Likelihood-based inference in temporal hierarchies

Author **Jan Moller**[1]

Discussion by **Dario Azzimonti**[2]

[1]Denmark Technical University, Denmark.

[2]Dalle Molle Institute for Artificial Intelligence (IDSIA), Switzerland.

IIF Workshop on Forecast Reconciliation

8th September 2023 - Prato, Italy

The problem

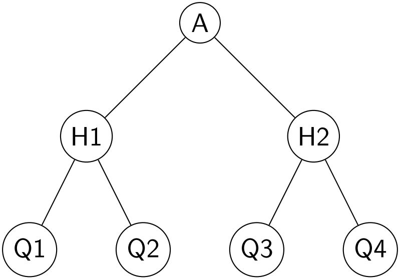
Forecast reconciliation in **temporal hierarchies**:

- improves forecast accuracy (Nystrup et al. 2020, Hollyman et al. 2021)
- impacted by covariance between/within levels. (Pritularga et al. 2021)

Example:

Temporal hierarchy with aggregation at

- **A**nnual;
- **H**alf year;
- **Q**uarterly.



The problem

Given the summing matrix S and base forecasts $\hat{y}$,
the optimal reconciled base point forecasts are

$$\tilde{b} = P\hat{y}, \qquad P = (S^T\Sigma^{-1}S)^{-1}S^T\Sigma^{-1}$$

How to estimate Σ, the covariance between/within aggregations?

Estimators for the covariance

- Diagonal approximation (Athanasopoulos et al., 2017)
- Shrinkage estimates (Wickramasuriya et al. 2019; Nystrup et al., 2020)
- ??

Proposed method: introduce structure

Note: data shared between levels through summation matrix

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

Idea: include (part of) the structure of S in our covariance estimation!

$$\boldsymbol{S}_{\mathrm{HQ}} = \begin{bmatrix} \beta_{11}^{\mathrm{HQ}} & \beta_{12}^{\mathrm{HQ}} & 0 & 0 \\ 0 & 0 & \beta_{23}^{\mathrm{HQ}} & \beta_{24}^{\mathrm{HQ}} \end{bmatrix},$$
$$\boldsymbol{S}_{\mathrm{AQ}} = \begin{bmatrix} \beta_{11}^{\mathrm{AQ}} & \beta_{12}^{\mathrm{AQ}} & \beta_{13}^{\mathrm{AQ}} & \beta_{14}^{\mathrm{AQ}} \end{bmatrix},$$
$$\boldsymbol{S}_{\mathrm{AH}} = \begin{bmatrix} \beta_{11}^{\mathrm{AH}} & \beta_{12}^{\mathrm{AH}} \end{bmatrix},$$

Residual errors structure

Covariances between aggregation levels
model residual error ϵ by including data sharing

$$
\begin{aligned}
\boldsymbol{\epsilon}_1 &= \boldsymbol{u}_1 && ; \quad \boldsymbol{u}_1 \sim N(\mathbf{0}, \boldsymbol{\Sigma}_1), \\
\boldsymbol{\epsilon}_2 &= \boldsymbol{S}_{21}\boldsymbol{u}_1 + \boldsymbol{u}_2 && ; \quad \boldsymbol{u}_2 \sim N(\mathbf{0}, \boldsymbol{\Sigma}_2), \\
\boldsymbol{\epsilon}_3 &= \boldsymbol{S}_{31}\boldsymbol{u}_1 + \boldsymbol{S}_{32}\boldsymbol{u}_2 + \boldsymbol{u}_3 && ; \quad \boldsymbol{u}_3 \sim N(\mathbf{0}, \boldsymbol{\Sigma}_3), \\
&\vdots \\
\boldsymbol{\epsilon}_K &= \sum_{j=1}^{K-1} \boldsymbol{S}_{Kj}\boldsymbol{u}_j + \boldsymbol{u}_K && ; \quad \boldsymbol{u}_K \sim N(\mathbf{0}, \boldsymbol{\Sigma}_K),
\end{aligned}
$$

where $Cov[\boldsymbol{u}_i, \boldsymbol{u}_j] = \mathbf{0}$, $(i \neq j)$.

