Probabilistic cross-temporal forecast reconciliation

Rob J Hyndman

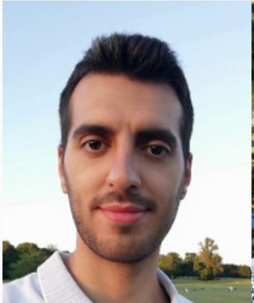
MONASH University

robjhyndman.com/ctprob

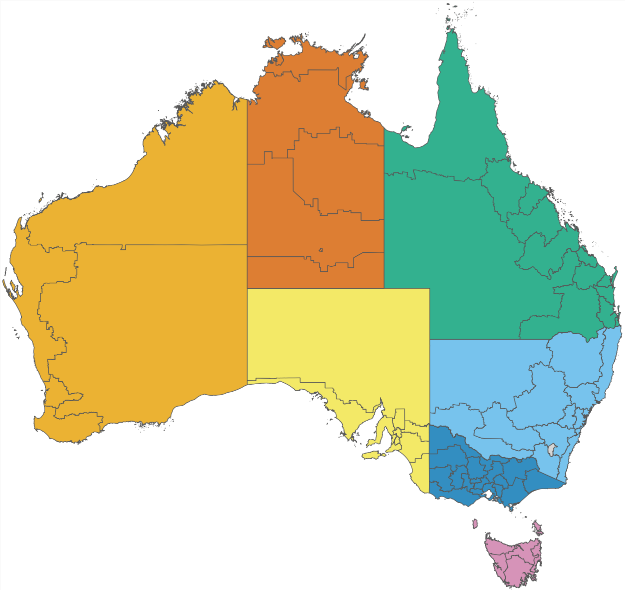
Photo by Edvard Alexander Rølvaag on Unsplash

Forthcoming paper

- Girolimetto, Athanasopoulos, Di Fonzo, Hyndman (2024) "Cross-temporal probabilistic forecast reconciliation: Methodological and practical issues".
- Preprint at `robjhyndman.com/ctprob`

