Forecast reconciliation

2. Perspectives on forecast reconciliation

Rob J Hyndman

MONASH University

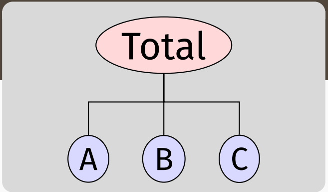
Photo by Edvard Alexander Rølvaag on

robjhyndman.com/fr2023

Outline

Outline

Notation reminder



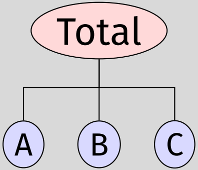
Every collection of time series with linear constraints can be written as

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

- $\boldsymbol{y}_t$ = 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



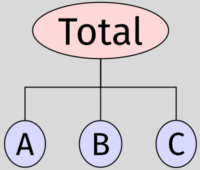
Every collection of time series with linear constraints can be written as

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

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

Zero-constraint representation

Aggregation matrix $\boldsymbol{A}$

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

Zero-constraint representation

Aggregation matrix A

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

Constraint matrix 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

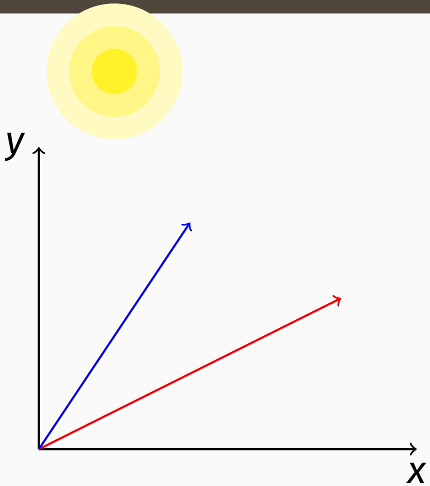
Assuming $\boldsymbol{C}$ is full rank

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

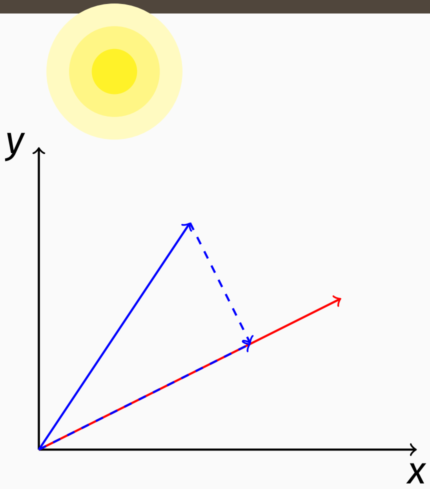
- Originally proved by Byron (1978, 1979) for reconciling data.
- Re-discovered by Wickramasuriya, Athanasopoulos, and Hyndman (2019) for reconciling forecasts.
- $\boldsymbol{M} = \boldsymbol{S}\boldsymbol{G}$ (the MinT solution)
- Leads to more efficient reconciliation than using $\boldsymbol{G}$.