Covariance matrix within aggregation levels
model the inverse of Σ_i with parametrized Cholesky: $\Sigma_{ii}^{-1} = S_{ii} S_{ii}^T$
(Sparse inverse covariance enforces conditional independece)

Residual errors structure

The overall model for the residuals is

$$\begin{bmatrix} \boldsymbol{\epsilon}_K \\ \boldsymbol{\epsilon}_{K-1} \\ \vdots \\ \boldsymbol{\epsilon}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & S_{K,K-1} & \cdots & S_{K,1} & \\ \mathbf{0} & \mathbf{I} & S_{K,K-1} & \cdots & S_{K,1} \\ & & \ddots & & \\ \mathbf{0} & \cdots & & & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{u}_K \\ \mathbf{u}_{K-1} \\ \vdots \\ \mathbf{u}_1 \end{bmatrix}$$

with $\mathbf{u} \sim N(0, \Sigma^u)$, Σ^u covariance within aggregating level. Therefore

$$\epsilon \sim N(0, S_u \Sigma^u S_u^T)$$

Note:

- S_u: covariances **between** aggregation levels
- Σ^u: covariances **within** aggregation levels

Maximum likelihood inference

$\epsilon \sim N(0, S_u \Sigma^u S_u^T) \Rightarrow (T-1)\mathbf{V} \sim W(\Sigma, T-1)$ (Wishart distribution), where $\mathbf{V}$ is the observed variance-covariance.

We can estimate the parameters in $\Sigma = S_u \Sigma^u S_u^T$ by maximum likelihood.

$$\begin{aligned} l(\boldsymbol{\beta}, \boldsymbol{\alpha}; \boldsymbol{V}) &\propto -\frac{1}{2} Tr(\boldsymbol{\Sigma}^{-1}(T-1)\boldsymbol{V}) - \frac{T-1}{2} \log|\boldsymbol{\Sigma}| \\ &= -\frac{T-1}{2} Tr\, \boldsymbol{\Sigma}^{-1}\boldsymbol{V} + \frac{T-1}{2} \log|\boldsymbol{\Sigma}^{-1}|. \end{aligned}$$

Check out the paper for extensive derivations of the gradients!

Inference, model reduction and shrinkage

The likelihood is optimized (Newton initialized with first order steps)

- **sequentially** over the aggregations (other levels params fixed)
- from the lowest aggregation level to the highest

Between each aggregation level: **Wald test** for model reduction

- to select only significant parameters;
- uses already computed Hessian matrix;
- also pairwise comparisons

Shrinkage of the covariance towards block-diagonal and diagonal is naturally available by adding weights to the likelihood.

"Operational" questions

Why is it interesting to do model reduction?

- fewer parameters might lead to more robust estimates
- however model reduction is done after optimization.
  Why not penalized likelihood with a LASSO-like regularization?

In which scenarios does this method lead to better performances?

- time series with small length?

“Structural” questions

Method requires normal assumption for errors, what about other distributions?

- If residuals are not normal, how robust is the method?
- Is there hope to use another likelihood-based method for other distributions? Count time series?

What about grouped time-series?

References

Hollyman, R., Petropoulos, F., & Tipping, M. E. (2021). Understanding forecast reconciliation. *European Journal of Operational Research*, 294(1), 149-160.

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.

Nystrup, P., Lindström, E., Pinson, P., & Madsen, H. (2020). Temporal hierarchies with autocorrelation for load forecasting. *European Journal of Operational Research*, 280(3), 876-888.

Pritularga, K. F., Svetunkov, I., & Kourentzes, N. (2021). Stochastic coherency in forecast reconciliation. *International Journal of Production Economics*, 240, Article 108221.

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