Forecast reconciliation

A brief overview

Rob J Hyndman

MONASH University

Photo by Edvard Alexander Rølvaag on Unsplash

Outline

1. Hierarchical time series notation
2. Optimal linear forecast reconciliation
3. Probabilistic forecast reconciliation
4. Cross-temporal probabilistic forecast reconciliation

Outline

1 Hierarchical time series notation

2 Optimal linear forecast reconciliation

3 Probabilistic forecast reconciliation

4 Cross-temporal probabilistic forecast reconciliation

Labour market participation

Australia & New Zealand Standard Classification of Occupations

- 8 major groups
  - 43 sub-major groups
    - 97 minor groups
      - 359 unit groups
        - 1023 occupations

Labour market participation

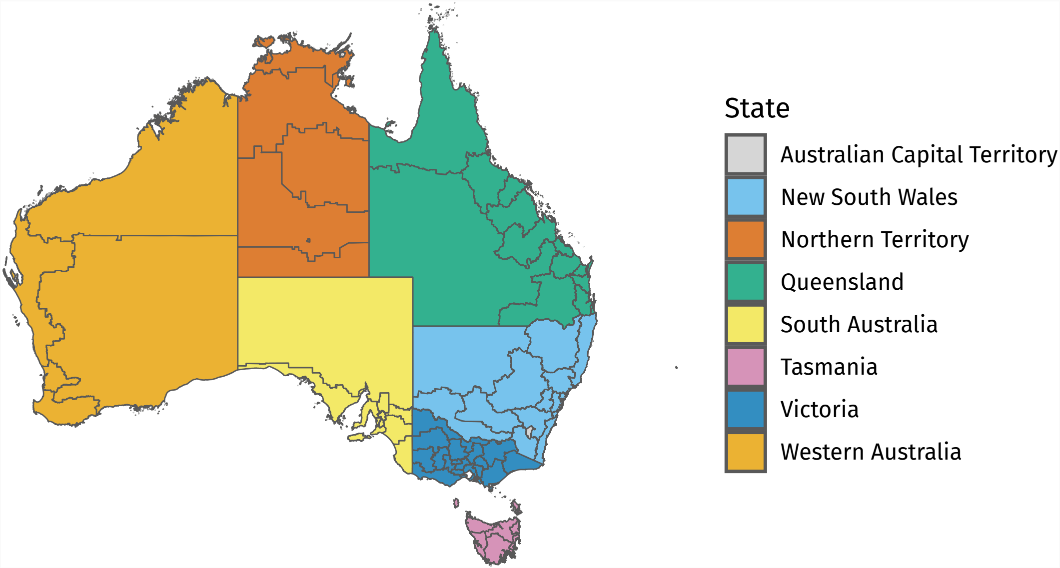
Australia & New Zealand Standard Classification of Occupations

- 8 major groups
  - 43 sub-major groups
    - 97 minor groups
      - 359 unit groups
        - 1023 occupations

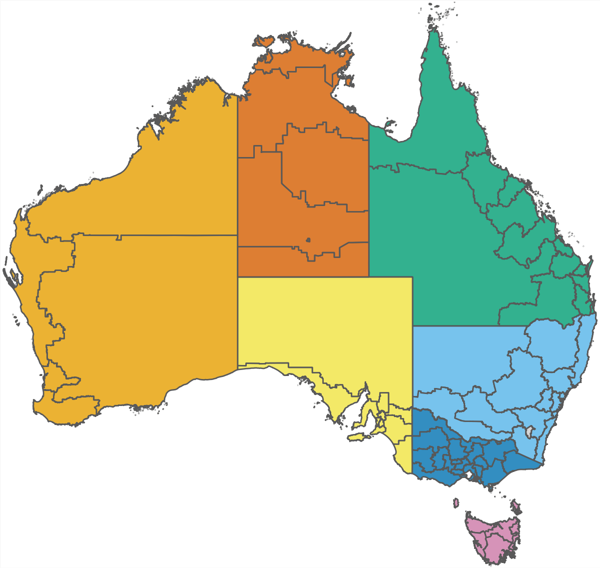
Example: statistician

2 Professionals
22 Business, Human Resource and Marketing Professionals
224 Information and Organisation Professionals
2241 Actuaries, Mathematicians and Statisticians
224113 Statistician

Australian tourism regions

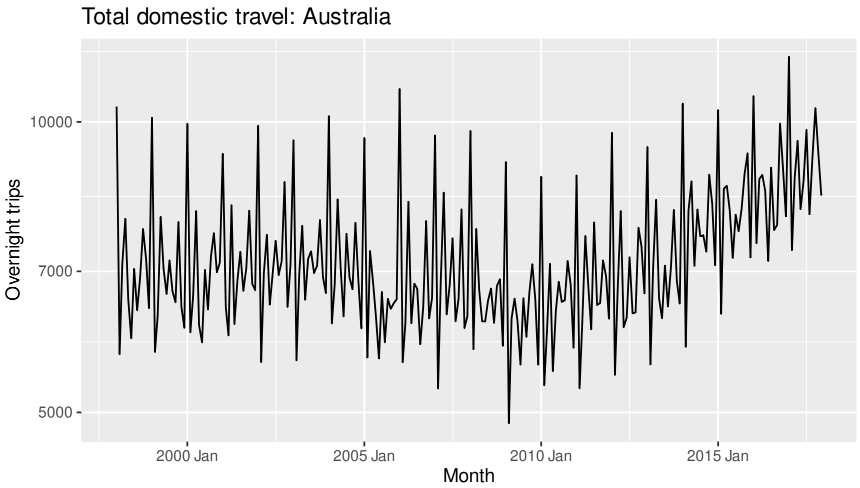


Australian tourism regions

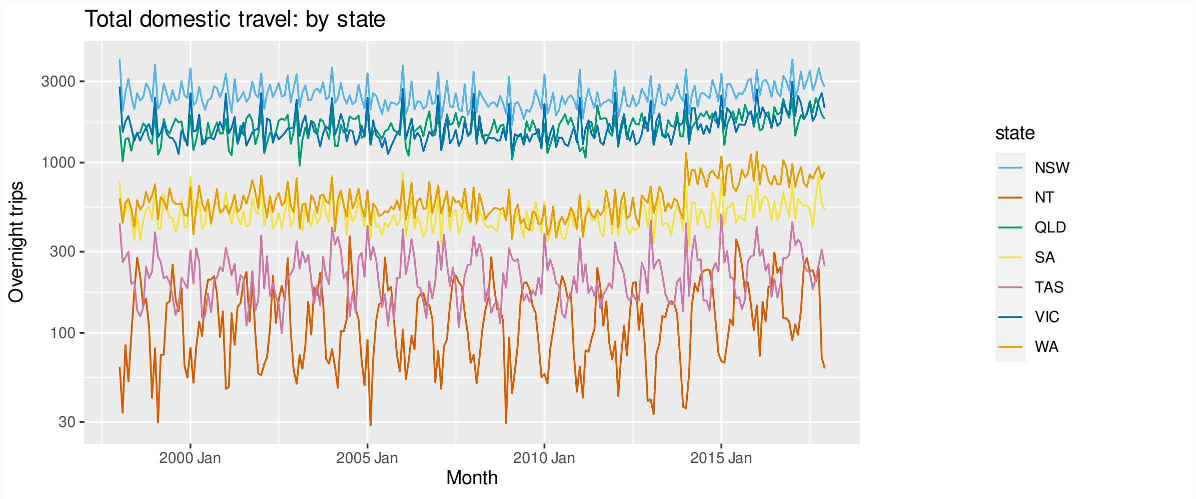


- Monthly data on visitor night from 1998 – 2017
- From *National Visitor Survey*, annual interviews of 120,000 Australians aged 15+.
- Geographical hierarchy split by
  - 7 states
  - 27 zones
  - 75 regions

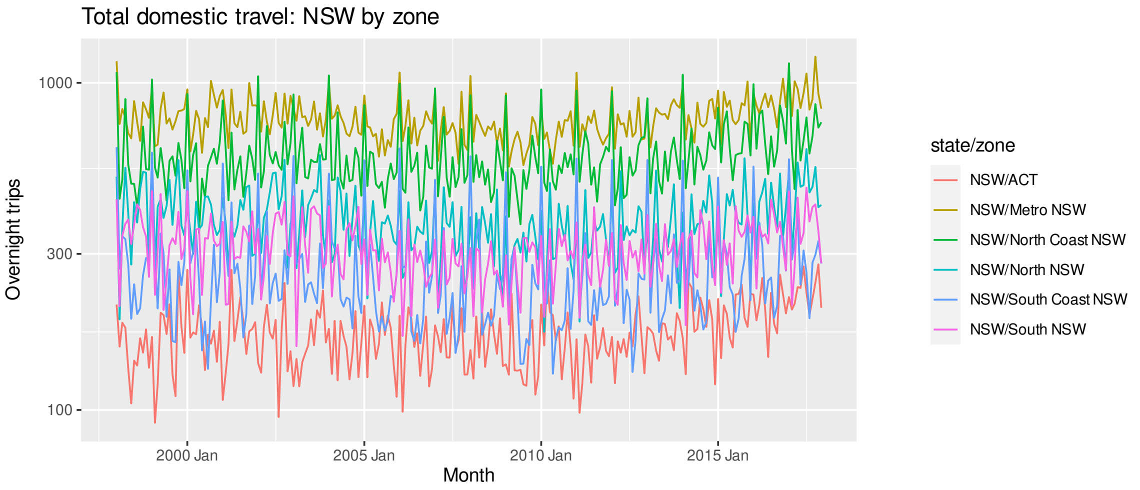
Australian tourism data



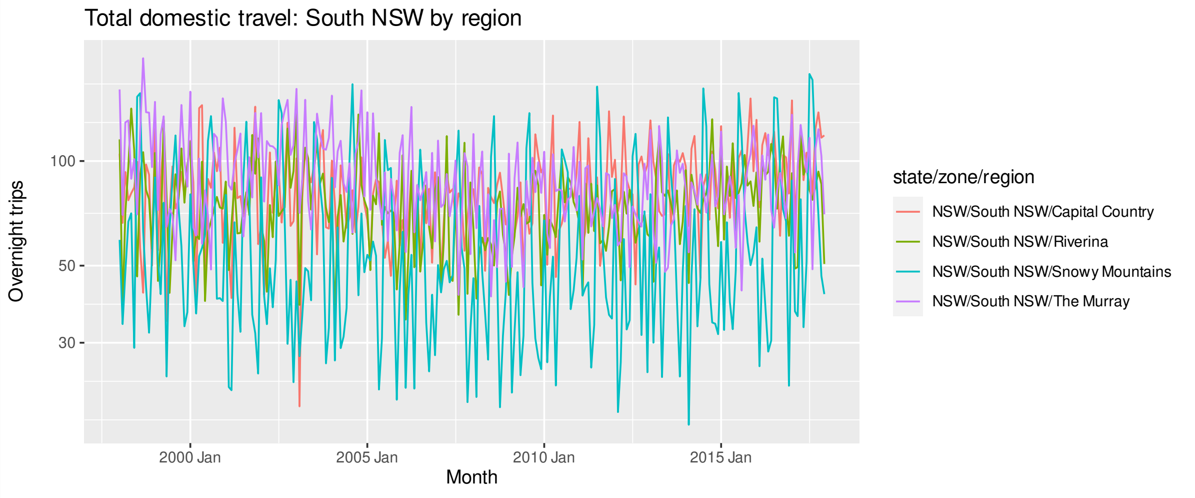
Australian tourism data



Australian tourism data



Australian tourism data

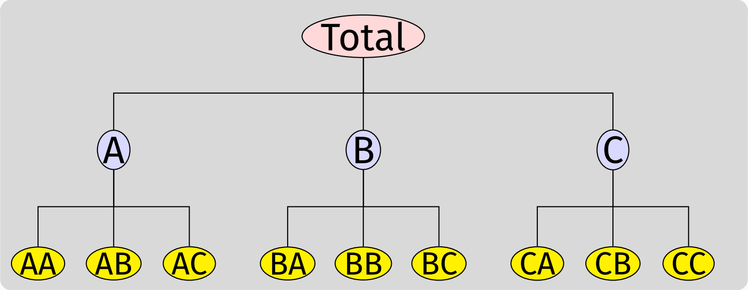


Spectacle sales

- Monthly UK sales data from 2000 – 2014
- Provided by a large spectacle manufacturer
- Split by brand (26), gender (3), price range (6), materials (4), and stores (600)
- About 1 million bottom-level series

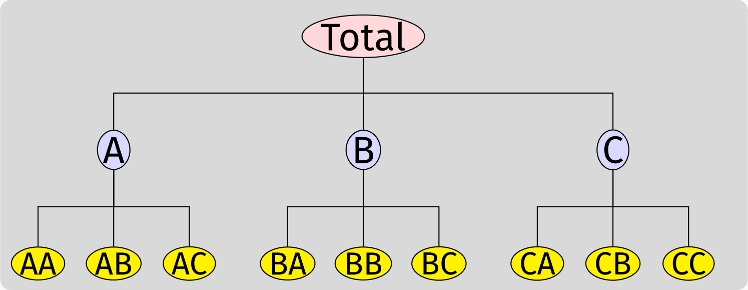
Hierarchical time series

A **hierarchical time series** is a collection of several time series that are linked together in a hierarchical structure.



