02427 lecture 8: Hierarchical forecast reconciliation

Jan Kloppenborg Møller

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

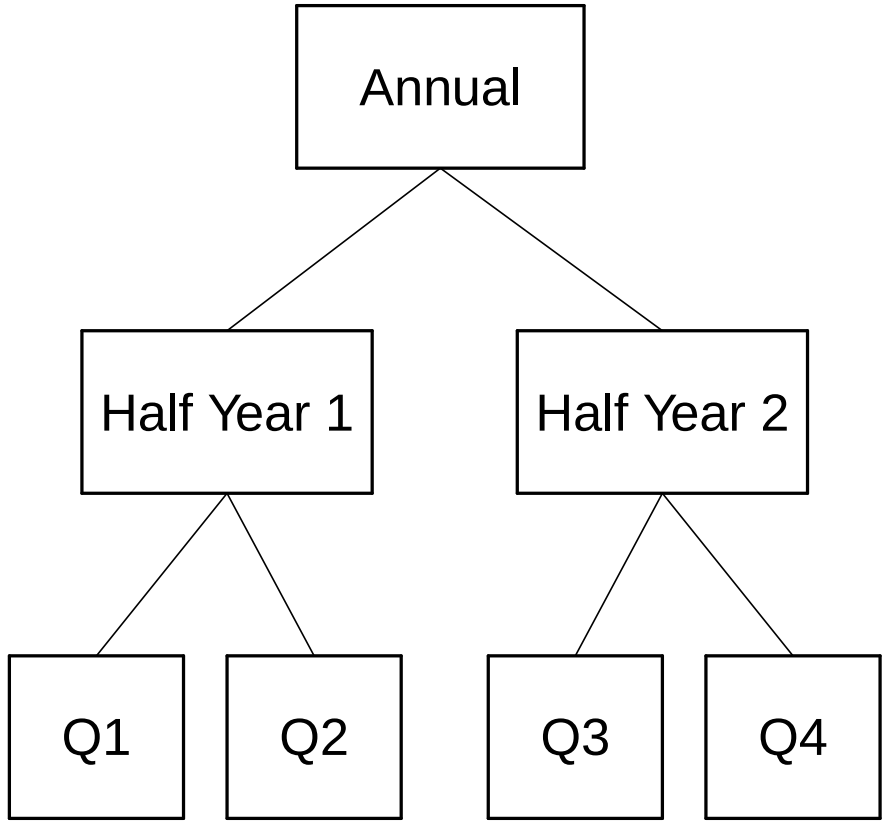
Advanced time series analysis – 2025

Outline

Outline

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



Mathematical formulation

The reconcilied forcast, $\tilde{\boldsymbol{y}}$, is found as the solutuon to the optimization problem

$$\text{minimize} \quad (\hat{\boldsymbol{y}} - \tilde{\boldsymbol{y}})^T \boldsymbol{\Sigma}^{-1} (\hat{\boldsymbol{y}} - \tilde{\boldsymbol{y}})$$

subject to the cohenrency constraints $\tilde{\boldsymbol{y}} = \boldsymbol{SP}\tilde{\boldsymbol{y}}$.

In a Gaussian setting the model can be formulated as

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

with an equivalent solution. Note that the dimension of $\tilde{\boldsymbol{y}}$ is different in the two settings.

Reconciliation

The reconciled forecast is calculated by

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

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

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

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

Choosing the correct variance

The "only" unknown part is the variance-covariance matrix $\mathbf{\Sigma}$, the optimal choice is (Wickramasuriya et al. 2019)

$$\mathbf{\Sigma} = V[\boldsymbol{y} - \hat{\boldsymbol{y}}].$$

Hence a natural choice is to use the observed counturpart, i.e.

$$\hat{\mathbf{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t)(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t)^T$$

In realistic settings this matrix is usually high dimensional and ill-conditioned, and hence other versions are usually used.

In sample properties

With $\hat{\boldsymbol{\Sigma}}$ as above, then:

For any column (and row) selection, $I \subseteq \{1, 2, ..., n\}$, of $\hat{\boldsymbol{Y}}$ (and $\boldsymbol{S}$) then

$$\boldsymbol{S}_I \left(\boldsymbol{Y} - \tilde{\boldsymbol{Y}}\right)^\top \left(\boldsymbol{Y} - \tilde{\boldsymbol{Y}}\right) \boldsymbol{S}_I^\top \leq \left(\boldsymbol{Y}\boldsymbol{S}_I^\top - \hat{\boldsymbol{Y}}_I\right)^\top \left(\boldsymbol{Y}\boldsymbol{S}_I^\top - \hat{\boldsymbol{Y}}_I\right), \quad (1)$$

and hence for the in-sample variance of the residuals are reduced on any level.

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.

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

Furter the optimal λ is $\hat{\lambda} = \frac{\sum_{i\neq j} \hat{var}[r_{ij}]}{r_{ij}^2}$, with r_{ij} elements of the correlation matrix.

Choosing the correct variance – simple examples

Some very simple examples are

$$\begin{aligned}
\boldsymbol{\Sigma}_{struc} &= \text{diag}(4, 2, 2, 1, 1, 1, 1) \\
\boldsymbol{\Sigma}_{svar} &= \text{diag}(\sigma_A^2, \sigma_H^2, \sigma_H^2, \sigma_Q^2, \sigma_Q^2, \sigma_Q^2, \sigma_Q^2) \\
\boldsymbol{\Sigma}_{hvar} &= \text{diag}(\sigma_A^2, \sigma_{H_1}^2, \sigma_{H_2}^2, \sigma_{Q_1}^2, \sigma_{Q_2}^2, \sigma_{Q_3}^2, \sigma_{Q_4}^2)
\end{aligned}$$

These all ignore cross covariance, and as we will see these might be important to obtain optimal reconciliation.

Choosing the correct variance – block structures

Also block diagonal structures have been suggested

$$\mathbf{\Sigma}_{acov} = \begin{bmatrix} \sigma_A^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_H & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Sigma}_Q \end{bmatrix}$$

which might further include some parametrized structures in the block diagonals. Here cross level covariance are ignores.

In conclusion the shrinkage estimator is usually preferred (Nystrup et al 2019).

Estimation of the precition matrix – dimensionally reduction

First observe that, while, it is usually not reasonable to assume that the variance–covariance matrix is sparse. It is often reasonable to assume that the precision matrix ($\mathbf{\Sigma}^{-1}$) is (e.g. AR(p) models).

Using this observation formulate the precision as

$$\mathbf{\Sigma}^{-1} = \mathbf{\Lambda}^{-1/2}\mathbf{\Theta}\mathbf{\Lambda}^{-1/2}$$

where $\mathbf{\Lambda}$ is a diagonal matrix with diagonal elements equal the marginal variances.

Estimation of the precition matrix – dimensionally reduction

Dimensionality reduction can be obtained by e.g.

- GLASSO, i.e. maximize the penalized log-likelihood: $\log|\boldsymbol{\Theta}| - tr(\boldsymbol{R}\boldsymbol{\Theta}) - \lambda \sum_{i \neq j} |\boldsymbol{\Theta}_{ij}|$
- λ is a regularization parameter and larger λ lead to simpler inverse correlation matrices.