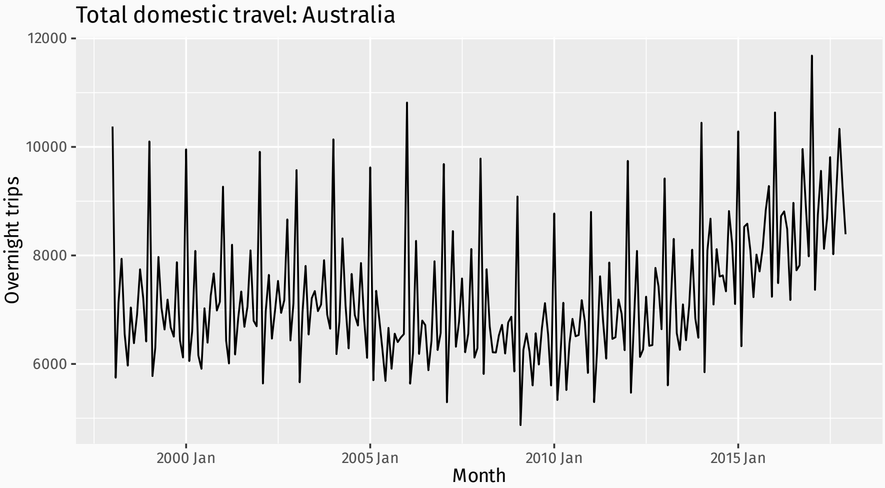


Australian tourism regions

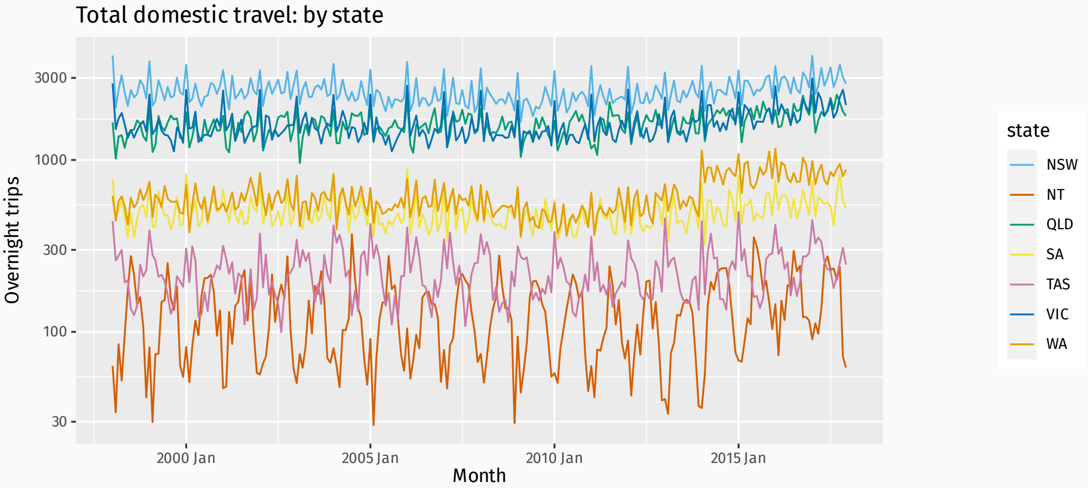


- Monthly data on visitor night from 1998 – 2017
- From *National Visitor Survey*, annual interviews of 120,000 Australians aged 15+.
- Geographical hierarchy split by
  - 7 states
  - 27 zones
  - 75 regions