Hierarchical time series

A **hierarchical time series** is a collection of several time series that are linked together in a hierarchical structure.

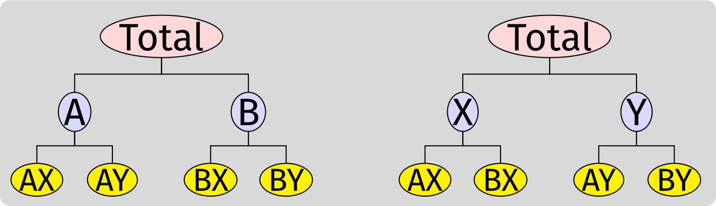


Examples

- Tourism by state and region
- Retail sales by product groups, sub groups, and SKUs

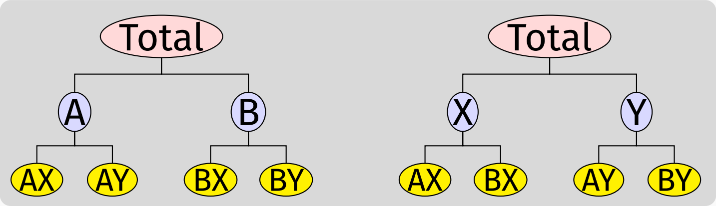
Grouped time series

A **grouped time series** is a collection of time series that can be grouped together in a number of non-hierarchical ways.



Grouped time series

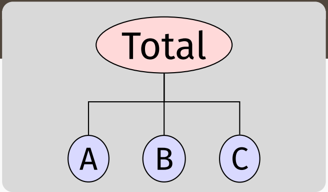
A **grouped time series** is a collection of time series that can be grouped together in a number of non-hierarchical ways.



Examples

- Tourism by state and purpose of travel
- Retail sales by product groups/sub groups, and by countries/regions

Hierarchical and grouped time series



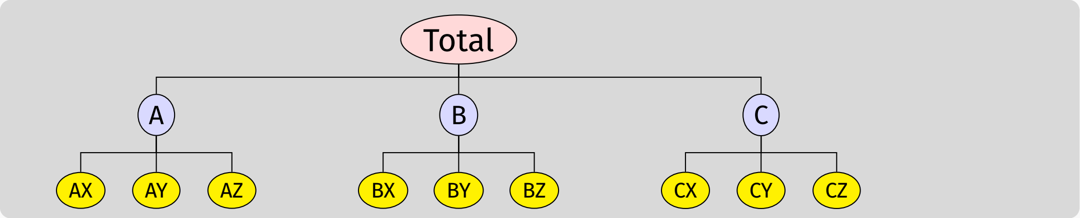
Almost all collections of time series with linear constraints can be written as

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

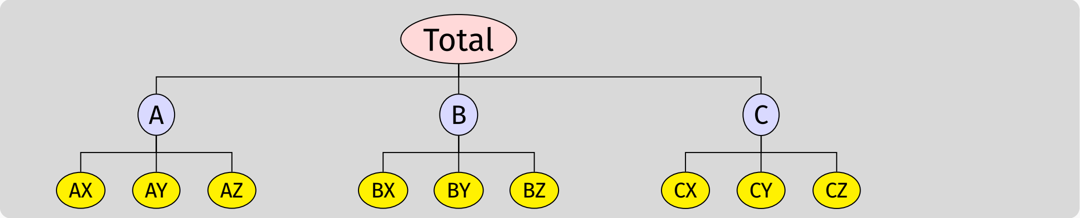
- $\boldsymbol{y}_t$ = vector of all series at time t
- $y_{\text{Total},t}$ = aggregate of all series at time t.
- $y_{X,t}$ = value of series X at time t.
- $\boldsymbol{b}_t$ = vector of most disaggregated series at time t
- $\boldsymbol{S}$ = "summing matrix" containing the linear constraints.

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

Hierarchical time series

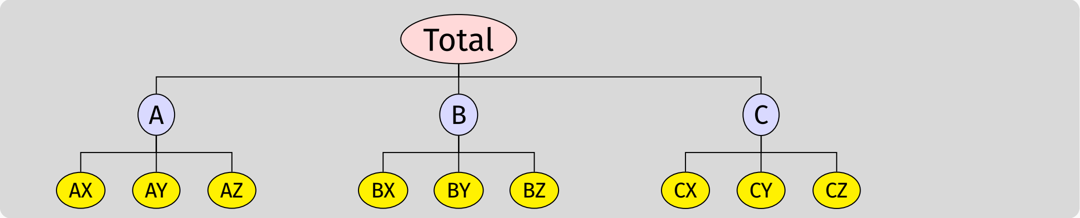


Hierarchical time series



$$
\boldsymbol{y}_t = \begin{pmatrix} y_{\text{Total},t} \\ y_{A,t} \\ y_{B,t} \\ y_{C,t} \\ y_{AX,t} \\ y_{AY,t} \\ y_{AZ,t} \\ y_{BX,t} \\ y_{BY,t} \\ y_{BZ,t} \\ y_{CX,t} \\ y_{CY,t} \\ y_{CZ,t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{AX,t} \\ y_{AY,t} \\ y_{AZ,t} \\ y_{BX,t} \\ y_{BY,t} \\ y_{BZ,t} \\ y_{CX,t} \\ y_{CY,t} \\ y_{CZ,t} \end{pmatrix}
$$

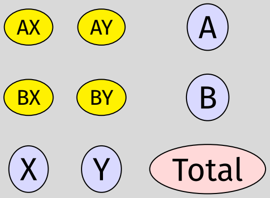
Hierarchical time series



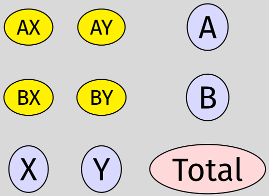
$$\boldsymbol{y}_t = \begin{pmatrix} y_{\text{Total},t} \\ y_{A,t} \\ y_{B,t} \\ y_{C,t} \\ y_{AX,t} \\ y_{AY,t} \\ y_{AZ,t} \\ y_{BX,t} \\ y_{BY,t} \\ y_{BZ,t} \\ y_{CX,t} \\ y_{CY,t} \\ y_{CZ,t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{AX,t} \\ y_{AY,t} \\ y_{AZ,t} \\ y_{BX,t} \\ y_{BY,t} \\ y_{BZ,t} \\ y_{CX,t} \\ y_{CY,t} \\ y_{CZ,t} \end{pmatrix}$$

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

Grouped data

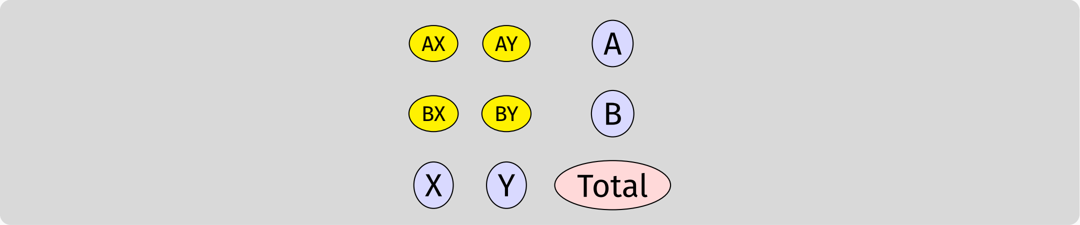


Grouped data



$$\boldsymbol{y}_t = \begin{pmatrix} y_{\text{Total},t} \\ y_{A,t} \\ y_{B,t} \\ y_{X,t} \\ y_{Y,t} \\ y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{pmatrix}$$

Grouped data



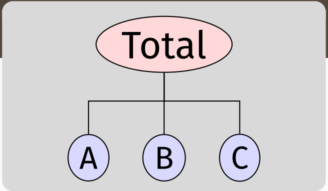
$$\boldsymbol{y}_t = \begin{pmatrix} y_{\mathrm{Total},t} \\ y_{A,t} \\ y_{B,t} \\ y_{X,t} \\ y_{Y,t} \\ y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{pmatrix}$$

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

The hierarchical forecasting problem

- We want forecasts at all levels of aggregation.
- If we model and forecast each series independently, the forecasts will almost certainly not add up.
- We need to impose constraints on the forecasts to ensure they are "coherent".
- We need to do this in a way that is computationally efficient.

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 $\boldsymbol{A}$

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

Constraint matrix $\boldsymbol{C}$

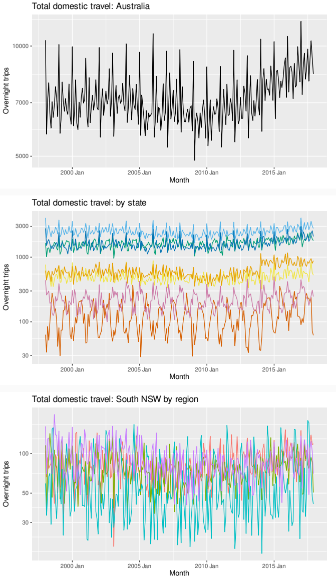
$$\boldsymbol{C}\boldsymbol{y}_t = \boldsymbol{0}$$

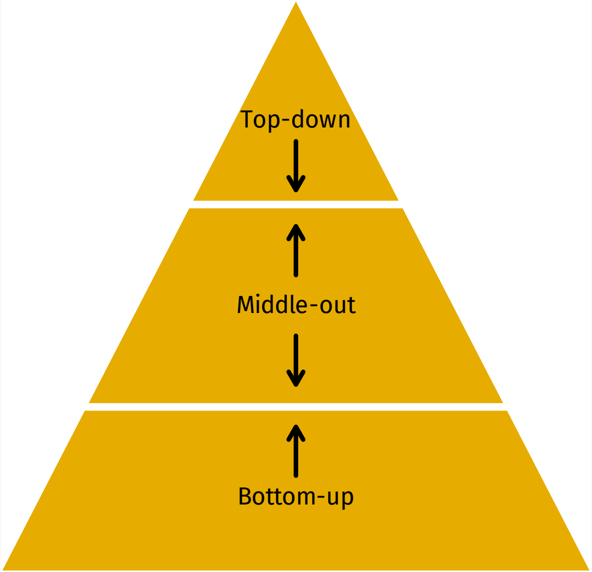
- Constraint matrix approach more general & more parsimonious.
- $\boldsymbol{C} = [\boldsymbol{I}_{n_a} \quad -\boldsymbol{A}]$.
- $\boldsymbol{S}$, $\boldsymbol{A}$ and $\boldsymbol{C}$ may contain any real values (not just 0s and 1s).

Outline

1. Hierarchical time series notation
2. Optimal linear forecast reconciliation
3. Probabilistic forecast reconciliation
4. Cross-temporal probabilistic forecast reconciliation

Hierarchical forecasting 20 years ago





Early history of forecast reconciliation

History

2001: Idea to use all available series to forecast Australia's labour market by occupation.

2004: PhD student Roman Ahmed begins, co-supervised with George Athanasopoulos.

2006: Presentation at ISF, Santander.

2007: Pre-print of "Optimal combination forecasts for hierarchical time series".

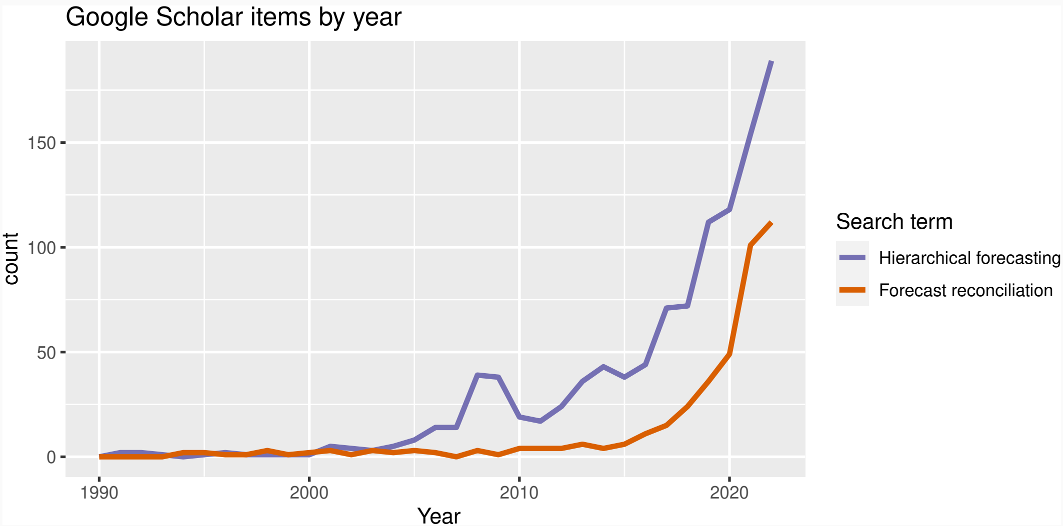
2009: Application to Australian tourism published in IJF.