Another option is spectral scaling (Nystrup et al 2021), where $\boldsymbol{\Theta}$ is chosen as

$$\boldsymbol{\Theta} = (\boldsymbol{W}\boldsymbol{A}\boldsymbol{W}^T + \sigma^2\boldsymbol{I})^{-1}$$

where $\boldsymbol{W}$ is the first n_e eigen vectors of the correlation matrix, and $\boldsymbol{A}$ is related to the eigenvalues. $\sigma^2\boldsymbol{I}$ is a regularization/shrinkage.

Example

Example: Assume that 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

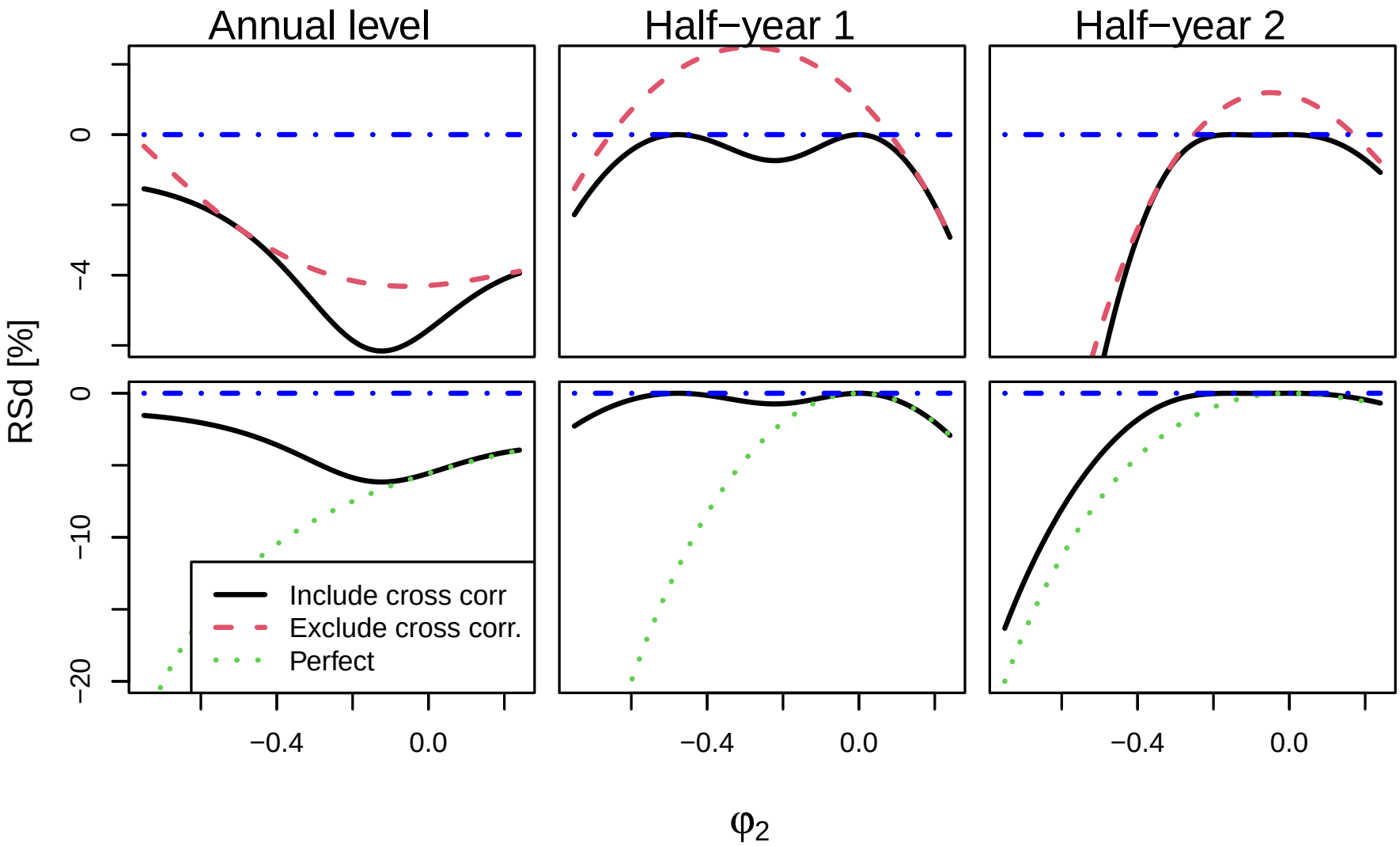
$$\begin{aligned} y_t^{\mathsf{A}} &= \phi_1^{\mathsf{A}} y_{t-1}^{\mathsf{A}} + \epsilon_t^{\mathsf{A}} \\ y_t^{\mathsf{H}} &= \phi_1^{\mathsf{H}} y_{t-1}^{\mathsf{H}} + \epsilon_t^{\mathsf{H}}. \end{aligned}$$

Full setup

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

and e.g. $\boldsymbol{\Sigma} = \mathrm{Var}[\boldsymbol{y}_{2t+2} - \hat{\boldsymbol{y}}_{2t+2}]$ can be calculated explicitly.

Choosing the correct variance-covariance



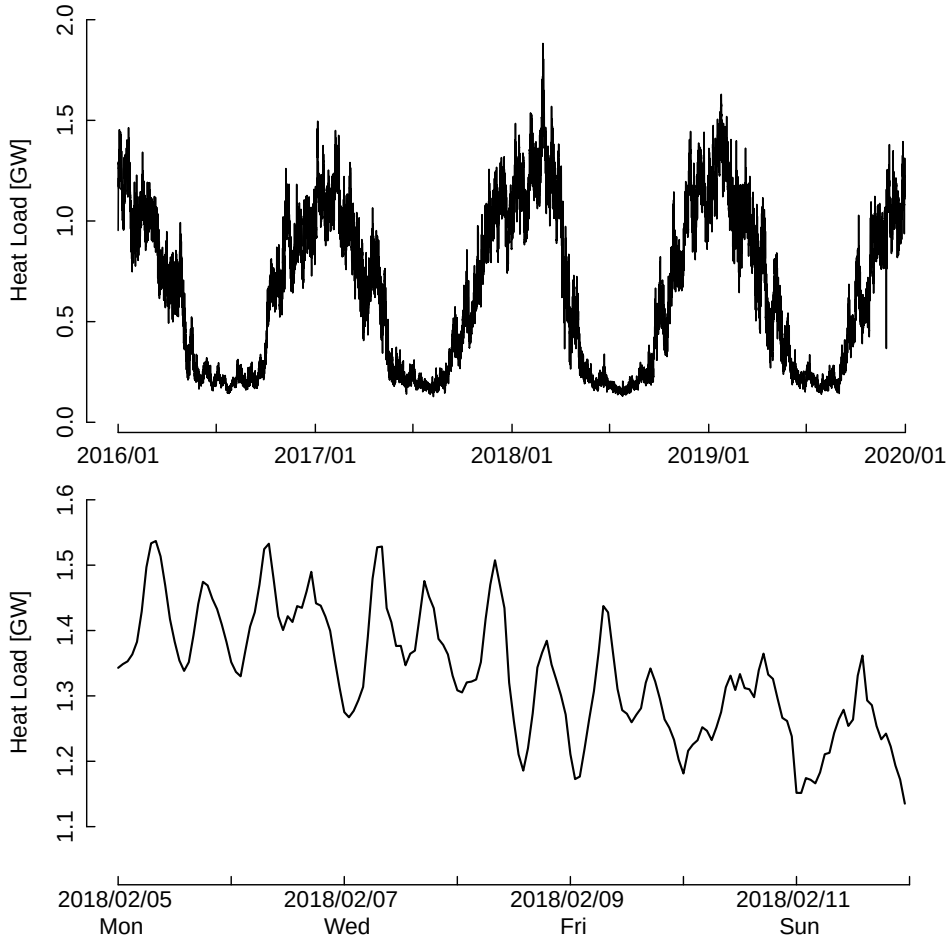
$\phi_1 = 0.75$

(from Møller et al. (2023))

