# Forecast reconciliation

## 2. Perspectives on forecast reconciliation

Rob J Hyndman

MONASH University

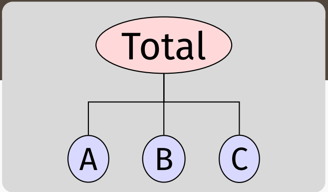
robjhyndman.com/fr2023

Photo by Edvard Alexander Rølvaag on

# Outline

# Outline

# Notation reminder



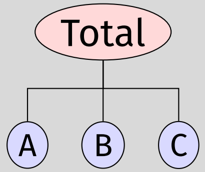

Every collection of time series with linear constraints can be written as

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

- $\boldsymbol{y}_t$ = 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 $t$.
- $\boldsymbol{b}_t$ = vector of most disaggregated series at time $t$
- $\boldsymbol{S}$ = "summing matrix" containing the linear constraints.

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

# Notation reminder



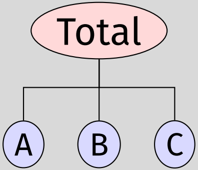

Every collection of time series with linear constraints can be written as

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

- $\boldsymbol{y}_t$ = 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 $t$.
- $\boldsymbol{b}_t$ = vector of most disaggregated series at time $t$
- $\boldsymbol{S}$ = "summing matrix" containing the linear constraints.

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

# Notation




## Aggregation matrix

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

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

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

## Constraint matrix

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

where

$$\begin{aligned} \boldsymbol{C} &= \begin{bmatrix} 1 & -1 & -1 & -1 \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{I}_{n_a} & -\boldsymbol{A} \end{bmatrix} \end{aligned}$$

# Zero-constraint representation

## Aggregation matrix $A$

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

# Zero-constraint representation

## Aggregation matrix $\boldsymbol{A}$

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

## Constraint matrix $\boldsymbol{C}$

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

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

# Zero-constraint representation

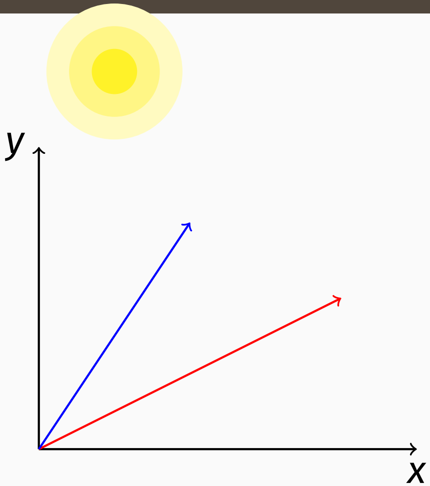
Assuming $\boldsymbol{C}$ is full rank

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

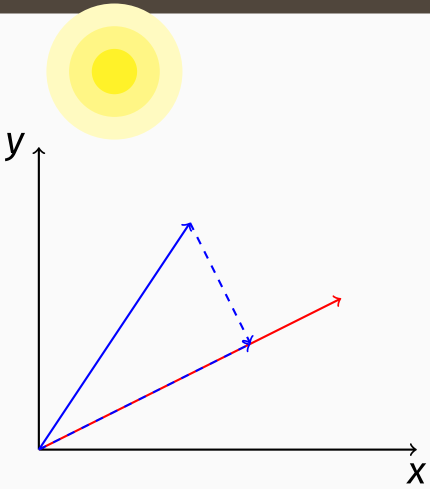
- Originally proved by Byron (1978, 1979) for reconciling data.
- Re-discovered by Wickramasuriya, Athanasopoulos, and Hyndman (2019) for reconciling forecasts.
- $\boldsymbol{M} = \boldsymbol{S}\boldsymbol{G}$ (the MinT solution)
- Leads to more efficient reconciliation than using $\boldsymbol{G}$.