2010: First version of hts package on CRAN.

2011: "Optimal combination forecasts for hierarchical time series" appears in CSDA.

2019: "Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization"' appears in JASA.

Forecast reconciliation research



The coherent subspace

Coherent subspace

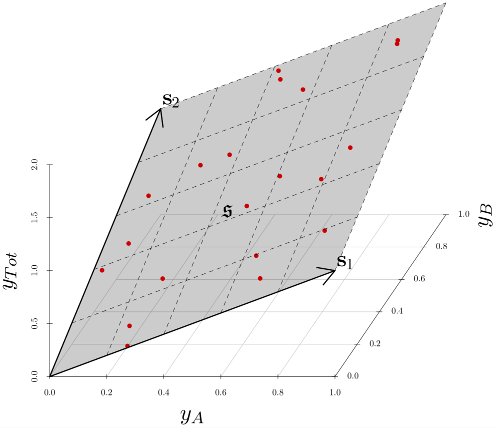
n_b-dimensional linear subspace $\mathfrak{s} \subset \chi^n$ for which linear constraints hold for all $\boldsymbol{y} \in \mathfrak{s}$.

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

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

Hierarchical time series

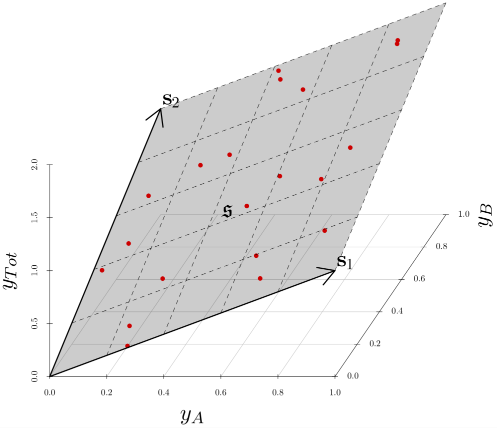
An n-dimensional multivariate time series such that $\boldsymbol{y}_t \in \mathfrak{s} \quad \forall t$.

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

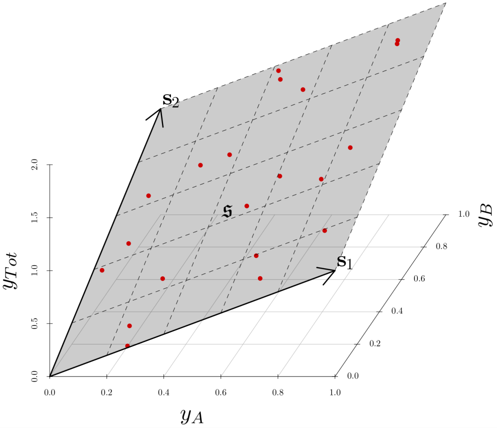
n_b-dimensional linear subspace $\mathfrak{s} \subset \chi^n$ for which linear constraints hold for all $\boldsymbol{y} \in \mathfrak{s}$.

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

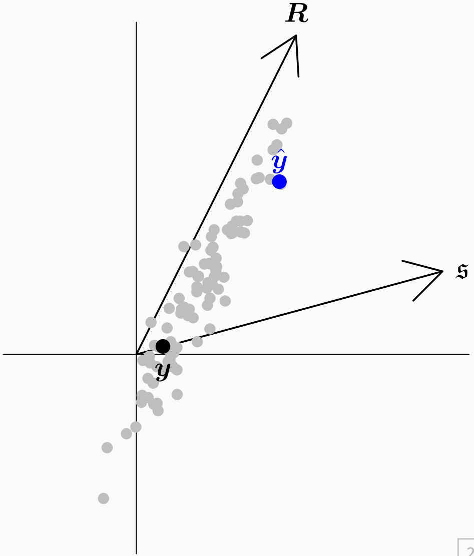
Let $\hat{\boldsymbol{y}}_{t+h|t}$ be vector of *incoherent* initial h-step forecasts.

Reconciled forecasts

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

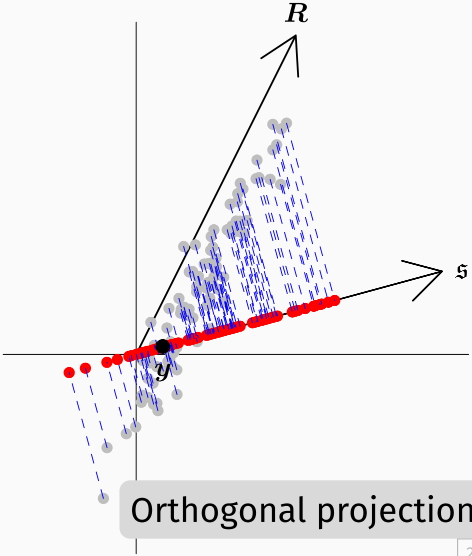
Linear projection reconciliation

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



Linear projection reconciliation

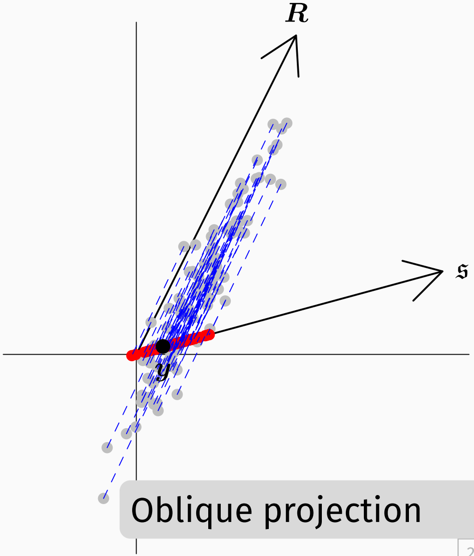
- R is the most likely direction of deviations from $\mathfrak{s}$.
- Grey: potential base forecasts
- Red: reconciled forecasts
- Orthogonal projections (i.e., OLS) lead to smallest possible adjustments of base forecasts.



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.



Oblique projection

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 $\boldsymbol{My} = \boldsymbol{y}$ 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 $\mathfrak{s}$.
- If $\boldsymbol{S}$ forms a basis set for $\mathfrak{s}$, then 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.
- The projection is orthogonal if and only if $\boldsymbol{M}' = \boldsymbol{M}$. Equivalently, $\Psi = \boldsymbol{I}$.

Linear projection reconciliation

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

$$\text{where} \quad \boldsymbol{M} = \boldsymbol{S}(\boldsymbol{S}'\Psi\boldsymbol{S})^{-1}\boldsymbol{S}'\Psi = \boldsymbol{I} - \Psi\boldsymbol{C}'(\boldsymbol{C}\Psi\boldsymbol{C}')^{-1}\boldsymbol{C}$$

OLS:	$\Psi = \boldsymbol{I}$
WLS:	$\Psi = \text{diagonal}$
MinT:	$\Psi = \boldsymbol{W}_h$

- $\boldsymbol{W}_h = \text{Var}[\boldsymbol{y}_{T+h} - \hat{\boldsymbol{y}}_{T+h|T} \mid \boldsymbol{y}_1, \dots, \boldsymbol{y}_T]$ 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 the covariance matrix of the reconciled forecast errors.

Minimum trace reconciliation

Minimum trace (MinT) reconciliation

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

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

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

Reconciled forecasts Base forecasts

- Trace of $\boldsymbol{V}_h$ is sum of forecast variances.
- MinT is L_2 optimal amongst linear unbiased forecasts.
- How to estimate $\boldsymbol{W}_h = \text{Var}[\boldsymbol{y}_{T+h} - \hat{\boldsymbol{y}}_{T+h|T} \mid \boldsymbol{y}_1, \dots, \boldsymbol{y}_T]$?

Linear projections

$$\tilde{\boldsymbol{y}}_{T+h|T} = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}_{T+h|T}$$

Reconciliation method	$\boldsymbol{G}$
OLS	$(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$
WLS(var)	$(\boldsymbol{S}'\Lambda_v\boldsymbol{S})^{-1}\boldsymbol{S}'\Lambda_v$
WLS(struct)	$(\boldsymbol{S}'\Lambda_s\boldsymbol{S})^{-1}\boldsymbol{S}'\Lambda_s$
MinT(sample)	$(\boldsymbol{S}'\hat{\boldsymbol{W}}_{\text{sam}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\hat{\boldsymbol{W}}_{\text{sam}}^{-1}$
MinT(shrink)	$(\boldsymbol{S}'\hat{\boldsymbol{W}}_{\text{shr}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\hat{\boldsymbol{W}}_{\text{shr}}^{-1}$

These approximate MinT by assuming $\boldsymbol{W}_h = k_h\boldsymbol{W}_1$.

- $\Lambda_v = \text{diag}(\boldsymbol{W}_1)^{-1}$
- $\Lambda_s = \text{diag}(\boldsymbol{S}\boldsymbol{1})^{-1}$
- $\hat{\boldsymbol{W}}_{\text{sam}}$ is sample estimate of the residual covariance matrix
- $\hat{\boldsymbol{W}}_{\text{shr}}$ is shrinkage estimator $\tau\text{diag}(\hat{\boldsymbol{W}}_{\text{sam}}) + (1-\tau)\hat{\boldsymbol{W}}_{\text{sam}}$ where τ selected optimally.
- Still need a good estimate of $\boldsymbol{W}_h$ for forecast variance.

Example: Australian tourism

```r
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

```r
tourism_agg <- tourism |>
  aggregate_key(state / zone / region, visitors = sum(visitors))
```

```
# A tsibble: 26,400 x 5 [1M]
# Key:       state, zone, region [110]
      month state        zone         region       visitors
      <mth> <chr*>       <chr*>       <chr*>          <dbl>
 1 1998 Jan <aggregated> <aggregated> <aggregated>   10376.
 2 1998 Feb <aggregated> <aggregated> <aggregated>    5746.
 3 1998 Mar <aggregated> <aggregated> <aggregated>    7129.
 4 1998 Apr <aggregated> <aggregated> <aggregated>    7939.
 5 1998 May <aggregated> <aggregated> <aggregated>    6552.
 6 1998 Jun <aggregated> <aggregated> <aggregated>    5969.
 7 1998 Jul <aggregated> <aggregated> <aggregated>    7041.
 8 1998 Aug <aggregated> <aggregated> <aggregated>    6382.
 9 1998 Sep <aggregated> <aggregated> <aggregated>    6907.
10 1998 Oct <aggregated> <aggregated> <aggregated>    7744.
# i 26,390 more rows
```

Example: Australian tourism

```
fit <- tourism_agg |>
  filter(year(month) <= 2015) |>
  model(ets = ETS(visitors))
```

```
# A mable: 110 x 4
# Key:     state, zone, region [110]
   state  zone            region                   ets
   <chr*> <chr*>          <chr*>               <model>
 1 NSW    ACT             Canberra        <ETS(M,N,A)>
 2 NSW    ACT             <aggregated>    <ETS(M,N,A)>
 3 NSW    Metro NSW       Central Coast   <ETS(M,N,M)>
 4 NSW    Metro NSW       Sydney          <ETS(M,N,A)>
 5 NSW    Metro NSW       <aggregated>    <ETS(M,N,A)>
 6 NSW    North Coast NSW Hunter          <ETS(M,N,M)>
 7 NSW    North Coast NSW North Coast NSW <ETS(M,N,M)>
 8 NSW    North Coast NSW <aggregated>    <ETS(M,N,M)>
 9 NSW    North NSW       Blue Mountains  <ETS(M,N,A)>
10 NSW    North NSW       Central NSW     <ETS(M,N,M)>
```

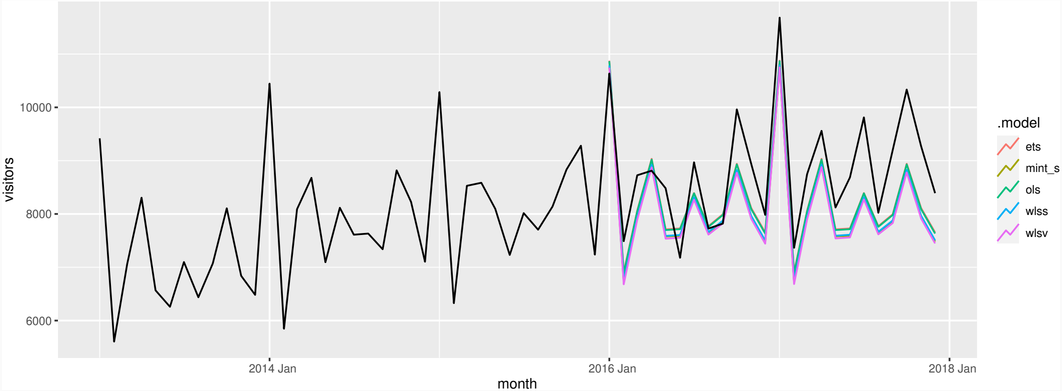
Example: Australian tourism

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