Outline

Data

- District heating from an area of greater Copenhagen
- Clear annual and diurnal variation
- State of the art commercial hourly forecast (1-24 hours ahead)
- 2016 used for initialization

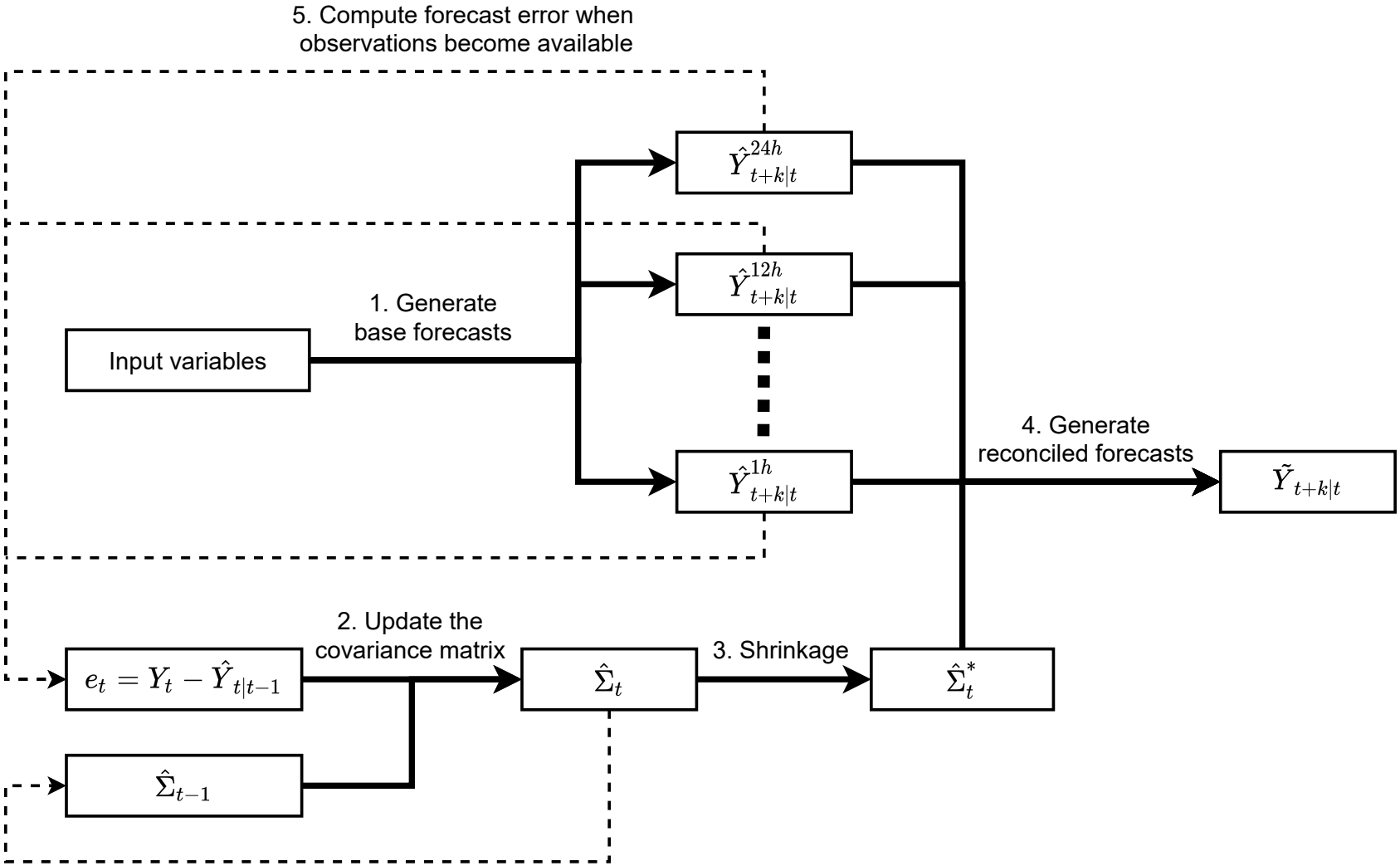


From Bergsteinsson et al. (2021).

A case study

- 1 hour level used for evaluation
- All other levels modeled in the study
- 2, 3, 4, 6, 8, 12, and 24 hours forecast:
  - Recursive Least Square
  - Forecast of ambient temperature
  - Diurnal variation
  - Auto-regressive parts
- $\boldsymbol{\Sigma}_t$ (60 by 60) estimated using the full variance covariance with shrinkage, and recursive updating.

Work flow of modeling



From Bergsteinsson et al. (2021).

Some results (RRMSE)

	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).

Outline

Motivation / Aim

The starting point

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

Comments and aim

- Observation appear only through the variance-covariance matrix $\boldsymbol{\Sigma}$
- The estimation of $\boldsymbol{\Sigma}$ include a large number of parameters
- Formulate a parameterized model for obtaining $\boldsymbol{\Sigma}$
- Reduce dimension of the parameter space by well-known likelihood techniques

