

Forecast reconciliation



Rob J Hyndman

 MONASH University



robjhyndman.com/fr2023

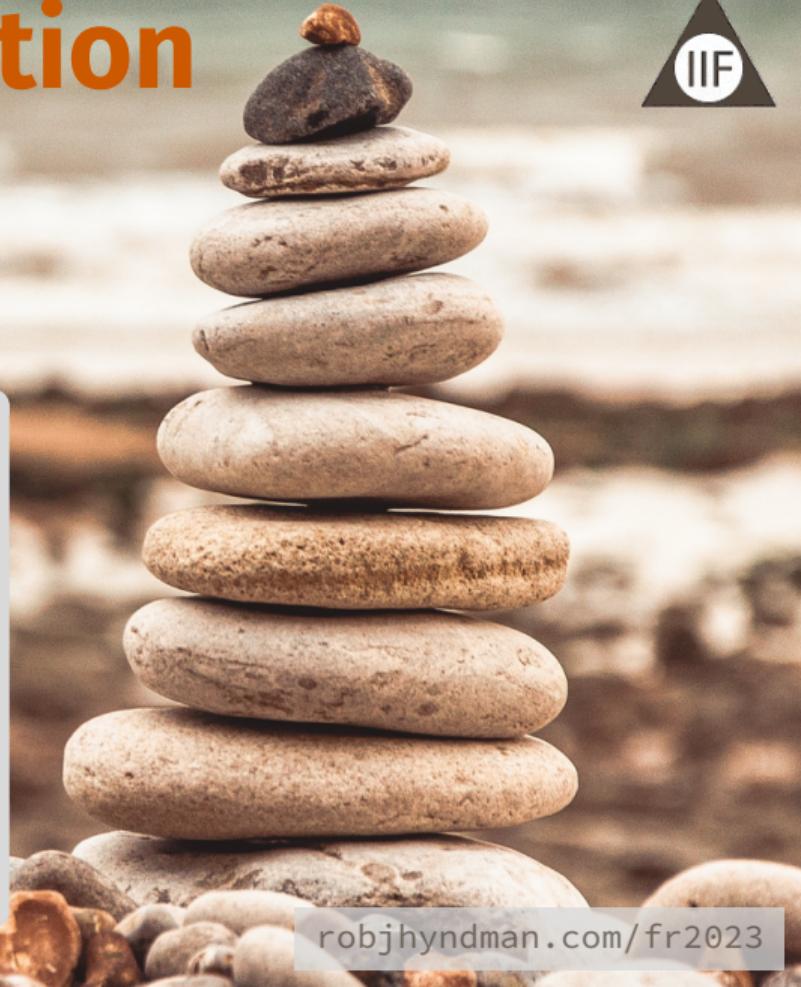
Forecast reconciliation



Rob J Hyndman

 MONASH University

- 1 Hierarchical time series and forecast reconciliation (Nov 6/7)
- 2 **Perspectives on forecast reconciliation** (Nov 9/10)
- 3 Probabilistic forecast reconciliation (Nov 13/14)
- 4 Temporal and cross-temporal forecast reconciliation (Nov 16/17)



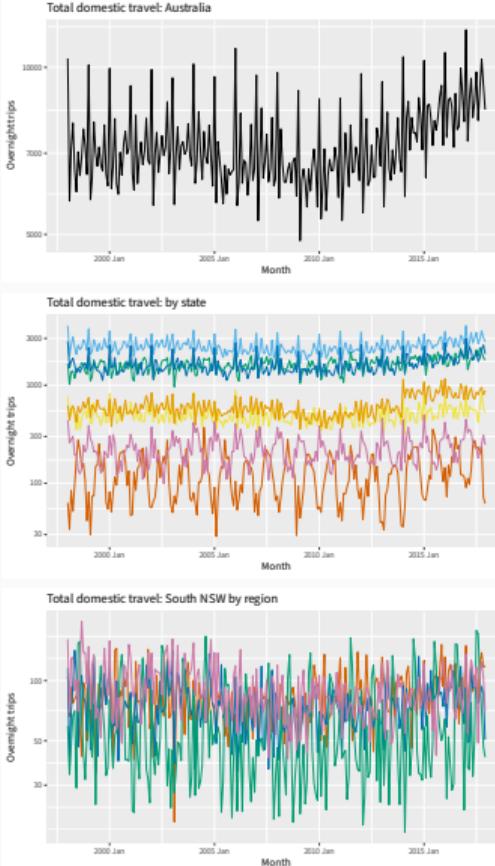
Outline

- 1 Recap
- 2 Time series reconciliation
- 3 Reconciliation via constraints
- 4 Example: reconciling GDP forecasts
- 5 The geometry of forecast reconciliation
- 6 Optimization and reconciliation
- 7 In-built coherence
- 8 Time series cross-validation

Outline

- 1 Recap
- 2 Time series reconciliation
- 3 Reconciliation via constraints
- 4 Example: reconciling GDP forecasts
- 5 The geometry of forecast reconciliation
- 6 Optimization and reconciliation
- 7 In-built coherence
- 8 Time series cross-validation

Recap



- We want to forecast many time series that are related due to aggregation (or linear) constraints.
- Forecast reconciliation involves forecasting all series, and then adjusting the forecasts to impose the constraints.
- The optimal (L_2) solution is known as "Minimum Trace" reconciliation.

Outline

1 Recap

2 Time series reconciliation

3 Reconciliation via constraints

4 Example: reconciling GDP forecasts

5 The geometry of forecast reconciliation

6 Optimization and reconciliation

7 In-built coherence

8 Time series cross-validation

Time series reconciliation

- Stone, Champernowne & Meade (1942): reconciling national economic accounts (disaggregated into production, income, outlay, capital transactions, etc.)
- Byron (1978): extended Stone's work using more computationally efficient methods.
- 1984: Stone wins Nobel Prize in Economics.
- Dagum & Cholette (2006): Same approach used for reconciling seasonally adjusted data.
- Chow & Lin (1971): Temporal reconciliation of monthly or quarterly estimates to sum to annual estimates.
- Di Fonzo (1990): Cross-temporal reconciliation of time series data.

Outline

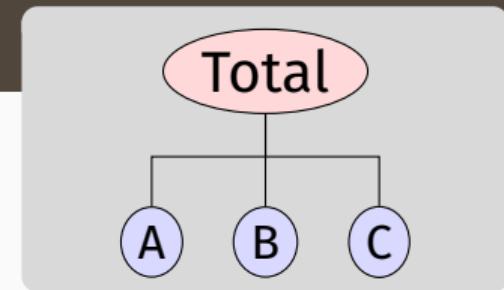
- 1 Recap
- 2 Time series reconciliation
- 3 Reconciliation via constraints
- 4 Example: reconciling GDP forecasts
- 5 The geometry of forecast reconciliation
- 6 Optimization and reconciliation
- 7 In-built coherence
- 8 Time series cross-validation

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.



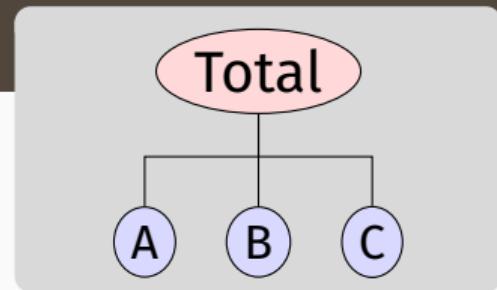
$$\begin{aligned} \mathbf{y}_t &= \begin{pmatrix} y_{\text{Total},t} \\ y_{A,t} \\ y_{B,t} \\ y_{C,t} \end{pmatrix} \\ &= \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_S \underbrace{\begin{pmatrix} y_{A,t} \\ y_{B,t} \\ y_{C,t} \end{pmatrix}}_{\mathbf{b}_t}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

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.

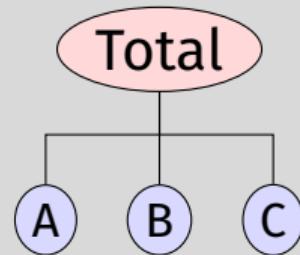
Notation

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

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

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

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, the optimal (MinT) solution is

$$\tilde{\mathbf{y}}_{T+h|T} = \mathbf{M}\hat{\mathbf{y}}_{T+h|T}$$

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

- Originally derived by Byron (1978) for reconciling data.
- Re-discovered by Wickramasuriya, Athanasopoulos, and Hyndman (2019) for reconciling forecasts.
- Leads to more efficient reconciliation than using

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

Zero-constraint representation

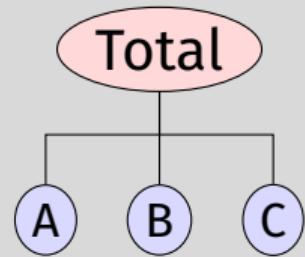
Suppose $\mathbf{W}_h = \mathbf{I}$. Then

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

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

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

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



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

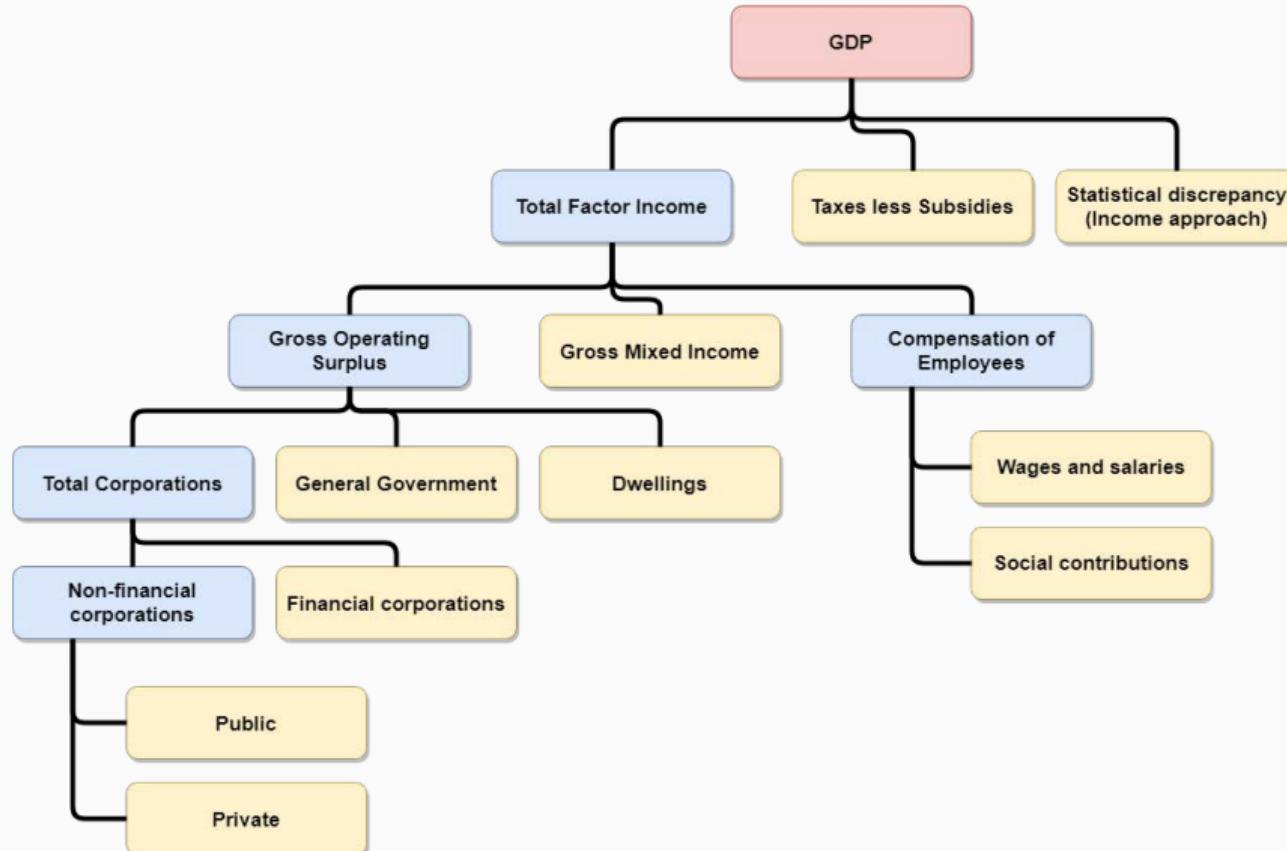
$$\mathbf{s} = \begin{pmatrix} \mathbf{A} \\ \mathbf{I}_{n_b} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

