

# Forecast reconciliation

## 2. Perspectives on forecast reconciliation

Rob J Hyndman



MONASH University

# Outline

- 1 Reconciliation via constraints
- 2 The geometry of forecast reconciliation
- 3 Optimization and reconciliation
- 4 ML and regularization
- 5 In-built coherence

# Outline

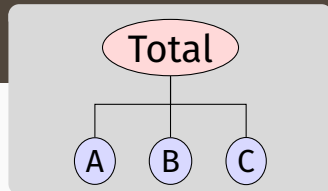
- 1 Reconciliation via constraints
- 2 The geometry of forecast reconciliation
- 3 Optimization and reconciliation
- 4 ML and regularization
- 5 In-built coherence

# Notation reminder

Every collection of time series with linear constraints can be written as

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t$$

- $\mathbf{y}_t$  = vector of all series at time  $t$
- $y_{\text{Total},t}$  = aggregate of all series at time  $t$ .
- $y_{X,t}$  = value of series  $X$  at time  $t$ .
- $\mathbf{b}_t$  = vector of most disaggregated series at time  $t$
- $\mathbf{S}$  = “summing matrix” containing the linear constraints.



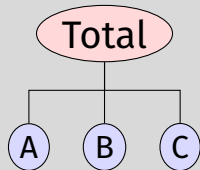
$$\mathbf{y}_t = \begin{pmatrix} y_{\text{Total},t} \\ y_{A,t} \\ y_{B,t} \\ y_{C,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\mathbf{S}} \underbrace{\begin{pmatrix} y_{A,t} \\ y_{B,t} \\ y_{C,t} \end{pmatrix}}_{\mathbf{b}_t}$$

# Notation reminder

Every collection of time series with linear constraints can be written as

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t$$

- $\mathbf{y}_t$  = vector of all series at time  $t$
- $y_{\text{Total},t}$  = aggregate of all series at time  $t$ .
- $y_{X,t}$  = value of series  $X$  at time  $t$ .
- $\mathbf{b}_t$  = vector of most disaggregated series at time  $t$
- $\mathbf{S}$  = “summing matrix” containing the linear constraints.



- Base forecasts:  $\hat{\mathbf{y}}_{T+h|T}$
- Reconciled forecasts:  
 $\tilde{\mathbf{y}}_{T+h|T} = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_{T+h|T}$
- MinT:  
 $\mathbf{G} = (\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h^{-1}$   
where  $\mathbf{W}_h$  is  
covariance matrix of  
base forecast errors.

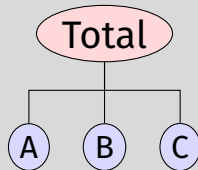
# Notation

## Aggregation matrix

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t$$

$$\begin{pmatrix} y_{\text{Total},t} \\ y_{A,t} \\ y_{B,t} \\ y_{C,t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{A,t} \\ y_{B,t} \\ y_{C,t} \end{pmatrix}$$

$$\begin{pmatrix} \mathbf{a}_t \\ \mathbf{b}_t \end{pmatrix} = \begin{pmatrix} \mathbf{A} \\ \mathbf{I}_{n_b} \end{pmatrix} \mathbf{b}_t$$



## Constraint matrix

$$\mathbf{C}\mathbf{y}_t = \mathbf{0}$$

$$\text{where } \mathbf{C} = \begin{bmatrix} 1 & -1 & -1 & -1 \end{bmatrix} \\ = \begin{bmatrix} \mathbf{I}_{n_a} & -\mathbf{A} \end{bmatrix}$$

# Zero-constraint representation

## Aggregation matrix $A$

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{a}_t \\ \mathbf{b}_t \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \mathbf{I}_{n_b} \end{bmatrix} \mathbf{b}_t = \mathbf{S} \mathbf{b}_t$$

# Zero-constraint representation

## Aggregation matrix $A$

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{a}_t \\ \mathbf{b}_t \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \mathbf{I}_{n_b} \end{bmatrix} \mathbf{b}_t = \mathbf{S} \mathbf{b}_t$$

## Constraint matrix $C$

$$\mathbf{C} \mathbf{y}_t = \mathbf{0}$$

- Constraint matrix approach more general & more parsimonious.
- $\mathbf{C} = [\mathbf{I}_{n_a} \quad -\mathbf{A}]$ .
- $\mathbf{S}$ ,  $\mathbf{A}$  and  $\mathbf{C}$  may contain any real values (not just 0s and 1s).



# Zero-constraint representation

Assuming  $\mathbf{C}$  is full rank

$$\text{where } \tilde{\mathbf{y}}_{T+h|T} = \mathbf{M}\hat{\mathbf{y}}_{T+h|T}$$
$$\mathbf{M} = \mathbf{I} - \mathbf{W}_h \mathbf{C}' (\mathbf{C} \mathbf{W}_h \mathbf{C}')^{-1} \mathbf{C}$$

- Originally proved by Byron (1978) & Byron (1979) for reconciling data.
- Re-discovered by Wickramasuriya, Athanasopoulos, and Hyndman (2019) for reconciling forecasts.
- $\mathbf{M} = \mathbf{SG}$  (the MinT solution)
- Leads to more efficient reconciliation than using  $\mathbf{G}$ .

# Zero-constraint representation

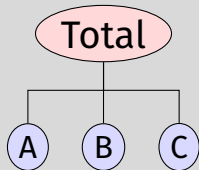
Suppose  $\mathbf{W}_h = \mathbf{I}$ . Then

$$\mathbf{M} = \mathbf{I} - \mathbf{W}_h \mathbf{C}' (\mathbf{C} \mathbf{W}_h \mathbf{C}')^{-1} \mathbf{C}$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 \\ -1 \\ -1 \\ -1 \end{pmatrix} \frac{1}{4} (1 \quad -1 \quad -1 \quad -1)$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{3}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \end{pmatrix}$$



$$\mathbf{A} = (1 \quad 1 \quad 1)$$

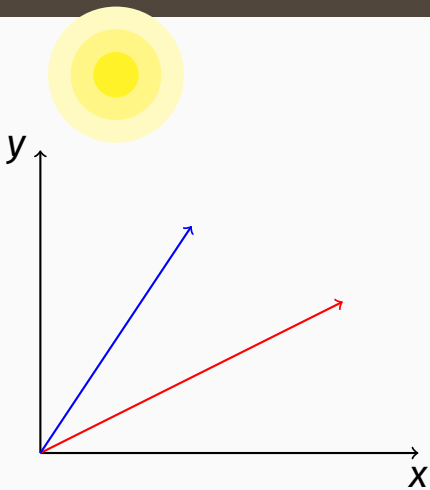
$$\mathbf{S} = \begin{pmatrix} \mathbf{A} \\ \mathbf{I}_{n_b} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{C} = (\mathbf{I}_{n_a} \quad -\mathbf{A}) = (1 \quad -1 \quad -1 \quad -1)$$

