Optimal combination forecasts for hierarchical time series

Rob J. Hyndman

Roman A. Ahmed

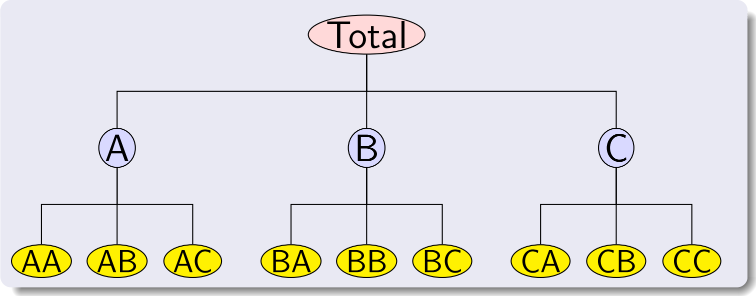
Department of Econometrics and Business Statistics

MONASH University

Outline

1. Review of hierarchical forecasting
2. A new approach
3. Simulation study
4. Summary

Introduction



Examples

- Manufacturing product hierarchies
- Pharmaceutical sales
- Net labour turnover

Hierarchical time series

- A **hierarchical time series** is a collection of several time series that are linked together in a hierarchical structure.
- Forecasts should be aggregate consistent, unbiased, minimum variance.
- Existing methods:
  - Bottom-up
  - Top-down
  - Middle-out
- How to compute forecast intervals?
- Most research is concerned about relative performance of existing methods.

Top-down method

Advantages

- Works well in presence of low counts.
- Single forecasting model easy to build
- Provides reliable forecasts for aggregate levels.

Disadvantages

- Loss of information, especially individual series dynamics.
- Distribution of forecasts to lower levels can be difficult
- No prediction intervals

Bottom-up method

Advantages

- No loss of information.
- Better captures dynamics of individual series.

Disadvantages

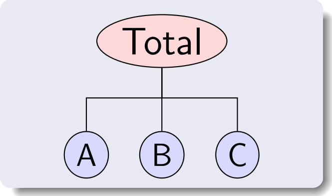
- Laborious because of number of series to be forecast.
- Constructing forecasting models is harder because of noisy data at bottom level.
- No prediction intervals

A new approach

We aim to develop a new statistical methodology for forecasting hierarchical time series which:

1. provides point forecasts that are consistent across the hierarchy;
2. allows for correlations and interaction between series at each level;
3. provides estimates of forecast uncertainty which are consistent across the hierarchy;
4. allows for ad hoc adjustments and inclusion of covariates at any level.

Notation



K : number of levels in the hierarchy (excl. Total).

Y_t : observed aggregate of all series at time t.

$Y_{X,t}$: observation on series X at time t.

$\mathbf{Y}_{i,t}$: vector of all series at level i in time t.

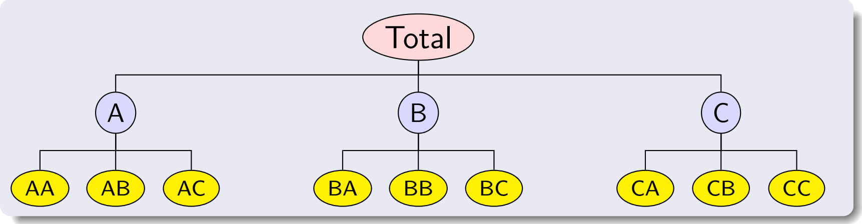
$\mathbf{Y}_t = [Y_t, \mathbf{Y}_{1,t}, \dots, \mathbf{Y}_{K,t}]'$

$$\mathbf{Y}_t = [Y_t, Y_{A,t}, Y_{B,t}, Y_{C,t}]' =$$

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{S} \begin{pmatrix} Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \end{pmatrix} \underbrace{\begin{pmatrix} Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \end{pmatrix}}_{\mathbf{Y}_{K,t}}$$

$$\boxed{\mathbf{Y}_t = S\mathbf{Y}_{K,t}}$$

Notation



$$\mathbf{Y}_t = \begin{pmatrix} Y_t \\ Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \\ Y_{AA,t} \\ Y_{AB,t} \\ Y_{AC,t} \\ Y_{BA,t} \\ Y_{BB,t} \\ Y_{BC,t} \\ Y_{CA,t} \\ Y_{CB,t} \\ Y_{CC,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}}_{S} \underbrace{\begin{pmatrix} Y_{AA,t} \\ Y_{AB,t} \\ Y_{AC,t} \\ Y_{BA,t} \\ Y_{BB,t} \\ Y_{BC,t} \\ Y_{CA,t} \\ Y_{CB,t} \\ Y_{CC,t} \end{pmatrix}}_{\mathbf{Y}_{K,t}}$$

Forecasts

$$\mathbf{Y}_t = S\mathbf{Y}_{K,t}$$

Let $\hat{\mathbf{Y}}_n(h)$ be vector of independent (base) forecasts for horizon h, stacked in same order as $\mathbf{Y}_t$.