Outline

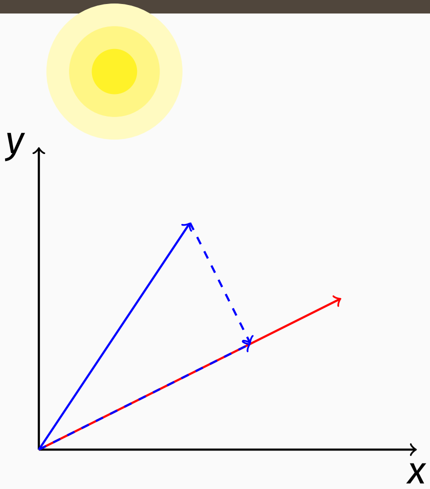
Projections in linear algebra

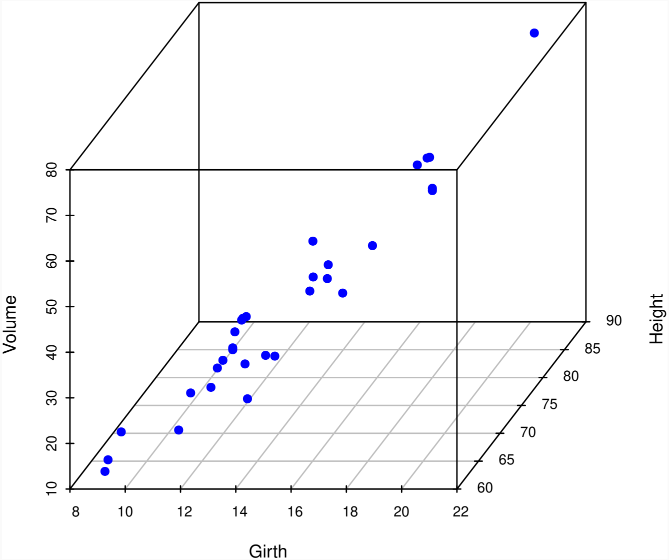


Projections in linear algebra

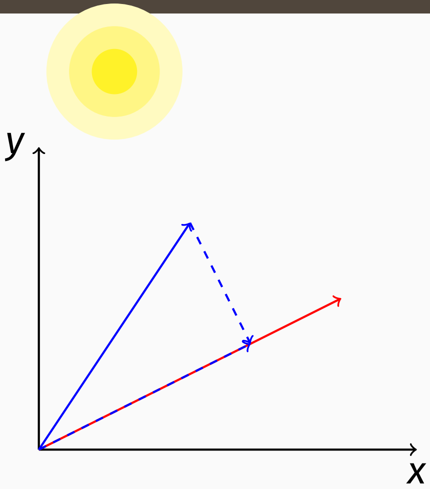


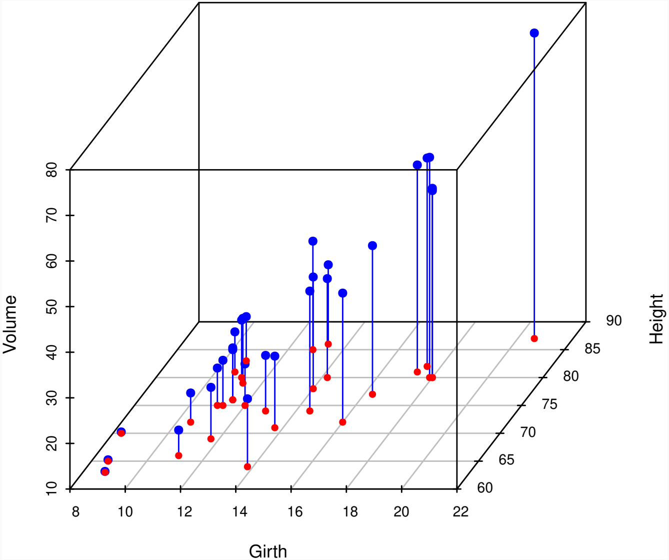
Projections in linear algebra



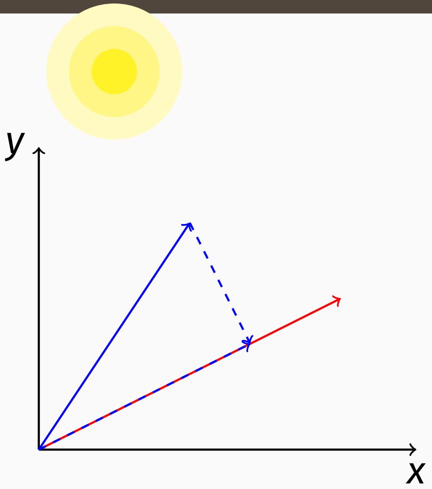


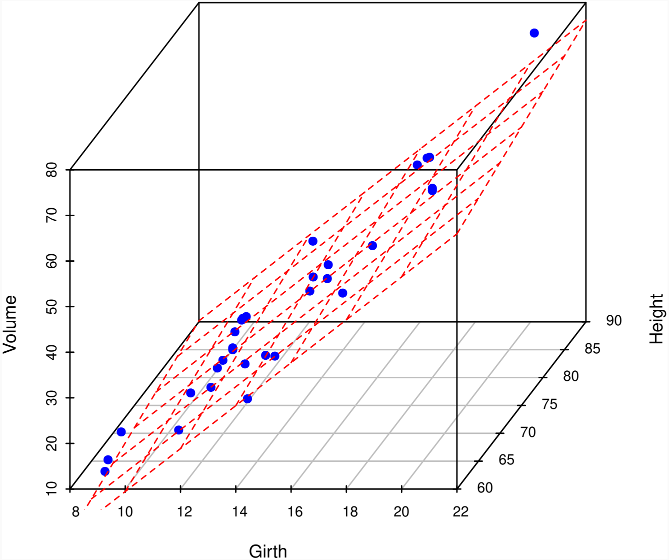
Projections in linear algebra



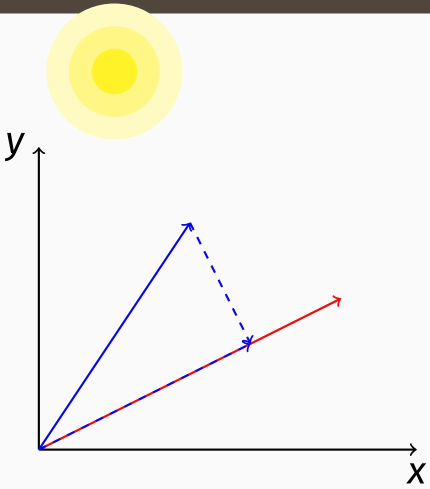


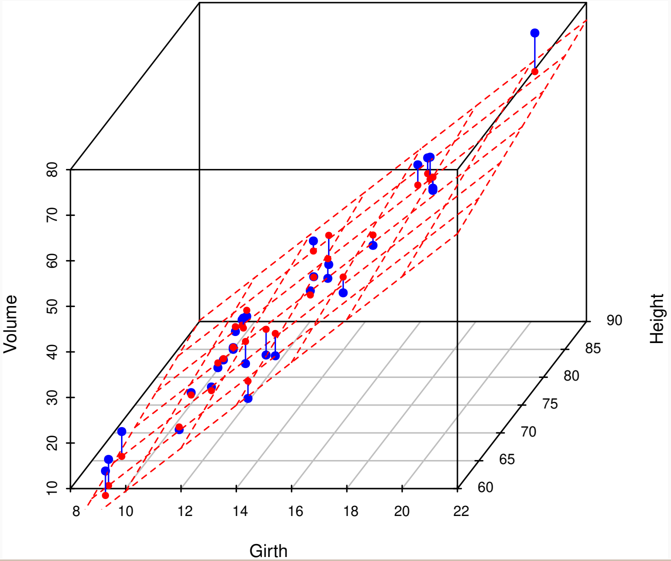
Projections in linear algebra





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 $\boldsymbol{M}\boldsymbol{y} = \boldsymbol{y}$ 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