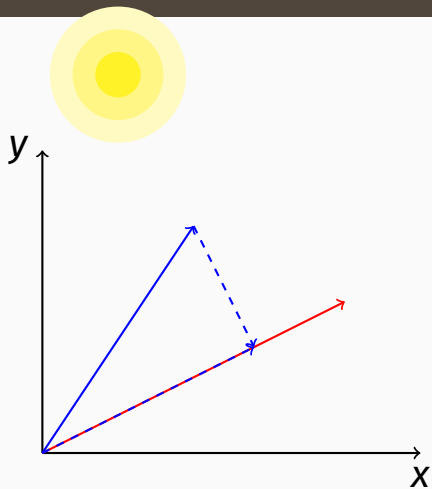
# Outline

- 1 Reconciliation via constraints
- 2 The geometry of forecast reconciliation
- 3 Optimization and reconciliation
- 4 ML and regularization
- 5 In-built coherence

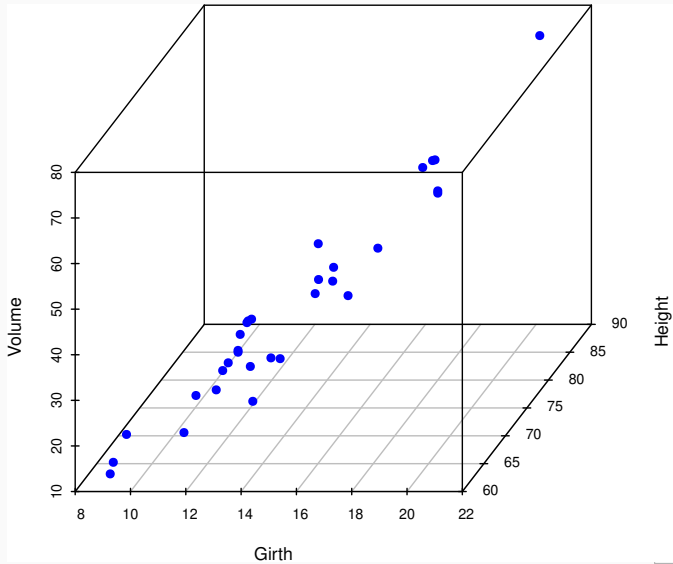
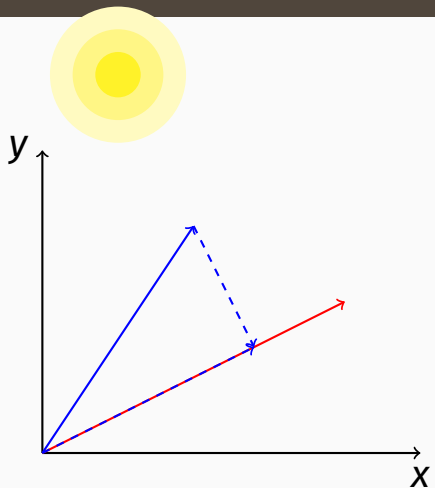
# Projections in linear algebra



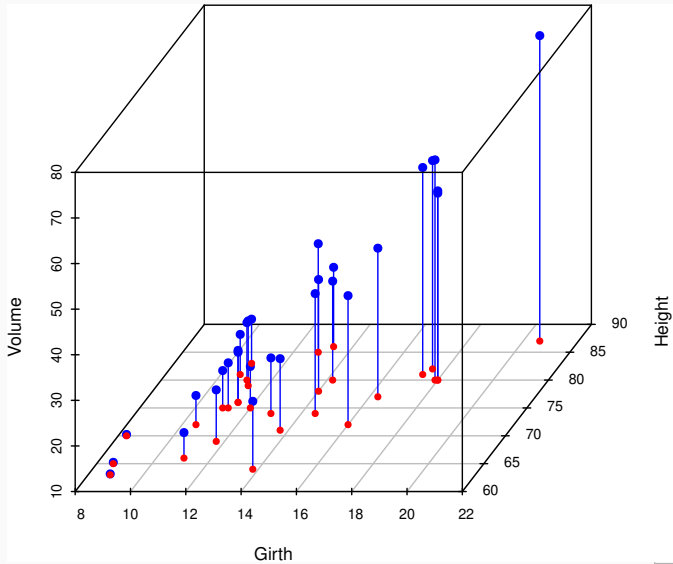
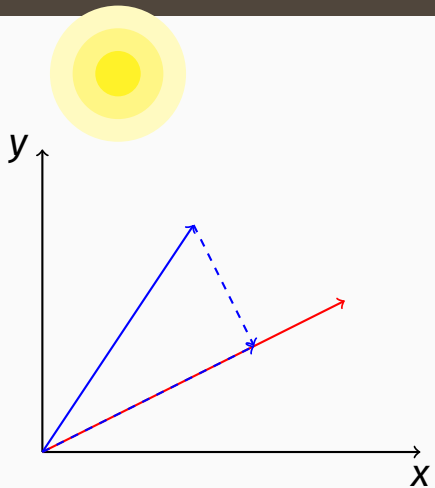
# Projections in linear algebra