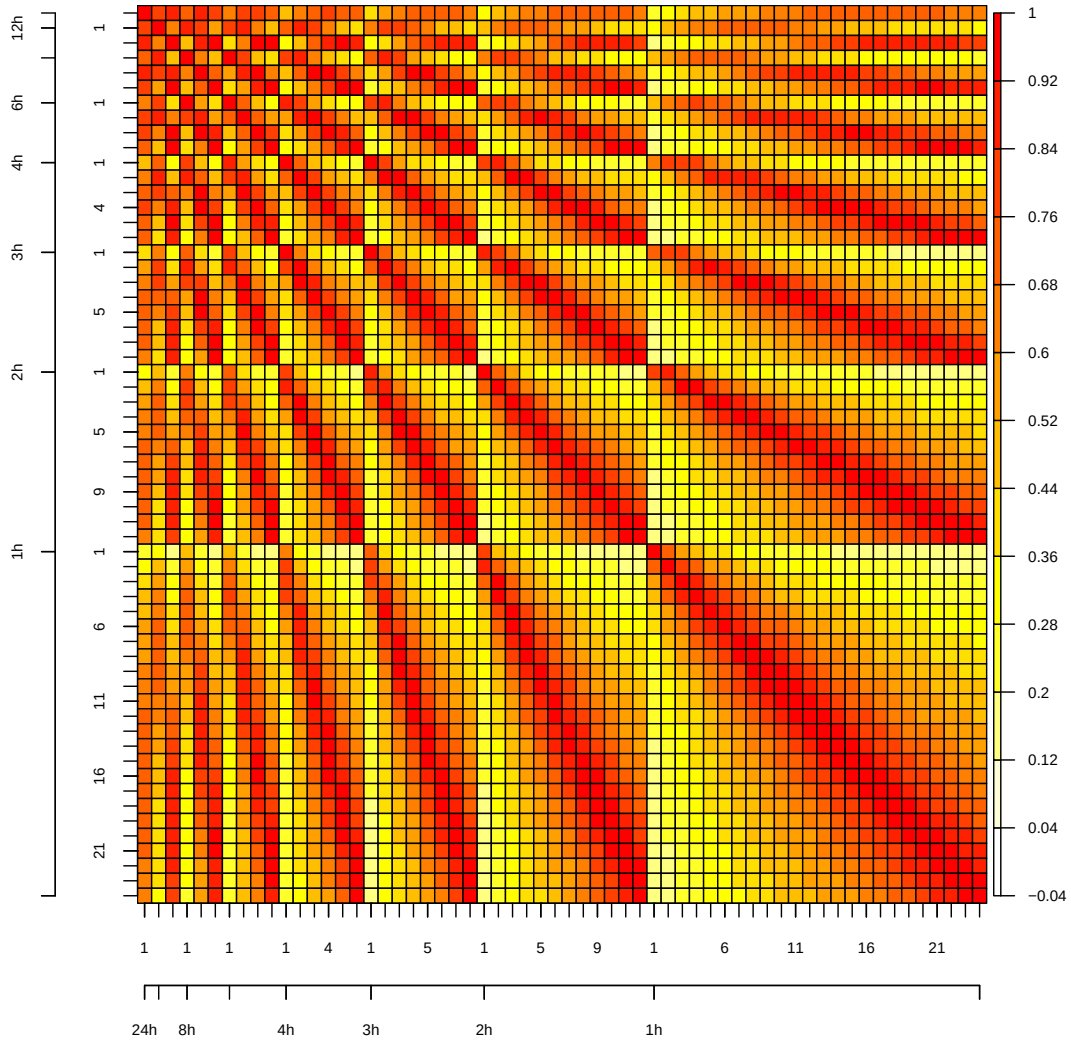
Modeling of variance-covariance matrix

Σ (some challenges):

- High-dimensional
- High correlations

Some suggestions

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



Graphics: Møller et al. (2023)

Technical University of Denmark

A parametric model

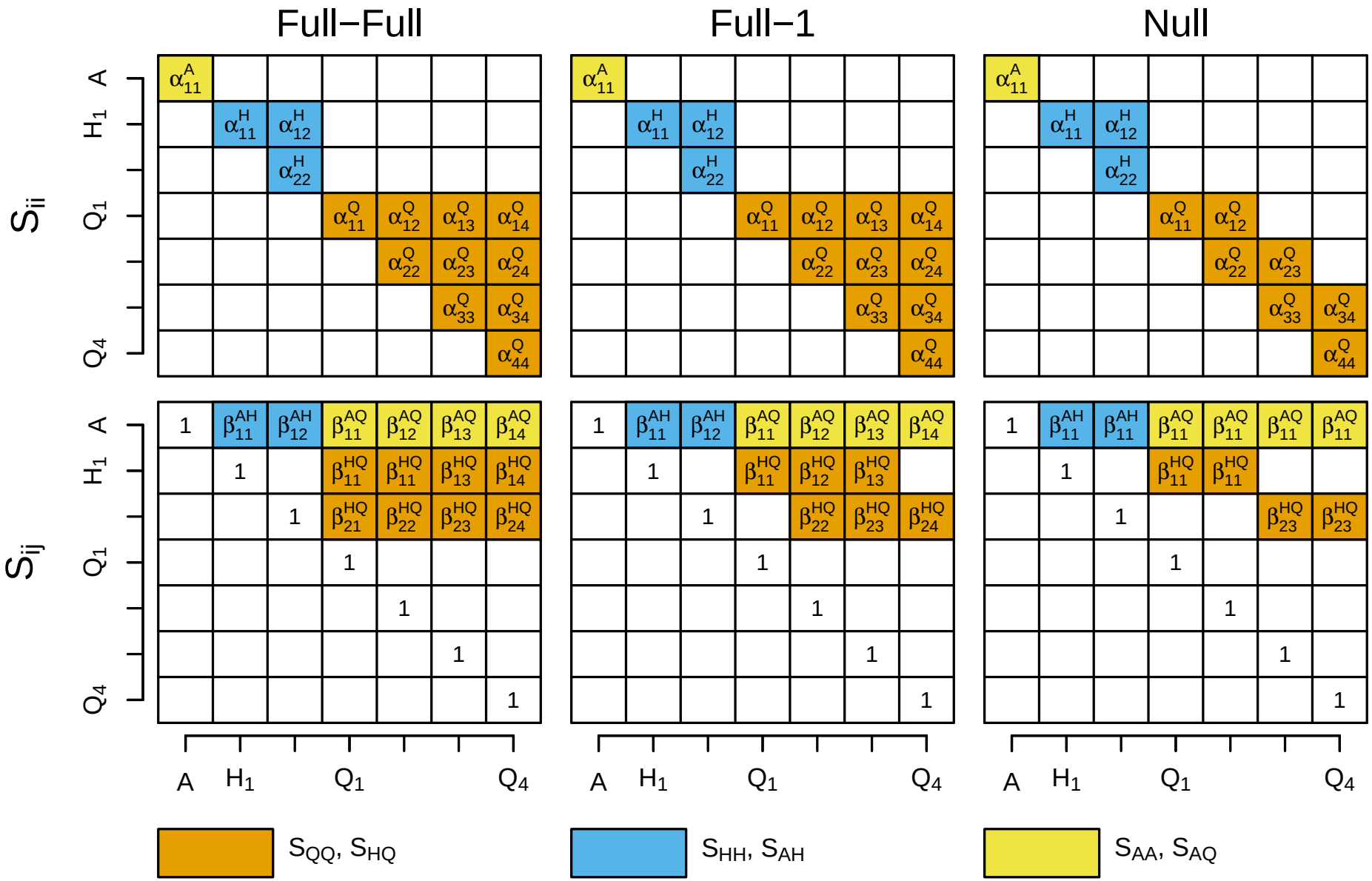
Starting from the bottom level define (Møller et al., 2023)

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

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

$$\boldsymbol{\Sigma}_i^{-1} = \boldsymbol{S}_{ii}\boldsymbol{S}_{ii}^T,$$

with $\boldsymbol{S}_{ii}$ is an upper triangular matrix.