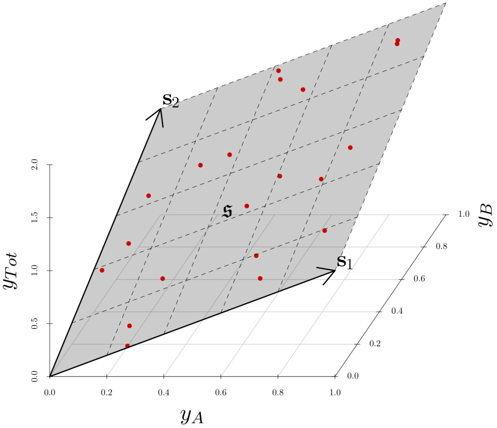
m-dimensional linear subspace $\mathfrak{s} \subset \mathbb{R}^n$ for which linear constraints hold for all $\boldsymbol{y} \in \mathfrak{s}$.

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

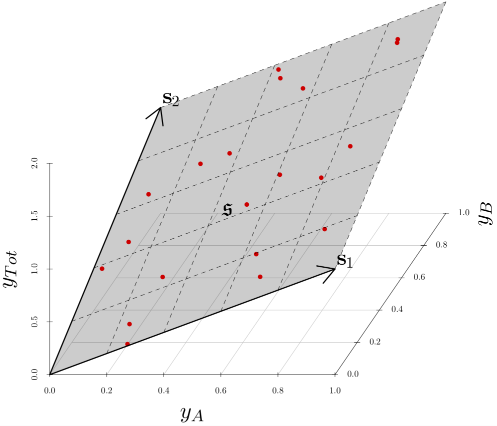
An n-dimensional multivariate time series such that $\boldsymbol{y}_t \in \mathfrak{s} \quad \forall t$.

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

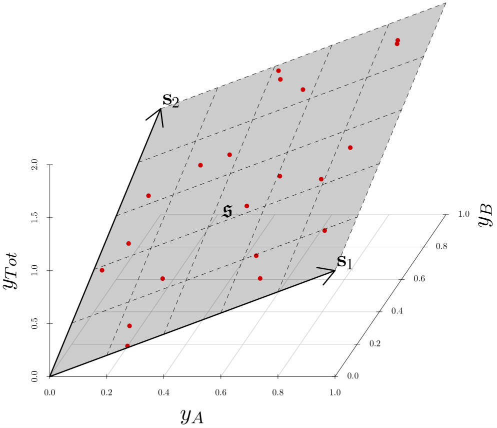
m-dimensional linear subspace $\mathfrak{s} \subset \mathbb{R}^n$ for which linear constraints hold for all $\boldsymbol{y} \in \mathfrak{s}$.