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

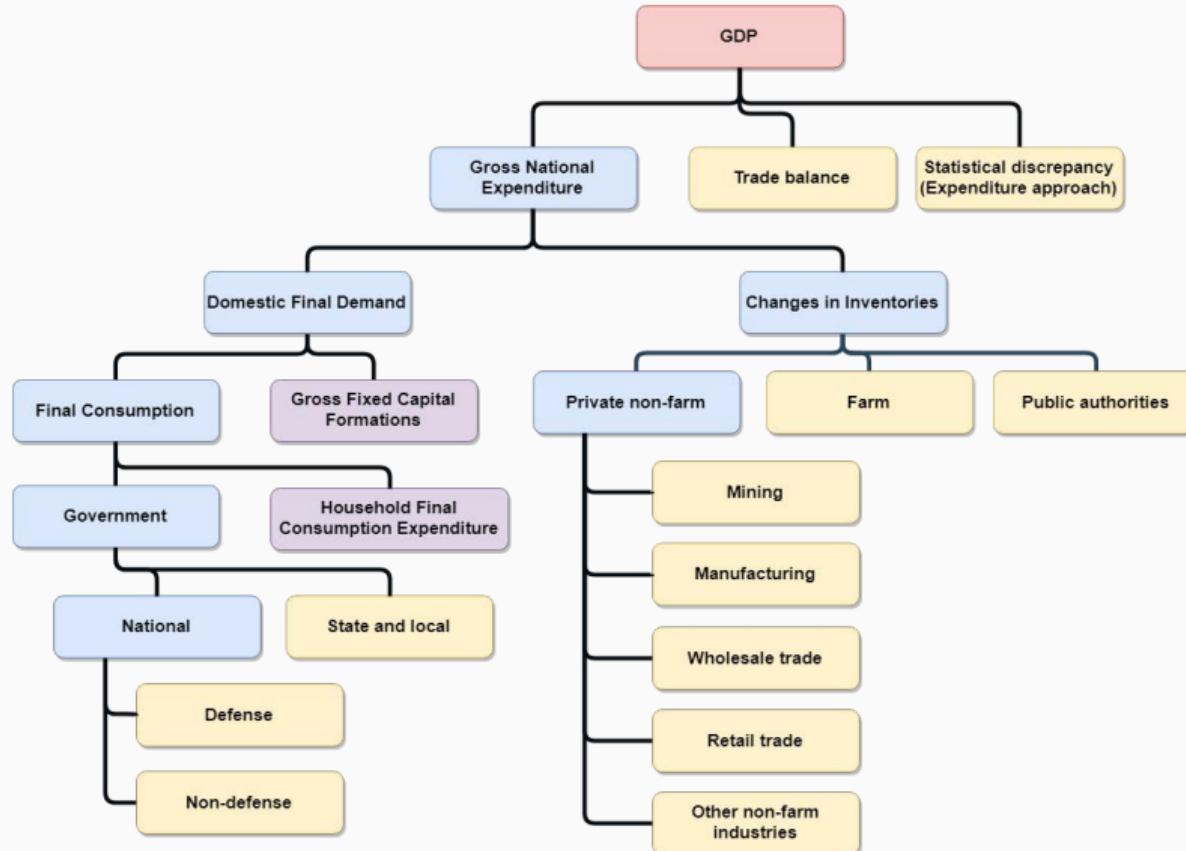
Outline

- 1 Recap
- 2 Time series reconciliation
- 3 Reconciliation via constraints
- 4 Example: reconciling GDP forecasts
- 5 The geometry of forecast reconciliation
- 6 Optimization and reconciliation
- 7 In-built coherence
- 8 Time series cross-validation

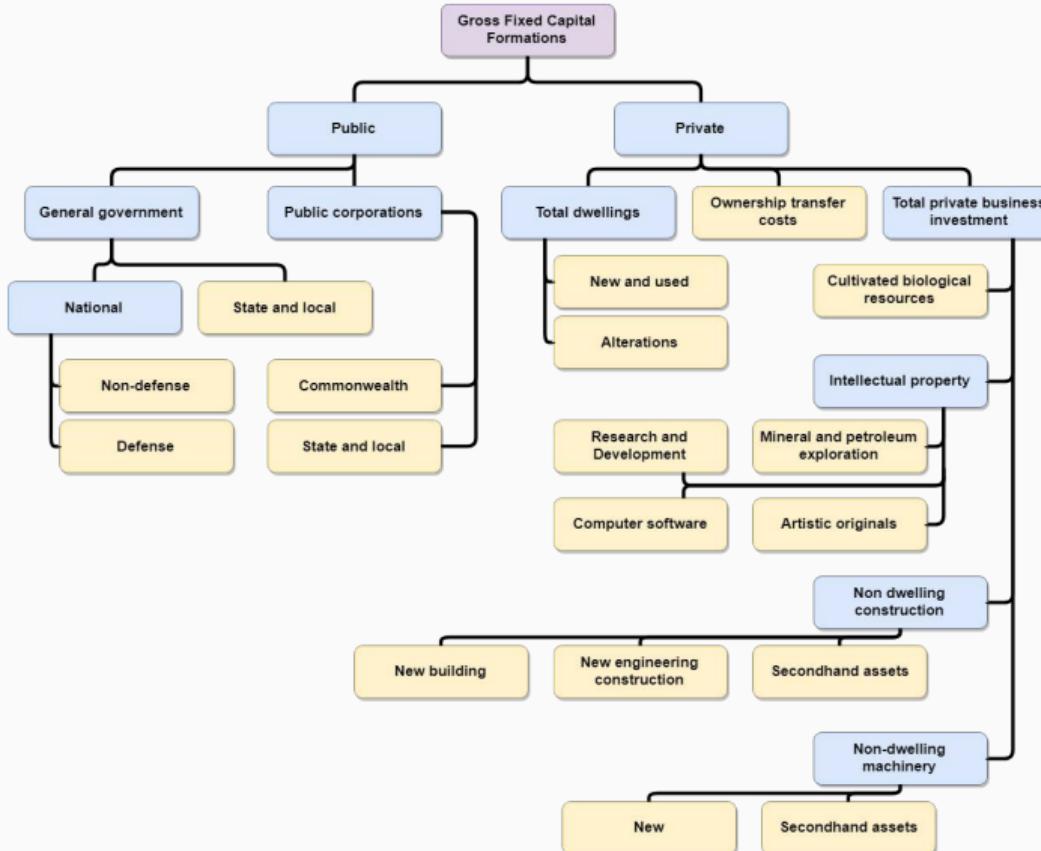
Example: reconciling GDP forecasts



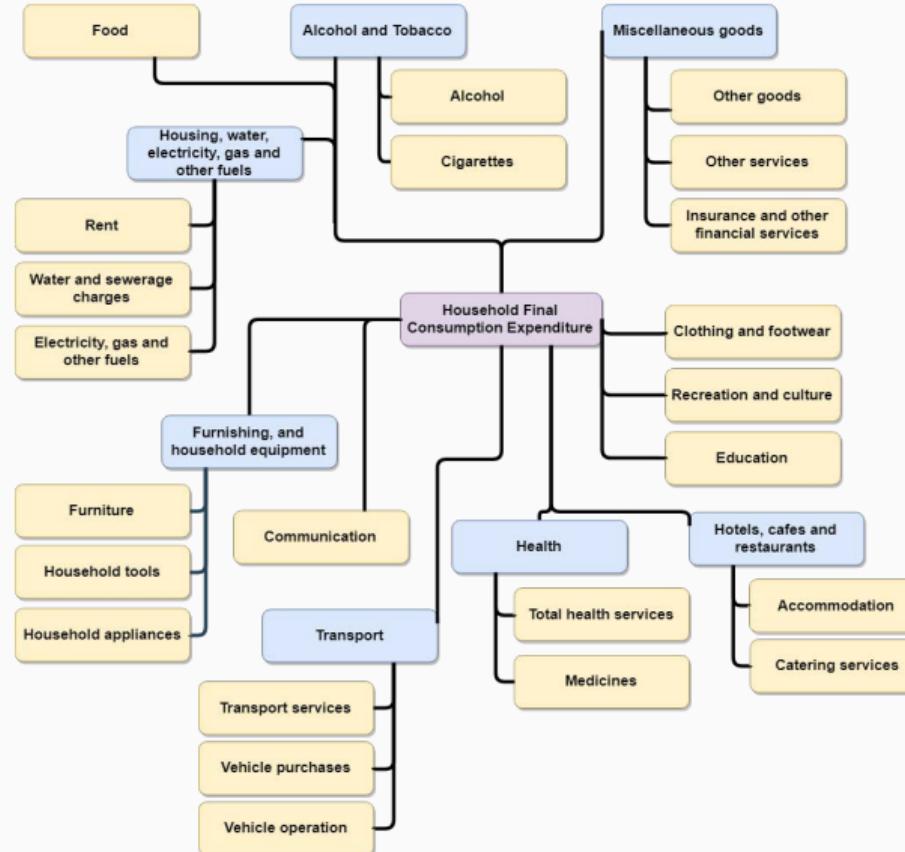
Example: reconciling GDP forecasts



Example: reconciling GDP forecasts



Example: reconciling GDP forecasts



Example: reconciling GDP forecasts

- No unique hierarchy.
- Several disaggregations with the same parent node
- Not possible to represent using structural **S** notation.
- Instead, we can use the constraint **C** notation.

Example: reconciling GDP forecasts

Using structural notation:

$$\mathbf{y}_t^I = \begin{bmatrix} X_t \\ \mathbf{a}_t^I \\ \mathbf{b}_t^I \end{bmatrix} = \mathbf{S}^I \mathbf{b}_t^I \quad \mathbf{y}_t^E = \begin{bmatrix} X_t \\ \mathbf{a}_t^E \\ \mathbf{b}_t^E \end{bmatrix} = \mathbf{S}^E \mathbf{b}_t^E$$

where

$$\mathbf{S}^I = \begin{bmatrix} \mathbf{1}'_{10} \\ \mathbf{A}' \\ \mathbf{I}_{10} \end{bmatrix} \quad \mathbf{S}^E = \begin{bmatrix} \mathbf{1}'_{53} \\ \mathbf{A}^E \\ \mathbf{I}_{53} \end{bmatrix}$$

- Can reconcile both trees, but the totals won't be equal.

Example: reconciling GDP forecasts

Using constraint notation:

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

where

$$\mathbf{y}_t = \begin{bmatrix} x_t \\ \mathbf{a}'_t \\ \mathbf{b}'_t \\ \mathbf{a}^E_t \\ \mathbf{b}^E_t \end{bmatrix} \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} 1 & \mathbf{0}'_5 & -\mathbf{1}'_{10} & \mathbf{0}'_{26} & \mathbf{0}'_{53} \\ 1 & \mathbf{0}'_5 & \mathbf{0}'_{10} & \mathbf{0}'_{26} & -\mathbf{1}'_{53} \\ \mathbf{0}_5 & \mathbf{I}_5 & -\mathbf{A}' & \mathbf{0}_{5 \times 26} & \mathbf{0}_{5 \times 53} \\ \mathbf{0}_{26} & \mathbf{0}_{26 \times 5} & \mathbf{0}_{26 \times 10} & \mathbf{I}_{26} & -\mathbf{A}^E \end{bmatrix}$$

Ref: Bisaglia, Di Fonzo, and Girolimetto (2020)

Outline

- 1 Recap
- 2 Time series reconciliation
- 3 Reconciliation via constraints
- 4 Example: reconciling GDP forecasts
- 5 The geometry of forecast reconciliation
- 6 Optimization and reconciliation
- 7 In-built coherence
- 8 Time series cross-validation

The coherent subspace

Coherent subspace

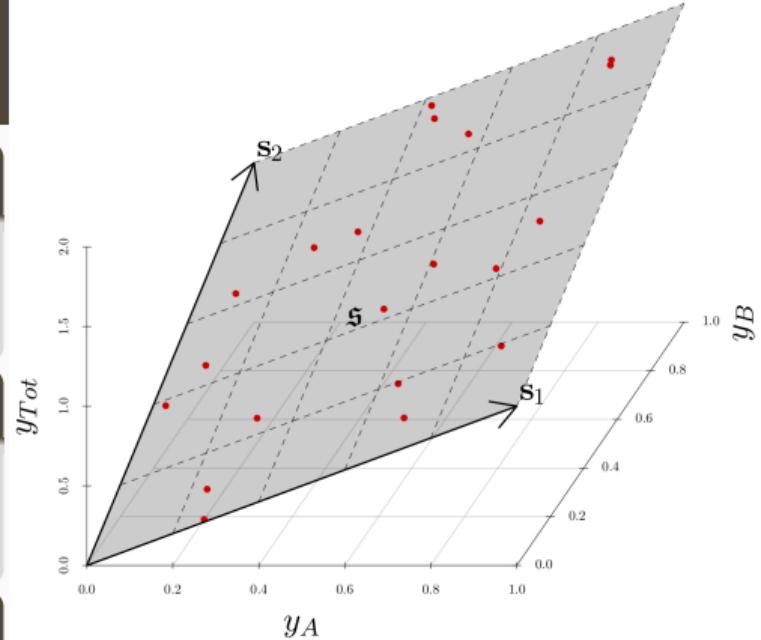
n_b -dimensional linear subspace $\mathfrak{s} \subset \chi^n$ for which linear constraints hold for all $\mathbf{y} \in \mathfrak{s}$.

