Optimal combination predictions for hierarchical time series

David Rivas Rodrigo Serrano Diego Vivallo

Federico Santa María Technical University
Department of Mathematics

September 27, 2023
Modeling Laboratory I [MAT282]

Index

Introduction

- In this work, two time series methods will be proposed to predict sales of the Walmart company.

- On the commonality of the hierarchical structure for a time series (For example, a tomato's sauce).

Objectives

- With the models proposed, make a prediction using the training data to compare it with the validation data.
- Make a comparison of both methods and see which is closer to the validation data.
- Specify an adjustment of the selected method in comparison to the validation data.
- Make a prediction for future events.

Mathematical tools

ARIMA

It is an integrated autoregressive moving average model. In In particular, a statistical model that uses variances and regressions of statistical data in order to find patterns for a prediction towards the future. future estimates are explained by past data and not by independent variables.

A seasonal ARIMA model is formed by including additional seasonal terms in the ARIMA models.

$$\text{ARIMA} \quad (p, d, q) \quad (P, D, Q)_m$$

where (p, d, q) is the non-seasonal part of the model and $(P, D, Q)_m$ is the seasonal part, with m the seasonal period (for example, the number of observations per year). We use upper case for the seasonal parts of the model and lower case for the non-seasonal parts.

Mathematical tools

Hierarchical time series

In many applications, there are multiple time series that are organized hierarchically. cally and can be aggregated at several different levels of groups based on products, geography or other characteristics.

Notation

- $\boldsymbol{AF}$ **:** F series at level 2 within A series at level 1.
- Observations are recorded in times $t = 1, 2, ..., n$ and we are interested in forecasting each series at each level in times $t = n + 1, n + 2, ..., n + h$.
- $\boldsymbol{X}$ **:** Generic series within the hierarchy.
- $\boldsymbol{Y_{X,t}}$ **:** Value of series X at time t.
- $\boldsymbol{Y_t}$ **:** Sum of all series at time t.
- $\boldsymbol{m_i}$ **:** The total number of series at level i.
- $\boldsymbol{m}$ **:** The total number of series in the hierarchy.
- $\boldsymbol{Y_{i,t}}$ **:** Vector of all observations at level i and time t.
- $\boldsymbol{S}$ **:** Summing matrix used to aggregate the lowest level series.

Example

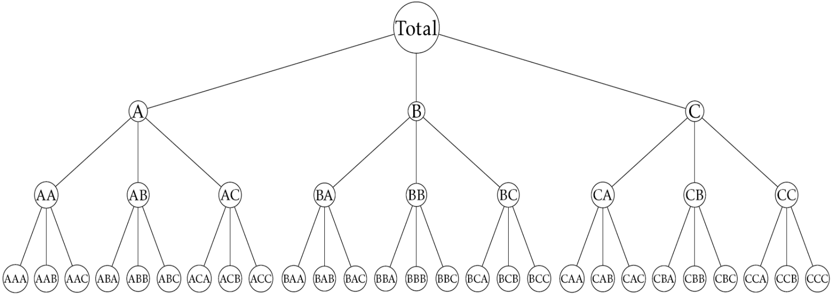


Figure: A 3-level hierarchical tree diagram

$$Y_t = \sum_i Y_{i,t}, \quad Y_{i,t} = \sum_j Y_{ij,t}, \quad Y_{ij,t} = \sum_k Y_{ijk,t}, \quad Y_{ijk,t} = \sum_\ell Y_{ijk\ell,t}$$

Equations

$$
\begin{aligned}
\boldsymbol{Y}_t &= [Y_t, Y_{1,t}, \ldots, Y_{K,t}]^T \\
\boldsymbol{Y}_t &= \boldsymbol{S}\boldsymbol{Y}_{K,t} \qquad (1) \\
\boldsymbol{S} &= \begin{bmatrix}
111111111111111111111111111 \\
111111111000000000000000000 \\
000000000111111111000000000 \\
\vdots \\
000000000000000000000000111 \\
100000000000000000000000000 \\
010000000000000000000000000 \\
\vdots \\
000000000000000000000000001
\end{bmatrix}
\end{aligned}
$$

General hierarchical forecasting

- $\hat{Y}_n(h)$: Base total forecast with h steps in advance.
- $\tilde{\boldsymbol{Y}}_n(h)$: Revised hierarchical forecasts.

$$\tilde{\boldsymbol{Y}}_n(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{Y}}_n(h) \tag{2}$$