Full−Full
Full−1
Null
S_{ii}
S_{ij}
A
H_1
Q_1
Q_4
α^{A}_{11}
α^{H}_{11} α^{H}_{12} α^{H}_{22}
α^{Q}_{11} α^{Q}_{12} α^{Q}_{13} α^{Q}_{14}
α^{Q}_{22} α^{Q}_{23} α^{Q}_{24}
α^{Q}_{33} α^{Q}_{34}
α^{Q}_{44}
1
β^{AH}_{11} β^{AH}_{12}
β^{AQ}_{11} β^{AQ}_{12} β^{AQ}_{13} β^{AQ}_{14}
β^{HQ}_{11} β^{HQ}_{13} β^{HQ}_{14}
β^{HQ}_{21} β^{HQ}_{22} β^{HQ}_{23} β^{HQ}_{24}
S_{QQ}, S_{HQ}
S_{HH}, S_{AH}
S_{AA}, S_{AQ}

Estimation

Likelihood

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

where $\boldsymbol{V}$ is the observed variance-covariance matrix, $\boldsymbol{\Sigma}^{-1}$ is parameterized through the model and estimation is done

- Sequentially starting from the bottom level
- Using:
  - a robust EM-like algorithm (solving normal equations) and
  - the Newton method (using the Hessian of the likelihood)

diagonal elements of $\boldsymbol{S}_{ii}$ is estimated in the log-domain.

Shrinkage

A simple modification of the likelihood by introducing weights in the following way:

$$l_s(\boldsymbol{\Sigma}; \boldsymbol{V}, \boldsymbol{w}) = l(\boldsymbol{\Sigma}; w_1 \boldsymbol{V} + w_2 \text{blockdiag}\boldsymbol{V} + w_3 \text{diag}\boldsymbol{V}),$$

where $\sum_i w_i = 1$.

Statistical tests

As the method is based on likelihood estimation we have access to

- Wald test for individual parameters or sets of parameters
- Likelihood ratio test for individual parameters or specific hypothesis

In the work we explore

- Wald test for testing if parameters should be zero or equal and confirm using LRT
- Effect of different initial structures

The algorithm, Bottom level

1. Estimation.
   1. Initialise a variance–covariance structure for the bottom level
   2. Iterate between $\boldsymbol{\alpha}_{ij}$ and $\boldsymbol{\alpha}_{ii}$ a fixed number of times
   3. Use the result from 2 as initial value for the Newton method.
   4. Report the parameter estimates, likelihood, and Hessian.
2. Model reduction.
   1. Choose in which order to test parameters for removal
   2. Calculate the test statistics for the increasing set of parameters until significant and confirm by LRT.
   3. Calculate the test statistics for all relevant pairwise comparison and individual parameters that can be set to zero
   4. Reduce model, and confirm by likelihood-ratio test.
   5. Report parameters and structure of $\boldsymbol{S}_{ii}$ and $\boldsymbol{S}_{ij}$.

The algorithm, Top levels

1. Iterate through higher aggregation levels with the lower levels fixed in the same way as the bottom level, now including estimation of β
2. Report the final result.

Case study

- Electricity load in Sweden 2016-2020
- 2016-2019 used for estimating mean value structure
- Linear model including annual and diurnal variation
- Double seasonal AR models of the residuals (daily and and weekly)
- Temporal reconciliation of residuals

Technical University of Denmark DTU

Some results

	SE		SE1		SE2		SE3		SE4	
	df	RRMSE	df	RRMSE	df	RRMSE	df	RRMSE	df	RRMSE
Obs-test	1830	-12.4	1830	-8.5	1830	-9.8	1830	-15.9	1830	-15.3
Obs-train	1830	0.1	1830	4.1	1830	2.1	1830	-2.3	1830	-5.2
Full-Full$_1$	402	-4.8	306	-1.7	413	**-4.1**	413	-5.6	464	-5.5
Full-Full$_2$	402	**-5.3**	306	0.8	413	-3.1	413	-5.8	464	-6.6
Full-2$_1$	219	-3.8	180	**-3.2**	237	-4.0	240	-3.9	264	-4.0
Full-2$_2$	219	-3.8	180	-2.2	237	-3.8	240	-4.9	264	-3.7
Null-Null$_1$	83	-4.3	77	-2.7	81	**-4.1**	85	-4.8	85	-4.6
Null-Null$_2$	83	-3.7	77	-0.7	81	-3.1	85	-4.0	85	-4.1
Shrink	0.015	**-5.3**	0.016	-1.3	0.035	-4.0	0.015	**-7.4**	0.016	**-7.0**
Auto-cov.	69	-2.2	49	-2.9	49	-3.5	71	-1.6	71	-2.2
Model-AR1	60	-2.4	60	-3.0	60	-3.2	60	-1.9	60	-2.2
Diag	60	-3.1	60	-2.4	60	-2.6	60	-2.9	60	-2.6

Outline

A general linear model formulation

Consider the general linear model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}; \quad \boldsymbol{\epsilon} \sim N(0, \boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_r).$$

where $\boldsymbol{y}^T = \left[\boldsymbol{y}_T^T, \boldsymbol{y}_{T-1}^T \cdots, \boldsymbol{y}_1^T\right]^T$.
Question: Can we find a formulation of $\boldsymbol{X}$ such that $\boldsymbol{\beta}$ and

$$\boldsymbol{P} = (\boldsymbol{S}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{\Sigma}^{-1}$$

are quivalent?

Theorem (Equivalence with hierarchical reconciliation)

The model,

$$\boldsymbol{y}_t = \boldsymbol{X}_{\cdot,t}\boldsymbol{\beta} + \boldsymbol{\epsilon}_t; \quad \boldsymbol{\epsilon}_t \sim N(0, \boldsymbol{\Sigma}_r),$$

with $\boldsymbol{X}_{\cdot,t} = \boldsymbol{I} \otimes \hat{\boldsymbol{y}}_t$*, and linear constraints* $vec^{-1}(\boldsymbol{\beta})^T \boldsymbol{S} = \boldsymbol{I}$*, is equivalent to the hierarchical reconciliation, formally*

$$\boldsymbol{P} = vec^{-1}(\hat{\boldsymbol{\beta}})^T,$$

where $\boldsymbol{P}$ *is the usual hierarchical reconciliation with* $\boldsymbol{\Sigma}_h = \frac{1}{T}(\boldsymbol{Y}\boldsymbol{S}^T - \hat{\boldsymbol{Y}})^T(\boldsymbol{Y}\boldsymbol{S}^T - \hat{\boldsymbol{Y}})$*. The constraint problem can further be formulated as*

$$\boldsymbol{y}_t - \hat{\boldsymbol{y}}_{bot,t} = [\boldsymbol{I}_m \otimes (\hat{\boldsymbol{y}}^T_{T,t} - \hat{\boldsymbol{y}}^T_{bot,t}\boldsymbol{S}^T_T)]\boldsymbol{\beta}_T + \boldsymbol{\epsilon}_t; \quad \boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_r).$$

Comments

Here

$$\hat{\boldsymbol{y}} = \begin{bmatrix} \hat{\boldsymbol{y}}_{\text{top},t} \\ \hat{\boldsymbol{y}}_{\text{bot},t} \end{bmatrix}, \quad \boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}_{\text{top}} \\ \boldsymbol{I}_m \end{bmatrix}.$$

The full model can be writte as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}; \quad \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_r)$$

and the MLE of $\boldsymbol{\beta}$ is independent of $\boldsymbol{\Sigma}_r$, and

$$\hat{\boldsymbol{\Sigma}}_{r,\text{REML}} = \frac{1}{T-(n-m)} \sum_{t=1}^{T} \boldsymbol{e}_t \boldsymbol{e}_t^\top,$$

where $\boldsymbol{e}_t = \boldsymbol{y}_t - \boldsymbol{X}_{\cdot,t}\hat{\boldsymbol{\beta}}$.