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

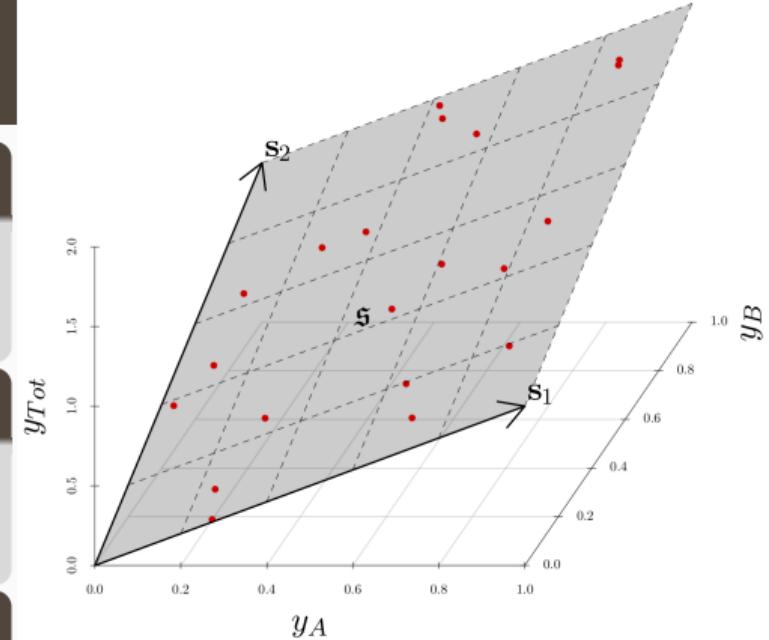
n_b -dimensional linear subspace $\mathfrak{s} \subset \chi^n$ for which linear constraints hold for all $\mathbf{y} \in \mathfrak{s}$.

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



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

Base forecasts

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

The coherent subspace

Coherent subspace

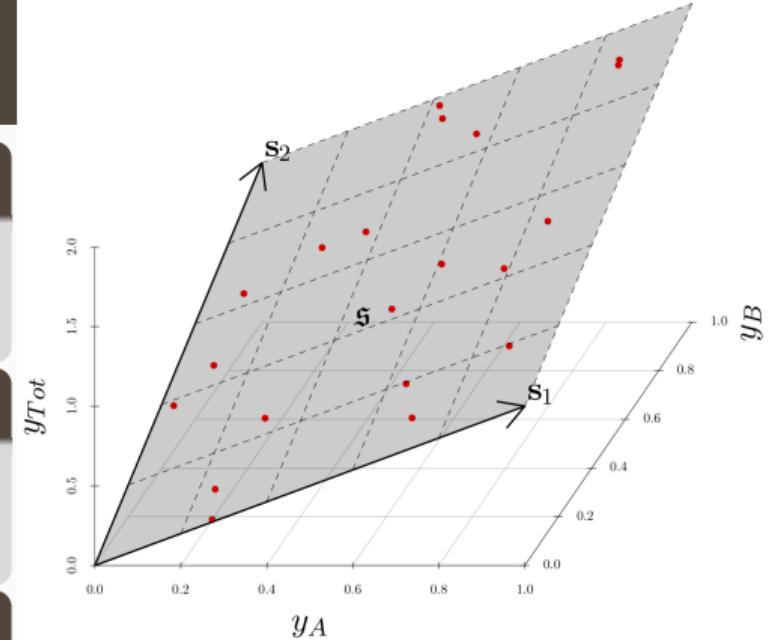
n_b -dimensional linear subspace $\mathfrak{s} \subset \chi^n$ for which linear constraints hold for all $\mathbf{y} \in \mathfrak{s}$.

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



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

Base forecasts

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

Reconciled forecasts

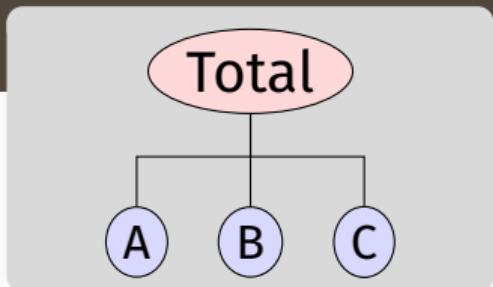
Let ψ be a mapping, $\psi : \chi^n \rightarrow \mathfrak{s}$.
 $\tilde{\mathbf{y}}_{t+h|t} = \psi(\hat{\mathbf{y}}_{t+h|t})$ “reconciles” $\hat{\mathbf{y}}_{t+h|t}$.

The coherent subspace

The columns of \mathbf{S} form a basis set for \mathfrak{s} .

They are not unique.

Each corresponds to different vector of “bottom-level” series.

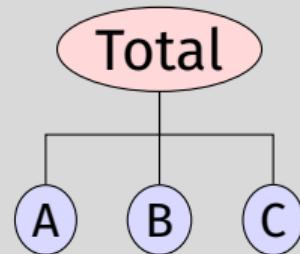


The coherent subspace

The columns of \mathbf{S} form a basis set for \mathfrak{s} .

They are not unique.

Each corresponds to different vector of “bottom-level” series.



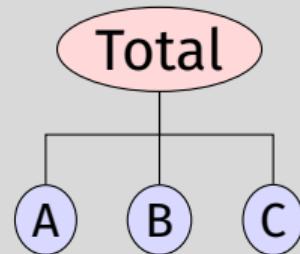
$$\mathbf{y} = \begin{pmatrix} \text{Total} \\ A \\ B \\ C \end{pmatrix} \quad \mathbf{S} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} A \\ B \\ C \end{pmatrix}$$

The coherent subspace

The columns of \mathbf{S} form a basis set for \mathfrak{s} .

They are not unique.

Each corresponds to different vector of “bottom-level” series.



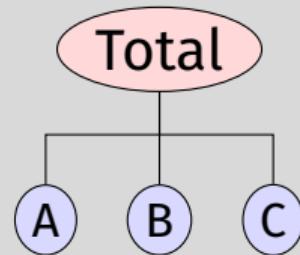
$$\mathbf{y} = \begin{pmatrix} \text{Total} \\ A \\ B \\ C \end{pmatrix} \quad \mathbf{S} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & -1 & -1 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} \text{Total} \\ B \\ A \end{pmatrix}$$

The coherent subspace

The columns of \mathbf{S} form a basis set for \mathfrak{s} .

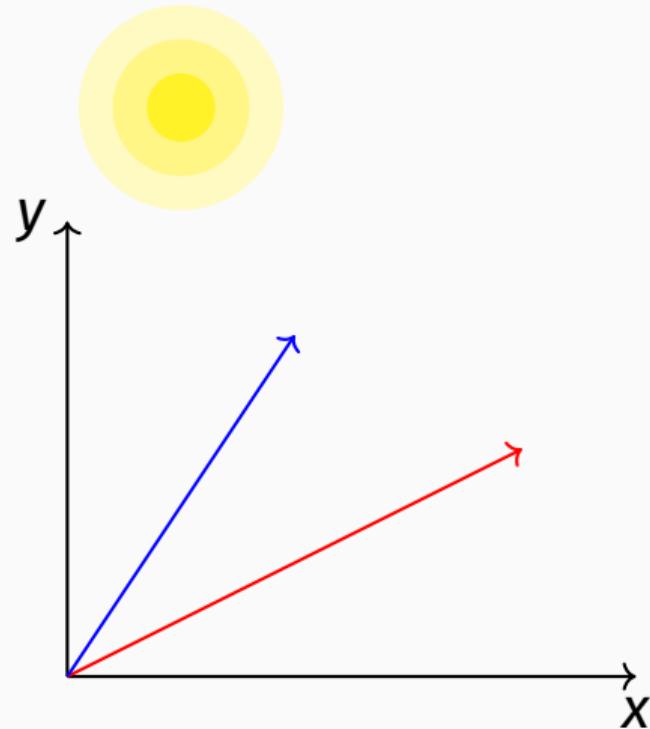
They are not unique.

Each corresponds to different vector of “bottom-level” series.

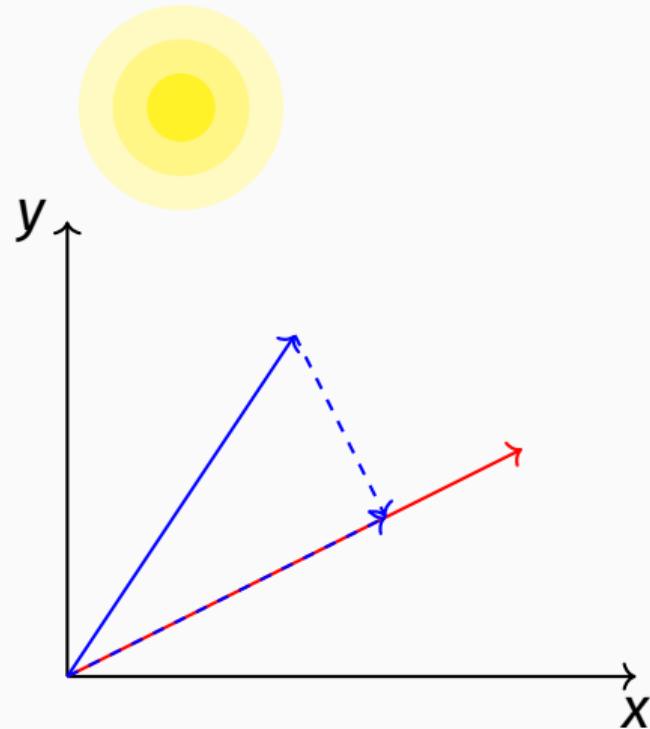


$$\mathbf{y} = \begin{pmatrix} \text{Total} \\ A \\ B \\ C \end{pmatrix} \quad \mathbf{S} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & -1 \\ -1 & 1 & 1 \\ 1 & -1 & 0 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} \text{Total} \\ B + A \\ C + B \end{pmatrix}$$