```
# A fable: 13,200 x 7 [1M]
# Key:     state, zone, region, .model [550]
   state  zone   region   .model    month       visitors .mean
   <chr*> <chr*> <chr*>   <chr>     <mth>         <dist> <dbl>
 1 NSW    ACT    Canberra ets    2016 Jan N(202, 1437)  202.
 2 NSW    ACT    Canberra ets    2016 Feb  N(160, 912)  160.
 3 NSW    ACT    Canberra ets    2016 Mar N(204, 1489)  204.
 4 NSW    ACT    Canberra ets    2016 Apr N(207, 1527)  207.
```

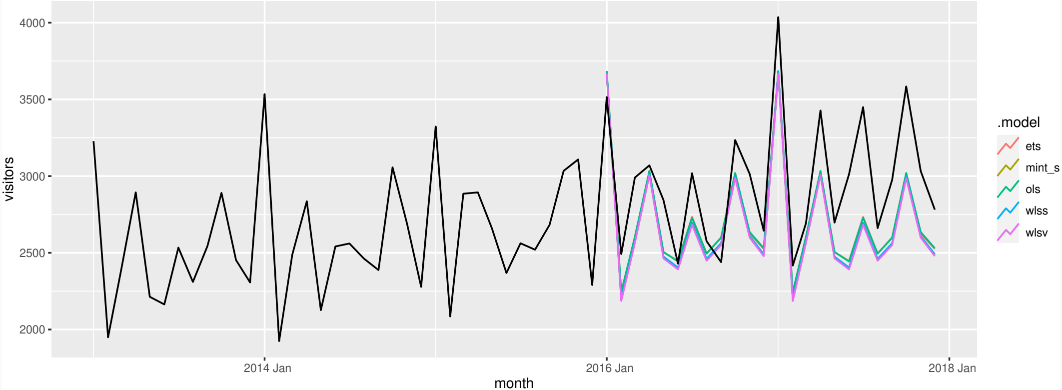
Example: Australian tourism

```
fc |>
  filter(is_aggregated(state)) |>
  autoplot(filter(tourism_agg, year(month) > 2012), level = NULL)
```



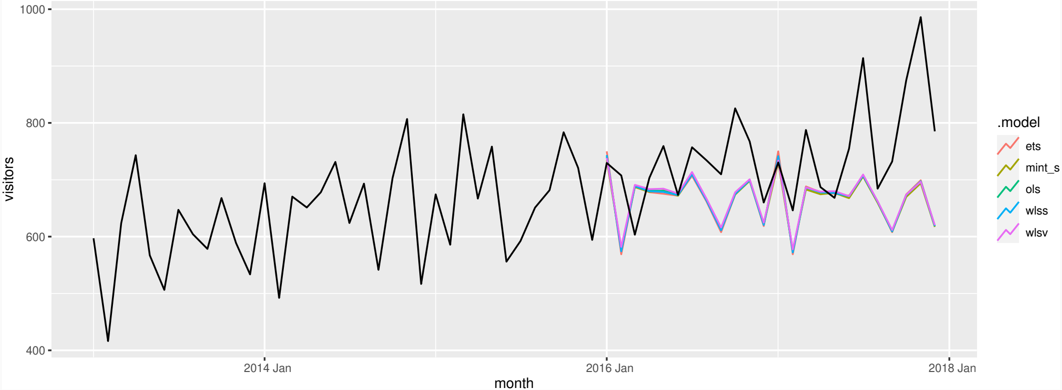
Example: Australian tourism

```
fc |>
  filter(state == "NSW" & is_aggregated(zone)) |>
  autoplot(filter(tourism_agg, year(month) > 2012), level = NULL)
```



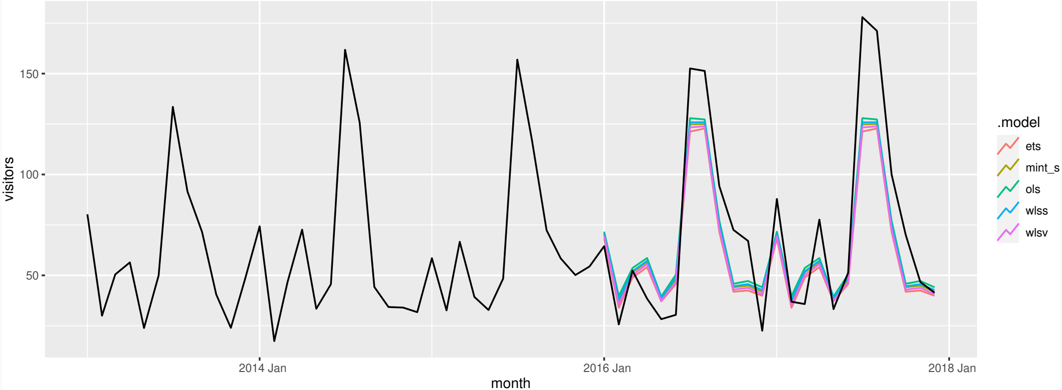
Example: Australian tourism

```r
fc |>
  filter(region == "Melbourne") |>
  autoplot(filter(tourism_agg, year(month) > 2012), level = NULL)
```



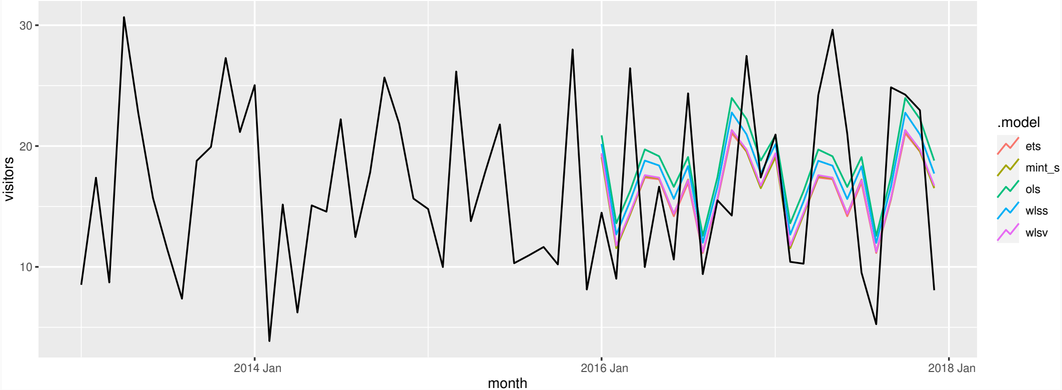
Example: Australian tourism

```
fc |>
  filter(region == "Snowy Mountains") |>
  autoplot(filter(tourism_agg, year(month) > 2012), level = NULL)
```



Example: Australian tourism

```
fc |>
  filter(region == "Barossa") |>
  autoplot(filter(tourism_agg, year(month) > 2012), level = NULL)
```



Performance evaluation

$$\text{MASE} = \text{mean}(|q_j|)$$

$$q_j = \frac{e_j}{\frac{1}{T-m}\sum_{t=m+1}^{T}|y_t - y_{t-m}|}$$

- y_t = observation for period t
- e_j = forecast error for forecast horizon j
- T = size of training set
- $m = 12$

Performance evaluation

$$\text{RMSSE} = \sqrt{\text{mean}(q_j^2)}$$

$$q_j^2 = \frac{e_j^2}{\frac{1}{T-m}\sum_{t=m+1}^{T}(y_t - y_{t-m})^2}$$

- y_t = observation for period t
- e_j = forecast error for forecast horizon j
- T = size of training set
- $m = 12$

Example: Australian tourism

```r
fc |>
  accuracy(tourism_agg, measures = list(mase = MASE, rmsse = RMSSE))
```

```
# A tibble: 550 x 7
   .model state  zone             region           .type  mase rmsse
   <chr>  <chr*> <chr*>           <chr*>           <chr> <dbl> <dbl>
 1 ets    NSW    ACT              Canberra         Test  0.866 0.835
 2 ets    NSW    ACT              <aggregated>     Test  0.866 0.835
 3 ets    NSW    Metro NSW        Central Coast    Test  0.777 0.747
 4 ets    NSW    Metro NSW        Sydney           Test  1.20  1.16
 5 ets    NSW    Metro NSW        <aggregated>     Test  1.18  1.18
 6 ets    NSW    North Coast NSW  Hunter           Test  1.33  1.21
 7 ets    NSW    North Coast NSW  North Coast NSW  Test  0.845 0.884
 8 ets    NSW    North Coast NSW  <aggregated>     Test  1.11  1.02
 9 ets    NSW    North NSW        Blue Mountains   Test  1.02  1.02
10 ets    NSW    North NSW        Central NSW      Test  1.30  1.18
# i 540 more rows
```

Example: Australian tourism

```
fc |>
  accuracy(tourism_agg, measures = list(mase = MASE, rmsse = RMSSE)) |>
  group_by(.model) |>
  summarise(mase = mean(mase), rmsse = sqrt(mean(rmsse^2))) |>
  arrange(rmsse)
```

```
# A tibble: 5 x 3
  .model  mase rmsse
  <chr>  <dbl> <dbl>
1 ols    0.930 0.926
2 wlss   0.949 0.948
3 mint_s 0.953 0.954
4 wlsv   0.964 0.965
5 ets    0.968 0.968
```

Example: Australian tourism

```
fc |>
  accuracy(tourism_agg, measures = list(mase = MASE, rmsse = RMSSE)) |>
  group_by(.model) |>
  summarise(mase = mean(mase), rmsse = sqrt(mean(rmsse^2))) |>
  arrange(rmsse)
```

```
# A tibble: 5 x 3
  .model  mase rmsse
  <chr>  <dbl> <dbl>
1 ols    0.930 0.926
2 wlss   0.949 0.948
3 mint_s 0.953 0.954
4 wlsv   0.964 0.965
5 ets    0.968 0.968
```

- Overall, every reconciliation method is better than the base ETS forecasts.

Example: Australian tourism

```
# A tibble: 20 x 4
# Groups:   .model [5]
   .model level     mase rmsse
   <chr>  <fct>    <dbl> <dbl>
 1 ets    National 1.44  1.27 
 2 ols    National 1.46  1.29 
 3 wlss   National 1.61  1.43 
 4 mint_s National 1.64  1.45 
 5 wlsv   National 1.69  1.49 
 6 ols    State    1.07  1.08 
 7 ets    State    1.10  1.11 
 8 wlss   State    1.13  1.14 
 9 mint_s State    1.15  1.15 
10 wlsv   State    1.18  1.17 
11 ols    Zone     0.954 0.948
12 wlss   Zone     0.987 0.980
13 mint_s Zone     0.995 0.988
14 ets    Zone     1.01  0.999
15 wlsv   Zone     1.01  1.00 
16 ols    Region   0.901 0.895
17 wlss   Region   0.910 0.907
18 mint_s Region   0.911 0.911
19 wlsv   Region   0.917 0.919
20 ets    Region   0.935 0.938
```

Example: Australian tourism