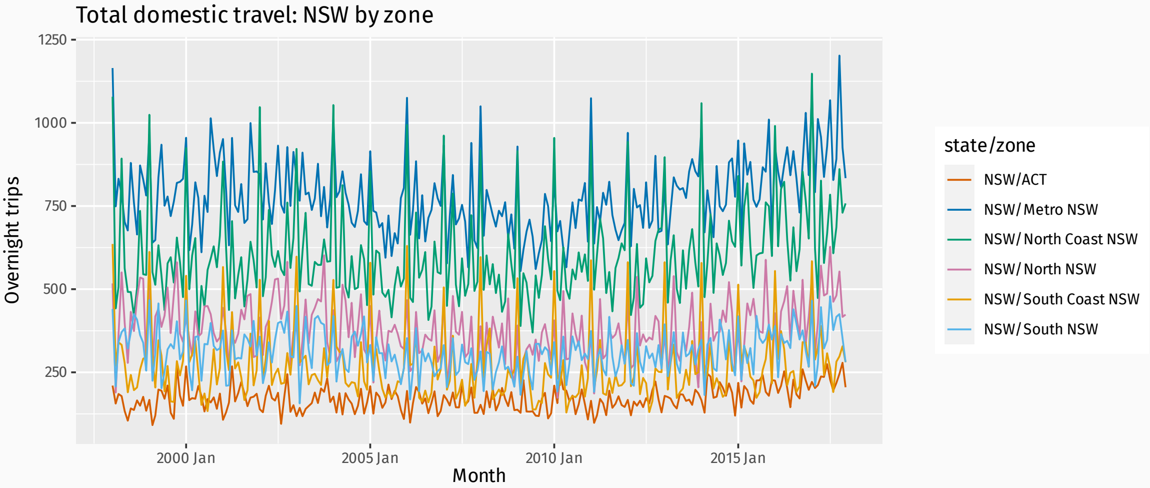
Australian tourism data



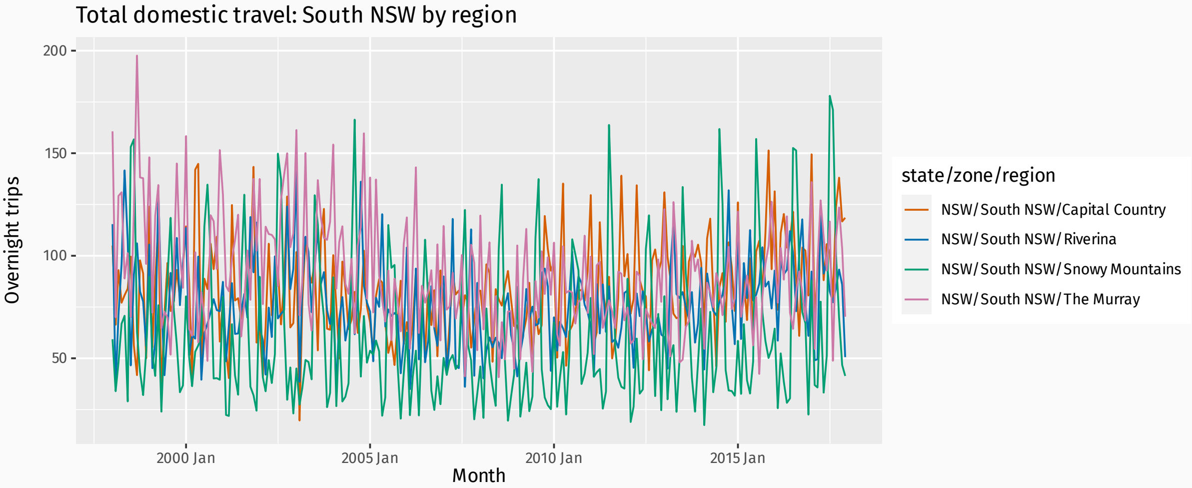
Australian tourism data



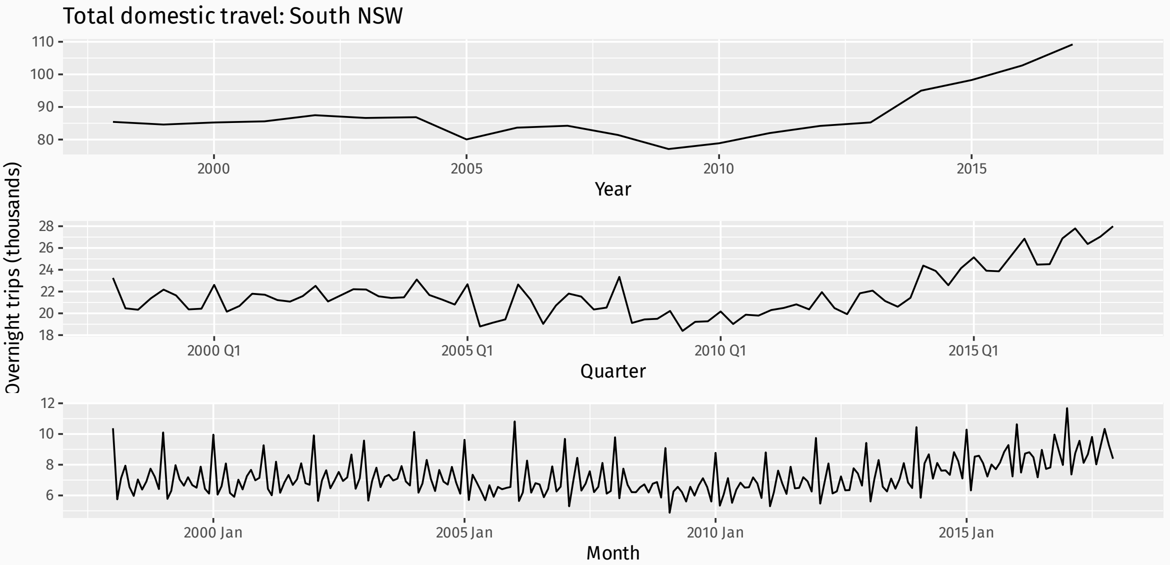
Australian tourism data



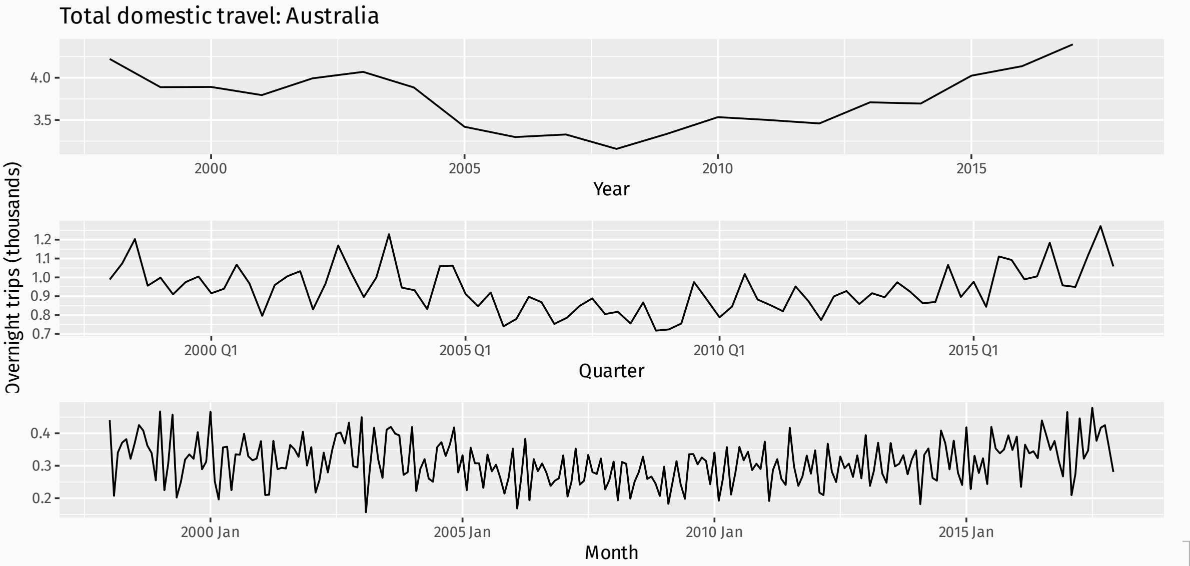
Australian tourism data



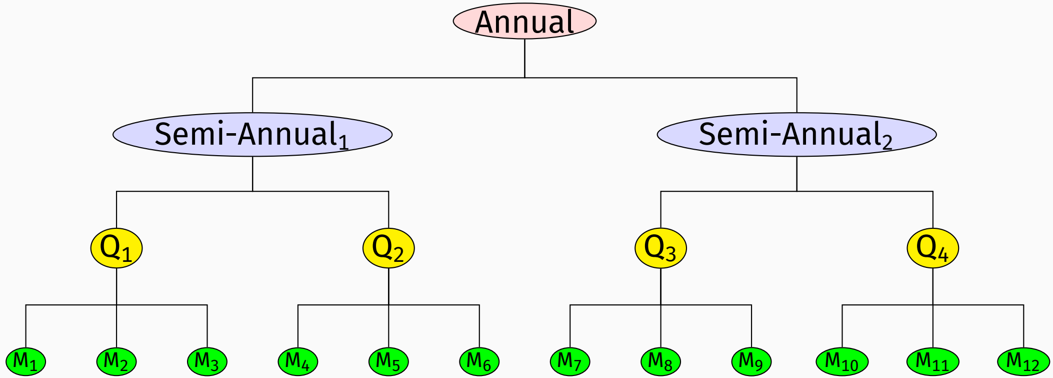
Australian tourism data



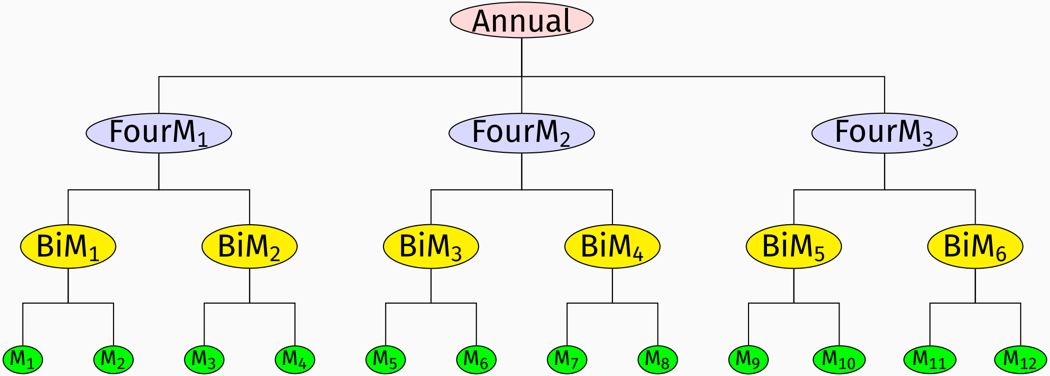
Australian tourism data



Temporal aggregations: monthly data



Temporal aggregations: monthly data



Australian tourism data

- national total
- 7 states & territories
- 21 zones
- 76 regions
- 4 purposes of travel (Business, Holidays, Visiting, Other)
- temporally aggregated into 2-month, 3-month, 4-month, 6-month and 12-month periods.

Most disaggregated series: $76 \times 4 = 304$ monthly series.

Total series: $(1 + 7 + 21 + 76) \times (1 + 4) \times 6 = 3150$ series.