# Outline

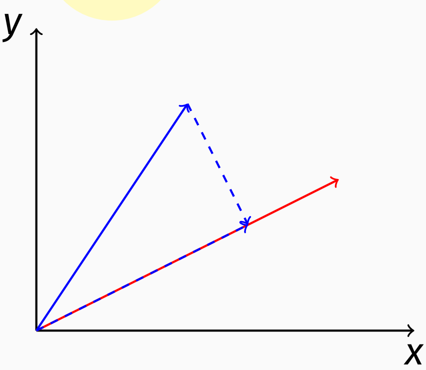
# Projections in linear algebra

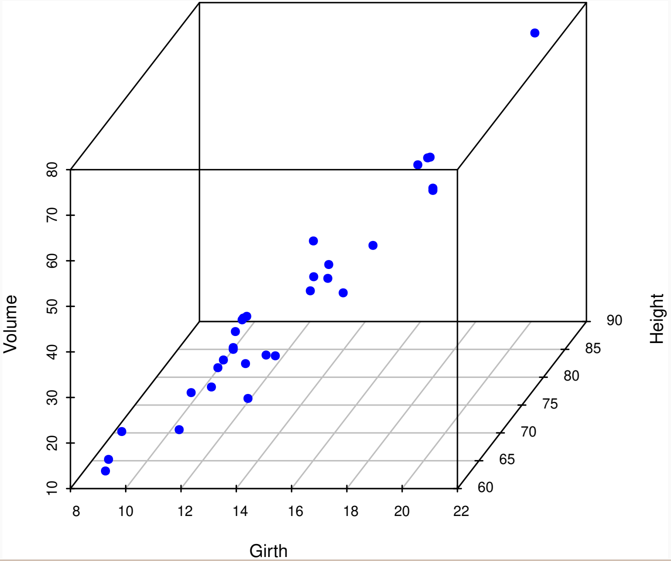


# Projections in linear algebra

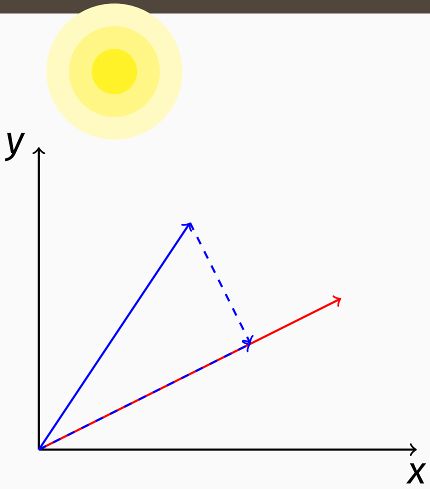


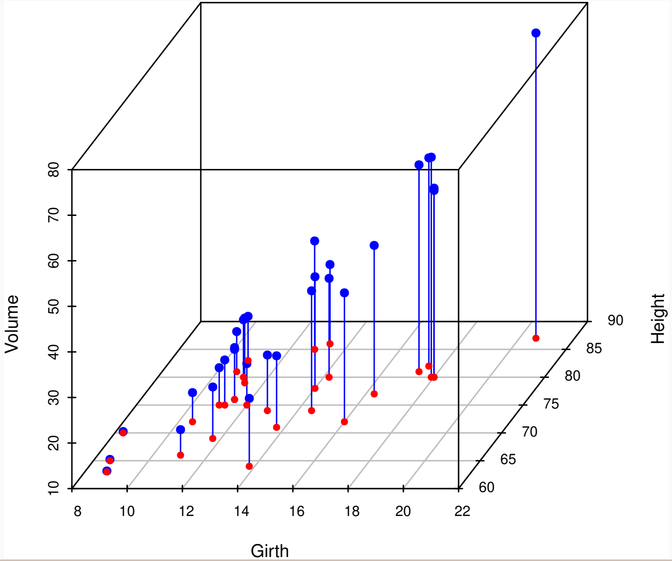
# Projections in linear algebra



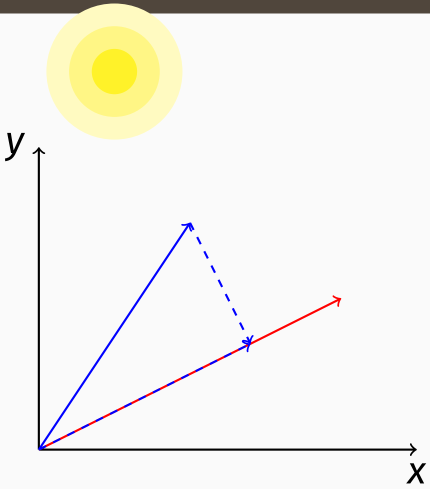



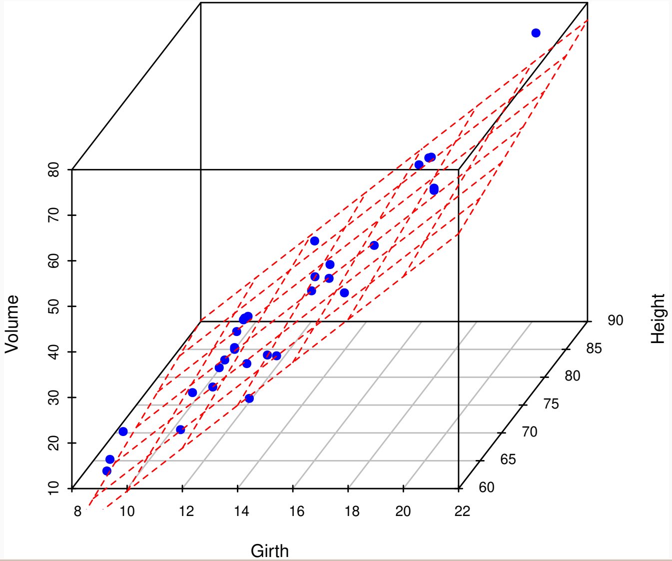
# Projections in linear algebra



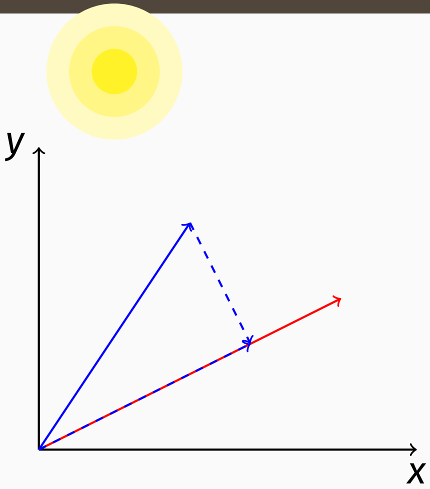



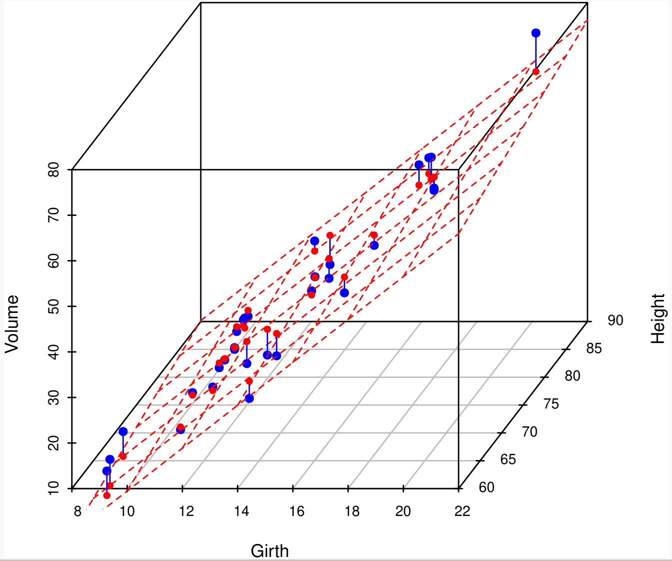
# Projections in linear algebra



# Projections in linear algebra



# Projections in linear algebra

- A projection is a linear transformation $\boldsymbol{M}$ such that $\boldsymbol{M}^2 = \boldsymbol{M}$.
- i.e., $M$ is idempotent: it leaves its image unchanged.
- $\boldsymbol{M}$ projects onto $\mathfrak{s}$ if $\boldsymbol{My} = \boldsymbol{y}$ for all $\boldsymbol{y} \in \mathfrak{s}$.
- All eigenvalues of $\boldsymbol{M}$ are either 0 or 1.
- All singular values of $\boldsymbol{M}$ are greater than or equal to 1 (with equality iff $\boldsymbol{M}$ is orthogonal).
- A projection is *orthogonal* if $\boldsymbol{M}' = \boldsymbol{M}$.
- 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