- $\beta_n(h) = \mathrm{E}\left[\boldsymbol{Y}_{K,n+h} \mid \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n\right]$: The unknown mean of the bottom level K.
- The unbiasedness of the revised forecast will hold provided $\boldsymbol{SPS} = \boldsymbol{S}$.
- Let the variance of the base forecasts, $\hat{Y}_n(h)$, be given by $\boldsymbol{\Sigma}_h$. Then the variance of the revised forecasts is given by:

$$\mathrm{Var}[\tilde{\boldsymbol{Y}}_n(h)] = \boldsymbol{S}\boldsymbol{P}\boldsymbol{\Sigma}_h\boldsymbol{P}^T\boldsymbol{S}^T \tag{3}$$

Optimal forecasts using regression

We can write the base forecasts as

$$\hat{\boldsymbol{Y}}_n(h) = \boldsymbol{S}\beta_n(h) + \varepsilon_h \tag{4}$$

- $\beta_n(h) = \mathrm{E}\left[\boldsymbol{Y}_{K,n+h} \mid \boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n\right]$: The unknown mean of the bottom level K.
- ε_h : Has zero mean and covariance matrix $\mathrm{Var}(\epsilon_h) = \Sigma_h$.

If Σ_h was known...

$$\mathrm{Var}[\tilde{\boldsymbol{Y}}_n(h)] = \boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{\Sigma}_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}^T \tag{5}$$

Optimal forecasts using regression

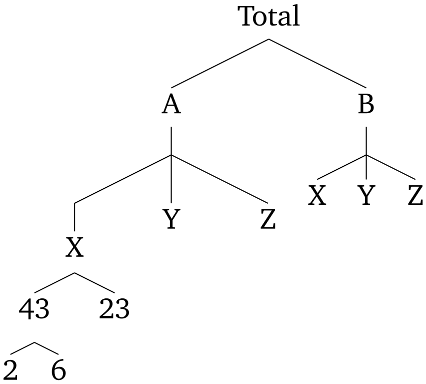
Theorem

Let $\boldsymbol{Y} = \boldsymbol{S}\boldsymbol{\beta}_h + \boldsymbol{\varepsilon}$ with $\mathrm{Var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}_h = \boldsymbol{S}\boldsymbol{\Omega}_h\boldsymbol{S}^T$ and $\boldsymbol{S}$ a "summing" matrix. Then the generalized least squares estimate of $\boldsymbol{\beta}$ obtained using the Moore-Penrose generalized inverse is independent of $\boldsymbol{\Omega}_h$:

$$\hat{\boldsymbol{\beta}}_h = \left(\boldsymbol{S}^T\boldsymbol{\Sigma}_h^{\dagger}\boldsymbol{S}\right)^{-1}\boldsymbol{S}^T\boldsymbol{\Sigma}_h^{\dagger}\boldsymbol{Y} = \left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}\boldsymbol{S}^T\boldsymbol{Y} \qquad (6)$$

with variance matrix $\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\Omega}_h$. Moreover, this is the minimum variance linear unbiased estimate.

Data



Implemetation

- The implementation of the models was in python
- for arima, the statsforecast library was used
- for optimal combination the hierarchicalforecast library was used

Data analysis and procedure

- Sales for the years 2015 and 2016 and 722 business days of sales
- 4118 series in level of categories
- 1977 series with more than 20% missing days, remaining 2141 series
- Series with missing data are filled with 0
- We wanted to work with 20% of the series that sold the most
- For time we work with the total and store A and B

Model execution time

When executing the SARIMA and optimal combination code, a time of answer of 6638,39 seconds which is approximately 110 minutes. where a great volume of time was allocated to the execution of the SARIMA method

SARIMA Results and S matrix

For the total series, the best model was ARIMA(1, 0, 2)(0, 1, 0), for the store series A was ARIMA(1, 0, 3)(0, 1, 0) and for store B was ARIMA(0, 1, 1)(0, 1, 0), with m = 362, which experimentally was the one that gave the best results.