```
# A tibble: 20 x 4
# Groups:   .model [5]
   .model level     mase rmsse
   <chr>  <fct>    <dbl> <dbl>
 1 ets    National 1.44  1.27
 2 ols    National 1.46  1.29
 3 wlss   National 1.61  1.43
 4 mint_s National 1.64  1.45
 5 wlsv   National 1.69  1.49
 6 ols    State    1.07  1.08
 7 ets    State    1.10  1.11
 8 wlss   State    1.13  1.14
 9 mint_s State    1.15  1.15
10 wlsv   State    1.18  1.17
11 ols    Zone     0.954 0.948
12 wlss   Zone     0.987 0.980
13 mint_s Zone     0.995 0.988
14 ets    Zone     1.01  0.999
15 wlsv   Zone     1.01  1.00
16 ols    Region   0.901 0.895
17 wlss   Region   0.910 0.907
18 mint_s Region   0.911 0.911
19 wlsv   Region   0.917 0.919
20 ets    Region   0.935 0.938
```

- OLS is best for all levels except national.
- Improvements due to reconciliation are greater at lower levels.

Mean square error bounds

Panagiotelis et al. (2021)

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.

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)

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

Using sums of variances:

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

Non-negative forecasts

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

Non-negative forecasts

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

Solve via quadratic programming:

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

(Wickramasuriya, Turlach, and Hyndman, 2020)

Non-negative forecasts

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

Solve via quadratic programming:

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

(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, 2023c)

Reconciliation and regularization

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

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

Reconciliation and regularization

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

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

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

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

Outline

1 Hierarchical time series notation

2 Optimal linear forecast reconciliation

3 Probabilistic forecast reconciliation

4 Cross-temporal probabilistic forecast reconciliation

The coherent subspace

Coherent subspace

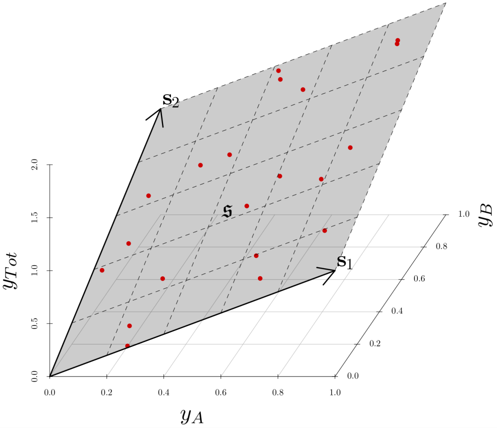
m-dimensional linear subspace $\mathfrak{s} \subset \chi^n$ for which linear constraints hold for all $\boldsymbol{y} \in \mathfrak{s}$.

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 $\mathfrak{s} \subset \chi^n$ for which linear constraints hold for all $\boldsymbol{y} \in \mathfrak{s}$.

Hierarchical time series

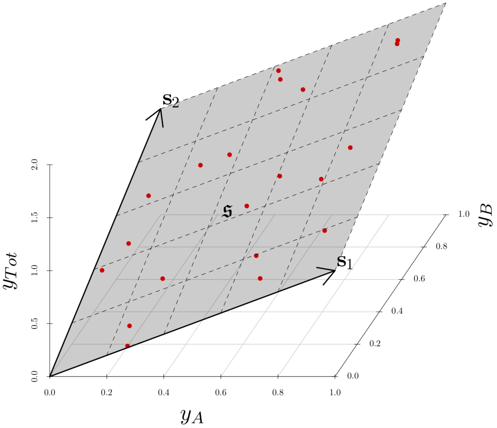
An n-dimensional multivariate time series such that $\boldsymbol{y}_t \in \mathfrak{s} \quad \forall t$.

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

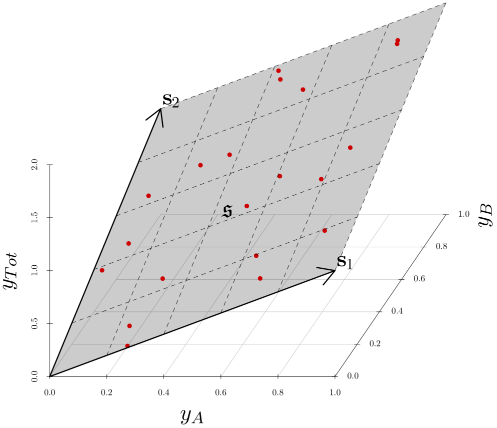
m-dimensional linear subspace $\mathfrak{s} \subset \chi^n$ for which linear constraints hold for all $\boldsymbol{y} \in \mathfrak{s}$.

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, $\psi : \chi^n \rightarrow \mathfrak{s}$. $\tilde{\boldsymbol{y}}_{t+h|t} = \psi(\hat{\boldsymbol{y}}_{t+h|t})$ "reconciles" $\hat{\boldsymbol{y}}_{t+h|t}$.

Coherent probabilistic forecasts

Coherent probabilistic forecasts

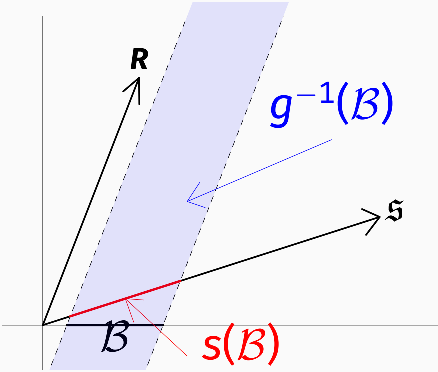
A probability triple $(\mathfrak{s}, \mathscr{F}_{\mathfrak{s}}, \breve{\nu})$ is coherent with the bottom probability triple $(\chi^m, \mathscr{F}_{\chi^m}, \nu)$, if

$$\breve{\nu}(s(\mathcal{B})) = \nu(\mathcal{B}) \quad \forall \mathcal{B} \in \mathscr{F}_{\chi^m}$$

i.e., probability of any point not on $\mathfrak{s}$ is zero.

Probabilistic forecast reconciliation

Let $(\chi^n, \mathscr{F}_{\chi^n}, \hat{\nu})$ be the base forecast. Then the reconciled probability distribution $\breve{\nu}$ is a transformation of $\hat{\nu}$ that is coherent on $\mathscr{F}_{\mathfrak{s}}$.



$\psi = s \circ g$

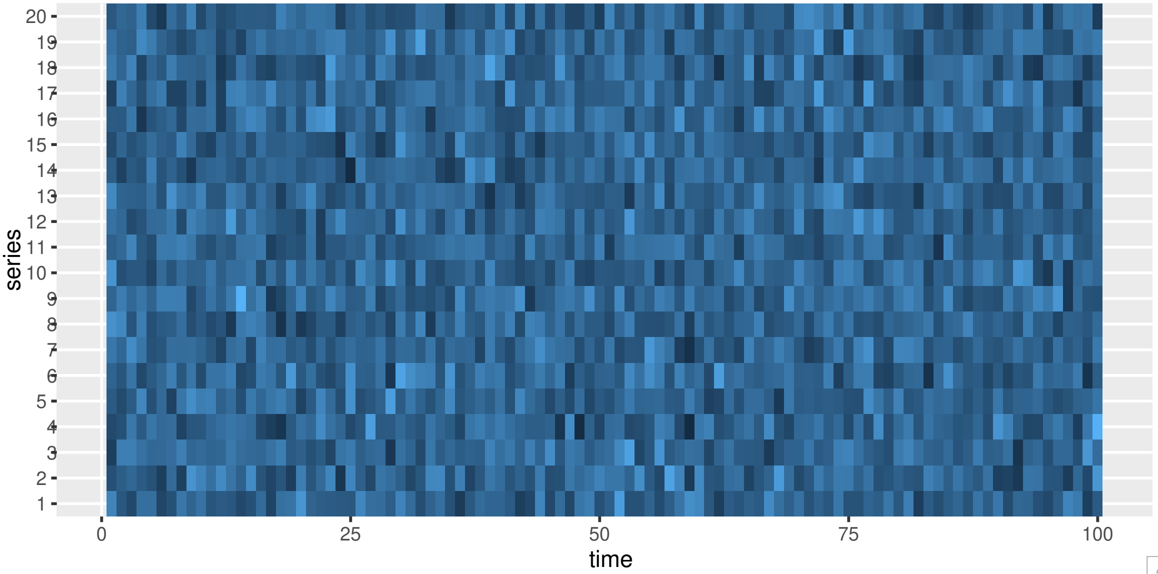
Panagiotelis et al. (2023) and Corani, Azzimonti, and Rubattu (2023)

Simulation from a reconciled distribution

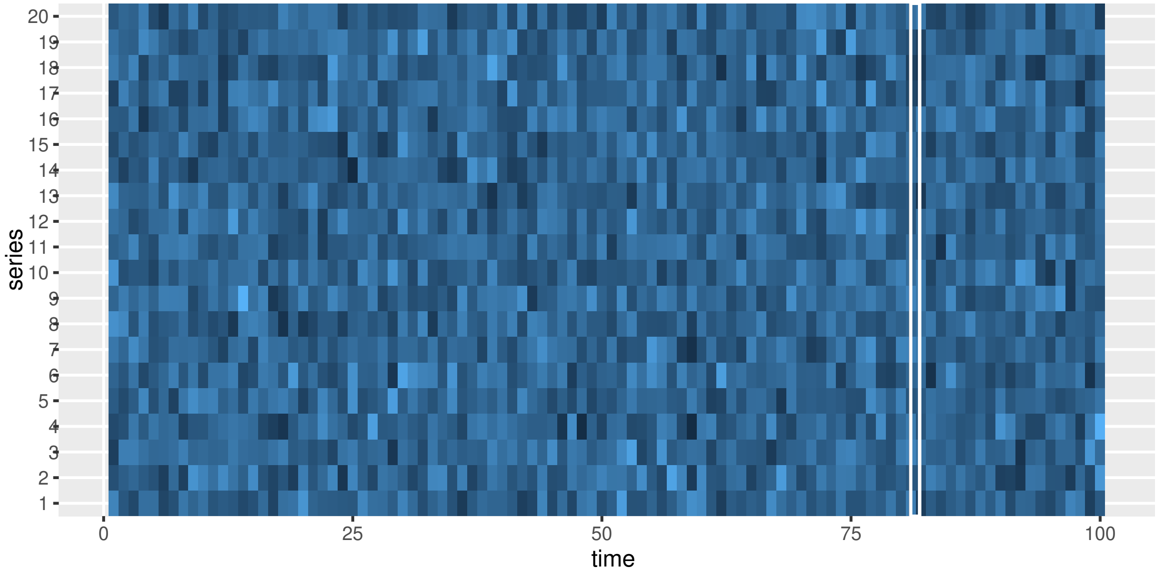
Suppose that $(\hat{\boldsymbol{y}}^{[1]},\dots,\hat{\boldsymbol{y}}^{[L]})$ is a sample drawn from an incoherent probability measure $\hat{\nu}$. Then $(\tilde{\boldsymbol{y}}^{[1]},\dots,\tilde{\boldsymbol{y}}^{[L]})$ where $\tilde{\boldsymbol{y}}^{[\ell]} := \psi(\hat{\boldsymbol{y}}^{[\ell]})$ for $\ell = 1,\dots,L$, is a sample drawn from the reconciled probability measure $\tilde{\nu}$.

- Simulate future sample paths for each series, by simulating from each model using a multivariate bootstrap of the residuals (to preserve cross-correlations).
- Reconcile the sample paths.
- The reconciled sample paths are a sample from the reconciled distribution.

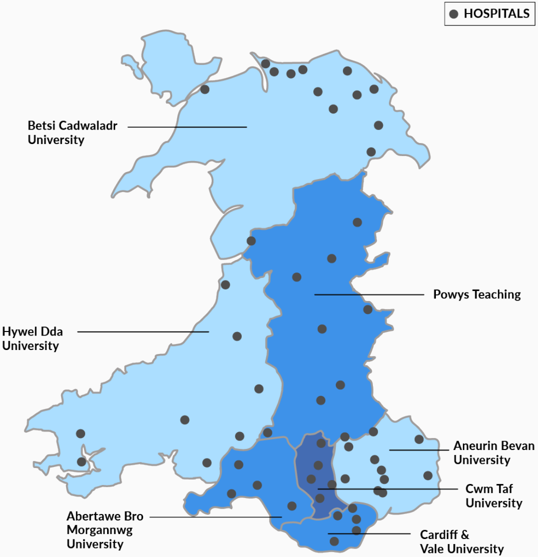
Simulation from a reconciled distribution



Simulation from a reconciled distribution



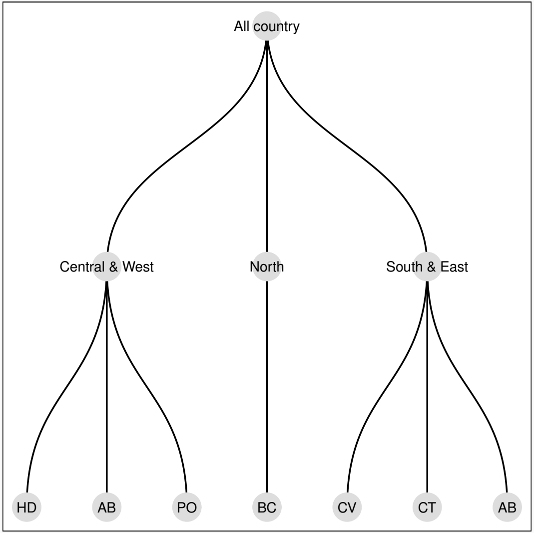
Wales Health Board Areas



Data

- Daily number of attended incidents:
  1 October 2015 – 31 July 2019
- Disaggregated by:
  - control area
  - health board
  - priority
  - nature of incidents
- 2,142,000 rows observations from 1,530 time series.

Data structure



*

Priority
Red
Amber
Green

*

Nature of incident
Chest pain
Stroke
Breathing problem
...
Abdominal pain

Data structure

Level	Number of series
All country	1
Control	3
Health board	7
Priority	3
Priority * Control	9
Priority * Health board	21
Nature of incident	35
Nature of incident * Control	105
Nature of incident * Health board	245
Priority * Nature of incident	104
Control * Priority * Nature of incident	306
Control * Health board * Priority * Nature of incident (Bottom level)	691
Total	1530

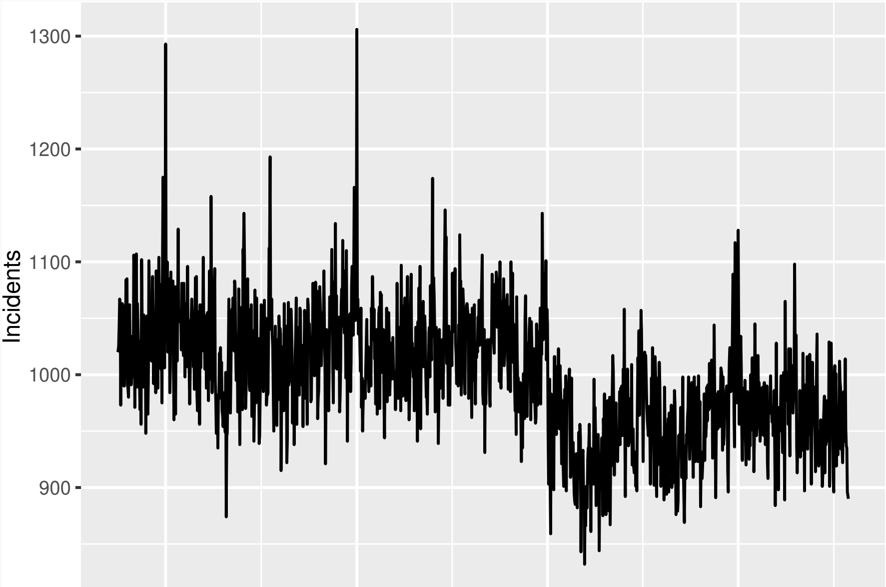
Data