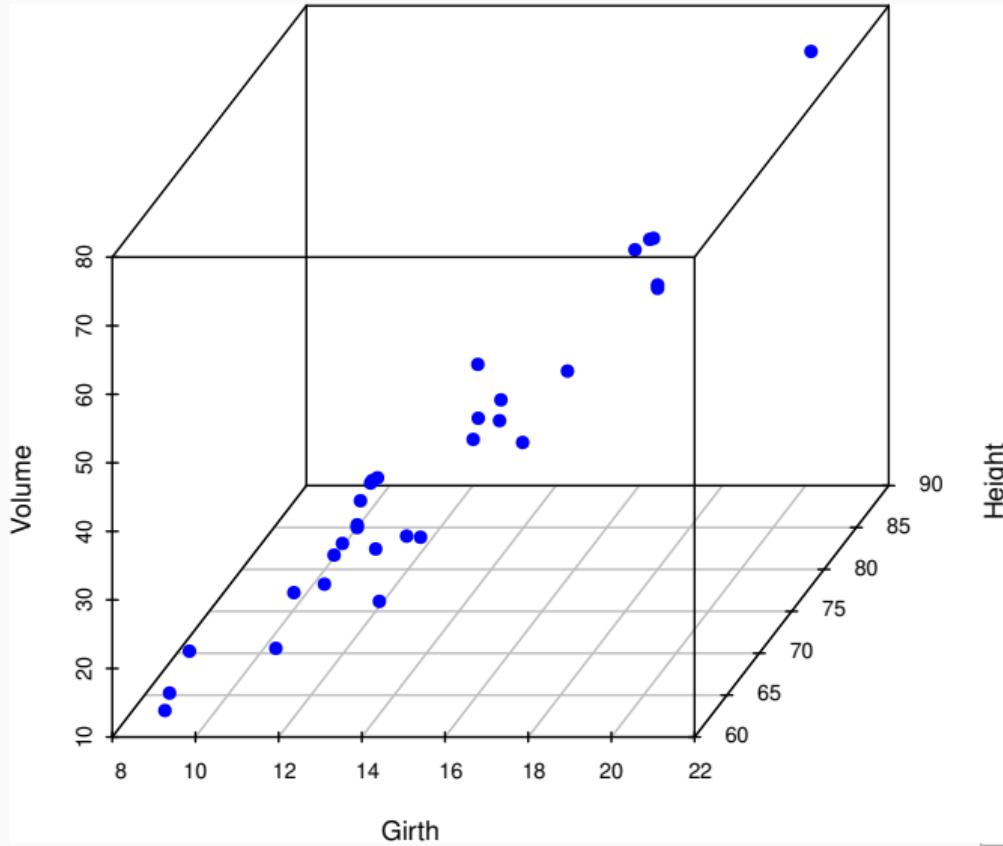
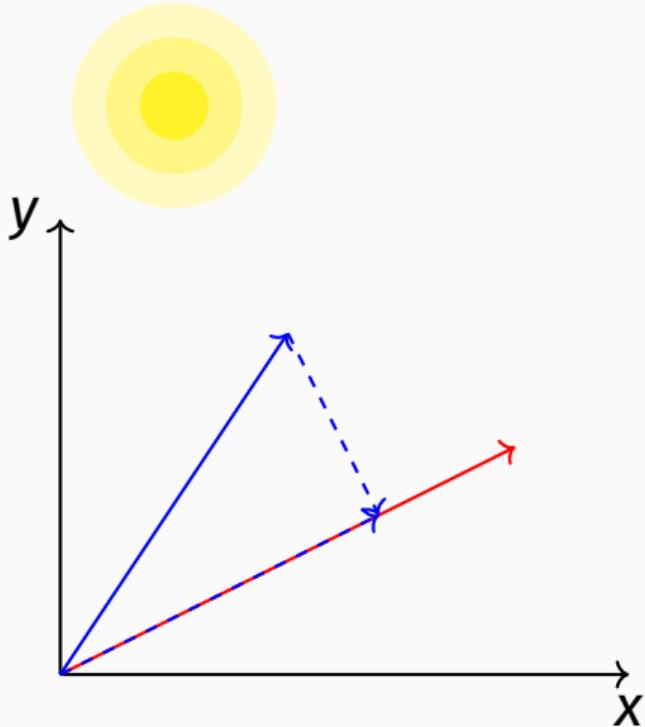
Projections in linear algebra



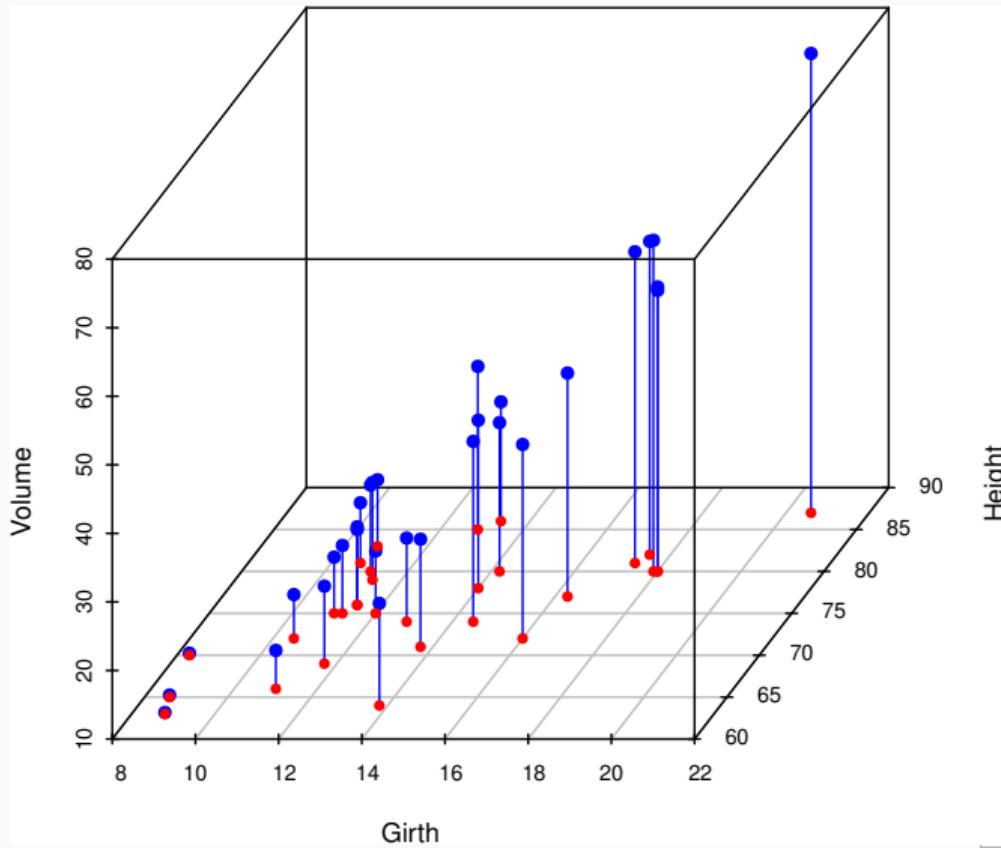
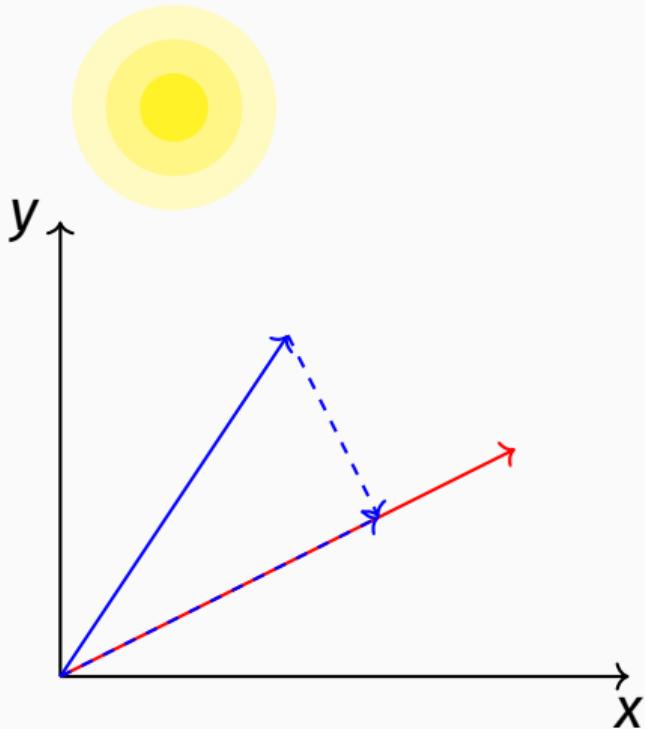
Projections in linear algebra



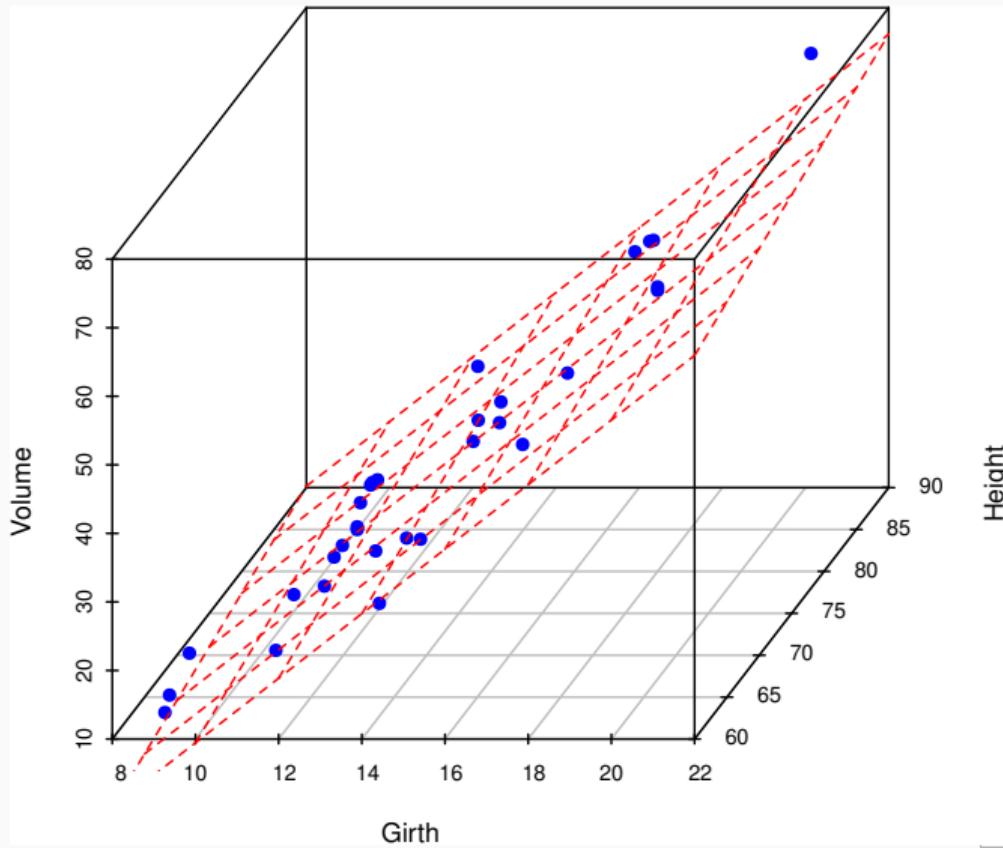
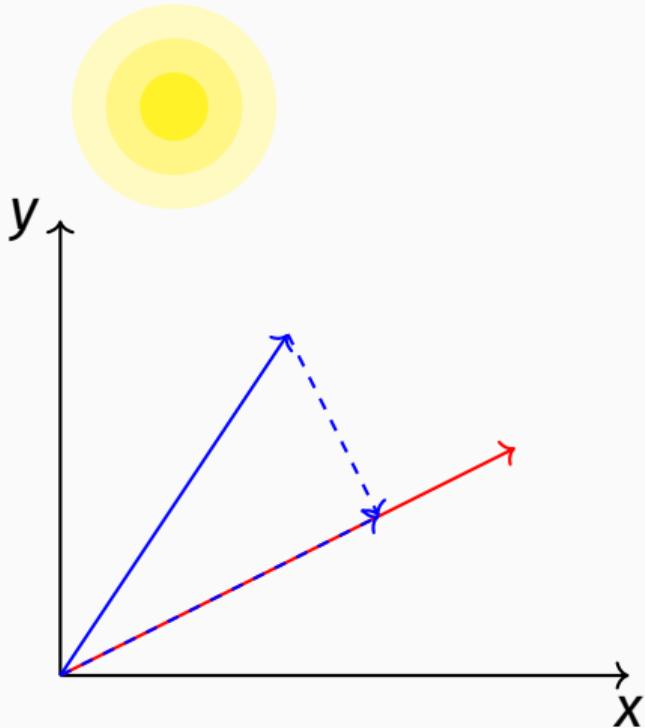
Projections in linear algebra



