Hierarchical forecast reconciliation

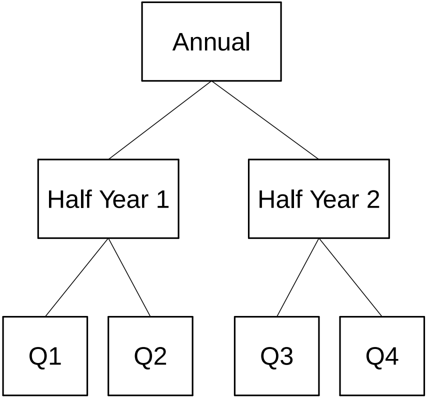
Jan Kloppenborg Møller

DTU-Compute - Technical University of Denmark
jkmo@dtu.dk

DIREC Seminar 2022

Motivation

- Models on different aggregation levels
- Models on each level may not agree
- Reconciliation ensure consistent forecasts
- Reconciliation often improve forecast accuracy on all levels



Reconciliation

The reconciled forecast is calculated by

$$\tilde{\boldsymbol{y}} = (\boldsymbol{S}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{S})^{-1} \boldsymbol{S}^T \boldsymbol{\Sigma}^{\boldsymbol{-1}} \hat{\boldsymbol{y}}$$

where:

- $\tilde{\boldsymbol{y}}$: reconciled forecast
- $\boldsymbol{S}$: the summation matrix
- $\boldsymbol{\Sigma}$: a variance-covariance matrix
- $\hat{\boldsymbol{y}}$: the base forecast

Regression setting

$$\hat{\boldsymbol{y}} = \boldsymbol{S}\tilde{\boldsymbol{y}} + \boldsymbol{\epsilon}; \quad \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$$

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

Technical University of Denmark DTU

Choosing the correct variance

Different options for $\mathbf{\Sigma}$:

- $\mathbf{\Sigma} = \boldsymbol{I}$
- Variance scaling (proportional to volume of level)
- Use observed variance-covariance of base forecast error
- Ignore cross level correlation
- Use shrinkage on the observed variance-covariance (usually preferred). I.e.

$$\hat{\mathbf{\Sigma}}_s = \lambda \hat{\mathbf{\Sigma}} + (1-\lambda)\mathrm{diag}\hat{\mathbf{\Sigma}}$$

Example

Example: Half year levels generated by

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t$$

AR(1) models half-year levels and annual levels, i.e. the models

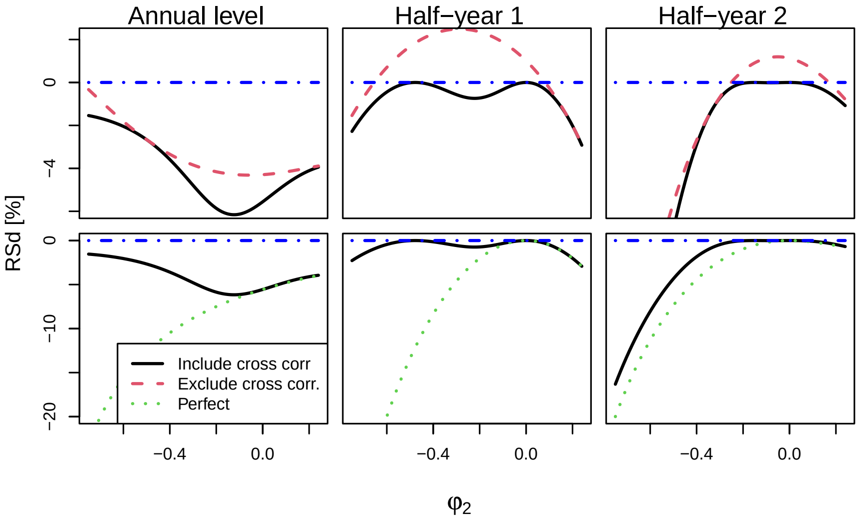
$$\begin{aligned} y_t^{\mathsf{A}} &= \phi_1^{\mathsf{A}} y_t^{\mathsf{A}} + \epsilon_t^{\mathsf{A}} \\ y_t^{\mathsf{H}} &= \phi_1^{\mathsf{H}} y_t^{\mathsf{H}} + \epsilon_t^{\mathsf{H}} \end{aligned}$$