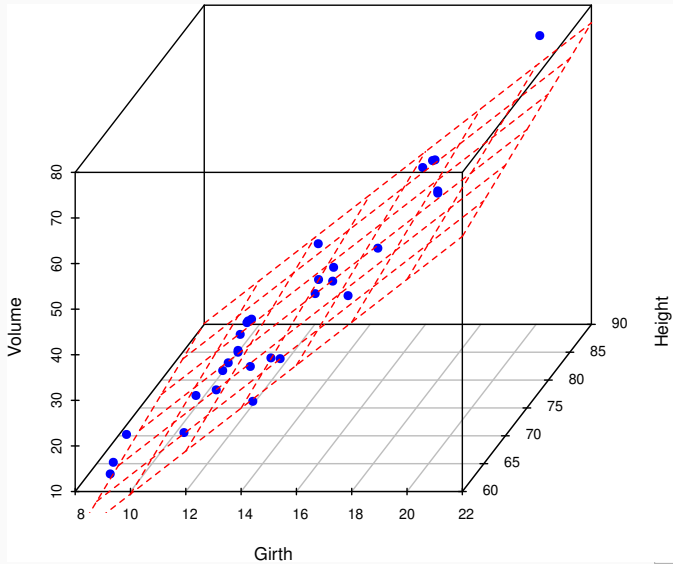
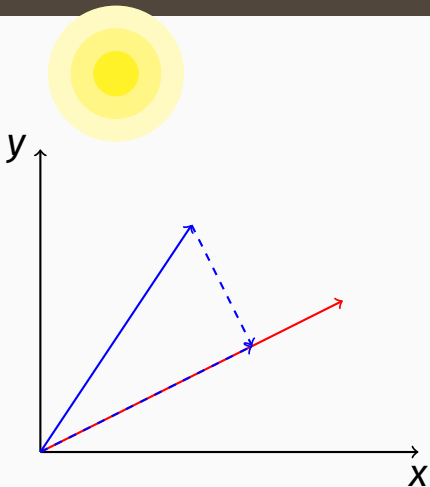
# Projections in linear algebra



# Projections in linear algebra

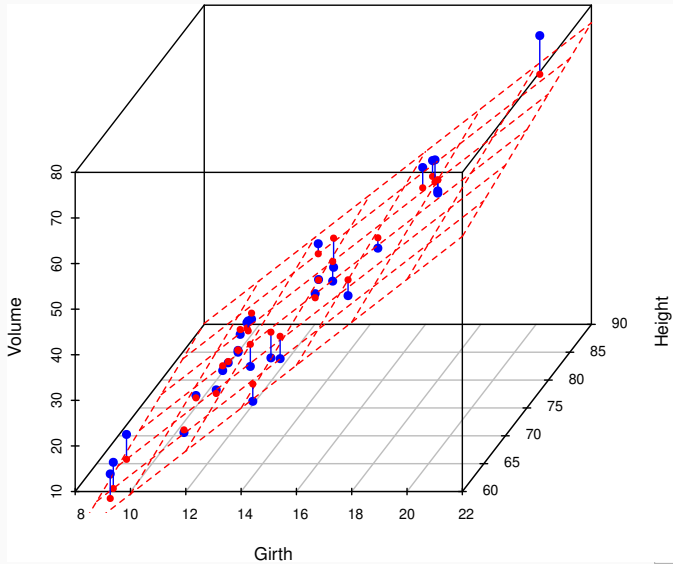
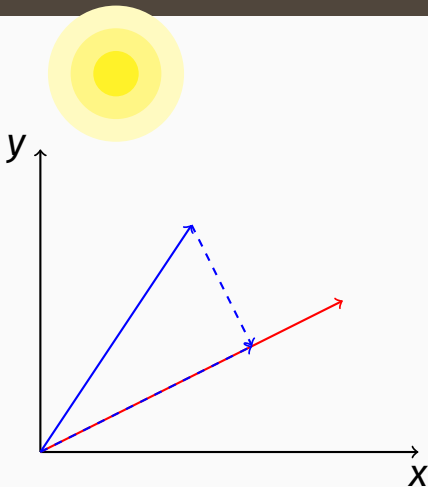


# Projections in linear algebra





# Projections in linear algebra



# Projections in linear algebra

- A projection is a linear transformation  $\mathbf{M}$  such that  $\mathbf{M}^2 = \mathbf{M}$ .
- i.e.,  $\mathbf{M}$  is idempotent: it leaves its image unchanged.
- $\mathbf{M}$  projects onto  $\mathfrak{s}$  if  $\mathbf{M}\mathbf{y} = \mathbf{y}$  for all  $\mathbf{y} \in \mathfrak{s}$ .
- All eigenvalues of  $\mathbf{M}$  are either 0 or 1.
- All singular values of  $\mathbf{M}$  are greater than or equal to 1 (with equality iff  $\mathbf{M}$  is orthogonal).
- A projection is *orthogonal* if  $\mathbf{M}' = \mathbf{M}$ .
- If a projection is not orthogonal, it is called *oblique*.
- In regression, OLS is an orthogonal projection onto space spanned by predictors.

# The coherent subspace

## Coherent subspace

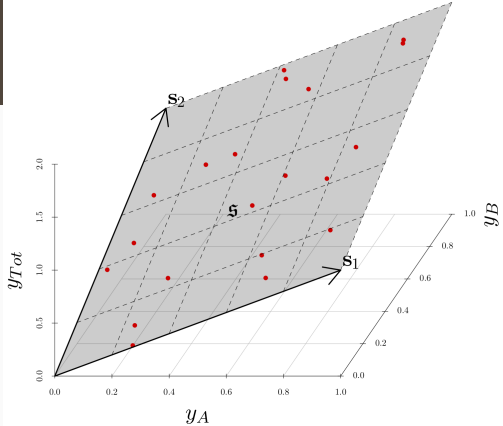
$m$ -dimensional linear subspace  $\mathfrak{s} \subset \mathbb{R}^n$  for which linear constraints hold for all  $\mathbf{y} \in \mathfrak{s}$ .

## Hierarchical time series

An  $n$ -dimensional multivariate time series such that  $\mathbf{y}_t \in \mathfrak{s} \quad \forall t$ .

## Coherent point forecasts

$\tilde{\mathbf{y}}_{t+h|t}$  is coherent if  $\tilde{\mathbf{y}}_{t+h|t} \in \mathfrak{s}$ .



$$y_{Tot} = y_A + y_B$$

# The coherent subspace

## Coherent subspace

$m$ -dimensional linear subspace  $\mathfrak{s} \subset \mathbb{R}^n$  for which linear constraints hold for all  $\mathbf{y} \in \mathfrak{s}$ .