Write

$$\hat{\mathbf{Y}}_n(h) = S\beta_n(h) + \varepsilon_h$$

where

- $\beta_n(h) = E[\mathbf{Y}_{K,n+h} \mid \mathbf{Y}_1, \dots, \mathbf{Y}_n]$ is unknown mean of bottom level K
- ε_h has zero mean and covariance matrix Σ_h.

Idea: Estimate $\beta_n(h)$ using regression.

Optimal combination forecasts

$$\hat{\beta}_n(h) = (S'\Sigma_h^\dagger S)^{-1} S'\Sigma_h^\dagger \hat{\mathbf{Y}}_n(h)$$

where $\Sigma_h^\dagger$ is generalized inverse of Σ_h.

$$\tilde{\mathbf{Y}}_n(h) = S\hat{\beta}_n(h) = S(S'\Sigma_h^\dagger S)^{-1} S'\Sigma_h^\dagger \hat{\mathbf{Y}}_n(h)$$

Revised forecasts Base forecasts

- Revised forecasts unbiased if $\hat{\mathbf{Y}}_n(h)$ unbiased:

$$\mathrm{E}[\tilde{\mathbf{Y}}_n(h)] = S(S'\Sigma_h^\dagger S)^{-1} S'\Sigma_h^\dagger S\beta_n(h) = S\beta_n(h).$$

- Minimum variance by construction:

$$\mathrm{Var}[\tilde{\mathbf{Y}}_n(h)] = S(S'\Sigma_h^\dagger S)^{-1} S'$$

Optimal combination forecasts

$$\tilde{\mathbf{Y}}_n(h) = S(S'\Sigma_h^\dagger S)^{-1}S'\Sigma_h^\dagger \hat{\mathbf{Y}}_n(h)$$

Revised forecasts Base forecasts

- **Problem:** Don't know Σ_h and hard to estimate.
- **Solution:** Assume $\varepsilon_h \approx S\varepsilon_{K,h}$ where $\varepsilon_{K,h}$ is the forecast error at bottom level.
  Then $\Sigma_h \approx S\Omega_h S'$ where $\Omega_h = \text{Var}(\varepsilon_{K,h})$.
  If Moore-Penrose generalized inverse used, then

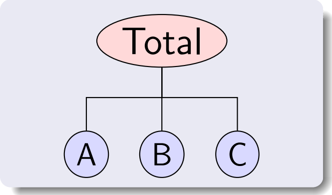
$$(S'\Sigma^\dagger S)^{-1}S'\Sigma^\dagger = (S'S)^{-1}S'.$$

$$\tilde{\mathbf{Y}}_n(h) = S(S'S)^{-1}S'\hat{\mathbf{Y}}_n(h)$$

Optimal combination forecasts

$$\tilde{\mathbf{Y}}_n(h) = S(S'S)^{-1}S'\hat{\mathbf{Y}}_n(h)$$

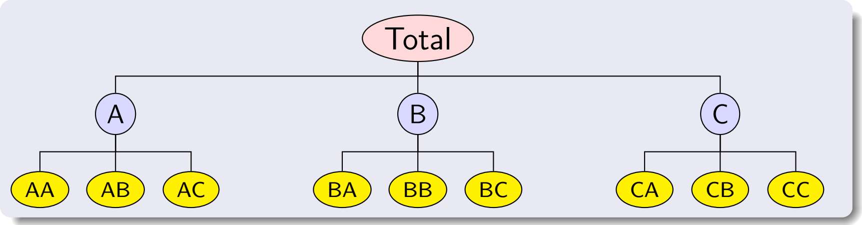
- OLS solution.
- Optimal weighted average of base forecasts.
- Computational difficulties in big hierarchies due to size of S matrix.
- Optimal weights are $S(S'S)^{-1}S'$
- Weights are independent of the data!