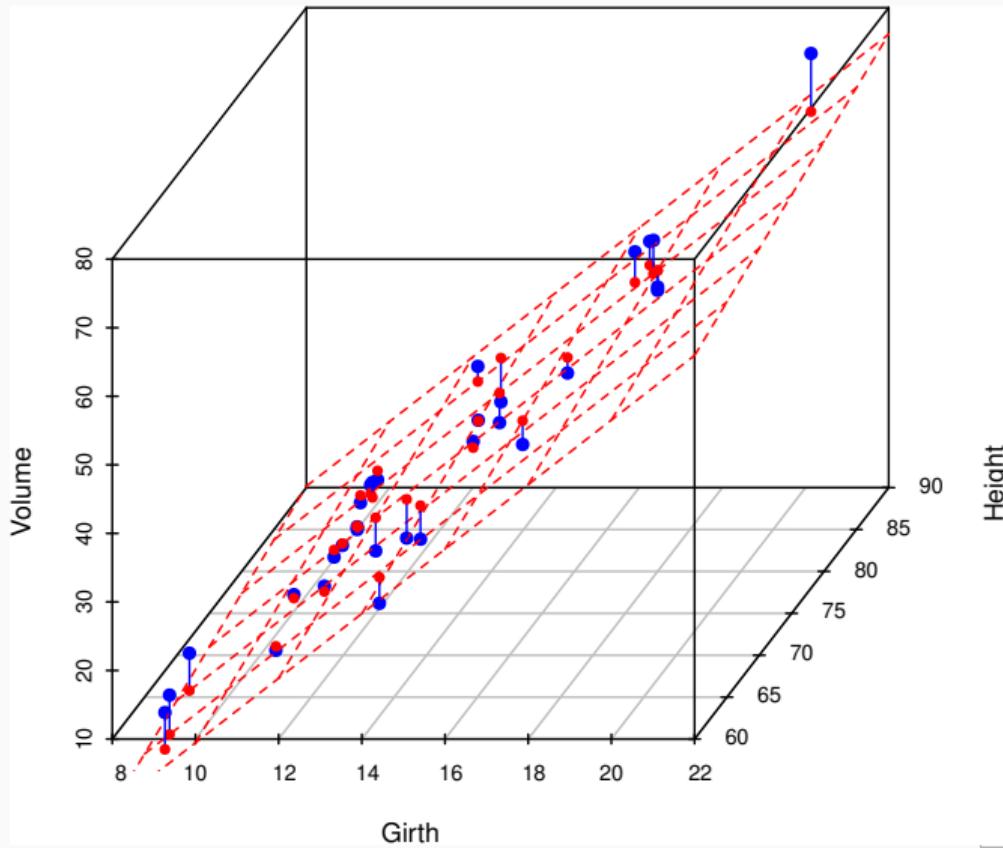
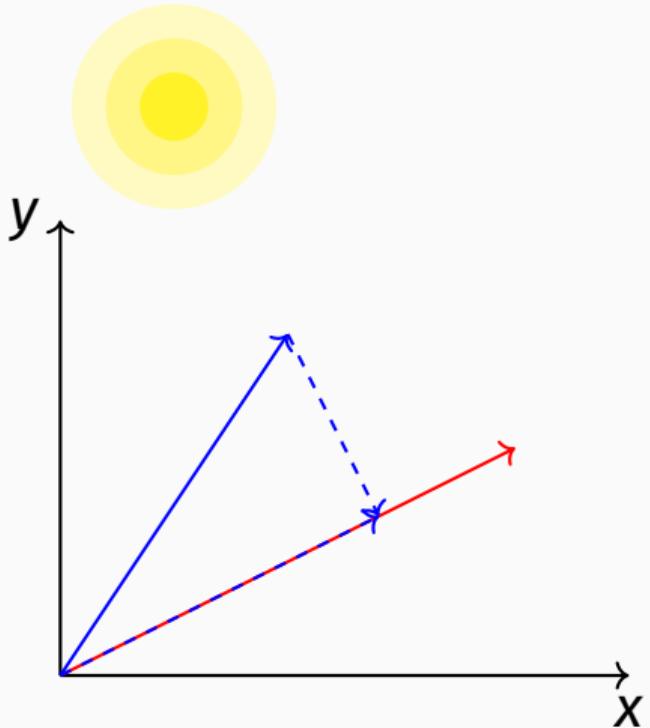
Projections in linear algebra



Projections in linear algebra



Projections in linear algebra

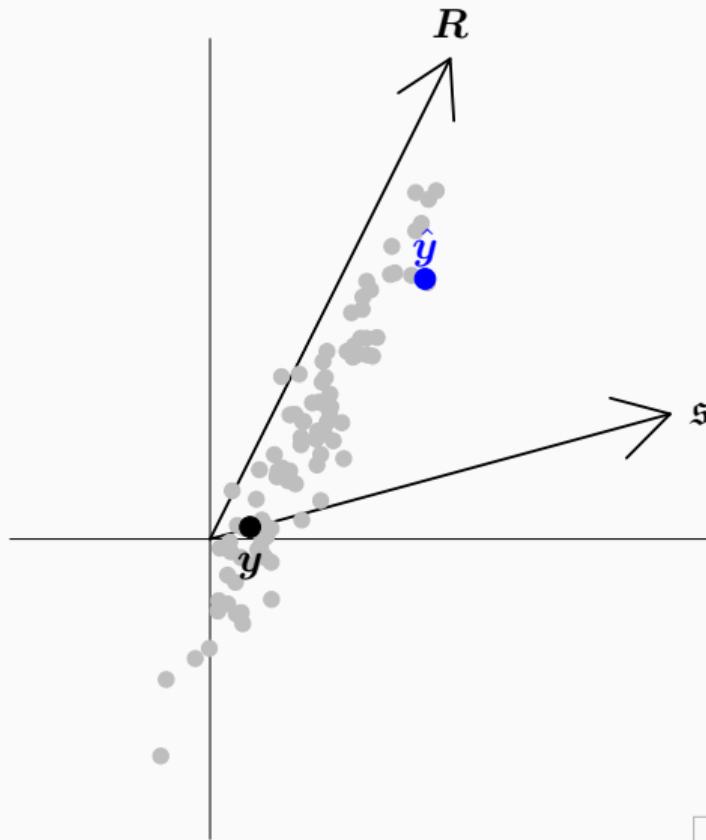


Projections in linear algebra

- 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.

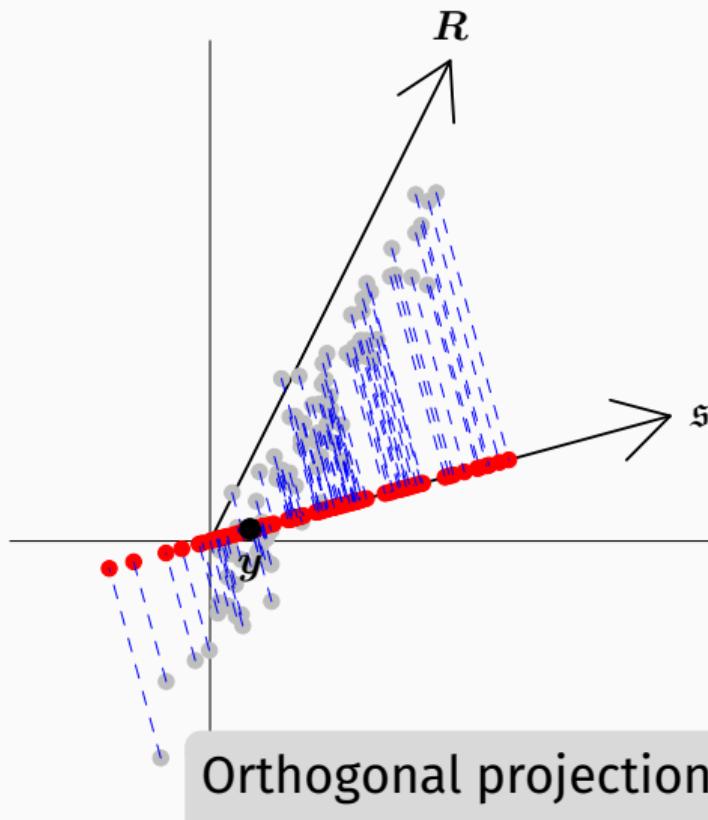
Linear projection reconciliation

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



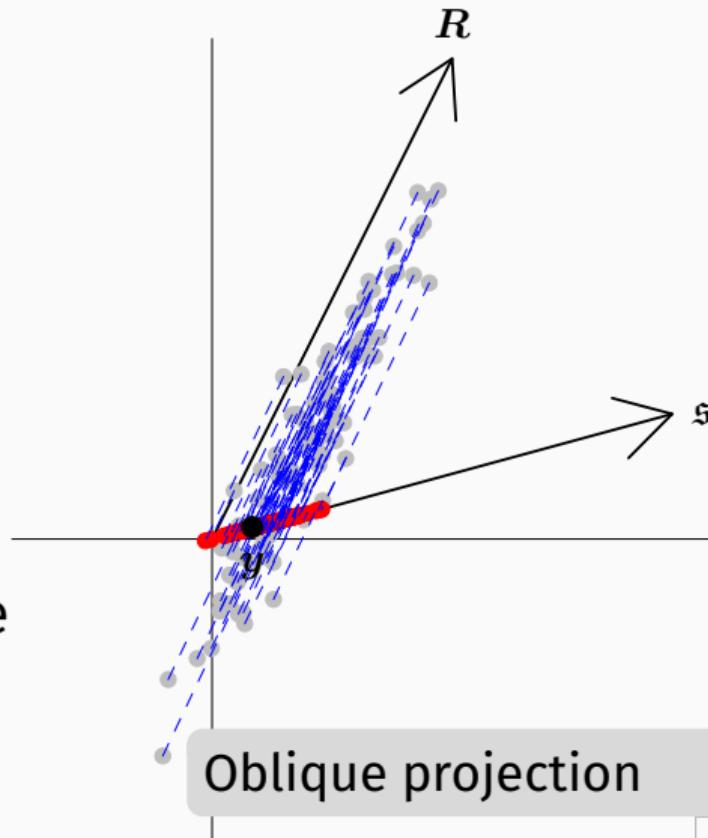
Linear projection reconciliation

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

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

- \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 Ψ is a positive definite matrix.

Linear projection reconciliation

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

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

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

- \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 $\Psi = \mathbf{W}_h$.

Distance reducing property

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

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

- 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.

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

- σ_{\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 Recap
- 2 Time series reconciliation
- 3 Reconciliation via constraints
- 4 Example: reconciling GDP forecasts
- 5 The geometry of forecast reconciliation
- 6 Optimization and reconciliation
- 7 In-built coherence
- 8 Time series cross-validation

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

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

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

- $V \leq 0$: reconciliation guaranteed to reduce loss.
- If $\ell(\mathbf{y}, \tilde{\mathbf{y}}) = \|\mathbf{y} - \tilde{\mathbf{y}}\|_\Psi = (\mathbf{y} - \tilde{\mathbf{y}})' \Psi (\mathbf{y} - \tilde{\mathbf{y}})$, where Ψ 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 Ψ .

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
- $\mathbf{Y} = [\mathbf{y}_{T_1+h}, \dots, \mathbf{y}_T]'$
- $\hat{\mathbf{Y}} = [\hat{\mathbf{y}}_{T_1+h|T_1}, \dots, \hat{\mathbf{y}}_{T|T-h}]'$
- λ is a regularization parameter
- When $\lambda = 0$: $\hat{\mathbf{G}} = \mathbf{B}'\hat{\mathbf{Y}}(\hat{\mathbf{Y}}'\hat{\mathbf{Y}})^{-1}$ where $\mathbf{B} = [\mathbf{b}_{T_1+h}, \dots, \mathbf{b}_T]'$.
- Theory assumes series are stationary.