```
# A tsibble: 2,142,000 x 6 [1D]
# Key:       region, category, nature, lhb [1,530]
   date       region       category     nature       lhb          incident
   <date>     <chr*>       <chr*>       <chr*>       <chr*>          <dbl>
 1 2015-10-01 <aggregated> <aggregated> <aggregated> <aggregated>     1020
 2 2015-10-02 <aggregated> <aggregated> <aggregated> <aggregated>     1021
 3 2015-10-03 <aggregated> <aggregated> <aggregated> <aggregated>     1025
 4 2015-10-04 <aggregated> <aggregated> <aggregated> <aggregated>     1043
 5 2015-10-05 <aggregated> <aggregated> <aggregated> <aggregated>     1067
 6 2015-10-06 <aggregated> <aggregated> <aggregated> <aggregated>     1063
 7 2015-10-07 <aggregated> <aggregated> <aggregated> <aggregated>      973
 8 2015-10-08 <aggregated> <aggregated> <aggregated> <aggregated>     1057
 9 2015-10-09 <aggregated> <aggregated> <aggregated> <aggregated>     1026
10 2015-10-10 <aggregated> <aggregated> <aggregated> <aggregated>     1063
# i 2,141,990 more rows
```

Data

```
# A tsibble: 2,142,000 x 6 [1D]
# Key:       region, category, nature, lhb [1,530]
   date       region category nature    lhb          incident
   <date>     <chr*> <chr*>   <chr*>    <chr*>          <dbl>
 1 2015-10-01 C      Amber    ABDOMINAL HD                  0
 2 2015-10-01 C      Amber    ABDOMINAL PO                  0
 3 2015-10-01 C      Amber    ABDOMINAL SB                  0
 4 2015-10-01 C      Amber    ABDOMINAL <aggregated>        0
 5 2015-10-01 C      Amber    ALLERGIES HD                  0
 6 2015-10-01 C      Amber    ALLERGIES PO                  1
 7 2015-10-01 C      Amber    ALLERGIES SB                  0
 8 2015-10-01 C      Amber    ALLERGIES <aggregated>        1
 9 2015-10-01 C      Amber    ANIMALBIT HD                  0
10 2015-10-01 C      Amber    ANIMALBIT PO                  0
# i 2,141,990 more rows
```

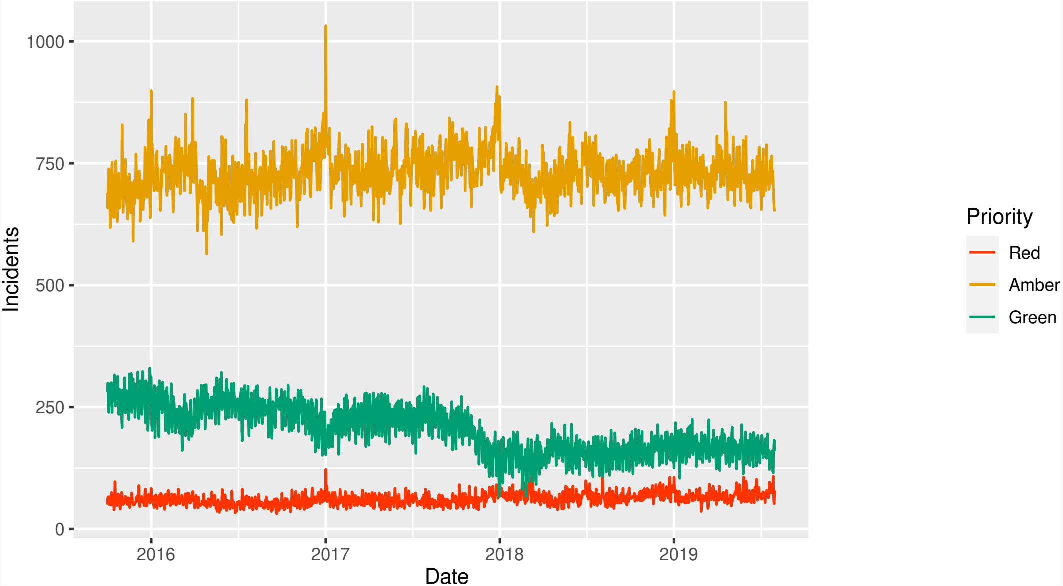
Aggregated daily incidents



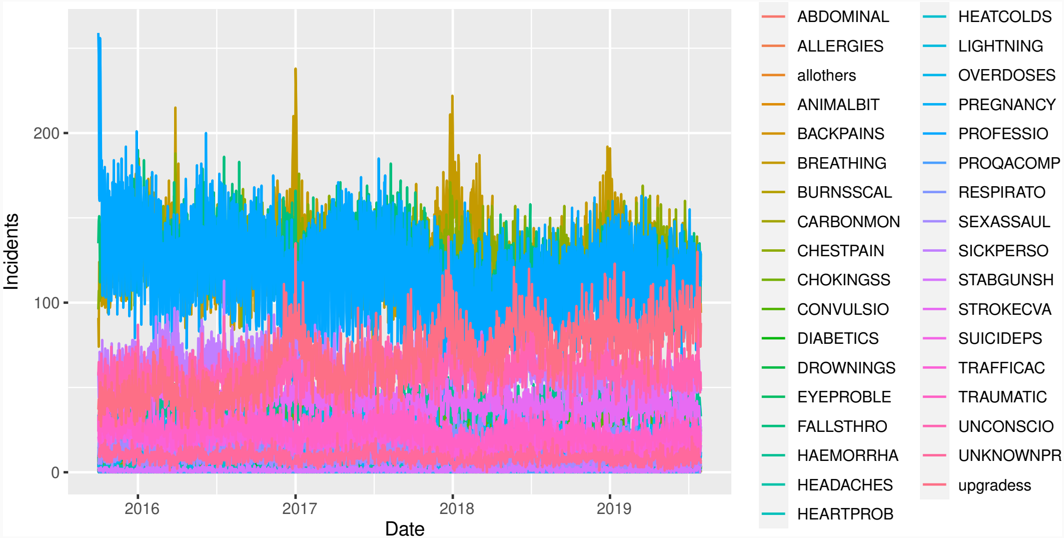
Daily incidents by control area

Data incidents by health board

Data incidents by priority



Data incidents by nature of incident



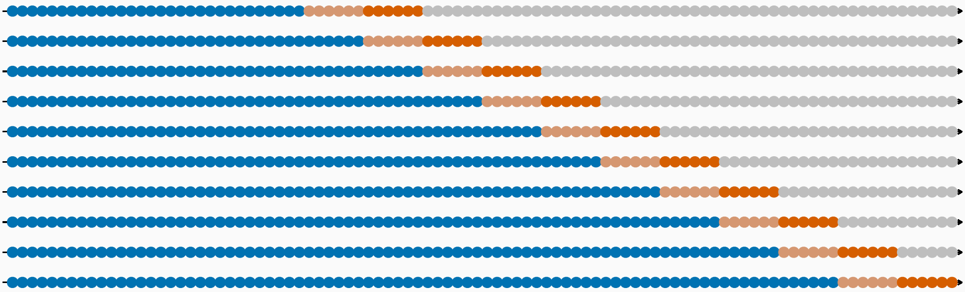
Data incidents by nature of incident

Forecasting methods

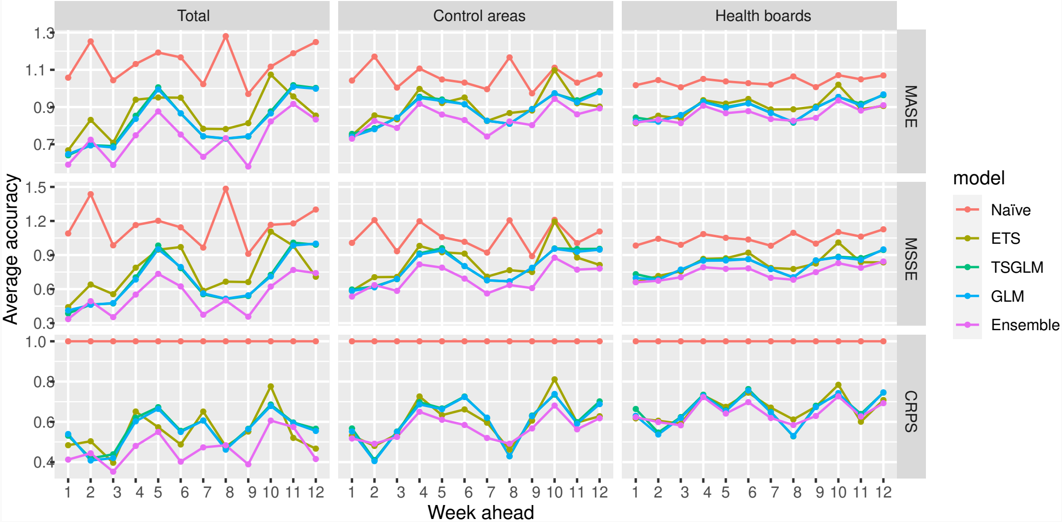
1. **Naïve**: Empirical distribution of past daily attended incidents.
2. **ETS**: Exponential Smoothing State Space models.
3. **GLM**: Poission Regression with spline trend, day of the week, annual Fourier seasonality, public holidays, school holidays, Christmas Day, New Year's Day.
4. **TSGLM**: Poisson Regression with same covariates plus three autoregressive terms.
5. **Ensemble**: Mixture distribution of 1–4.

Performance evaluation

- Ten-fold time series cross-validation
- Forecast horizon of 1–84 days
- Each training set contains an additional 42 days.
- Forecasts at 43–84 days correspond to planning horizon.



Forecast accuracy



Forecast accuracy: 43–84 days ahead

		MSSE			
Method	Model	Total	Control areas	Health boards	Bottom
Base	Naïve	1.169	1.056	1.062	1.031
Base	ETS	0.979	0.875	0.816	**0.975**
Base	GLM	0.813	0.897	0.875	1.009
Base	TSGLM	0.822	0.901	0.875	1.050
Base	Ensemble	0.599	0.729	0.774	0.993
MinT	Naïve	1.168	1.057	1.062	2.095
MinT	ETS	0.785	0.852	0.845	0.994
MinT	GLM	0.720	0.827	0.837	1.803
MinT	TSGLM	0.722	0.833	0.839	1.851
MinT	Ensemble	**0.560**	**0.706**	**0.765**	1.557

Forecast accuracy: 43–84 days ahead

		MASE			
Method	Model	Total	Control areas	Health boards	Bottom
Base	Naïve	1.139	1.059	1.047	1.019
Base	ETS	0.963	0.930	0.899	1.038
Base	GLM	0.910	0.940	0.923	**1.002**
Base	TSGLM	0.911	0.939	0.924	1.005
Base	Ensemble	0.782	0.856	0.876	1.008
MinT	Naïve	1.138	1.059	1.047	2.651
MinT	ETS	0.877	0.916	0.915	1.289
MinT	GLM	0.848	0.901	0.902	2.493
MinT	TSGLM	0.852	0.903	0.903	2.513
MinT	Ensemble	**0.753**	**0.844**	**0.872**	2.260

Forecast accuracy: 43–84 days ahead

		CRPS			
Method	Model	Total	Control areas	Health boards	Bottom
Base	Naïve	30.387	10.882	5.500	0.302
Base	ETS	14.309	6.074	3.476	0.244
Base	GLM	15.396	6.253	3.576	0.244
Base	TSGLM	15.316	6.227	3.575	0.245
Base	Ensemble	12.978	**5.727**	3.430	0.243
MinT	Naïve	30.368	10.902	5.498	0.313
MinT	ETS	13.515	5.967	3.547	**0.243**
MinT	GLM	13.839	5.917	3.453	0.246
MinT	TSGLM	14.000	5.947	3.455	0.248
MinT	Ensemble	**12.585**	5.728	**3.426**	0.247

Forecast accuracy

- Ensemble mixture distributions give better forecasts than any component methods.
- Forecast reconciliation improves forecast accuracy, even when some component methods are quite poor.
- The ensemble without the Naïve method was worse.
- Forecast reconciliation allows coordinated planning and resource allocation.

Outline

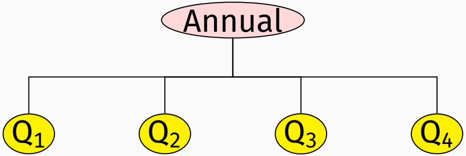
1 Hierarchical time series notation

2 Optimal linear forecast reconciliation

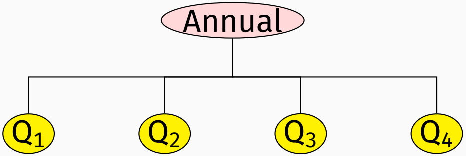
3 Probabilistic forecast reconciliation

4 Cross-temporal probabilistic forecast reconciliation