More comments

Notice that

1. Estimation of the $n \times n$ ($\frac{n(n+1)}{2}$ parameters) variance covariance matrix is replaced by estimation of $m(n-m)$ mean value parameters.
2. The result is independent of the forecast variance $\mathbf{\Sigma}_r$, which may be parameterized by some model (e.g. AR(1) structures)

Shrinkage

Usually a skrinkage estimator for $\mathbf{\Sigma}$ is used, in order to improve the condition number of $\mathbf{\Sigma}$.

- In the regression setting an ill conditioned covarince matrix will show up as multicollinarity.
- Strong multicollinarity is usually handled by regularizations (e.g. Ridge regression) or equivalent by introducing parameter priors.
- Specific choices of priors (depending on the shrinkage) lead to equivalence between the regression setting and the usual (called MinT) approach.

Theorem

The weight matrix $\boldsymbol{P}(\lambda)$ *is equivalent to the linear regression model*

$$\boldsymbol{y}_t - \hat{\boldsymbol{y}}_{B,t} = [\boldsymbol{I}_m \otimes (\hat{\boldsymbol{Y}}_{T,t} - \hat{\boldsymbol{Y}}_{B,t}\boldsymbol{S}_T^T]\boldsymbol{\beta}_T + \boldsymbol{\epsilon}_t; \quad \boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_r),$$

with prior distribution $\boldsymbol{\beta}_T \sim N(\boldsymbol{\beta}_{0,T}, \boldsymbol{\Sigma}_r \otimes \boldsymbol{\Sigma}_{\beta,0})$, *and*

$$\begin{aligned}
\boldsymbol{\beta}_{0,T} =& vec\left[\left(\boldsymbol{\Sigma}_{h,T}^d + \boldsymbol{S}_T\boldsymbol{\Sigma}_{h,B}^d\boldsymbol{S}_T^T\right)^{-1}\boldsymbol{S}_T\boldsymbol{\Sigma}_{h,B}^d\right] \\
\boldsymbol{\Sigma}_{0,\beta} =& \frac{1-\lambda}{\lambda T}(\boldsymbol{\Sigma}_{h,T}^d + \boldsymbol{S}_T\boldsymbol{\Sigma}_{h,B}^d\boldsymbol{S}_T^T)^{-1},
\end{aligned}$$

Residual variance–covariance

To obtain complete equivalence priors for the residual variance $\boldsymbol{\Sigma}_r$ is also needed (related to the Inverse Wishart distribution).

Using the regression formulation (including priors), we arrive at two variance estimators, one with a REML corretion term and one without

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_{r,\text{shrink}} &= \boldsymbol{P}(\lambda)\hat{\boldsymbol{\Sigma}}_h\boldsymbol{P}^\top(\lambda)\\ \hat{\boldsymbol{\Sigma}}_{r,\text{sREML}} &= \frac{T}{T-(n-m)(1-\lambda)}\boldsymbol{P}(\lambda)\hat{\boldsymbol{\Sigma}}_s\boldsymbol{P}^\top(\lambda).\end{aligned}$$

Forecast variance

In general the forecast varaince can be written as

$$\hat{V}[\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_{t+h}] = \hat{V}[\boldsymbol{X}_{\cdot,t+h}\hat{\boldsymbol{\beta}} + \boldsymbol{\epsilon}_{t+h}] = \boldsymbol{X}_{\cdot,t+h}\hat{V}[\hat{\boldsymbol{\beta}}]\boldsymbol{X}_{\cdot,t+h}^{\top} + \hat{\boldsymbol{\Sigma}}_{r\cdot}$$

with

$$\hat{V}[\hat{\boldsymbol{\beta}}_{\text{top}}] = \hat{\boldsymbol{\Sigma}}_r \otimes \left((\boldsymbol{X}_{1,\cdot}^{\top}\boldsymbol{X}_{1,\cdot} + \boldsymbol{\Sigma}_{\beta,0})^{-1}\boldsymbol{X}_{1,\cdot}^{\top}\boldsymbol{X}_{1,\cdot}(\boldsymbol{X}_{1,\cdot}^{\top}\boldsymbol{X}_{1,\cdot} + \boldsymbol{\Sigma}_{\beta,0})^{-1} \right).$$

The parameters (weights) unceartinty and the parameters correlation can be assessed from the using the above, this may be used for testing and further unceartainty may propagate through to predictions variances.

