709	S300709
1	SR	bu	24,68	8398413,85	13,72	9,24	78,47	27,01	39451310,28	5379153,91	12,24	18,77	12,12	11,55	75,22	82,72	27,78	40,36
2	SR	tdgsa	19,84	65920898,54	8,77	8,97	64,96	64,92	261821114,38	##########	10,21	13,84	10,69	8,67	63,70	66,48	66,96	56,21
3	SR	tdgsf	19,84	68646684,04	8,73	8,90	66,12	68,92	272606735,94	##########	10,27	13,87	10,64	8,60	64,85	67,66	70,78	61,08
4	SR	tdfp	19,84	11836888,55	11,82	12,19	75,72	31,10	49805849,11	7581516,47	11,52	15,98	13,62	14,95	72,59	79,78	30,68	43,80
5	SR	mo	23,53	11289442,11	13,52	9,35	76,81	27,01	47447050,61	7230876,16	12,12	18,60	12,16	11,75	73,80	80,70	27,78	40,36
6	SR	comb-ols	20,78	13616326,83	12,51	10,18	75,20	30,93	53196060,47	##########	11,65	15,76	12,34	13,65	72,60	78,30	29,63	46,76
7	SR	comb-wls	22,58	10297042,84	13,15	9,64	75,08	27,52	46162794,11	6406054,64	11,80	18,43	12,30	12,09	69,92	82,60	28,15	40,71
8	SR	comb-nseries	22,68	11385067,94	13,04	9,56	76,66	28,83	46944767,18	8365787,68	11,89	17,20	12,22	12,56	73,79	80,24	28,65	43,35
9	SR	comb-mint-sam	21,17	9828013,05	14,33	9,11	72,24	36,50	53911932,43	3119862,87	12,51	19,75	11,93	12,69	70,07	75,11	34,73	51,74
10	SR	comb-mint-shr	22,32	10444841,12	13,27	9,69	74,89	28,28	47422593,74	6263535,86	11,85	18,55	12,29	12,48	69,80	82,32	28,53	41,68

Picture 1. MAPE for the hierarchy SR, horizon 8 weeks.

Conclusion

To sum up, the state-of-art methods for hierarchical forecasting such as optimal reconciliation approaches seem to outperform the other simple forecasting techniques. Top-down approach with proportions based on history seems to be not sensitive to trend changes in practice. The decision for any of the models is based on many factors: for example, objectives of forecasting or complexity, including computational complexity.

References

Abolghasemi, M., Hyndman, R.J., Spiliotis, E., Bergmeir, C. (2022). Model selection in reconciling hierarchical time series. Machine Learning (2022), 111, 739–789; URL: https://doi.org/10.1007/s10994-021-06126-z

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

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. URL: https://doi.org/10.1016/j.csda.2011.03.006

Hyndman, R.J., & Athanasopoulos, G. (2018). Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. URL: otexts.com/fpp2

Oliveira, J. M., & Ramos, P. (2019). Assessing the performance of hierarchical forecasting methods on the retail sector. Entropy, 21(4), 436. URL: https://doi.org/10.3390/e21040436

RDocumentation. URL: https://www.rdocumentation.org/

Schäfer, J.; Strimmer, K. A. (2005). Shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Statistical Applications in Genetics and Molecular Biology (2005), 4, 151–163. URL: https://doi.org/10.2202/1544-6115.1175

Shumway, R. H. Stoffer, D. S. Time Series Analysis and Its Applications. (2010). Springer, 4 Edition. URL: https://doi.org/10.1007/978-3-319-52452-8

Villegas, M. A., Pedregal, D. J. (2018). Supply chain decision support systems based on a novel hierarchical forecasting approach. Decision Support Systems, 114 (2018), 29-36. URL: https://doi.org/10.1016/j.dss.2018.08.003

Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J. (2019). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. Journal of the American Statistical Association, 114(526), 804–819. URL: https://doi.org/10.1080/01621459.2018.1448825