Temporal reconciliation: quarterly data

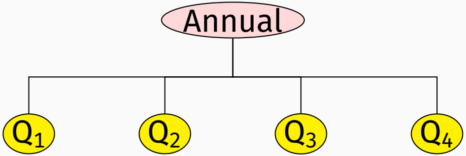


Temporal reconciliation: quarterly data



- Forecast series at each available frequency.
- Optimally combine forecasts within the same year.

Temporal reconciliation: quarterly data

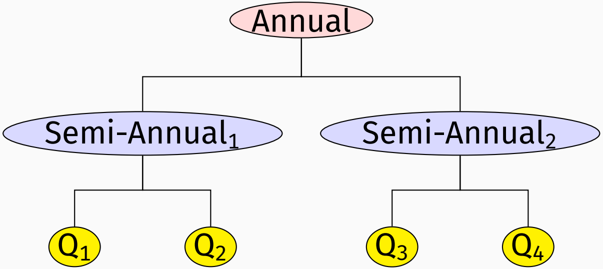


$$\boldsymbol{y}_\tau = \begin{bmatrix} x_\tau^{[4]} \\ x_{\tau,1}^{[1]} \\ x_{\tau,2}^{[1]} \\ x_{\tau,3}^{[1]} \\ x_{\tau,4}^{[1]} \end{bmatrix} \qquad \boldsymbol{S} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- Forecast series at each available frequency.
- Optimally combine forecasts within the same year.

τ = index of largest temporal aggregation level.

Temporal reconciliation: quarterly data

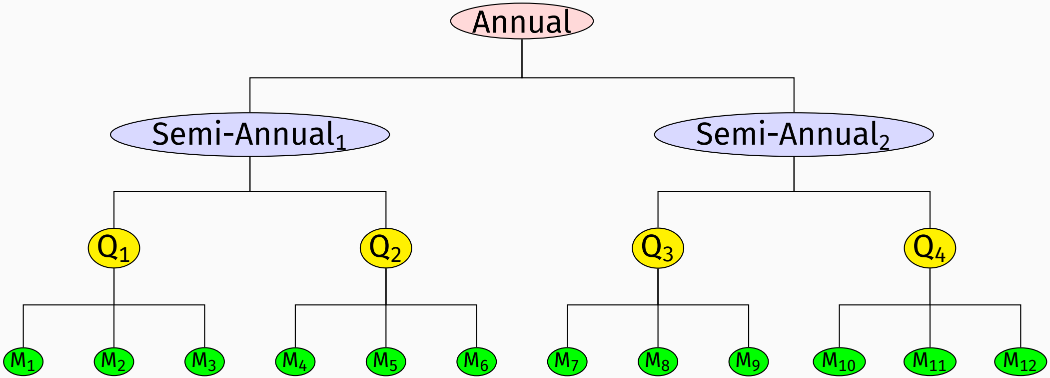


- Forecast series at each available frequency.
- Optimally combine forecasts within the same year.

$$\boldsymbol{y}_\tau = \begin{bmatrix} x_\tau^{[4]} \\ x_{\tau,1}^{[2]} \\ x_{\tau,2}^{[2]} \\ x_{\tau,1}^{[1]} \\ x_{\tau,2}^{[1]} \\ x_{\tau,3}^{[1]} \\ x_{\tau,4}^{[1]} \end{bmatrix} \qquad \boldsymbol{S} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

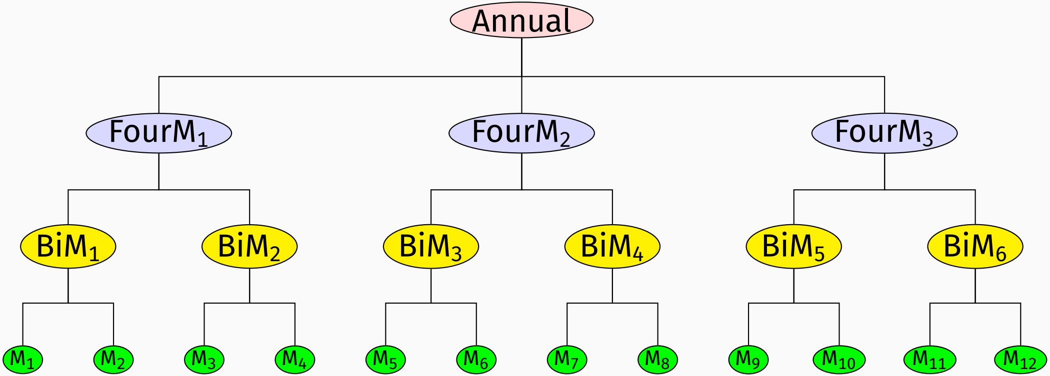
τ = index of largest temporal aggregation level.

Temporal reconciliation: monthly data



- Forecast series at each available frequency.
- Optimally combine forecasts within the same year.

Temporal reconciliation: monthly data



- Forecast series at each available frequency.
- Optimally combine forecasts within the same year.

Temporal reconciliation: monthly data

$$
\boldsymbol{y}_\tau = \begin{bmatrix} x_\tau^{[12]} \\ \boldsymbol{x}_\tau^{[6]} \\ \boldsymbol{x}_\tau^{[4]} \\ \boldsymbol{x}_\tau^{[3]} \\ \boldsymbol{x}_\tau^{[2]} \\ \boldsymbol{x}_\tau^{[1]} \end{bmatrix}
\qquad
\boldsymbol{S} = \begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & & & & & & \vdots & & & & & \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & & & & & & \vdots & & & & & \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
 & & & & & & & & & & & \\
 & & & & & & \boldsymbol{I}_{12} & & & & &
\end{bmatrix}
$$

Temporal reconciliation

For a time series $y_1,\dots,y_T$, observed at frequency m:

$$x_j^{[k]} = \sum_{t=(j-1)k+1}^{jk} y_t \qquad \text{for } j = 1,\dots,\lfloor T/k \rfloor$$

- $k \in \mathcal{K} = \{k_1,\dots,k_p\}$ denote the p factors of m in ascending order, where $k_1 = 1$ and $k_p = m$
- $x_j^{[1]} = y_t$
- A single unique hierarchy is only possible when there are no coprime pairs in $\mathcal{K}$.
- $M_k = m/k$ is seasonal period of aggregated series.

Temporal reconciliation

$$\boldsymbol{x}_\tau = \boldsymbol{S}\boldsymbol{x}_\tau^{[1]}, \qquad \boldsymbol{S} = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{I} \end{bmatrix}$$

where

$$\boldsymbol{x}_\tau = \begin{bmatrix} x_\tau^{[k_p]} \\ \boldsymbol{x}_\tau^{[k_{p-1}]} \\ \vdots \\ \boldsymbol{x}_\tau^{[k_1]} \end{bmatrix} \qquad \boldsymbol{x}_\tau^{[k]} = \begin{bmatrix} x_{M_k(\tau-1)+1}^{[k]} \\ x_{M_k(\tau-1)+2}^{[k]} \\ \vdots \\ x_{M_k\tau}^{[k]} \end{bmatrix} \qquad \boldsymbol{A} = \begin{bmatrix} \boldsymbol{1}'_m \\ \boldsymbol{I}_{m/k_{p-1}} \otimes \boldsymbol{1}'_{k_{p-1}} \\ \vdots \\ \boldsymbol{I}_{m/k_2} \otimes \boldsymbol{1}'_{k_2} \end{bmatrix}$$

τ is time index for most aggregated series,

$k \in \mathcal{K} = \{k_1, \dots, k_p\}, \quad k_1 = 1, \quad k_p = m, \quad \tau = 1, \dots, T/m.$

Cross-temporal forecast reconciliation

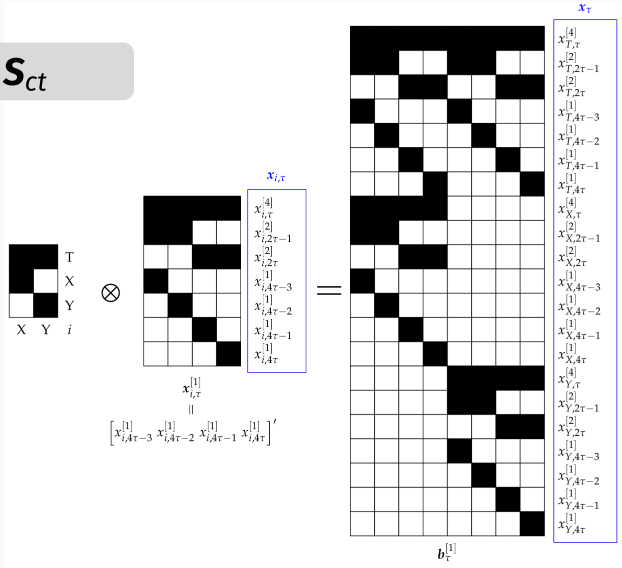
Structural matrix approach

- $\boldsymbol{S}_{cs}$ = structural cross-sectional matrix
- $\boldsymbol{S}_{te}$ = structural temporal matrix
- $\boldsymbol{S}_{ct} = \boldsymbol{S}_{cs} \otimes \boldsymbol{S}_{te}$

$$\boldsymbol{x}_\tau = \boldsymbol{S}_{ct}\boldsymbol{b}_\tau, \qquad \text{where} \qquad \boldsymbol{b}_\tau = \begin{bmatrix} \boldsymbol{x}_{1,\tau}^{[1]} \\ \vdots \\ \boldsymbol{x}_{n,\tau}^{[1]} \end{bmatrix}.$$

Cross-temporal forecast reconciliation

$$\boldsymbol{S}_{cs} \otimes \boldsymbol{S}_{te} = \boldsymbol{S}_{ct}$$



Cross-temporal forecast reconciliation

Constraint matrix approach

- $\boldsymbol{C}_{cs}$ = cross-sectional constraint matrix
- $\boldsymbol{C}_{te}$ = temporal constraint matrix

$$\boldsymbol{C}_{ct}\boldsymbol{x}_{\tau} = \boldsymbol{0} \qquad \text{where} \qquad \boldsymbol{C}_{ct} = \begin{bmatrix} (\boldsymbol{0}_{(n_a m \times nk^*)} \quad \boldsymbol{I}_m \otimes \boldsymbol{C}_{cs})\boldsymbol{P}' \\ \boldsymbol{I}_n \otimes \boldsymbol{C}_{te} \end{bmatrix}$$

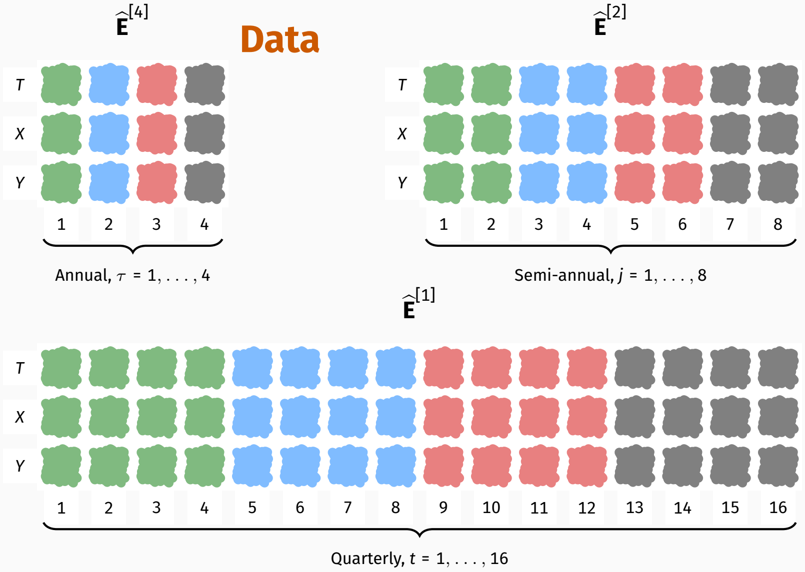
- $k^* = \displaystyle\sum_{k \in \mathcal{K} \setminus \{k_1\}} \frac{m}{k}$
- $\boldsymbol{P}$ = the permutation matrix such that $\boldsymbol{P}\text{vec}(\boldsymbol{Y}_{\tau}) = \text{vec}(\boldsymbol{Y}'_{\tau})$.

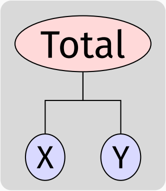
Cross-temporal probabilistic forecast reconciliation

Nonparametric bootstrap

- Simulate future sample paths from all models using bootstrapped residuals, then reconcile them to obtain coherent sample paths.