Coherent cross-temporal forecasts

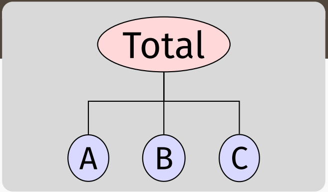
What we want

- We want forecasts of all series at all levels of cross-sectional aggregation.
- We want forecasts at monthly, quarterly, annual and other temporal aggregations.
- We want “coherent” probabilistic forecasts.

Solution

- We model and forecast all series independently.
- We “reconcile” the forecasts to make them coherent.

Notation



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

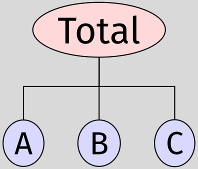
- $\boldsymbol{y}_t$ = vector of all series at time t
- $\boldsymbol{b}_t$ = vector of most disaggregated series at time t
- $\boldsymbol{S}$ = “structural matrix” containing the linear constraints.

$$\boldsymbol{y}_t = \begin{pmatrix} y_{\text{Total},t} \\ y_{A,t} \\ y_{B,t} \\ y_{C,t} \end{pmatrix}$$

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

Notation

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



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

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

- $\boldsymbol{S}_{cs}$ = cross-sectional aggregations.
- $\boldsymbol{S}_{te}$ = temporal aggregations.
- $\boldsymbol{S}_{ct} = \boldsymbol{S}_{cs} \otimes \boldsymbol{S}_{te}$
  = all cross-temporal aggregations.

Temporal constraints: monthly data

If $y_1, \dots, y_T$ observed at frequency m, and k is a factor of m, let:

- $x_j^{[k]}$ = sum of k consecutive observations from time $(j-1)k+1$.
- $\boldsymbol{x}_\tau^{[k]} = (x_\tau^{[k]}, \dots, x_{\tau+m/k-1}^{[k]})'$.

Temporal constraints: monthly data

If $y_1, \dots, y_T$ observed at frequency m, and k is a factor of m, let:

- $x_j^{[k]}$ = sum of k consecutive observations from time $(j-1)k+1$.
- $\boldsymbol{x}_\tau^{[k]} = (x_\tau^{[k]}, \dots, x_{\tau+m/k-1}^{[k]})'$.

$$\boldsymbol{y}_\tau = \begin{bmatrix} x_\tau^{[12]} \\ \boldsymbol{x}_\tau^{[6]} \\ \boldsymbol{x}_\tau^{[4]} \\ \boldsymbol{x}_\tau^{[3]} \\ \boldsymbol{x}_\tau^{[2]} \\ \boldsymbol{x}_\tau^{[1]} \end{bmatrix} \qquad \boldsymbol{S}_{te} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & & & \vdots & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ & & & & & & & & & & & \\ & & & & & \boldsymbol{I}_{12} & & & & & & \end{bmatrix} \qquad \boldsymbol{b}_\tau = \boldsymbol{x}_\tau^{[1]} = \begin{bmatrix} y_{12\tau-11} \\ y_{12\tau-10} \\ \vdots \\ y_{12\tau} \end{bmatrix}$$

The coherent subspace

Coherent subspace

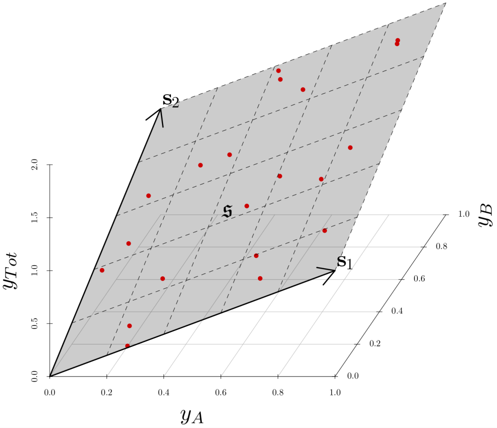
n_b-dimensional linear subspace $\mathfrak{s} \subset \chi^n$ for which linear constraints hold for all $\boldsymbol{y} \in \mathfrak{s}$.