## Hierarchical time series

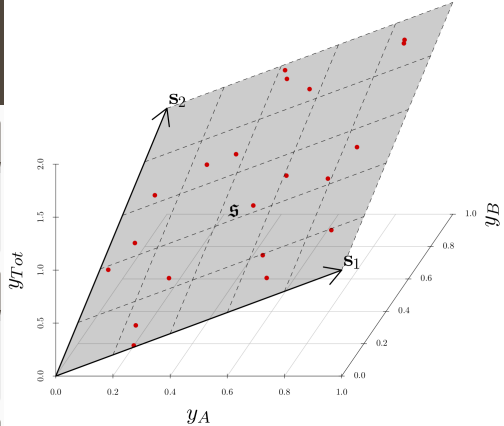
An  $n$ -dimensional multivariate time series such that  $\mathbf{y}_t \in \mathfrak{s} \quad \forall t$ .

## Coherent point forecasts

$\tilde{\mathbf{y}}_{t+h|t}$  is *coherent* if  $\tilde{\mathbf{y}}_{t+h|t} \in \mathfrak{s}$ .

## Base forecasts

Let  $\hat{\mathbf{y}}_{t+h|t}$  be vector of *incoherent* initial  $h$ -step forecasts.



$$y_{Tot} = y_A + y_B$$

# The coherent subspace

## Coherent subspace

$m$ -dimensional linear subspace  $\mathfrak{s} \subset \mathbb{R}^n$  for which linear constraints hold for all  $\mathbf{y} \in \mathfrak{s}$ .

## Hierarchical time series

An  $n$ -dimensional multivariate time series such that  $\mathbf{y}_t \in \mathfrak{s} \quad \forall t$ .

## Coherent point forecasts

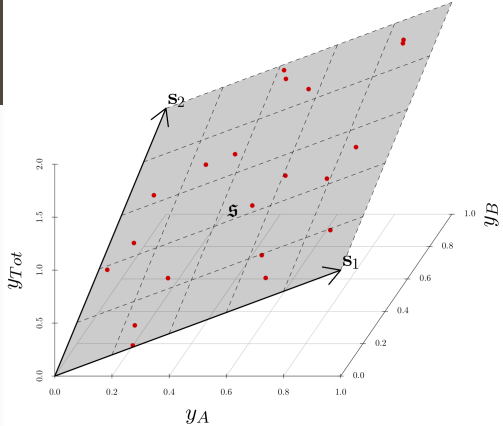
$\tilde{\mathbf{y}}_{t+h|t}$  is *coherent* if  $\tilde{\mathbf{y}}_{t+h|t} \in \mathfrak{s}$ .

## Base forecasts

Let  $\hat{\mathbf{y}}_{t+h|t}$  be vector of *incoherent* initial  $h$ -step forecasts.

## Reconciled forecasts

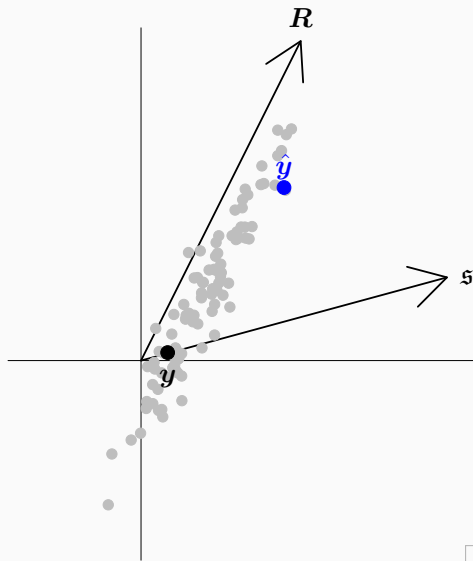
Let  $\psi$  be a mapping,  $\psi : \mathbb{R}^n \rightarrow \mathfrak{s}$ .  
 $\tilde{\mathbf{y}}_{t+h|t} = \psi(\hat{\mathbf{y}}_{t+h|t})$  “reconciles”  $\hat{\mathbf{y}}_{t+h|t}$ .



$$\mathbf{y}_{Tot} = \mathbf{y}_A + \mathbf{y}_B$$

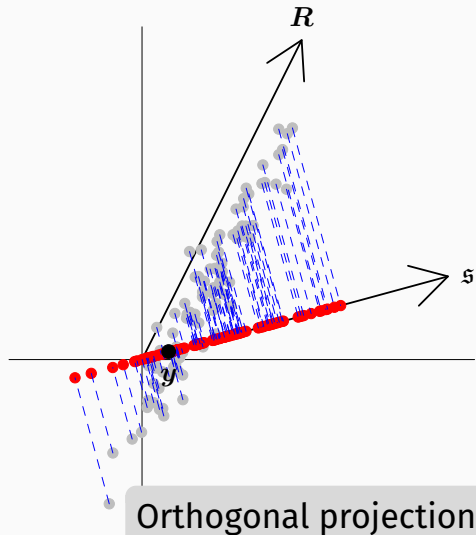
# Linear projection reconciliation

- $R$  is the most likely direction of deviations from  $\varsigma$ .
- Grey: potential base forecasts



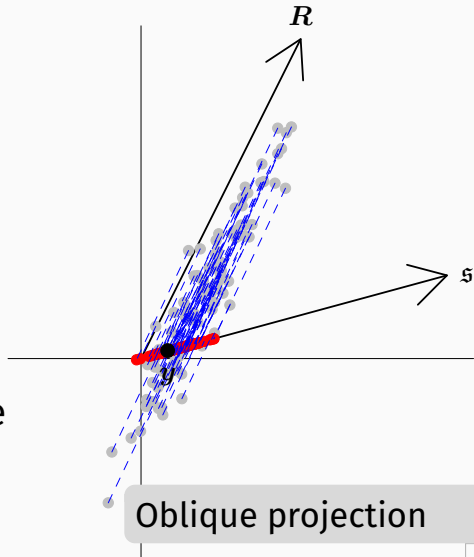
# Linear projection reconciliation

- $R$  is the most likely direction of deviations from  $\xi$ .
- Grey: potential base forecasts
- Red: reconciled forecasts
- Orthogonal projections (i.e., OLS) lead to smallest possible adjustments of base forecasts.