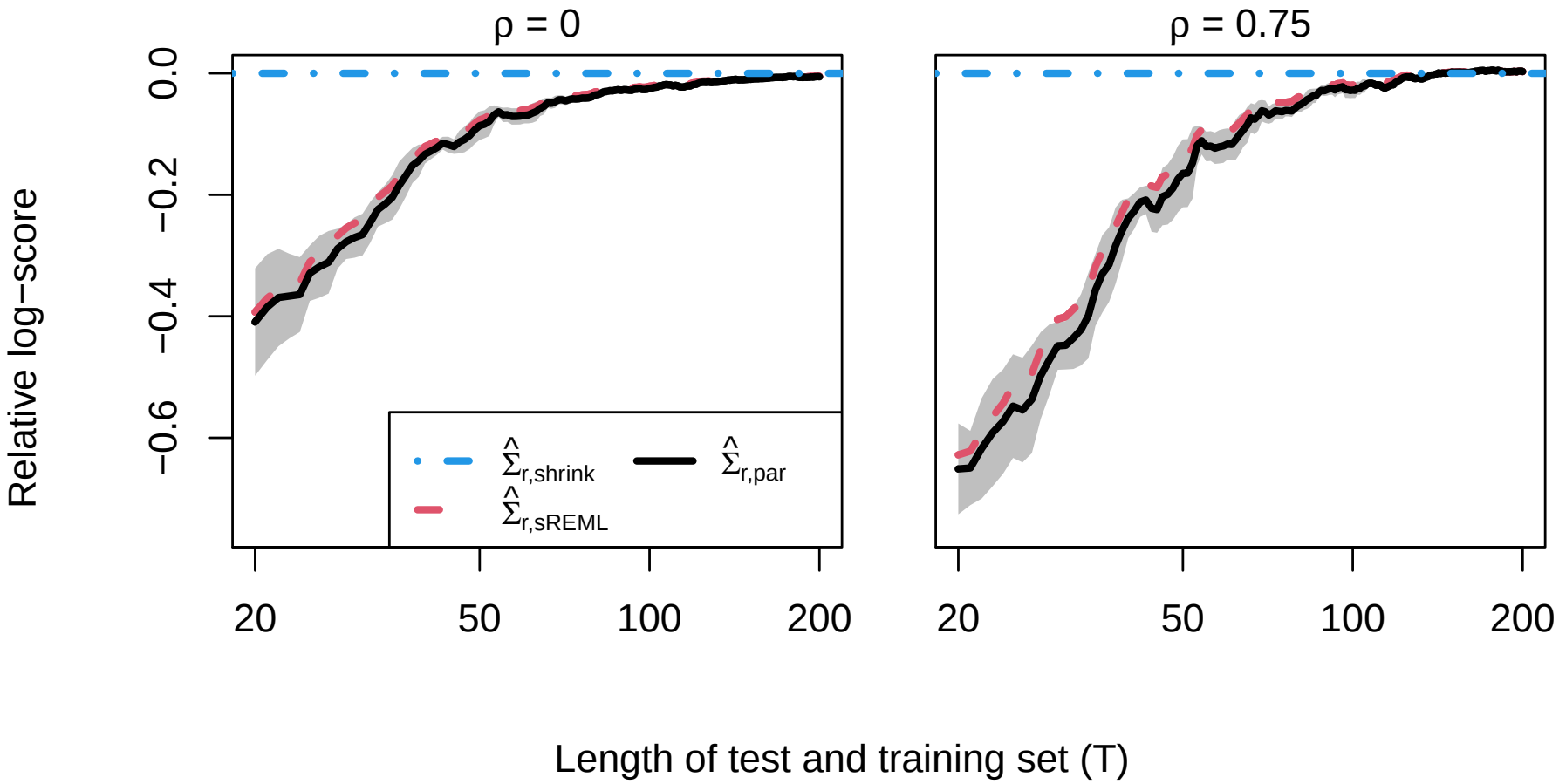
Simulation study

We simulate the data-generating process at the bottom level ($\boldsymbol{y}_t \in \mathbb{R}^4$) using a multivariate AR(1) process

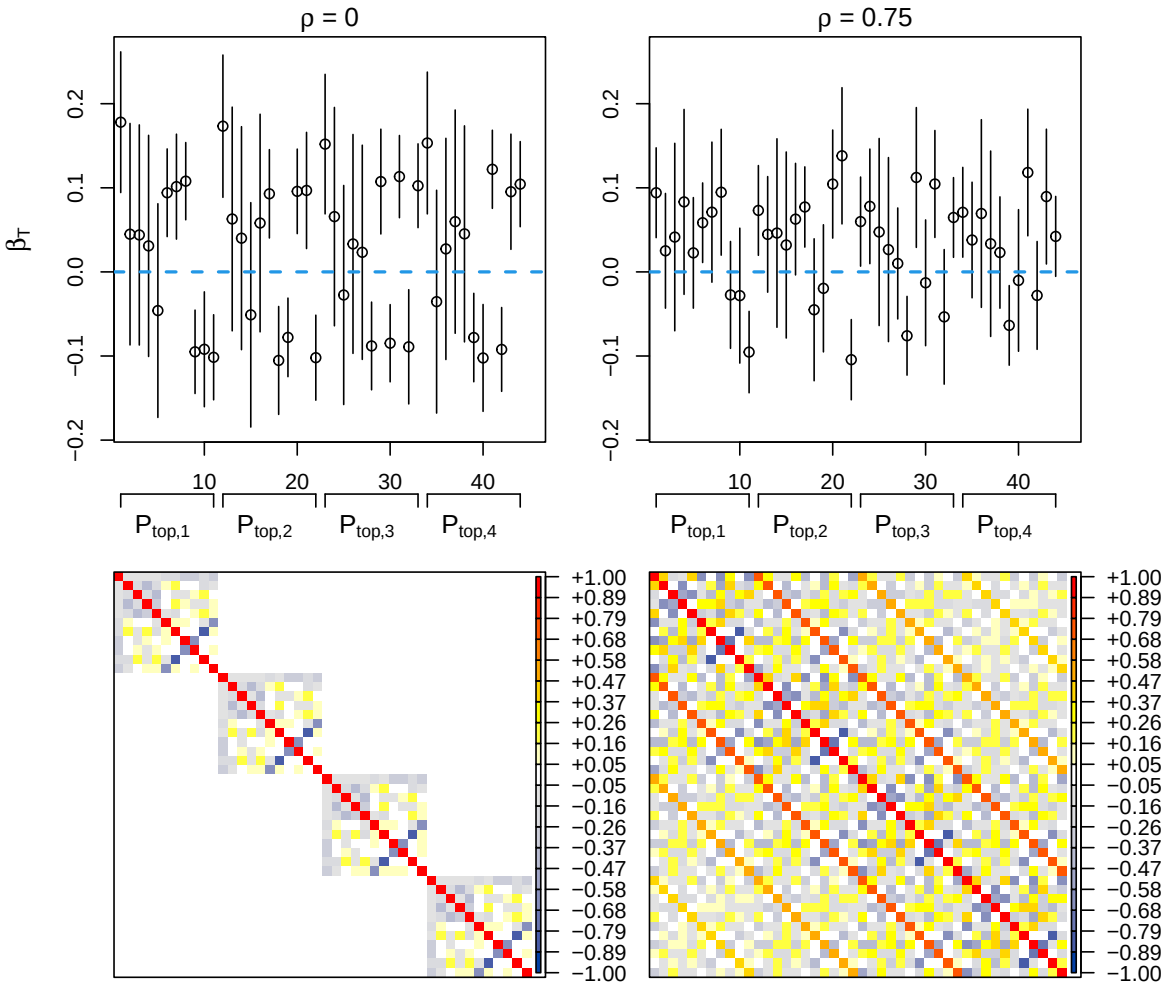
$$\boldsymbol{y}_t = \boldsymbol{A}\boldsymbol{y}_{t-1} + \boldsymbol{\epsilon}_t; \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon) \quad \text{and iid.}, \tag{2}$$

with $\boldsymbol{A}_{ii} = 0.7$, $\boldsymbol{A}_{ij} = -0.25$ $(i \neq j)$, and $(\boldsymbol{\Sigma}_\epsilon)_{ij} = \rho^{|i-j|}$, here we present simulation studies using $\rho = 0$ and $\rho = 0.75$.
Univariate AR(1) models for all levels in the hierarchy (all combinations of summations of 1, 2, 3, or 4 variables).

Simulation results



Parameter standard error and correlation[1]