MinT expressed as a regression

Since $\tilde{\mathbf{b}}_{t+h|t} = (\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h^{-1}\hat{\mathbf{y}}_{t+h|t}$, we can write the MinT solution as a regression problem:

$$\begin{aligned}\tilde{\mathbf{b}}_{t+h|t} &= \arg \min_{\mathbf{b}} [\hat{\mathbf{y}}_{t+h|t} - \mathbf{S}\mathbf{b}]' \mathbf{W}_h^{-1} [\hat{\mathbf{y}}_{t+h|t} - \mathbf{S}\mathbf{b}] \\ &= \arg \min_{\mathbf{b}} [\mathbf{b}' \mathbf{S}' \mathbf{W}_h^{-1} \mathbf{S} \mathbf{b} - 2 \mathbf{b}' \mathbf{S}' \mathbf{W}_h^{-1} \hat{\mathbf{y}}_{t+h|t} + \hat{\mathbf{y}}_{t+h|t}' \mathbf{W}_h^{-1} \hat{\mathbf{y}}_{t+h|t}] \\ &= \arg \min_{\mathbf{b}} [\mathbf{b}' \mathbf{S}' \mathbf{W}_h^{-1} \mathbf{S} \mathbf{b} - 2 \mathbf{b}' \mathbf{S}' \mathbf{W}_h^{-1} \hat{\mathbf{y}}_{t+h|t}]\end{aligned}$$

- MinT solution is equivalent to a GLS regression of $\hat{\mathbf{y}}_{t+h|t}$ on \mathbf{S} with covariance weights \mathbf{W}_h^{-1} .
- The estimated coefficients are the forecasts of the bottom level series.

Non-negative forecasts

$$\begin{aligned} \min_{\mathbf{G}_h} \text{tr} & \left(\mathbb{E}[\mathbf{y}_{t+h} - \mathbf{S}\mathbf{G}_h \hat{\mathbf{y}}_{t+h|t}]' [\mathbf{y}_{t+h} - \mathbf{S}\mathbf{G}_h \hat{\mathbf{y}}_{t+h|t}] \right) \\ \text{such that } & \mathbf{b}_{t+h|t} = \mathbf{G}_h \hat{\mathbf{y}}_{t+h|t} \geq 0 \end{aligned}$$

Non-negative forecasts

$$\begin{aligned} \min_{\mathbf{G}_h} \text{tr} & \left(\mathbb{E}[\mathbf{y}_{t+h} - \mathbf{S}\mathbf{G}_h \hat{\mathbf{y}}_{t+h|t}]' [\mathbf{y}_{t+h} - \mathbf{S}\mathbf{G}_h \hat{\mathbf{y}}_{t+h|t}] \right) \\ \text{such that } & \mathbf{b}_{t+h|t} = \mathbf{G}_h \hat{\mathbf{y}}_{t+h|t} \geq 0 \end{aligned}$$

Solve via quadratic programming:

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

(Wickramasuriya, Turlach, and Hyndman, 2020)

Non-negative forecasts

$$\begin{aligned} \min_{\mathbf{G}_h} \text{tr} & \left(\mathbb{E}[\mathbf{y}_{t+h} - \mathbf{S}\mathbf{G}_h \hat{\mathbf{y}}_{t+h|t}]' [\mathbf{y}_{t+h} - \mathbf{S}\mathbf{G}_h \hat{\mathbf{y}}_{t+h|t}] \right) \\ \text{such that } & \mathbf{b}_{t+h|t} = \mathbf{G}_h \hat{\mathbf{y}}_{t+h|t} \geq 0 \end{aligned}$$

Solve via quadratic programming:

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

(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)

$$\hat{\mathbf{y}}_{t+h|t} = \begin{bmatrix} \hat{\mathbf{a}}_{t+h|t} \\ \hat{\mathbf{v}}_{t+h|t} \\ \hat{\mathbf{u}}_{t+h|t} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{I}_{n_b-k} & \mathbf{O} \\ \mathbf{O} & \mathbf{I}_k \end{bmatrix} \begin{bmatrix} \hat{\mathbf{v}}_{t+h|t} \\ \hat{\mathbf{u}}_{t+h|t} \end{bmatrix}$$

Suppose $\hat{\mathbf{u}}_{t+h|t}$ are fixed and let $\hat{\mathbf{w}}_{t+h|t} = \begin{bmatrix} \hat{\mathbf{a}}_{t+h|t} - \mathbf{A}_2 \hat{\mathbf{u}}_{t+h|t} \\ \hat{\mathbf{v}}_{t+h|t} \end{bmatrix}$.

Optimization problem

$$\min_{\mathbf{v}} [\hat{\mathbf{w}}_{t+h|t} - \mathbf{A}_3 \mathbf{v}]' \mathbf{W}_{\mathbf{v}}^{-1} [\hat{\mathbf{w}}_{t+h|t} - \mathbf{A}_3 \mathbf{v}] \quad \text{where} \quad \mathbf{A}_3 = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{I}_{n_b-k} \end{bmatrix}$$

and $\mathbf{W}_{\mathbf{v}}$ contains elements of \mathbf{W}_h corresponding to $\hat{\mathbf{v}}_{t+h|t}$.

$$\hat{\mathbf{y}}_{t+h|t} = \begin{bmatrix} \hat{\mathbf{a}}_{t+h|t} \\ \hat{\mathbf{v}}_{t+h|t} \\ \hat{\mathbf{u}}_{t+h|t} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{I}_{n_b-k} & \mathbf{O} \\ \mathbf{O} & \mathbf{I}_k \end{bmatrix} \begin{bmatrix} \hat{\mathbf{v}}_{t+h|t} \\ \hat{\mathbf{u}}_{t+h|t} \end{bmatrix}$$

Suppose $\hat{\mathbf{u}}_{t+h|t}$ are fixed and let $\hat{\mathbf{w}}_{t+h|t} = \begin{bmatrix} \hat{\mathbf{a}}_{t+h|t} - \mathbf{A}_2 \hat{\mathbf{u}}_{t+h|t} \\ \hat{\mathbf{v}}_{t+h|t} \end{bmatrix}$.

Solve with non-negativity constraint

$$\min_{\mathbf{v}} [\hat{\mathbf{w}}_{t+h|t} - \mathbf{A}_3 \mathbf{v}]' \mathbf{W}_{\mathbf{v}}^{-1} [\hat{\mathbf{w}}_{t+h|t} - \mathbf{A}_3 \mathbf{v}] \quad \text{where} \quad \mathbf{A}_3 = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{I}_{n_b-k} \end{bmatrix}$$

such that $\mathbf{A}_3 \mathbf{v} \geq \begin{bmatrix} -\mathbf{A}_2 \hat{\mathbf{u}}_{t+h|t} \\ \mathbf{0} \end{bmatrix}$

Reconciliation and regularization

Mishchenko, Montgomery, and Vaggi (2019):
Optimize all forecasts with an incoherence penalty

$$\min_{\hat{\mathbf{y}}_t} \sum_{t=1}^T \|\mathbf{y}_t - \hat{\mathbf{y}}_t\|_2 + \lambda \sum_{t=1}^T \|\hat{\mathbf{y}}_t - \mathbf{s}_t \hat{\mathbf{b}}_t\|_2$$

Reconciliation and regularization

Mishchenko, Montgomery, and Vaggi (2019):
Optimize all forecasts with an incoherence penalty

$$\min_{\hat{\mathbf{y}}_t} \sum_{t=1}^T \|\mathbf{y}_t - \hat{\mathbf{y}}_t\|_2 + \lambda \sum_{t=1}^T \|\hat{\mathbf{y}}_t - \mathbf{s}_t \hat{\mathbf{b}}_t\|_2$$

Shiratori, Kobayashi, and Takano (2020):
Optimize bottom level forecasts with an incoherence penalty

$$\min_{\hat{\mathbf{b}}_t} \sum_{t=1}^T \|\hat{\mathbf{b}}_t - \mathbf{b}_t\|_2 + \sum_{t=1}^T \Lambda \|\mathbf{a}_t - \mathbf{A}_t \hat{\mathbf{b}}_t\|_2$$

Outline

- 1 Recap
- 2 Time series reconciliation
- 3 Reconciliation via constraints
- 4 Example: reconciling GDP forecasts
- 5 The geometry of forecast reconciliation
- 6 Optimization and reconciliation
- 7 In-built coherence
- 8 Time series cross-validation

In-built coherence

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

One-step approaches: compute coherent \tilde{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 $\mathbf{x}_{t,i} = (1, x_{t,1,i}, \dots, x_{t,p,i})$ & $\hat{\mathbf{y}}_i = (\hat{y}_{1,i}, \dots, \hat{y}_{T,i})$.

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

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad \hat{\mathbf{y}}_{t+h} = \mathbf{X}_{t+h}^* \hat{\mathbf{B}}$$

In-built coherence using linear models

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

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

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

In-built coherence using linear models

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

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

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

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

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

Reference: Ashouri, Hyndman, and Shmueli (2022)

In-built coherence using state space models

Pennings and Dalen (2017) propose the state space model

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

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

- Coherent forecasts arise naturally using the Kalman filter
- Covariance matrices difficult to estimate except for small hierarchies.
- Requires the same model for all series
- A related approach proposed by Villegas and Pedregal (2018)

Outline

- 1 Recap
- 2 Time series reconciliation
- 3 Reconciliation via constraints
- 4 Example: reconciling GDP forecasts
- 5 The geometry of forecast reconciliation
- 6 Optimization and reconciliation
- 7 In-built coherence
- 8 Time series cross-validation