# Linear projection reconciliation

- $R$  is the most likely direction of deviations from  $\xi$ .
- Grey: potential base forecasts
- Red: reconciled forecasts
- Orthogonal projections (i.e., OLS) lead to smallest possible adjustments of base forecasts.
- Oblique projections (i.e., MinT) give reconciled forecasts with smallest variance.





# Linear projection reconciliation

$$\tilde{\mathbf{y}}_{t+h|t} = \psi(\hat{\mathbf{y}}_{t+h|t}) = \mathbf{M}\hat{\mathbf{y}}_{t+h|t}$$

- $\mathbf{M}$  is a projection onto  $\mathfrak{s}$  if and only if  $\mathbf{M}\mathbf{y} = \mathbf{y}$  for all  $\mathbf{y} \in \mathfrak{s}$ .
- Coherent base forecasts are unchanged since  $\mathbf{M}\hat{\mathbf{y}} = \hat{\mathbf{y}}$
- If  $\hat{\mathbf{y}}$  is unbiased, then  $\tilde{\mathbf{y}}$  is also unbiased since

$$E(\tilde{\mathbf{y}}_{t+h|t}) = E(\mathbf{M}\hat{\mathbf{y}}_{t+h|t}) = \mathbf{M}E(\hat{\mathbf{y}}_{t+h|t}) = E(\hat{\mathbf{y}}_{t+h|t}),$$

and unbiased estimates must lie on  $\mathfrak{s}$ .

- The projection is orthogonal if and only if  $\mathbf{M}' = \mathbf{M}$ .
- If  $\mathbf{S}$  forms a basis set for  $\mathfrak{s}$ , then projections are of the form  $\mathbf{M} = \mathbf{S}(\mathbf{S}'\Psi\mathbf{S})^{-1}\mathbf{S}'\Psi$  where  $\Psi$  is a positive definite matrix.

# Linear projection reconciliation

$$\tilde{\mathbf{y}}_{t+h|t} = \psi(\hat{\mathbf{y}}_{t+h|t}) = \mathbf{M}\hat{\mathbf{y}}_{t+h|t}, \quad \text{where} \quad \mathbf{M} = \mathbf{S}(\mathbf{S}'\Psi\mathbf{S})^{-1}\mathbf{S}'\Psi$$

$$\text{OLS: } \Psi = \mathbf{I} \quad \mathbf{M} = \mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}' \quad = \mathbf{I} - \mathbf{C}'(\mathbf{C}\mathbf{C}')^{-1}\mathbf{C}$$

$$\text{MinT: } \Psi = \mathbf{W}_h \quad \mathbf{M} = \mathbf{S}(\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h^{-1} \quad = \mathbf{I} - \mathbf{W}_h\mathbf{C}'(\mathbf{C}\mathbf{W}_h\mathbf{C}')^{-1}\mathbf{C}$$

- $\mathbf{M}$  is orthogonal iff  $\Psi = \mathbf{I}$ .
- $\mathbf{W}_h = \text{Var}[\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h|T} \mid \mathbf{y}_1, \dots, \mathbf{y}_T]$  is the covariance matrix of the base forecast errors.
- $\mathbf{V}_h = \text{Var}[\mathbf{y}_{T+h} - \tilde{\mathbf{y}}_{T+h|T} \mid \mathbf{y}_1, \dots, \mathbf{y}_T] = \mathbf{M}\mathbf{W}_h\mathbf{M}'$  is minimized when  $\Psi = \mathbf{W}_h$ .

# Mean square error bounds

Panagiotelis, Gamakumara,  
Athanasopoulos, and  
Hyndman (2021)

## Distance reducing property

Let  $\|\mathbf{u}\|_{\Psi} = \mathbf{u}'\Psi\mathbf{u}$ . Then

$$\|\mathbf{y}_{t+h} - \tilde{\mathbf{y}}_{t+h|t}\|_{\Psi} \leq \|\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h|t}\|_{\Psi}$$

- $\Psi$ -projection is guaranteed to improve forecast accuracy over base forecasts *using this distance measure*.
- Distance reduction holds for any realisation and any forecast.
- OLS reconciliation minimizes Euclidean distance.
- Other measures of forecast accuracy may be worse.

$$\begin{aligned}\|\mathbf{y}_{t+h} - \tilde{\mathbf{y}}_{t+h}\|_2^2 &= \|\mathbf{M}(\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h})\|_2^2 \\ &\leq \|\mathbf{M}\|_2^2 \|\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h}\|_2^2 \\ &= \sigma_{\max}^2 \|\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h}\|_2^2\end{aligned}$$

- $\sigma_{\max}$  is the largest eigenvalue of  $\mathbf{M}$
- $\sigma_{\max} \geq 1$  as  $\mathbf{M}$  is a projection matrix.
- Every projection reconciliation is better than base forecasts using Euclidean distance.

$$\begin{aligned} & \text{tr}\left(\mathbb{E}[\mathbf{y}_{t+h} - \tilde{\mathbf{y}}_{t+h|t}^{\text{MinT}}]'[\mathbf{y}_{t+h} - \tilde{\mathbf{y}}_{t+h|t}^{\text{MinT}}]\right) \\ & \leq \text{tr}\left(\mathbb{E}[\mathbf{y}_{t+h} - \tilde{\mathbf{y}}_{t+h|t}^{\text{OLS}}]'[\mathbf{y}_{t+h} - \tilde{\mathbf{y}}_{t+h|t}^{\text{OLS}}]\right) \\ & \leq \text{tr}\left(\mathbb{E}[\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h|t}]'[\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h|t}]\right) \end{aligned}$$

Using sums of variances:

- MinT reconciliation is better than OLS reconciliation
- OLS reconciliation is better than base forecasts

# Outline

- 1 Reconciliation via constraints
- 2 The geometry of forecast reconciliation
- 3 Optimization and reconciliation
- 4 ML and regularization
- 5 In-built coherence

# Minimum trace reconciliation

## Minimum trace (MinT) reconciliation

If  $\mathbf{SG}$  is a projection, then the trace of  $\mathbf{V}_h = \text{Var}(\tilde{\mathbf{y}}_{t+h|t} - \mathbf{y}_{t+h})$  is **minimized** when

$$\mathbf{G} = (\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h^{-1}$$

$$\tilde{\mathbf{y}}_{T+h|T} = \mathbf{S}(\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h^{-1}\hat{\mathbf{y}}_{T+h|T}$$

Reconciled forecasts

Base forecasts

- Trace of  $\mathbf{V}_h$  is sum of forecast variances.
- MinT solution is  $L_2$  **optimal** amongst linear unbiased forecasts.