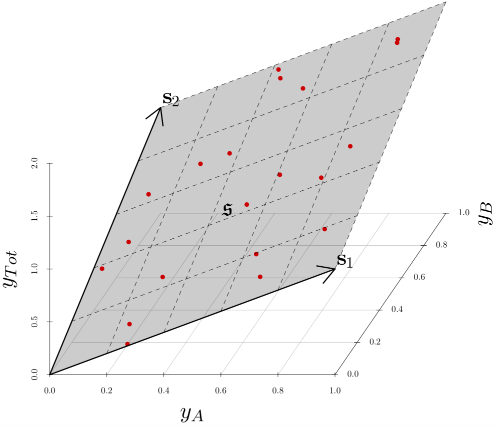
$m$-dimensional linear subspace $\mathfrak{s} \subset \mathbb{R}^n$ for which linear constraints hold for all $\boldsymbol{y} \in \mathfrak{s}$.

## Hierarchical time series

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

## Coherent point forecasts

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




$y_{Tot} = y_A + y_B$

# The coherent subspace

## Coherent subspace

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

## Hierarchical time series

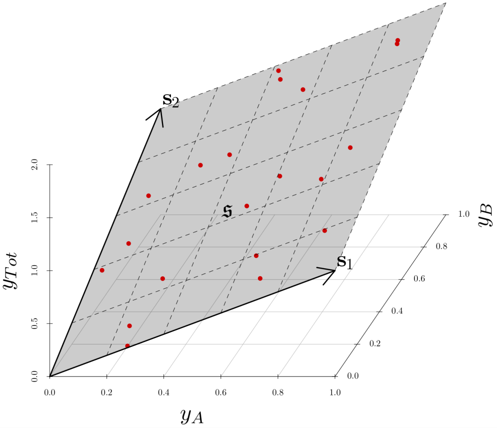
An $n$-dimensional multivariate time series such that $\boldsymbol{y}_t \in \mathfrak{s} \quad \forall t$.

## Coherent point forecasts

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

## Base forecasts

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




$y_{Tot} = y_A + y_B$

# The coherent subspace

## Coherent subspace

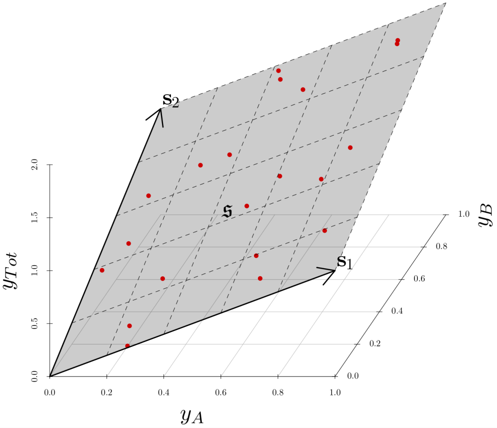
$m$-dimensional linear subspace $\mathfrak{s} \subset \mathbb{R}^n$ for which linear constraints hold for all $\boldsymbol{y} \in \mathfrak{s}$.

## Hierarchical time series

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

## Coherent point forecasts

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




$y_{Tot} = y_A + y_B$

## Base forecasts

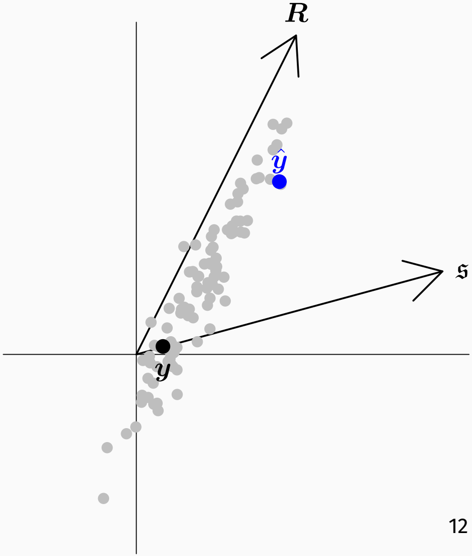
Let $\hat{\boldsymbol{y}}_{t+h|t}$ be vector of *incoherent* initial $h$-step forecasts.