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

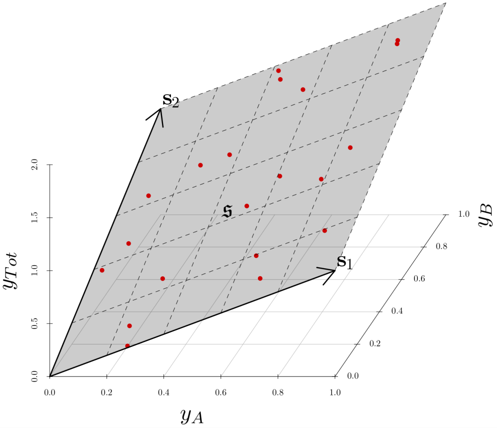
An n-dimensional multivariate time series such that $\boldsymbol{y}_t \in \mathfrak{s} \quad \forall t$.

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

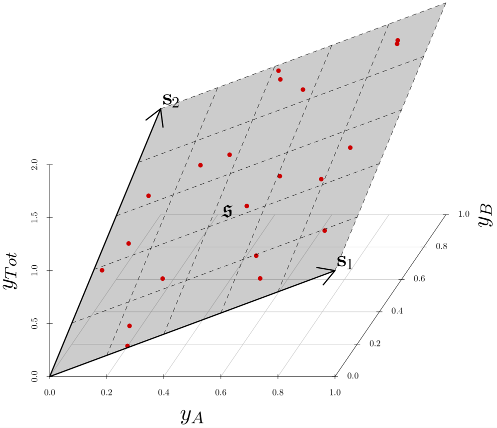
n_b-dimensional linear subspace $\mathfrak{s} \subset \chi^n$ for which linear constraints hold for all $\boldsymbol{y} \in \mathfrak{s}$.

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 $\boldsymbol{M}$ be a projection matrix.
$\tilde{\boldsymbol{y}}_{t+h|t} = \boldsymbol{M}\hat{\boldsymbol{y}}_{t+h|t}$ “reconciles” $\hat{\boldsymbol{y}}_{t+h|t}$.

Minimum trace reconciliation

Wickramasuriya et al (2019)

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

Reconciled forecasts Base forecasts

$$\boldsymbol{W}_h = \text{Var}[\boldsymbol{y}_{T+h} - \hat{\boldsymbol{y}}_{T+h|T} \mid \boldsymbol{y}_1, \dots, \boldsymbol{y}_T]$$

$$\boldsymbol{V}_h = \text{Var}[\boldsymbol{y}_{T+h} - \tilde{\boldsymbol{y}}_{T+h|T} \mid \boldsymbol{y}_1, \dots, \boldsymbol{y}_T] = \boldsymbol{M}\boldsymbol{W}_h\boldsymbol{M}'$$

Minimum trace (MinT) reconciliation

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

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

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

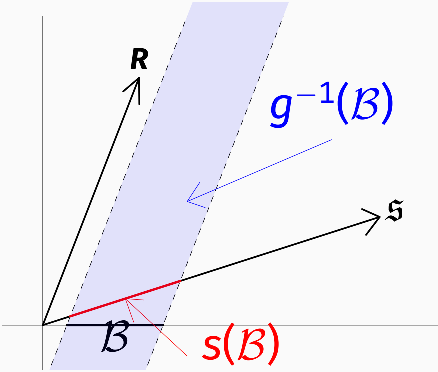
Coherent probabilistic forecasts

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

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

- Random draws from coherent distribution must lie on $\mathfrak{s}$.
- The probability of points not on $\mathfrak{s}$ is zero.
- The reconciled distribution is a transformation of the base forecast distribution that is coherent on $\mathfrak{s}$.



Panagiotelis et al. (2023) and Corani, Azzimonti, and Rubattu (2024)

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.

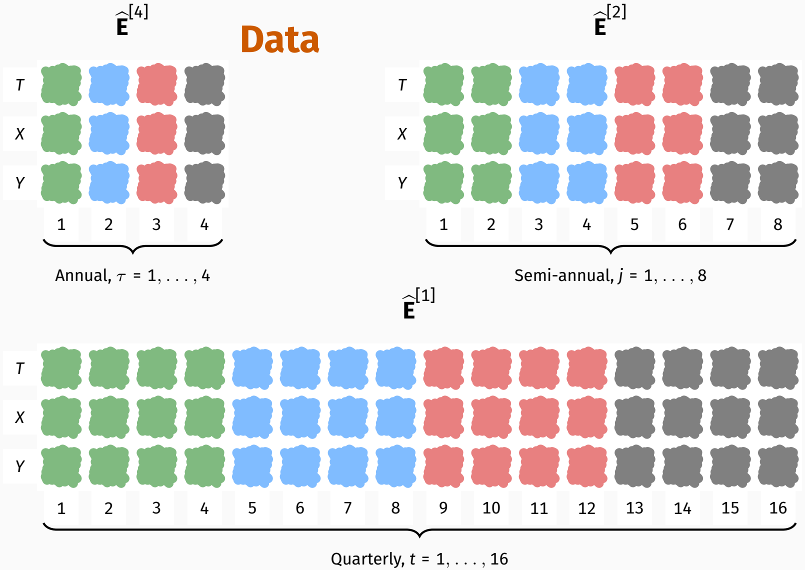
Cross-temporal probabilistic forecast reconciliation