Find the solution to the minimax problem

$$V = \min_{\tilde{\mathbf{y}} \in \mathfrak{s}} \max_{\mathbf{y} \in \mathfrak{s}} \{ \ell(\mathbf{y}, \tilde{\mathbf{y}}) - \ell(\mathbf{y}, \hat{\mathbf{y}}) \},$$

where  $\ell$  is a loss function, and  $\mathfrak{s}$  is the coherent subspace.

- $V \leq 0$ : reconciliation guaranteed to reduce loss.
- If  $\ell(\mathbf{y}, \tilde{\mathbf{y}}) = \|\mathbf{y} - \tilde{\mathbf{y}}\|_{\Psi} = (\mathbf{y} - \tilde{\mathbf{y}})' \Psi (\mathbf{y} - \tilde{\mathbf{y}})$ , where  $\Psi$  is any symmetric pd matrix, then:
  - 1  $\tilde{\mathbf{y}} = \mathbf{S}(\mathbf{S}'\Psi\mathbf{S})^{-1}\mathbf{S}'\Psi\hat{\mathbf{y}}$  will always improve upon the base forecasts;
  - 2 The MinT solution  $\tilde{\mathbf{y}} = \mathbf{S}(\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h^{-1}\hat{\mathbf{y}}$  will optimise loss in expectation over any choice of  $\Psi$ .



Regularized empirical risk minimization problem:

$$\min_{\mathbf{G}} \frac{1}{Nn} \|\mathbf{Y} - \hat{\mathbf{Y}}\mathbf{G}'\mathbf{S}'\|_F + \lambda \|\text{vec}\mathbf{G}\|_1,$$

- $N = T - T_1 - h + 1$ ,  $T_1$  is minimum training sample size
- $\|\cdot\|_F$  is the Frobenius norm
- $\mathbf{Y} = [\mathbf{y}_{T_1+h}, \dots, \mathbf{y}_T]'$
- $\hat{\mathbf{Y}} = [\hat{\mathbf{y}}_{T_1+h|T_1}, \dots, \hat{\mathbf{y}}_{T|T-h}]'$
- $\lambda$  is a regularization parameter.

When  $\lambda = 0$ :  $\hat{\mathbf{G}} = \mathbf{B}'\hat{\mathbf{Y}}(\hat{\mathbf{Y}}'\hat{\mathbf{Y}})^{-1}$  where  $\mathbf{B} = [\mathbf{b}_{T_1+h}, \dots, \mathbf{b}_T]'$ .

# Non-negative forecasts

$$\min_{\mathbf{G}_h} \text{tr} \left( \mathbb{E}[\mathbf{y}_{t+h} - \mathbf{S}\mathbf{G}_h\hat{\mathbf{y}}_{t+h|t}]' [\mathbf{y}_{t+h} - \mathbf{S}\mathbf{G}_h\hat{\mathbf{y}}_{t+h|t}] \right)$$

such that  $\mathbf{b}_{t+h|t} = \mathbf{G}_h\hat{\mathbf{y}}_{t+h|t} \geq 0$

# Non-negative forecasts

$$\min_{\mathbf{G}_h} \text{tr} \left( E[\mathbf{y}_{t+h} - \mathbf{S}\mathbf{G}_h\hat{\mathbf{y}}_{t+h|t}]' [\mathbf{y}_{t+h} - \mathbf{S}\mathbf{G}_h\hat{\mathbf{y}}_{t+h|t}] \right)$$

such that  $\mathbf{b}_{t+h|t} = \mathbf{G}_h\hat{\mathbf{y}}_{t+h|t} \geq 0$

**Solve via quadratic programming:**

$$\min_{\mathbf{b}} \frac{1}{2} \mathbf{b}' \mathbf{S}' \mathbf{W}_h^{-1} \mathbf{S} \mathbf{b} - \mathbf{b}' \mathbf{S}' \mathbf{W}_h^{-1} \hat{\mathbf{y}}_{T+h|T} \quad \text{s.t. } \mathbf{b} \geq 0$$

(Wickramasuriya, Turlach, and Hyndman, 2020)

# Non-negative forecasts

$$\min_{\mathbf{G}_h} \text{tr} \left( \mathbb{E}[\mathbf{y}_{t+h} - \mathbf{S}\mathbf{G}_h\hat{\mathbf{y}}_{t+h|t}]' [\mathbf{y}_{t+h} - \mathbf{S}\mathbf{G}_h\hat{\mathbf{y}}_{t+h|t}] \right)$$

such that  $\mathbf{b}_{t+h|t} = \mathbf{G}_h\hat{\mathbf{y}}_{t+h|t} \geq 0$

**Solve via quadratic programming:**

$$\min_{\mathbf{b}} \frac{1}{2} \mathbf{b}' \mathbf{S}' \mathbf{W}_h^{-1} \mathbf{S} \mathbf{b} - \mathbf{b}' \mathbf{S}' \mathbf{W}_h^{-1} \hat{\mathbf{y}}_{T+h|T} \quad \text{s.t. } \mathbf{b} \geq 0$$

(Wickramasuriya, Turlach, and Hyndman, 2020)

**Set-negative-to-zero heuristic solution**

- Negative reconciled forecasts at bottom level set to zero
- Remaining forecasts computed via aggregation

(Di Fonzo and Girolimetto, 2023)

$$\min_{\mathbf{G}_h} \text{tr} \left( \mathbb{E}[\mathbf{y}_{t+h} - \mathbf{S}\mathbf{G}_h\hat{\mathbf{y}}_{t+h|t}][\mathbf{y}_{t+h} - \mathbf{S}\mathbf{G}_h\hat{\mathbf{y}}_{t+h|t}]' \right) \\ \text{such that } \mathbf{C}\mathbf{S}\mathbf{G}_h\hat{\mathbf{y}}_{t+h|t} = \mathbf{d}$$

- Differs from top-down approaches in that it can be done while also preserving the unbiasedness of base forecasts.