## Reconciled forecasts

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

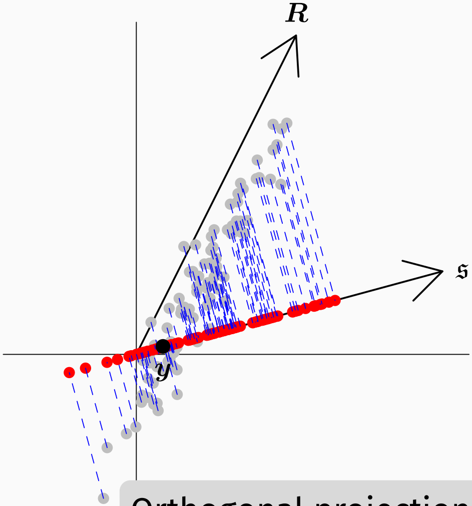
# Linear projection reconciliation

- $R$ is the most likely direction of deviations from $\mathfrak{s}$.
- Grey: potential base forecasts



# Linear projection reconciliation

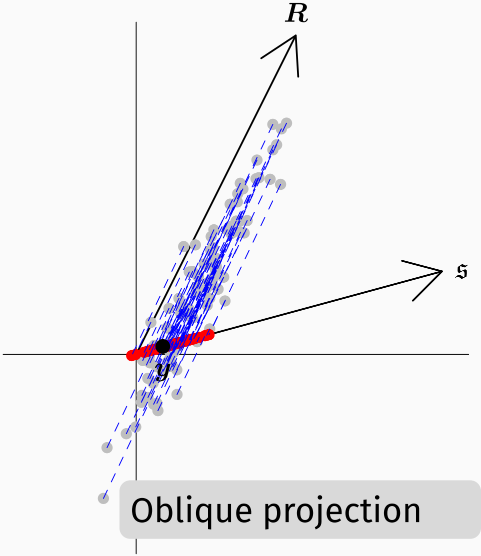
- $R$ is the most likely direction of deviations from $\mathfrak{s}$.
- Grey: potential base forecasts
- Red: reconciled forecasts
- Orthogonal projections (i.e., OLS) lead to smallest possible adjustments of base forecasts.




Orthogonal projection

# Linear projection reconciliation

- $R$ is the most likely direction of deviations from $\mathfrak{s}$.
- 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

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

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

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

  and unbiased estimates must lie on $\mathfrak{s}$.
- The projection is orthogonal if and only if $\boldsymbol{M}' = \boldsymbol{M}$.
- $\boldsymbol{S}$ forms a basis set for $\mathfrak{s}$.
- Projections are of the form $\boldsymbol{M} = \boldsymbol{S}(\boldsymbol{S}'\Psi\boldsymbol{S})^{-1}\boldsymbol{S}'\Psi$ where $\Psi$ is a positive definite matrix.

# Linear projection reconciliation

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

| | | | |
|---|---|---|---|
| OLS: | $\Psi = \boldsymbol{I}$ | $\boldsymbol{M} = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$ | $= \boldsymbol{I} - \boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C}$ |
| MinT: | $\Psi = \boldsymbol{W}_h$ | $\boldsymbol{M} = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}$ | $= \boldsymbol{I} - \boldsymbol{W}_h\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C}$ |

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

# Mean square error bounds

## Distance reducing property

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

$$\|\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_{t+h|t}\|_\Psi \le \|\boldsymbol{y}_{t+h} - \hat{\boldsymbol{y}}_{t+h|t}\|_\Psi$$

- $\Psi$-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.