- Need to generate samples that preserve cross-temporal relationships.
- Draw residual samples of all series at same time from most temporally aggregated level.
- Residuals for other levels obtained using the corresponding time indices.

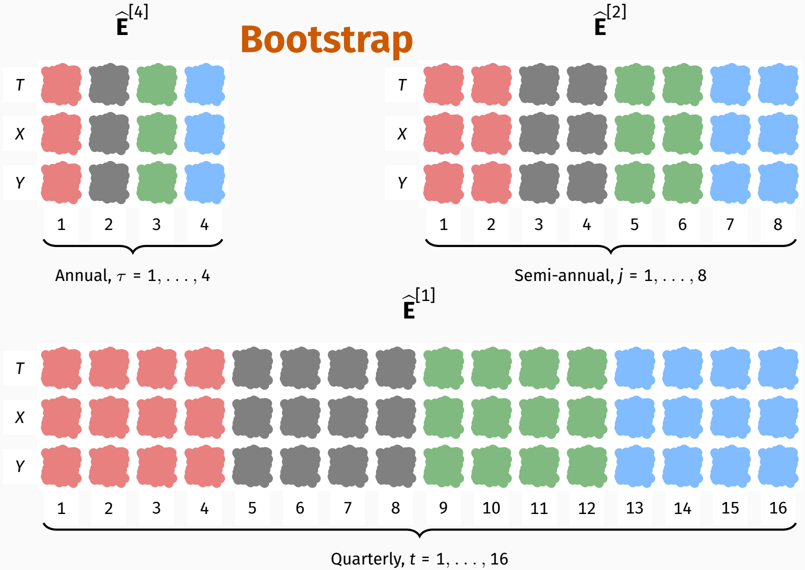
Cross-temporal probabilistic forecast reconciliation

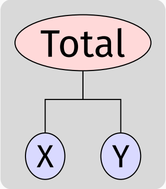




Cross-temporal probabilistic forecast reconciliation

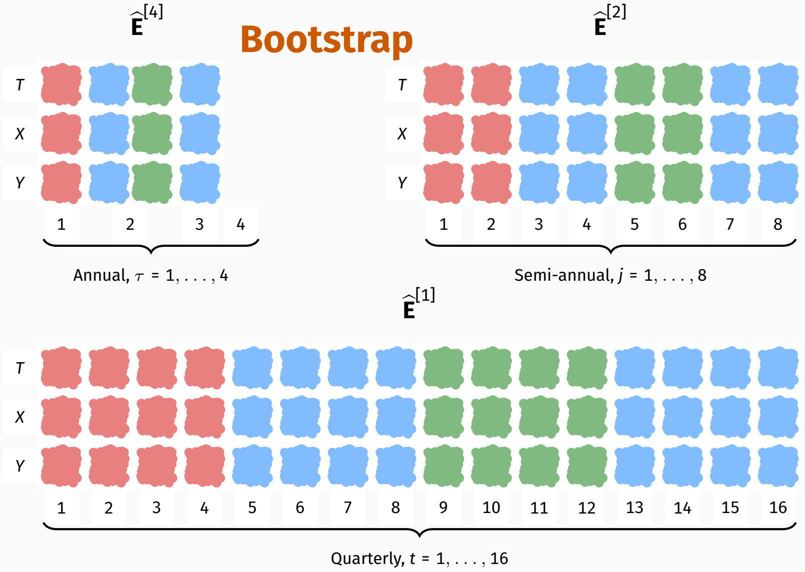
Bootstrap

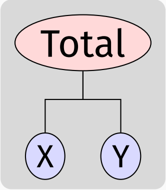




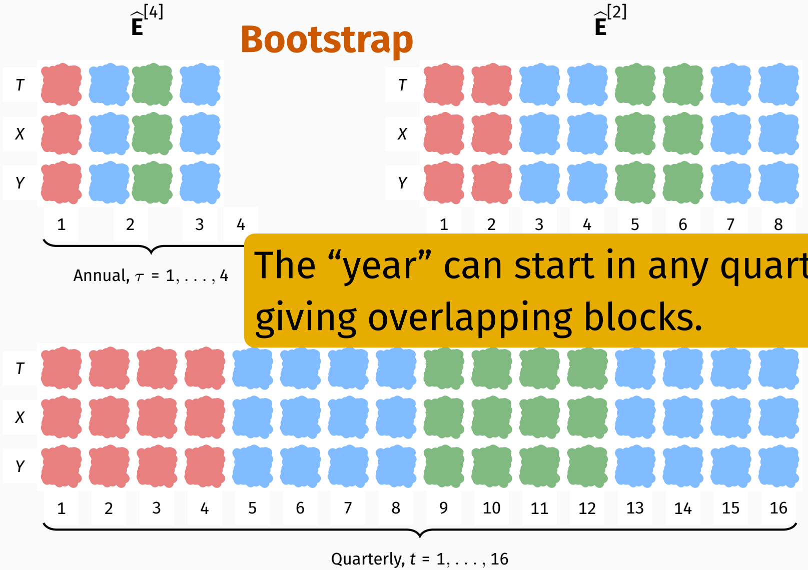
Cross-temporal probabilistic forecast reconciliation

Bootstrap

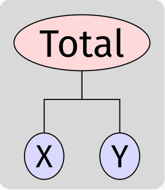




Cross-temporal probabilistic forecast reconciliation



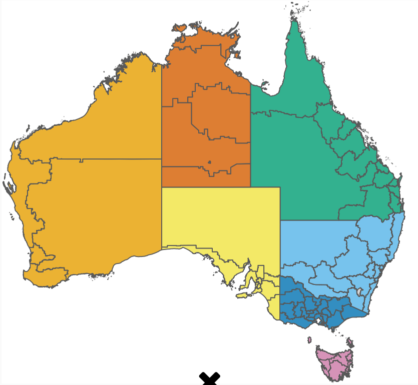
The "year" can start in any quarter, giving overlapping blocks.



Year 1
Year 2
Year 3
Year 4

Monthly Australian Tourism Demand

Geographical division



×

Purpose of travel

Holiday, Visiting friends & relatives, Business, Other

- **Grouped ts**
  (geographical divisions × purpose of travel)

	AUS	States	Zones*	Regions	Tot
geographical	1	7	21	76	105
purpose	4	28	84	304	420
Tot	5	35	105	380	**525**

$n_a = 221$, $n_b = 304$, and $\mathbf{n = 525}$

- **Temporal framework**, frequencies:
  - Monthly
  - Bi-Monthly
  - Quarterly
  - Four-Monthly
  - Semi-Annual
  - Annual

Monthly Australian Tourism Demand

- Monthly data: January 1998 to December 2016
- Time series cross-validation; initial training set 10 years.
- One-month increase in each training set
- For each training set, compute temporally aggregated series for $k \in \{1, 2, 3, 4, 6, 12\}$, and produce forecasts up to $h_2 = 6$, $h_3 = 4$, $h_4 = 3$, $h_6 = 2$ and $h_{12} = 1$ steps ahead.
- Automatic ETS forecasts on log-transformed data

Monthly Australian Tourism Demand

Reconciliation approaches

- Cross-temporal **bottom-up** and **partly bottom-up**

ct(bu)	ct(shr_{cs}, bu_{te})	ct($wlsv_{te}, bu_{cs}$)

- Optimal forecast reconciliation with **one-step residuals**

oct(ols)	oct($struc$)	oct($wlsv$)	oct($bdshr$)

- Optimal forecast reconciliation with **multi-step residuals**

oct$_h$($hbshr$)	oct$_h$($bshr$)	oct$_h$($hshr$)	oct$_h$(shr)

Monthly Australian tourism data – CRPS skill scores

Worse than benchmark **Best**

	$\forall k \in \{12, 6, 4, 3, 2, 1\}$	$k = 1$
base	1.000	1.000
ct(bu)	1.321	1.077
ct(shr_{cs}, bu_{te})	1.057	0.976
ct($wlsv_{te}, bu_{cs}$)	1.062	0.976
oct(ols)	0.989	0.982
oct($struc$)	0.982	0.970
oct($wlsv$)	0.987	0.952
oct($bdshr$)	0.975	**0.949**
oct$_h$($hbshr$)	0.989	0.982
oct$_h$($bshr$)	0.994	0.988
oct$_h$($hshr$)	**0.969**	0.953
oct$_h$(shr)	1.007	1.000

Software

Package	Language	Cross-sectional	Temporal	Cross-temporal	Probabilistic
`hts`	R	✓			
`thief`	R		✓		
`fable`	R	✓			✓
`FoReco`	R	✓	✓	✓	✓
`pyhts`	Python	✓	✓		
`hierarchicalforecast`	Python	✓			✓

- `hts`, `thief`, and `FoReco` use `ts` objects
- `fable` uses `tsibble` objects
- `fable` has plans to implement temporal and cross-temporal reconciliation

Thanks!

More information

- robjhyndman.com
- aus.social/@robjhyndman
- @robjhyndman
- rob.hyndman@monash.edu

References

- Ashouri, M, RJ Hyndman, and G Shmueli (2022). Fast forecast reconciliation using linear models. *J Computational & Graphical Statistics* **31**(1), 263–282.
- Athanasopoulos, G, RJ Hyndman, N Kourentzes, and F Petropoulos (2017). Forecasting with temporal hierarchies. *European J Operational Research* **262**(1), 60–74.
- Corani, G, D Azzimonti, and N Rubattu (2023). Probabilistic reconciliation of count time series. *International Journal of Forecasting* **forthcoming**.
- Di Fonzo, T and D Girolimetto (2023a). Cross-temporal forecast reconciliation: Optimal combination method and heuristic alternatives. *International Journal of Forecasting* **39**(1), 39–57.
- Di Fonzo, T and D Girolimetto (2023b). Forecast combination-based forecast reconciliation: Insights and extensions. *International Journal of Forecasting* **forthcoming**.

References

- Di Fonzo, T and D Girolimetto (2023c). Spatio-temporal reconciliation of solar forecasts. *Solar Energy* **251**, 13–29.
- Girolimetto, D, G Athanasopoulos, T Di Fonzo, and RJ Hyndman (2023). “Cross-temporal probabilistic forecast reconciliation”. robjhyndman.com/publications/ctprob.html.
- 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.
- Panagiotelis, A, P Gamakumara, G Athanasopoulos, and RJ Hyndman (2023). Probabilistic forecast reconciliation: properties, evaluation and score optimisation. *European J Operational Research* **306**(2), 693–706.

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.
- Wickramasuriya, SL (2021). Properties of point forecast reconciliation approaches. *arXiv preprint arXiv:2103.11129*.
- Wickramasuriya, SL, BA Turlach, and RJ Hyndman (2020). Optimal non-negative forecast reconciliation. *Statistics & Computing* **30**(5), 1167–1182.