See also Di Fonzo and Girolimetto (2022). Zhang, Kang, Panagiotelis, and Li (2022).

# Outline

- 1 Reconciliation via constraints
- 2 The geometry of forecast reconciliation
- 3 Optimization and reconciliation
- 4 ML and regularization**
- 5 In-built coherence

# ML and regularization

- Replace the linear regression formulation with a less restrictive method to obtain combinations of forecasts from the various hierarchical levels.
- Coherence is achieved via a bottom-up approach, or by embedding coherence in the ML training.



Gleason (2020) attempts to overcome the lack of focus on coherence by adjusting the objective function. Using neural network forecasts, he includes a regularisation term that penalises incoherences in the generated forecasts. This

# Outline

- 1 Reconciliation via constraints
- 2 The geometry of forecast reconciliation
- 3 Optimization and reconciliation
- 4 ML and regularization
- 5 In-built coherence



# In-built coherence

**Two-step approach:** compute base forecasts  $\hat{\mathbf{y}}_h$ , and then reconcile them to produce  $\tilde{\mathbf{y}}_h$ .

**One-step approaches:** compute coherent  $\tilde{\mathbf{y}}_h$  directly.

- Ashouri, Hyndman, and Shmueli (2022): linear regression models
- Pennings and Dalen (2017): state space models
- Villegas and Pedregal (2018): state space models

# In-built coherence using linear models

Suppose  $\hat{y}_{t,i} = \hat{\beta}_i' \mathbf{x}_{t,i}$  with  $\mathbf{x}_{t,i} = (1, x_{t,1,i}, \dots, x_{t,p,i})$  &  $\hat{\mathbf{y}}_i = (\hat{y}_{1,i}, \dots, \hat{y}_{T,i})$ .

# In-built coherence using linear models

Suppose  $\hat{y}_{t,i} = \hat{\beta}_i' \mathbf{x}_{t,i}$  with  $\mathbf{x}_{t,i} = (1, x_{t,1,i}, \dots, x_{t,p,i})$  &  $\hat{\mathbf{y}}_i = (\hat{y}_{1,i}, \dots, \hat{y}_{T,i})$ .

$$\begin{pmatrix} \hat{\mathbf{y}}_1 \\ \hat{\mathbf{y}}_2 \\ \vdots \\ \hat{\mathbf{y}}_n \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & 0 & \dots & 0 \\ 0 & \mathbf{X}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \mathbf{X}_n \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_n \end{pmatrix}, \quad \mathbf{X}_i = \begin{pmatrix} 1 & x_{1,i,1} & x_{1,i,2} & \dots & x_{1,i,p} \\ 1 & x_{2,i,1} & x_{2,i,2} & \dots & x_{2,i,p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{T,i,1} & x_{T,i,2} & \dots & x_{T,i,p} \end{pmatrix}$$

# In-built coherence using linear models

Suppose  $\hat{y}_{t,i} = \hat{\beta}_i' \mathbf{x}_{t,i}$  with  $\mathbf{x}_{t,i} = (1, x_{t,1,i}, \dots, x_{t,p,i})$  &  $\hat{\mathbf{y}}_i = (\hat{y}_{1,i}, \dots, \hat{y}_{T,i})$ .

$$\begin{pmatrix} \hat{\mathbf{y}}_1 \\ \hat{\mathbf{y}}_2 \\ \vdots \\ \hat{\mathbf{y}}_n \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & 0 & \dots & 0 \\ 0 & \mathbf{X}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \mathbf{X}_n \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_n \end{pmatrix}, \quad \mathbf{X}_i = \begin{pmatrix} 1 & x_{1,i,1} & x_{1,i,2} & \dots & x_{1,i,p} \\ 1 & x_{2,i,1} & x_{2,i,2} & \dots & x_{2,i,p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{T,i,1} & x_{T,i,2} & \dots & x_{T,i,p} \end{pmatrix}$$

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

# In-built coherence using linear models

Suppose  $\hat{y}_{t,i} = \hat{\beta}_i' \mathbf{x}_{t,i}$  with  $\mathbf{x}_{t,i} = (1, x_{t,1,i}, \dots, x_{t,p,i})$  &  $\hat{\mathbf{y}}_i = (\hat{y}_{1,i}, \dots, \hat{y}_{T,i})$ .

$$\begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & 0 & \dots & 0 \\ 0 & \mathbf{X}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \mathbf{X}_n \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_n \end{pmatrix}, \quad \mathbf{X}_i = \begin{pmatrix} 1 & x_{1,i,1} & x_{1,i,2} & \dots & x_{1,i,p} \\ 1 & x_{2,i,1} & x_{2,i,2} & \dots & x_{2,i,p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{T,i,1} & x_{T,i,2} & \dots & x_{T,i,p} \end{pmatrix}$$

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

$$\hat{\mathbf{y}}_{t+h} = \mathbf{X}_{t+h}^* \hat{\mathbf{B}}$$

# In-built coherence using linear models

Suppose  $\hat{y}_{t,i} = \hat{\beta}_i' \mathbf{x}_{t,i}$  with  $\mathbf{x}_{t,i} = (1, x_{t,1,i}, \dots, x_{t,p,i})$  &  $\hat{\mathbf{y}}_i = (\hat{y}_{1,i}, \dots, \hat{y}_{T,i})$ .

$$\begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & 0 & \dots & 0 \\ 0 & \mathbf{X}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \mathbf{X}_n \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_n \end{pmatrix}, \quad \mathbf{X}_i = \begin{pmatrix} 1 & x_{1,i,1} & x_{1,i,2} & \dots & x_{1,i,p} \\ 1 & x_{2,i,1} & x_{2,i,2} & \dots & x_{2,i,p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{T,i,1} & x_{T,i,2} & \dots & x_{T,i,p} \end{pmatrix}$$