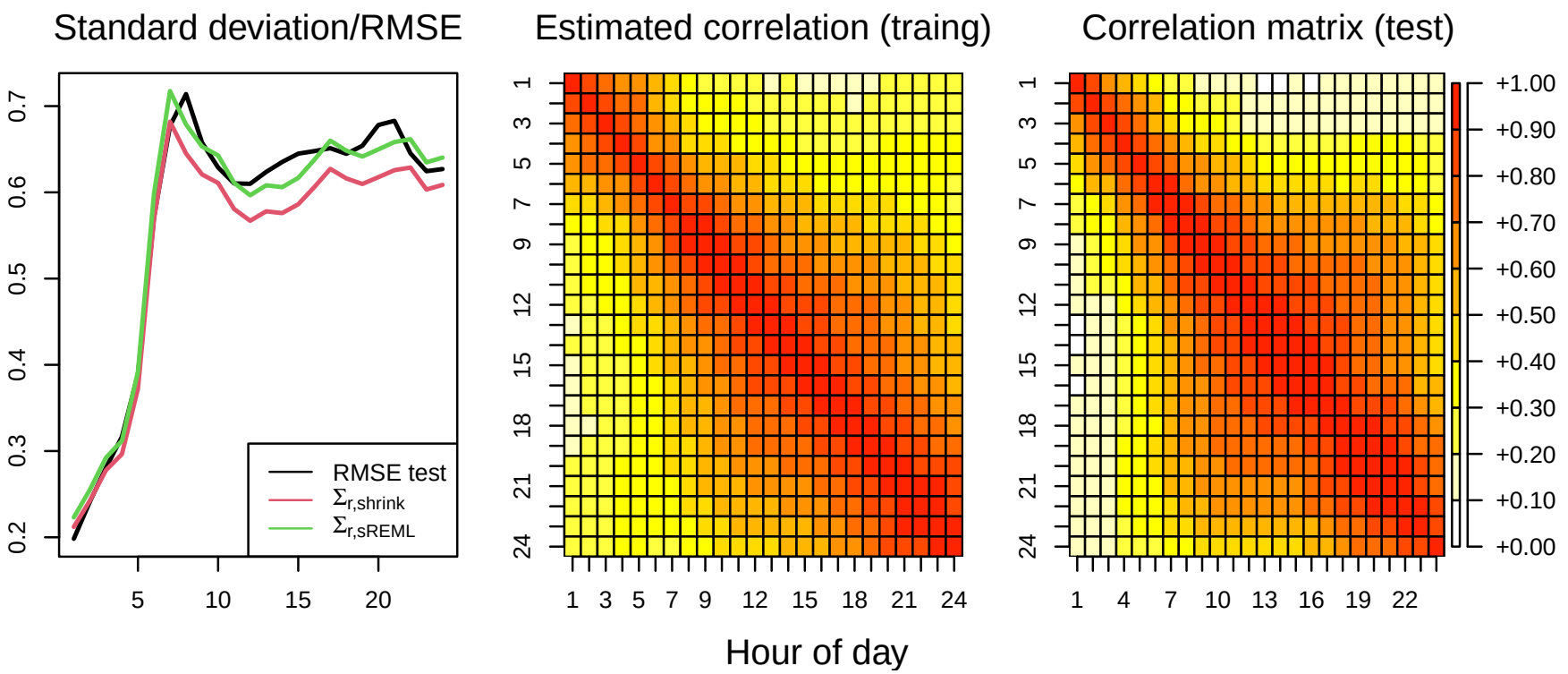


[1] moller2024

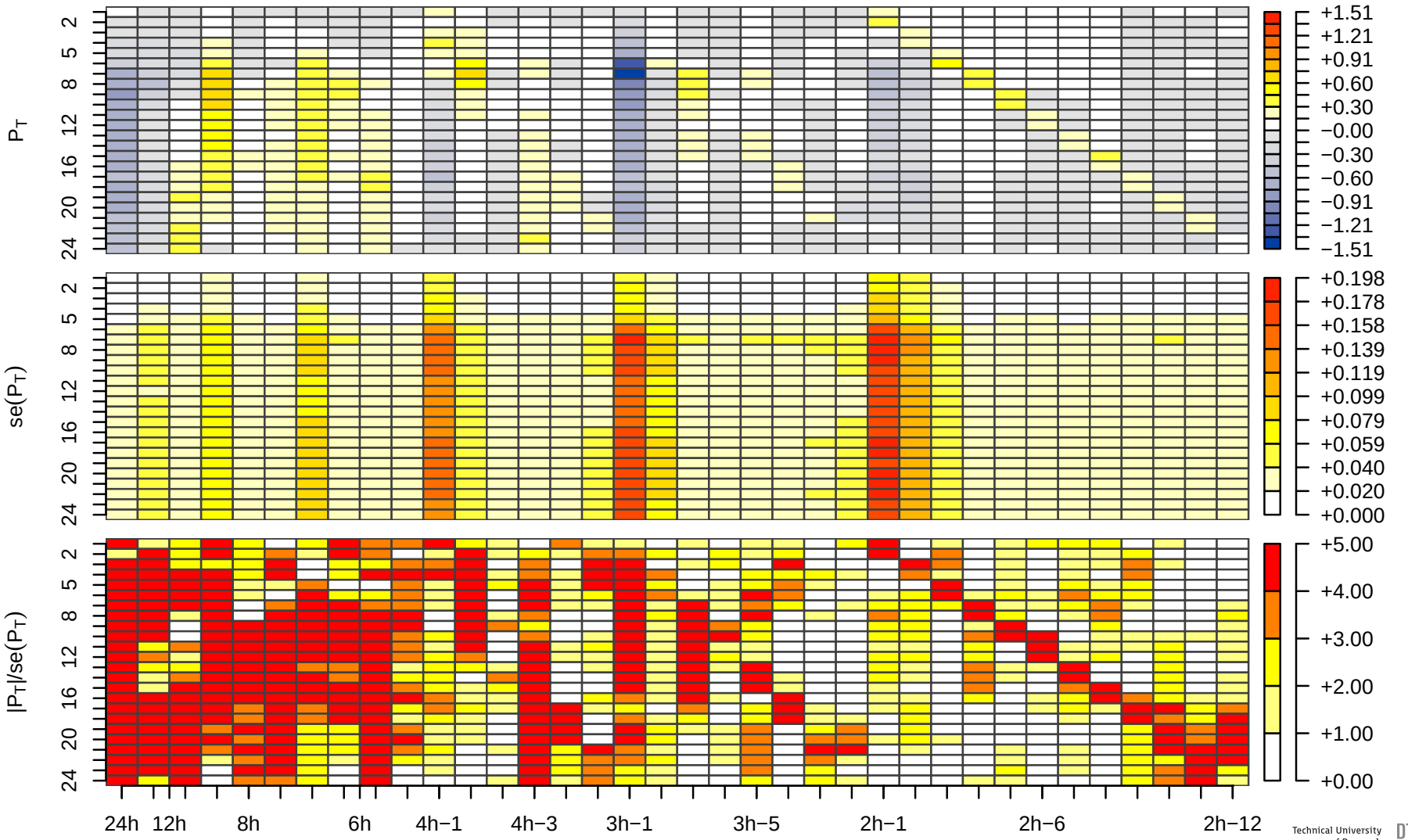
Case study

- Electricity load in Sweden 2016-2020
- Linear model including annual and diurnal variation
- Double seasonal AR models of the residuals (daily and and weekly)
- Temporal reconciliation of residuals

Case: variance–covariance



Case: parameters



Outline

Conclusion / Future

The are many articles on the devellopment of better estimators comming out, and all is not finalized as of yet. Some points are

- Reconciliation ensure coherent forecasts on all levels
- Reconciliation very often improve forecast skills
- The covariance matrix is often ill-conditioned and hence shrinkage should be applied (in the regression setting this is multicollinarity)

My peronal interest for future devellopment is

- Parametrized (time varying) residual variance covariance matrices
- Include auto–correlation between time steps
- Devellop test strategies to reduce model complexity

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., Hjort P.G., and Madsen H. (2024). Optimal Forecast Reconciliation with Uncertainty Quantification `arXiv:2402.06480`.

[3] Møller J.K., Nystrup P., and Madsen H. (2023). Likelihood-based inference in temporal hierarchies. *International Journal of Forecasting, in press*.

[4] Møller JK, Nystrup P, and Madsen H. (2022) Supplementary material for "Likelihood-based inference in temporal hierarchies. http://people.compute.dtu.dk/jkmo/

[5] Nystrup, P., Lindström, E., Møller, J. K., and Madsen, H. (2021). Dimensionality reduction in forecasting with temporal hierarchies. *International Journal of Forecasting, 37(3)*

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