# Mean square error bounds

Wickramasuriya (2021)

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

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

# Mean square error bounds

Wickramasuriya (2021)

$$
\begin{aligned}
&\operatorname{tr}\Big(\mathrm{E}[\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}^{\text{MinT}}_{t+h|t}]'[\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}^{\text{MinT}}_{t+h|t}]\Big) \\
\leq\ &\operatorname{tr}\Big(\mathrm{E}[\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}^{\text{OLS}}_{t+h|t}]'[\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}^{\text{OLS}}_{t+h|t}]\Big) \\
\leq\ &\operatorname{tr}\Big(\mathrm{E}[\boldsymbol{y}_{t+h} - \hat{\boldsymbol{y}}_{t+h|t}]'[\boldsymbol{y}_{t+h} - \hat{\boldsymbol{y}}_{t+h|t}]\Big)
\end{aligned}
$$

Using sums of variances:

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

# Outline

# Minimum trace reconciliation

## Minimum trace (MinT) reconciliation

If $\boldsymbol{SG}$ is a projection, then the trace of $\boldsymbol{V}_h$ is **minimized** when

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

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

Reconciled forecasts　　　　Base forecasts

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

# A game theoretic perspective

van Erven and Cugliari (2015)

Find the solution to the minimax problem

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

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

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

# Biased reconciliation

Ben Taieb and Koo (2019)

Regularized empirical risk minimization problem:

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

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

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

# Non-negative forecasts

Wickramasuriya, Turlach, and Hyndman (2020)

- How to ensure all forecasts are positive?

# Non-negative forecasts

Wickramasuriya, Turlach, and Hyndman (2020)

- How to ensure all forecasts are positive?

$$\min_{\boldsymbol{G}_h} \operatorname{tr}\Big(\mathrm{E}[\boldsymbol{y}_{t+h} - \boldsymbol{S}\boldsymbol{G}_h\hat{\boldsymbol{y}}_{t+h|t}]'[\boldsymbol{y}_{t+h} - \boldsymbol{S}\boldsymbol{G}_h\hat{\boldsymbol{y}}_{t+h|t}]\Big)$$
$$\text{such that } \boldsymbol{S}\boldsymbol{G}_h\hat{\boldsymbol{y}}_{t+h|t} \geq 0$$

# Non-negative forecasts

Wickramasuriya, Turlach, and Hyndman (2020)

- How to ensure all forecasts are positive?

$$\min_{\boldsymbol{G}_h} \operatorname{tr}\Big(\mathrm{E}[\boldsymbol{y}_{t+h} - \boldsymbol{S}\boldsymbol{G}_h\hat{\boldsymbol{y}}_{t+h|t}]'[\boldsymbol{y}_{t+h} - \boldsymbol{S}\boldsymbol{G}_h\hat{\boldsymbol{y}}_{t+h|t}]\Big)$$
$$\text{such that } \boldsymbol{S}\boldsymbol{G}_h\hat{\boldsymbol{y}}_{t+h|t} \geq 0$$

# Non-negative forecast reconciliation

**Minimum trace (MinT) reconciliation**

The trace of $\boldsymbol{V}_h$ is minimized when

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

subject to unbiasedness preservation ($\boldsymbol{SGS} = \boldsymbol{S}$).

# Non-negative forecast reconciliation

## Minimum trace (MinT) reconciliation

The trace of $\boldsymbol{V}_h$ is minimized when

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

subject to unbiasedness preservation ($\boldsymbol{SGS} = \boldsymbol{S}$).

Wickramasuriya, Turlach and Hyndman (*S&C*, 2020) replace the unbiased constraint by a non-negative constraint:

$$\tilde{\boldsymbol{b}}_{T+h|T} \geq 0$$

and show that it can be solved via quadratic programming:

# Immutable forecasts

Hollyman, Petropoulos, and Tipping (2021)

- How to ensure some forecasts are unchanged?