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

$$\hat{\mathbf{y}}_{t+h} = \mathbf{X}_{t+h}^* \hat{\mathbf{B}}$$

$$\mathbf{X}_{t+h}^* = \text{diag}(\mathbf{x}'_{t+h,1}, \dots, \mathbf{x}'_{t+h,n})$$

# In-built coherence using linear models

Suppose  $\hat{y}_{t,i} = \hat{\beta}'_i \mathbf{x}_{t,i}$  with  $\mathbf{x}_{t,i} = (1, x_{t,1,i}, \dots, x_{t,p,i})$  &  $\hat{\mathbf{y}}_i = (\hat{y}_{1,i}, \dots, \hat{y}_{T,i})$ .

$$\begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & 0 & \dots & 0 \\ 0 & \mathbf{X}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \mathbf{X}_n \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_n \end{pmatrix}, \quad \mathbf{X}_i = \begin{pmatrix} 1 & x_{1,i,1} & x_{1,i,2} & \dots & x_{1,i,p} \\ 1 & x_{2,i,1} & x_{2,i,2} & \dots & x_{2,i,p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{T,i,1} & x_{T,i,2} & \dots & x_{T,i,p} \end{pmatrix}$$

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad \hat{\mathbf{y}}_{t+h} = \mathbf{X}_{t+h}^* \hat{\mathbf{B}} \quad \mathbf{X}_{t+h}^* = \text{diag}(\mathbf{x}'_{t+h,1}, \dots, \mathbf{x}'_{t+h,n})$$

$$\tilde{\mathbf{y}}_{t+h} = \mathbf{S}(\mathbf{S}'\mathbf{W}_h\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h\hat{\mathbf{y}}_{t+h} = \mathbf{S}(\mathbf{S}'\mathbf{W}_h\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h\mathbf{X}_{t+h}^*(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

$$\mathbf{V}_h = \sigma^2 \mathbf{S}(\mathbf{S}'\mathbf{W}_h\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h \left[ 1 + \mathbf{X}_{T+h}^* (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}_{T+h}^*)' \right] \mathbf{W}_h \mathbf{S}' (\mathbf{S}'\mathbf{W}_h\mathbf{S})^{-1} \mathbf{S}'$$

Reference: Ashouri, Hyndman, and Shmueli (2022)

# In-built coherence

Pennings and Dalen (2017) propose the state space model

$$\mathbf{y}_t = \mathbf{S}\mu_t + \mathbf{Z}_t\beta + \varepsilon_t, \quad \varepsilon_t \sim N(\mathbf{0}, \Sigma_\varepsilon), \quad (1)$$

$$\mu_t = \mu_{t-1} + \eta_t, \quad \eta_t \sim N(\mathbf{0}, \Sigma_\eta). \quad (2)$$






- Coherent forecasts arise naturally using the Kalman filter
- Covariance matrices difficult to estimate except for small hierarchies.



# In-build coherence

A related state space approach was proposed by Villegas and Pedregal (2018), who show that their formulation subsumes bottom-up, top-down, and some forms of forecast reconciliation and combination forecasting.






# References

-  Ashouri, M, RJ Hyndman, and G Shmueli (2022). Fast forecast reconciliation using linear models. *J Computational & Graphical Statistics* **31**(1), 263–282.
-  Ben Taieb, S and B Koo (2019). Regularized regression for hierarchical forecasting without unbiasedness conditions. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. KDD '19. Anchorage, AK, USA: Association for Computing Machinery, pp.1337–1347.
-  Byron, RP (1978). The estimation of large social account matrices. *Journal of the Royal Statistical Society, Series A* **141**(3), 359–367.
-  Byron, RP (1979). Corrigenda: The estimation of large social account matrices. *Journal of the Royal Statistical Society. Series A* **142**(3), 405.
-  Di Fonzo, T and D Girolimetto (2022). Forecast combination-based forecast reconciliation: Insights and extensions. *International Journal of Forecasting* **forthcoming**.

# References

-  Di Fonzo, T and D Girolimetto (2023). Spatio-temporal reconciliation of solar forecasts. *Solar Energy* **251**, 13–29.
-  Gleason, JL (2020). Forecasting hierarchical time series with a regularized embedding space. In: *MileTS '20: 6th KDD Workshop on Mining and Learning from Time Series 2020*. San Diego, California, USA, pp.883–894.
-  Hollyman, R, F Petropoulos, and ME Tipping (2021). Understanding forecast reconciliation. *European Journal of Operational Research* **294**(1), 149–160.
-  Panagiotelis, A, P Gamakumara, G Athanasopoulos, and RJ Hyndman (2021). Forecast reconciliation: A geometric view with new insights on bias correction. *International J Forecasting* **37**(1), 343–359.
-  Pennings, CL and J van Dalen (2017). Integrated hierarchical forecasting. *European Journal of Operational Research* **263**(2), 412–418.

# References

-  van Erven, T and J Cugliari (2015). “Game-theoretically optimal reconciliation of contemporaneous hierarchical time series forecasts”. In: *Modeling and Stochastic Learning for Forecasting in High Dimension*. Ed. by A Antoniadis, JM Poggi, and X Brossat. Cham: Springer International Publishing, pp.297–317.
-  Villegas, MA and DJ Pedregal (2018). Supply chain decision support systems based on a novel hierarchical forecasting approach. *Decision Support Systems* **114**, 29–36.
-  Wickramasuriya, SL (2021). Properties of point forecast reconciliation approaches. *arXiv preprint arXiv:2103.11129*.
-  Wickramasuriya, SL, G Athanasopoulos, and RJ Hyndman (2019). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *J American Statistical Association* **114**(526), 804–819.
-  Wickramasuriya, SL, BA Turlach, and RJ Hyndman (2020). Optimal non-negative forecast reconciliation. *Statistics & Computing* **30**(5), 1167–1182.

# References



Zhang, B, Y Kang, A Panagiotelis, and F Li (2022). Optimal reconciliation with immutable forecasts. *European Journal of Operational Research* **forthcoming**.