Summing Matrix

$$\boldsymbol{S} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

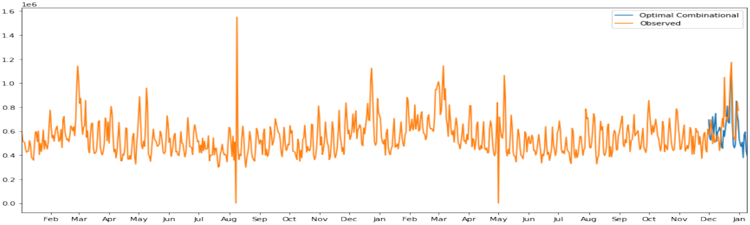


Figure: Sales of total series 2015-2016

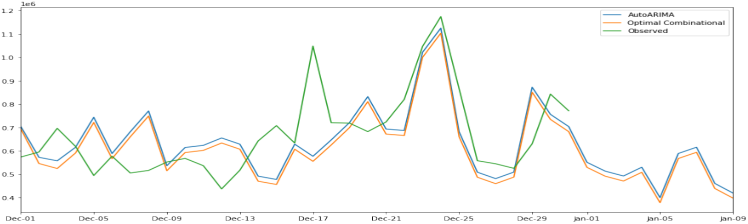


Figure: Sales of total series December 2016 and prediction

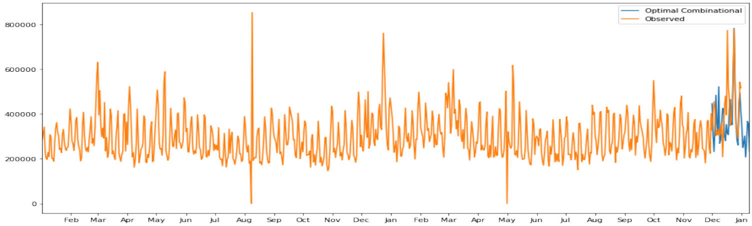


Figure: Sales of store A 2015-2016

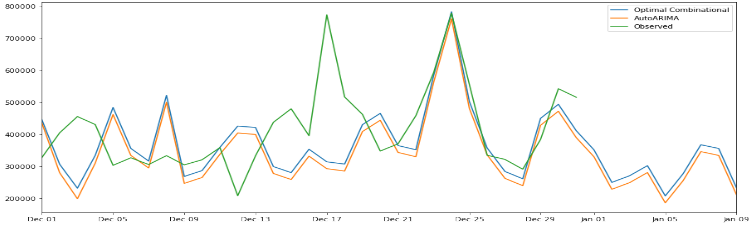


Figure: Sales of store A December 2016 and prediction

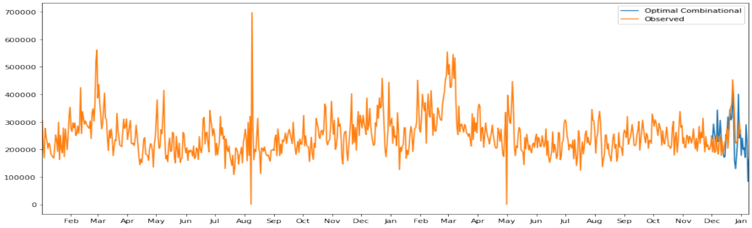


Figure: Sales of store B 2015-2016

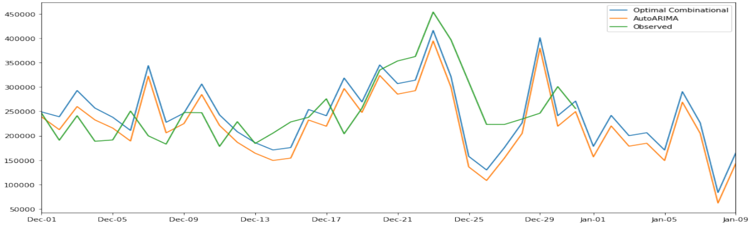


Figure: Sales of store B December 2016 and prediction

Error analisys

For the analysis of the error, Root Mean Squared Error (RMSE) was used, which is calculated using the formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{7}$$

Which when applied to the model, yielded the following results:

RMSE	total	A	B
SARIMA	147690,35	62762	77333,25
optimal combination	150526,9	84347,41	55747,8

Hours spent at work

- Our **initial** proposal vs what happened.

- Why the delay...?

Problems during the course

- **Records not taken or no sales.**
  The solution? create a filter for the data

- **The low similarity obtained between the data.**
  The solution? Using a method enters the process called SARIMA

References

Rob J.Hyndman, Roman A.Ahmed, George Athanasopoulos and Han LinShang.
≪*Optimal combination forecasts for hierarchical time series*≫.
Computational Statistics Data Analysis, Volume 55, Theme 9, Pages 2579-2589, 2001.

Bernstein, D.S..
≪*Matrix Mathematics: Theory, Facts, and Formulas with Applications to Linear Systems Theory*≫.
Princeton University Press, Princeton, Nueva Jersey., 2005.

Tian, Y., Wiens, D.P..
≪*On equality and proportionality of ordinary least squares, weighted least squares and best linear unbiased estimators in the general linear model*≫.
Statistics and Probability Letters 76 (12), Pages 1265-1722, 2006.

Thank you.
Questions?