Weights: $S(S'S)^{-1}S' =$

$$\begin{bmatrix} 0.75 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.75 & -0.25 & -0.25 \\ 0.25 & -0.25 & 0.75 & -0.25 \\ 0.25 & -0.25 & -0.25 & 0.75 \end{bmatrix}$$

Optimal combination forecasts



Weights: $S(S'S)^{-1}S' =$

$$\begin{bmatrix}
0.69 & 0.23 & 0.23 & 0.23 & 0.08 & 0.08 & 0.08 & 0.08 & 0.08 & 0.08 & 0.08 & 0.08 & 0.08 \\
0.23 & 0.58 & -0.17 & -0.17 & 0.19 & 0.19 & 0.19 & -0.06 & -0.06 & -0.06 & -0.06 & -0.06 & -0.06 \\
0.23 & -0.17 & 0.58 & -0.17 & -0.06 & -0.06 & -0.06 & 0.19 & 0.19 & 0.19 & -0.06 & -0.06 & -0.06 \\
0.23 & -0.17 & -0.17 & 0.58 & -0.06 & -0.06 & -0.06 & -0.06 & -0.06 & -0.06 & 0.19 & 0.19 & 0.19 \\
0.08 & 0.19 & -0.06 & -0.06 & 0.73 & -0.27 & -0.27 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 \\
0.08 & 0.19 & -0.06 & -0.06 & -0.27 & 0.73 & -0.27 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 \\
0.08 & 0.19 & -0.06 & -0.06 & -0.27 & -0.27 & 0.73 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 \\
0.08 & -0.06 & 0.19 & -0.06 & -0.02 & -0.02 & -0.02 & 0.73 & -0.27 & -0.27 & -0.02 & -0.02 & -0.02 \\
0.08 & -0.06 & 0.19 & -0.06 & -0.02 & -0.02 & -0.02 & -0.27 & 0.73 & -0.27 & -0.02 & -0.02 & -0.02 \\
0.08 & -0.06 & 0.19 & -0.06 & -0.02 & -0.02 & -0.02 & -0.27 & -0.27 & 0.73 & -0.02 & -0.02 & -0.02 \\
0.08 & -0.06 & -0.06 & 0.19 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & 0.73 & -0.27 & -0.27 \\
0.08 & -0.06 & -0.06 & 0.19 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.27 & 0.73 & -0.27 \\
0.08 & -0.06 & -0.06 & 0.19 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.27 & -0.27 & 0.73
\end{bmatrix}$$

Variances

$$\tilde{\mathbf{Y}}_n(h) = S(S'S)^{-1}S'\hat{\mathbf{Y}}_n(h)$$

$$\text{Var}[\tilde{\mathbf{Y}}_n(h)] = S\Omega_h S' = \Sigma_h.$$

- Revised forecasts have the same variance matrix as original forecasts.
- Revised forecasts have the same variance matrix as bottom-up forecasts.
- The assumption $\Sigma_h \approx S\Omega_h S'$ results in loss of minimum variance property.
- Need to estimate Ω_h to produce prediction intervals.

Adjustments and covariates

- Adjustments can be made to base forecasts at any level. These will automatically propagate through to other levels of the hierarchy.
- Covariates can be included in the base forecasts at any level.
- Point forecasts are always consistent across the hierarchy.
- Relationships between series at any level are handled automatically. However, some simplifying assumptions have been made.
- Estimates of forecast uncertainty are consistent across the hierarchy, but we need an estimate of Ω_h.

Simulation design

- Hierarchy of four levels with each disaggregation into four series.
- Bottom level: 64 series.
- Total series: 85.
- Series length: $n = 100$ (90 used for fitting, 10 for evaluation)
- Data generated for each series by ARIMA(p, d, q) process with random order and coefficients.
- Aggregate and three components generated. Fourth series obtained by subtraction.
- 600 hierarchies
- Base forecasts using Hyndman et al. (IJF 2002)

Simulation results

Average MAE by level

Level	Top-down	Bottom-up	**Optimal combination**
0 (Top)	2.5	21.0	2.8
1	20.4	12.5	7.4
2	27.4	9.2	8.1
3 (Bottom)	27.4	8.2	8.2
Average	19.4	12.7	**6.6**

Top-down uses disaggregation by historical proportions

Summary

- Very flexible method. Can work with *any* base forecasts.
- Method outperforms bottom-up and top-down, especially for middle levels.
- Conceptually easy to implement: OLS on base forecasts.
- Weights independent of the data. So calculate once and reapply when new data available.

Remaining problems

- How to explain variance results compared to simulation?
- How to compute Ω_h in order to obtain prediction intervals?
- Computationally it may be difficult for large hierarchies. (But solution coming up!)

For more information

- Paper and computer code will appear at
  http://www.robhyndman.info

- Paper will also appear on RePEc and in NEP-FOR report
  http://ideas.repec.org/n/nep-for/