Full setup

$$\boldsymbol{y}_{2t+2|2t} = \begin{bmatrix} y_{2t+2|2t}^{\mathsf{A}} \\ y_{2t+1|2t}^{\mathsf{H}} \\ y_{2t+2|2t}^{\mathsf{H}} \end{bmatrix}; \quad \hat{\boldsymbol{y}}_{2t+2|2t} = \begin{bmatrix} \hat{y}_{2t+2|2t}^{\mathsf{A}} \\ \hat{y}_{2t+1|2t}^{\mathsf{H}} \\ \hat{y}_{2t+2|2t}^{\mathsf{H}} \end{bmatrix}; \quad \boldsymbol{S} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and e.g. $\boldsymbol{\Sigma} = \mathsf{Var}[\boldsymbol{y}_{2t+2} - \hat{\boldsymbol{y}}_{2t+2}]$

Choosing the correct variance



$\phi_1 = 0.75$

(from Møller et al. (2022))

A case study

- District heating from an area of greater Copenhagen.
- State of the art commercial hourly forecast (1-24 hours ahead)
- 2, 3, 4, 6, 8, 12, and 24 hours forecast:
  - Recursive Least Square
  - Forecast of ambient temperature
  - Diurnal variation
  - Auto-regressive parts
- Σ_t (60 by 60) estimated using the full variance covariance with (optimal) shrinkage, and different updating schemes.

Some results

	2017–2019			
	Base RMSE	Expanding Window	Rolling Window	Exponential Smoothing
Daily	0.5960	-23.75	-22.49	**-23.93**
Twelve-hourly	0.3516	-24.08	-22.83	**-24.2**
Eight-hourly	0.3538	-43.51	-42.72	**-43.69**
Six-hourly	0.2876	-44.64	-43.75	**-44.76**
Four-hourly	0.1765	-36.06	-35.19	**-36.37**
Three-hourly	0.1334	-33.03	-32.05	**-33.26**
Two-hourly	0.0884	-30.09	-29.07	**-30.36**
Hourly	0.0383	-14.75	-13.46	**-15.07**

Adapted from Bergsteinsson et al. (2021).

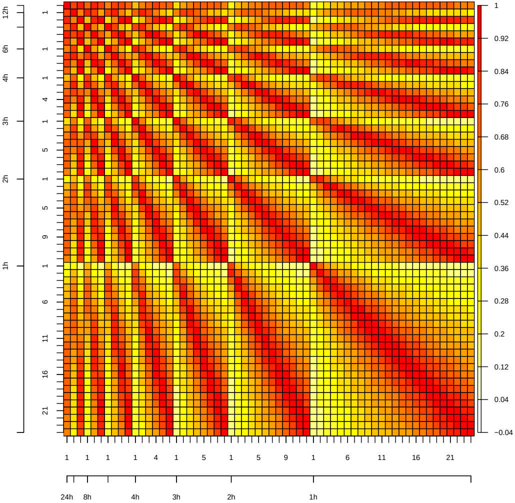
Modelling of variance-covariance matrix

$\boldsymbol{\Sigma}$ (some challenges):

- High-dimensional
- High correlations

Some suggestions

- Parametric models for the correlation
- Use statistical methods for reduction



Graphics: Møller et al. (2022)

A parametric model

Starting from the bottom level define (Møller et al., 2022)

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

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

References

[1] Bergsteinsson H.G., Møller J.K., Nystrup P., Pálsson Ó.P., Guericke D., and Madsen H. (2021). Heat load forecasting using adaptive temporal hierarchies. *Applied Energy*, **292**.

[2] Møller J.K., Nystrup P., and Madsen H. (2022). Likelihood-based inference in temporal hierarchies. *Under review*.

Technical University of Denmark DTU

Thank You!

Questions?