Time series cross-validation

Traditional evaluation

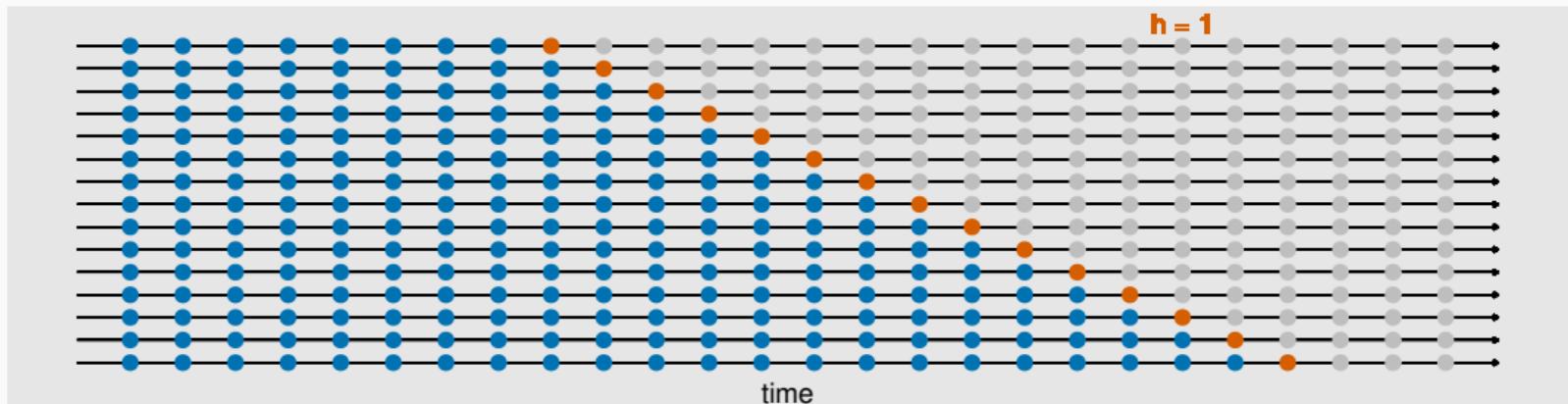


Time series cross-validation

Traditional evaluation



Time series cross-validation

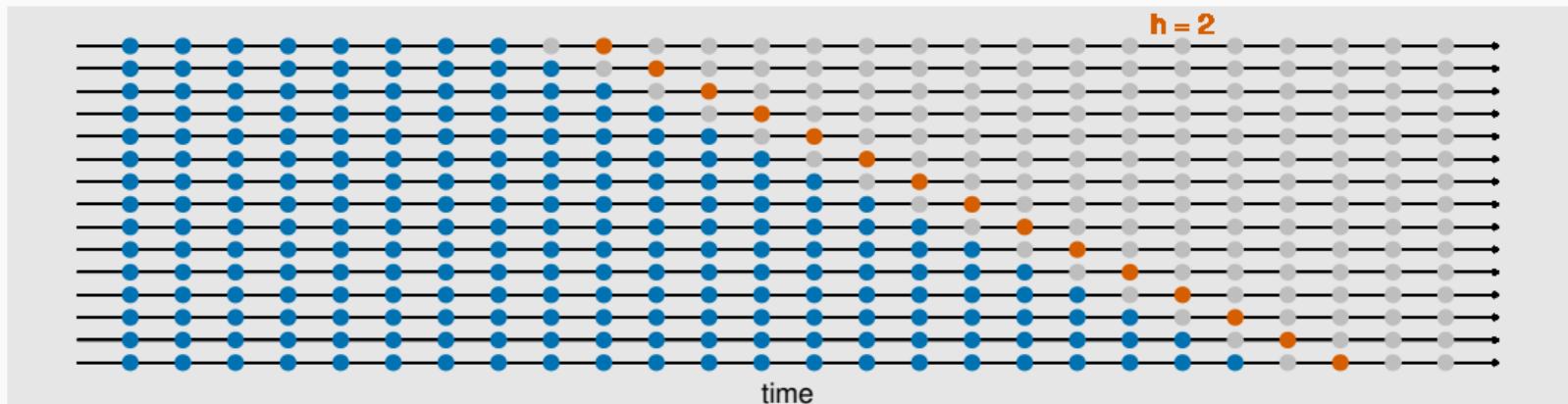


Time series cross-validation

Traditional evaluation



Time series cross-validation

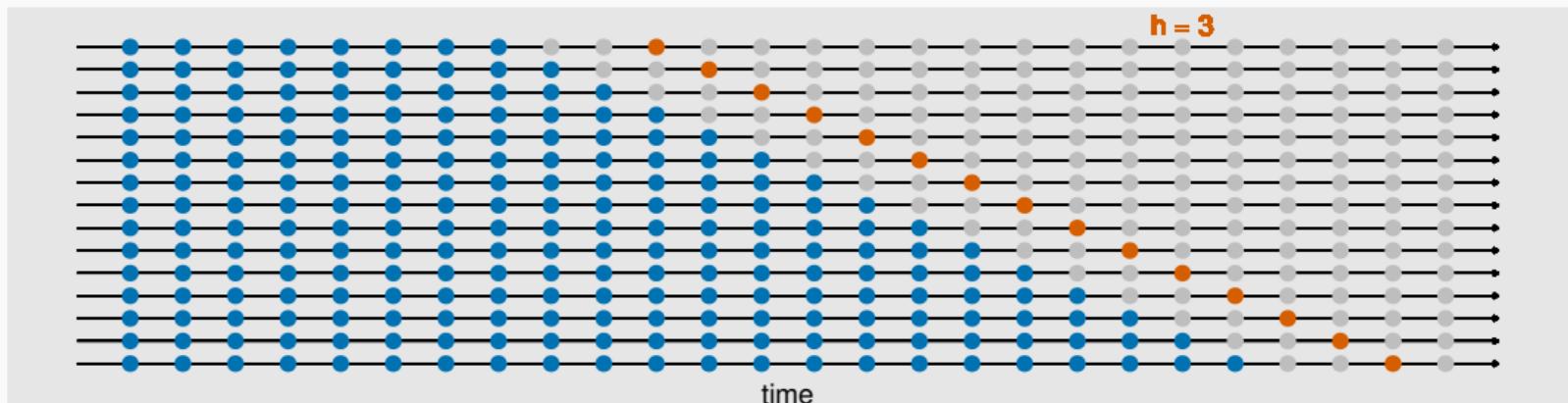


Time series cross-validation

Traditional evaluation



Time series cross-validation

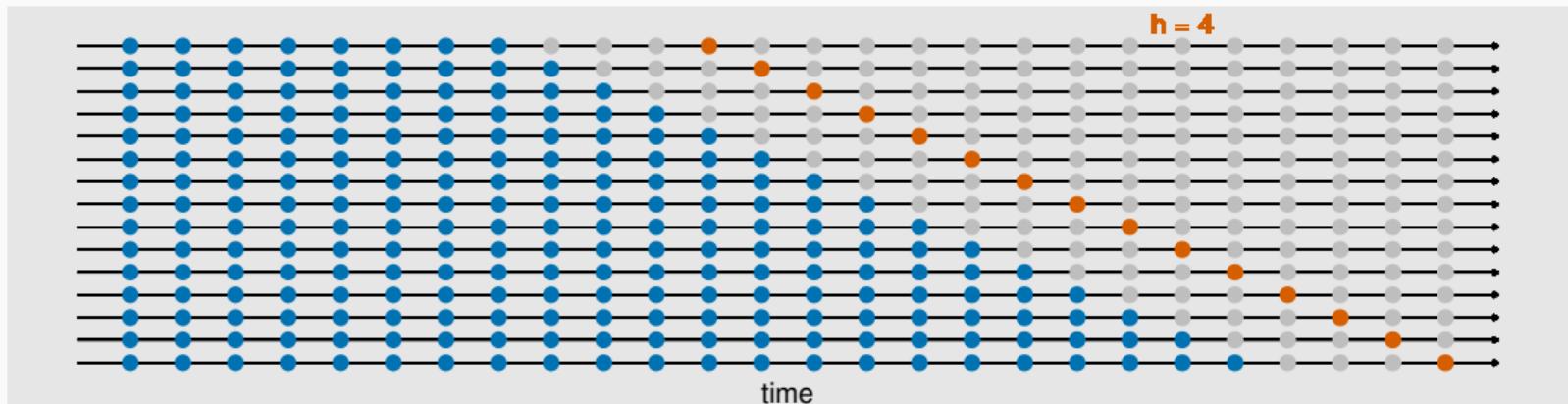


Time series cross-validation

Traditional evaluation

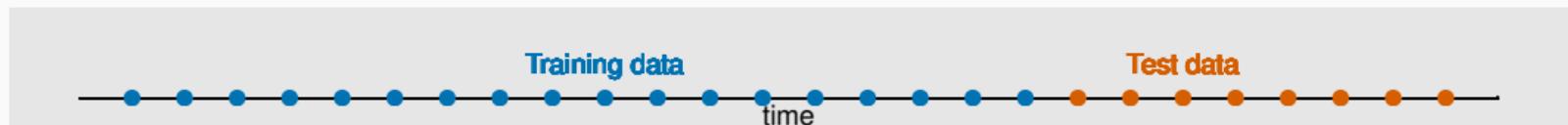


Time series cross-validation

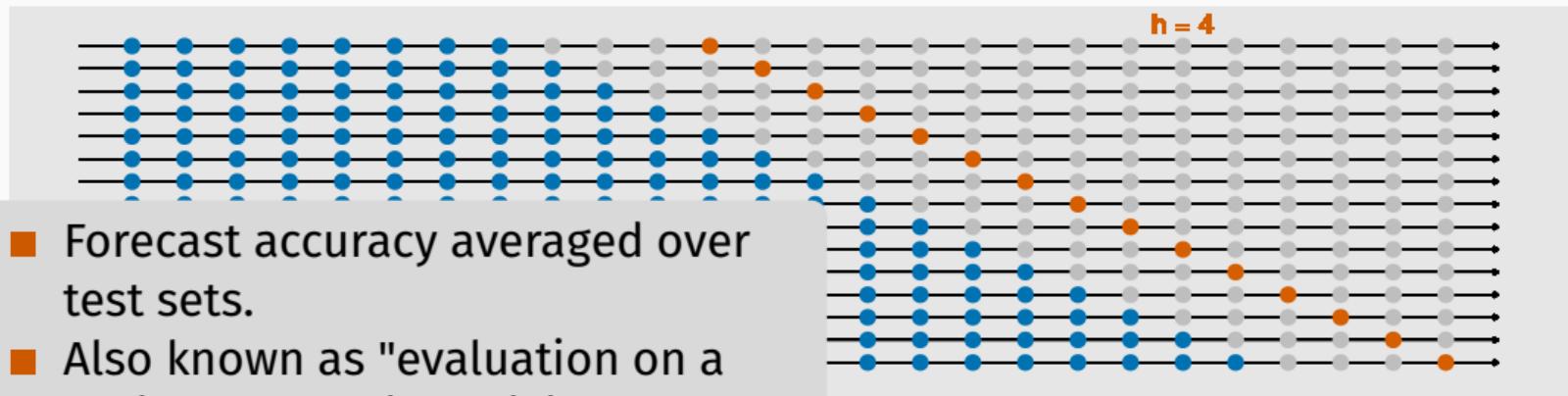


Time series cross-validation

Traditional evaluation



Time series cross-validation



Example: Australian tourism

tourism

```
# A tsibble: 18,000 x 5 [1M]
# Key:      state, zone, region [75]
  month state zone      region visitors
  <mth> <chr> <chr>      <chr>     <dbl>
1 1998  Jan NSW  Metro NSW Sydney     926.
2 1998  Feb NSW  Metro NSW Sydney     647.
3 1998  Mar NSW  Metro NSW Sydney     716.
4 1998  Apr NSW  Metro NSW Sydney     621.
5 1998  May NSW  Metro NSW Sydney     598.
6 1998  Jun NSW  Metro NSW Sydney     601.
7 1998  Jul NSW  Metro NSW Sydney     720.
8 1998  Aug NSW  Metro NSW Sydney     645.
9 1998  Sep NSW  Metro NSW Sydney     633.
10 1998 Oct NSW  Metro NSW Sydney    771.
# i 17,990 more rows
```

Example: Australian tourism

```
tourism_agg <- tourism |>
  aggregate_key(state/zone/region, visitors = sum(visitors))
tourism_stretch <- tourism_agg |>
  stretch_tsibble(.init = 48, .step = 1)
tourism_stretch
```

```
# A tsibble: 3,057,120 x 6 [1M]
# Key:      .id, state, zone, region [21,230]
  month state  zone   region visitors   .id
  <mth> <chr*> <chr*> <chr*>     <dbl> <int>
1 1998 Jan NSW ACT Canberra  210.     1
2 1998 Feb NSW ACT Canberra  156.     1
3 1998 Mar NSW ACT Canberra  185.     1
4 1998 Apr NSW ACT Canberra  178.     1
5 1998 May NSW ACT Canberra  134.     1
6 1998 Jun NSW ACT Canberra  105.     1
7 1998 Jul NSW ACT Canberra  142.     1
8 1998 Aug NSW ACT Canberra  137.     1
```

Example: Australian tourism

```
fit <- tourism_stretch |>
  model(ets = ETS(visitors))
fit
```