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

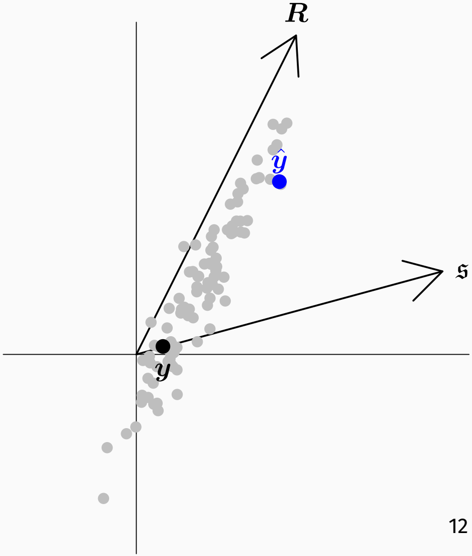
Let $\hat{\boldsymbol{y}}_{t+h|t}$ be vector of *incoherent* initial h-step forecasts.

Reconciled forecasts

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

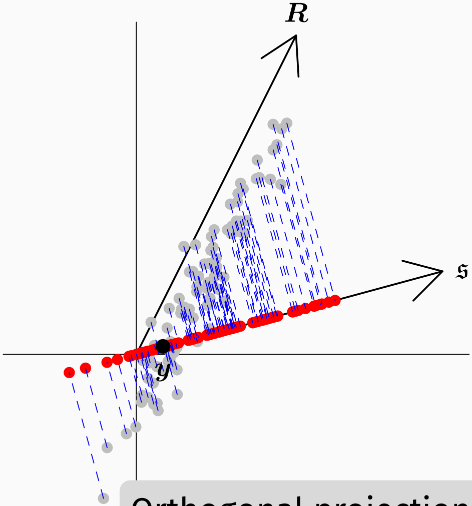
Linear projection reconciliation

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



Linear projection reconciliation

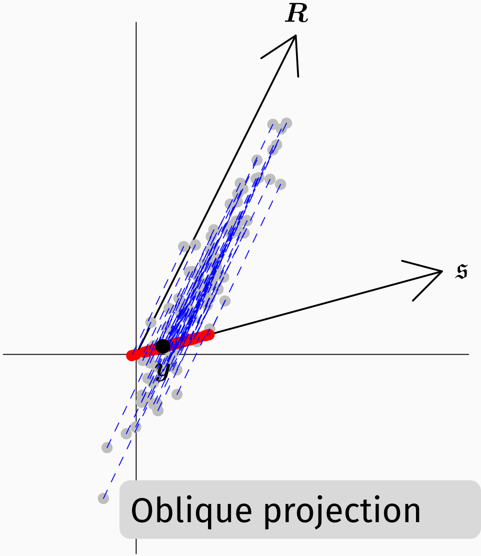
- R is the most likely direction of deviations from $\mathfrak{s}$.
- Grey: potential base forecasts
- Red: reconciled forecasts
- Orthogonal projections (i.e., OLS) lead to smallest possible adjustments of base forecasts.



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 Ψ 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 $\Psi = \boldsymbol{W}_h$.

Outline

Mean square error bounds

Distance reducing property

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

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

- Ψ-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)

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

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

Mean square error bounds

Wickramasuriya (2021)

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

Using sums of variances:

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

Outline

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 ℓ 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 Ψ 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 Ψ.

Biased reconciliation

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
- $\boldsymbol{Y} = [\boldsymbol{y}_{T_1+h}, \dots, \boldsymbol{y}_T]'$
- $\hat{\boldsymbol{Y}} = [\hat{\boldsymbol{y}}_{T_1+h|T_1}, \dots, \hat{\boldsymbol{y}}_{T|T-h}]'$
- λ is a regularization parameter.

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

Unconstrained MinT

Wickramasuriya (2021)

Include?

Outline

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?

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

with **WicEtAl2020** providing an early example for forecast reconciliation, and **di2023spatio** a more recent example.

di2023spatio also discuss an effective nonnegative heuristic called “set-negative-to-zero”, whereby the negative reconciled forecasts at the bottom level are set to zero, and the remaining forecasts computed via aggregation.

Immutble forecasts

HolEtAl2021

- How to ensure some forecasts are unchanged?

Immutble forecasts

HolEtAl2021

- How to ensure some forecasts are unchanged?

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

- Differs from top-down approaches in that it can be done while also preserving the unbiasedness of base forecasts.
- To briefly illustrate the main idea, for a three variable hierarchy where $y_{Tot,t} = y_{A,t} + y_{B,t}$, either setting

$$\begin{pmatrix}\tilde{y}_{Tot,t}\end{pmatrix} \quad \begin{pmatrix}\hat{y}_{Tot,t}\end{pmatrix} \qquad \begin{pmatrix}\tilde{y}_{Tot,t}\end{pmatrix} \quad \begin{pmatrix}\hat{y}_{Tot,t}\end{pmatrix}$$

Outline

ML and regularization

Outline

Bayesian versions

Outline

In-built coherence

References

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

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.
- 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.