# Immutable forecasts

Hollyman, Petropoulos, and Tipping (2021)

- How to ensure some forecasts are unchanged?

$$\min_{\boldsymbol{G}_h} \operatorname{tr}\Big(\mathrm{E}[\boldsymbol{y}_{t+h} - \boldsymbol{S}\boldsymbol{G}_h\hat{\boldsymbol{y}}_{t+h|t}]'[\boldsymbol{y}_{t+h} - \boldsymbol{S}\boldsymbol{G}_h\hat{\boldsymbol{y}}_{t+h|t}]\Big)$$
$$\text{such that } \boldsymbol{C}\boldsymbol{S}\boldsymbol{G}_h\hat{\boldsymbol{y}}_{t+h|t} = \boldsymbol{d}$$

- 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

# 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

# Bayesian versions

Novak, McGarvie, and Garcia (2017)

Another strain of the literature brings a Bayesian approach to the regression model interpretation of forecast reconciliation. Novak, McGarvie, and Garcia (2017) recognise that the posterior of $\beta_h$ can act as a probabilistic forecast for the bottom-level series. Using Markov chain Monte Carlo to obtain a sample from this posterior, and then aggregating, gives a probabilistic forecast for the entire hierarchy.

# Bayesian versions

Eckert, Hyndman, and Panagiotelis (2021) also obtain a posterior on $\beta_h$, but their focus is on augmenting the reconciliation regression equation with a vector of intercepts that allow for base forecasts to be biased and evolve according to a state space representation.

Judgement can be incorporated via the prior, in the latter case via an explicit empirical example where prior information about a structural break in data classification can be exploited. Also, while both papers recognise the potential of Bayesian inference to obtain probabilistic forecasts, neither paper

# Bayesian versions

Corani, Azzimonti, Augusto, and Zaffalon (2021) In particular, a prior is placed on the bottom-level series with the mean set to point forecasts obtained in the first step of forecast reconciliation and a variance given by the variance-covariance matrix of one-step ahead errors. This prior is updated using the top-level forecasts obtained in the first stage of forecast reconciliation via Bayes' rule. The method generalises MinT in the sense that the posterior mean is equivalent to the usual MinT approach. The necessary updates via Bayes' rule have parallels with the Kalman filter since the reconciliation problem is recast as a linear Gaussian model. The empirical

# Outline

# In-built coherence

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

as shown by Ashouri, Hyndman, and Shmueli (2022), if the base forecasts $\hat{\boldsymbol{y}}$ are produced using a linear regression model, the base forecasts and reconciliation can be combined, giving coherent forecasts directly in a single closed form equation. Further, the computation is extremely fast provided sparse matrix algebra is used.

# In-built coherence

Pennings and Dalen (2017) propose the state space model

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{\mu}_t + \boldsymbol{Z}_t\beta + \varepsilon_t, \qquad \varepsilon_t \sim N(\boldsymbol{0}, \Sigma_\varepsilon), \tag{1}$$

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \eta_t, \qquad \eta_t \sim N(\boldsymbol{0}, \Sigma_\eta). \tag{2}$$

- 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. *Journal of 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.
- Corani, G, D Azzimonti, JPSC Augusto, and M Zaffalon (2021). Probabilistic reconciliation of hierarchical forecast via Bayes' rule. In: *Machine Learning and Knowledge Discovery in Databases*. Ed. by F Hutter, K Kersting, J Lijffijt, and I Valera. Cham: Springer International Publishing, pp.211–226.
- Corani, G, N Rubattu, D Azzimonti, and A Antonucci (2022). "Probabilistic reconciliation of count time series". https://arxiv.org/abs/2207.09322.

# References


- Di Fonzo, T and D Girolimetto (2022). Forecast combination-based forecast reconciliation: Insights and extensions. *International Journal of Forecasting* **forthcoming**.
- Di Fonzo, T and D Girolimetto (2023). Spatio-temporal reconciliation of solar forecasts. *Solar Energy* **251**, 13–29.
- Eckert, F, RJ Hyndman, and A Panagiotelis (2021). Forecasting Swiss exports using Bayesian forecast reconciliation. *European J Operational Research* **291**(2), 693–710.
- 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.

# References


- Novak, J, S McGarvie, and BE Garcia (2017). "A Bayesian model for forecasting hierarchically structured time series". https://arxiv.org/abs/1711.04738.
- 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.
- 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.

# References


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