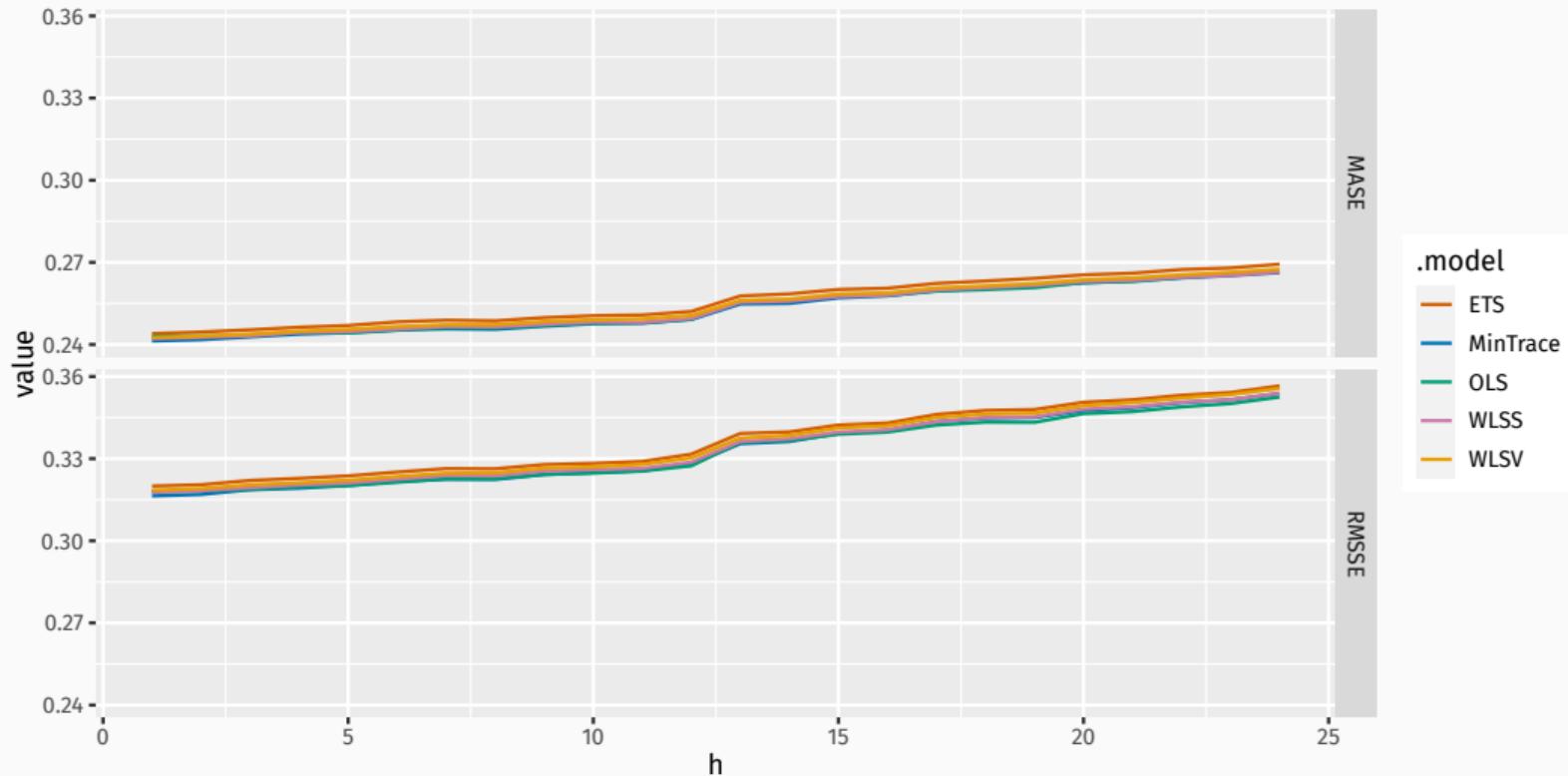
```
# A mable: 21,230 x 5
# Key:   .id, state, zone, region [21,230]
  .id state zone          region           ets
  <int> <chr*> <chr*>       <chr*>       <model>
1    1 NSW  ACT        Canberra      <ETS(A,N,N)>
2    1 NSW  ACT        <aggregated> <ETS(A,N,N)>
3    1 NSW Metro NSW Central Coast <ETS(A,N,A)>
4    1 NSW Metro NSW  Sydney      <ETS(A,N,N)>
5    1 NSW Metro NSW <aggregated> <ETS(M,N,M)>
6    1 NSW North Coast NSW Hunter <ETS(A,N,N)>
7    1 NSW North Coast NSW North Coast NSW <ETS(M,N,M)>
8    1 NSW North Coast NSW <aggregated> <ETS(M,N,M)>
9    1 NSW North NSW   Blue Mountains <ETS(M,N,N)>
10   1 NSW North NSW  Central NSW   <ETS(A,N,N)>
# i 21,220 more rows
# i Use `print(n = ...)` to see more rows
```

Example: Australian tourism

```
fc <- fit |>
  reconcile(
    ols = min_trace(ets, method = "ols"),
    wlsv = min_trace(ets, method = "wls_var"),
    wlss = min_trace(ets, method = "wls_struct"),
    mint_s = min_trace(ets, method = "mint_shrink"),
  ) |>
  forecast(h = "2 years")
fc
```

```
# A fable: 2,547,600 x 9 [1M]
# Key:     .id, state, zone, region, .model [106,150]
  .id state zone region .model month   visitors .mean     h
  <int> <chr*> <chr*> <chr*> <chr>   <mth>     <dist> <dbl> <int>
 1     1 NSW  ACT  Canberra ets  2002 Jan N(169, 1553) 169.     1
 2     1 NSW  ACT  Canberra ets  2002 Feb N(169, 1553) 169.     2
 3     1 NSW  ACT  Canberra ets  2002 Mar N(169, 1553) 169.     3
 4     1 NSW  ACT  Canberra ets  2002 Apr N(169, 1553) 169.     4
```

Example: Australian tourism



Example: Australian tourism

Overall

| Model | MASE | RMSSE |
|----------|--------------|--------------|
| ETS | 0.256 | 0.337 |
| MinTrace | 0.253 | 0.333 |
| OLS | 0.253 | 0.333 |
| WLSS | 0.253 | 0.334 |
| WLSV | 0.254 | 0.335 |

National

| Model | MASE | RMSSE |
|----------|--------------|--------------|
| ETS | 1.020 | 1.010 |
| MinTrace | 1.072 | 1.061 |
| OLS | 1.026 | 1.016 |
| WLSS | 1.060 | 1.052 |
| WLSV | 1.078 | 1.069 |

State

| Model | MASE | RMSSE |
|----------|--------------|--------------|
| ETS | 0.109 | 0.122 |
| MinTrace | 0.109 | 0.121 |
| OLS | 0.107 | 0.120 |
| WLSS | 0.108 | 0.121 |
| WLSV | 0.110 | 0.122 |

Zone

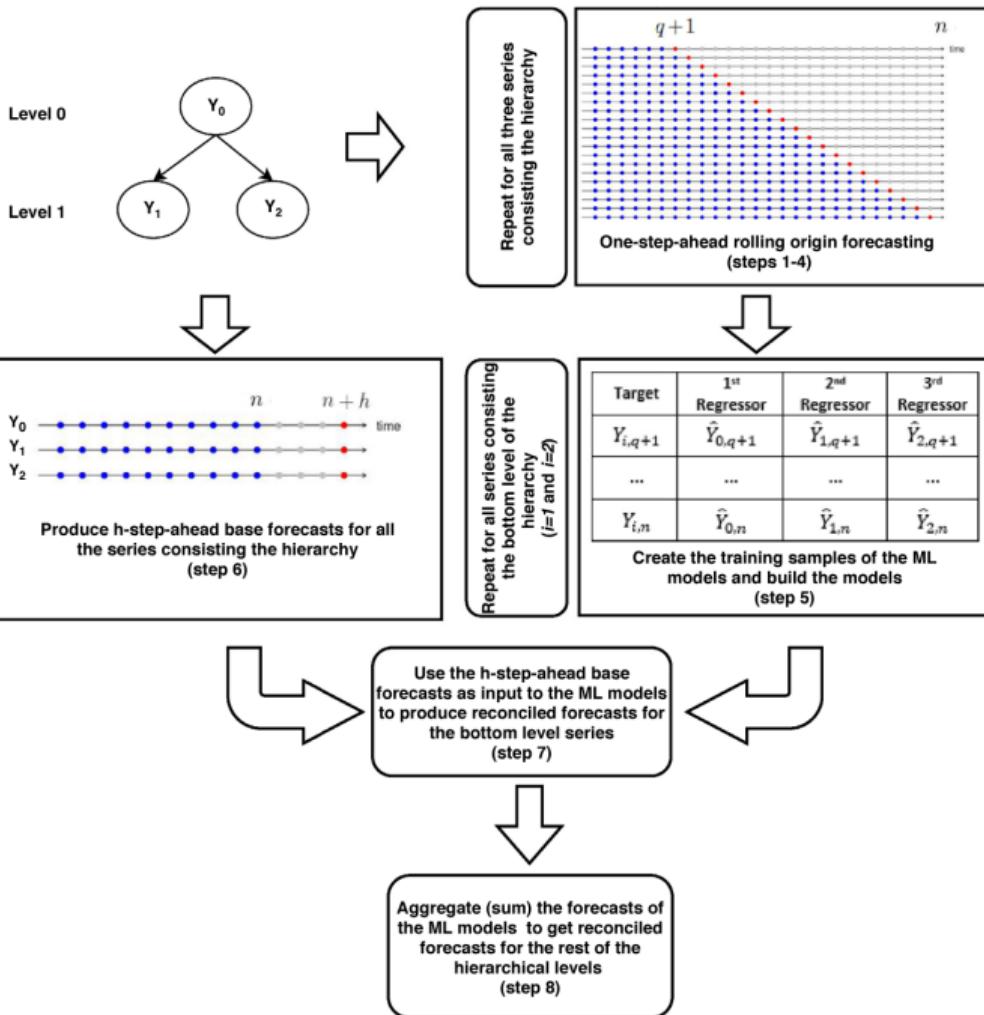
| Model | MASE | RMSSE |
|----------|--------------|--------------|
| ETS | 0.147 | 0.191 |
| MinTrace | 0.145 | 0.189 |
| OLS | 0.144 | 0.188 |
| WLSS | 0.145 | 0.189 |
| WLSV | 0.146 | 0.190 |

Region

| Model | MASE | RMSSE |
|----------|--------------|--------------|
| ETS | 0.298 | 0.372 |
| MinTrace | 0.294 | 0.366 |
| OLS | 0.296 | 0.367 |
| WLSS | 0.295 | 0.367 |
| WLSV | 0.296 | 0.368 |

ML reconciliation

- 1 Split all series using time series cross-validation
- 2 For each training set, compute one-step-ahead forecasts for all series
- 3 For each bottom-level series, use RF or XGB to predict values using forecasts of all series as inputs
- 4 Forecast all series
- 5 For each bottom-level series, apply ML model to improve forecasts
- 6 Aggregate bottom-level forecasts to obtain forecasts for other series.



ML reconciliation: tourism data

| Method | Total | States | Zones | Regions | Average |
|-------------|--------------|--------------|--------------|--------------|--------------|
| MASE | | | | | |
| MinT-Struct | 1.094 | 0.968 | 0.887 | 0.843 | 0.948 |
| MinT-Shrink | 1.047 | 0.956 | 0.872 | 0.824 | 0.925 |
| ML-RF | 1.045 | 0.964 | 0.859 | 0.812 | 0.920 |
| ML-XGB | 1.043 | 0.965 | 0.859 | 0.812 | 0.920 |
| RMSSE | | | | | |
| MinT-Struct | 1.308 | 1.225 | 1.137 | 1.109 | 1.195 |
| MinT-Shrink | 1.265 | 1.214 | 1.120 | 1.086 | 1.171 |
| ML-RF | 1.261 | 1.208 | 1.104 | 1.066 | 1.159 |
| ML-XGB | 1.255 | 1.208 | 1.101 | 1.064 | 1.157 |
| AMSE | | | | | |
| MinT-Struct | 0.988 | 0.611 | 0.426 | 0.349 | 0.593 |
| MinT-Shrink | 0.935 | 0.599 | 0.417 | 0.337 | 0.572 |
| ML-RF | 0.780 | 0.526 | 0.366 | 0.319 | 0.498 |
| ML-XGB | 0.779 | 0.526 | 0.365 | 0.317 | 0.497 |

- ML methods not significantly different.
- MinT methods significantly different from each other and from ML methods.

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.](#)
-  [Bisaglia, L, T Di Fonzo, and D Girolimetto \(2020\). Fully reconciled GDP forecasts from income and expenditure sides. Ed. by A Pollice, N Salvati, and F Schirripa Spagnolo, 951–956.](#)
-  [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.
-  Mishchenko, K, M Montgomery, and F Vaggi (2019). A self-supervised approach to hierarchical forecasting with applications to groupwise synthetic controls. In: *Proceedings of the Time Series Workshop at the 36th International Conference on Machine Learning, Long Beach, California*. PMLR 97.
-  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

-  Shiratori, T, K Kobayashi, and Y Takano (2020). Prediction of hierarchical time series using structured regularization and its application to artificial neural networks. *PLOS ONE* **15**(11), e0242099.
-  Spiliotis, E, M Abolghasemi, RJ Hyndman, F Petropoulos, and V Assimakopoulos (2021). Hierarchical forecast reconciliation with machine learning. *Applied Soft Computing* **112**, 107756.
-  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.

References

-  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 (2023). Optimal reconciliation with immutable forecasts. *European Journal of Operational Research* **308**(2), 650–660.