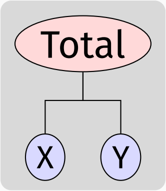
Nonparametric bootstrap

- 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

Data

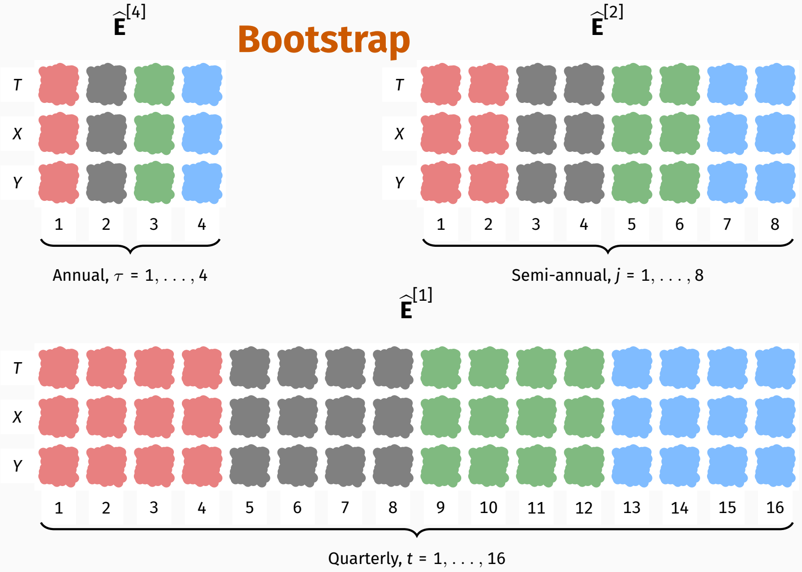


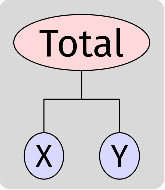


Year 1
Year 2
Year 3
Year 4

Cross-temporal probabilistic forecast reconciliation

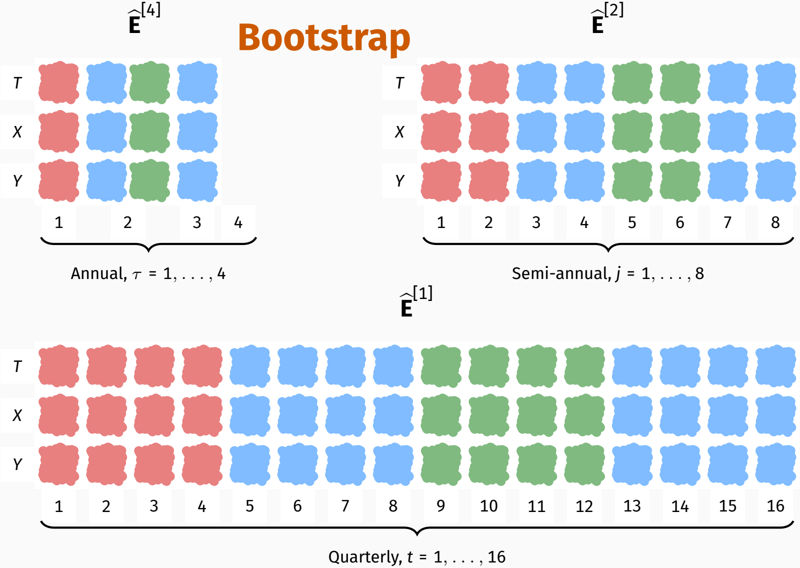
Bootstrap

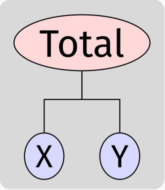




Year 1
Year 2
Year 3
Year 4

Cross-temporal probabilistic forecast reconciliation

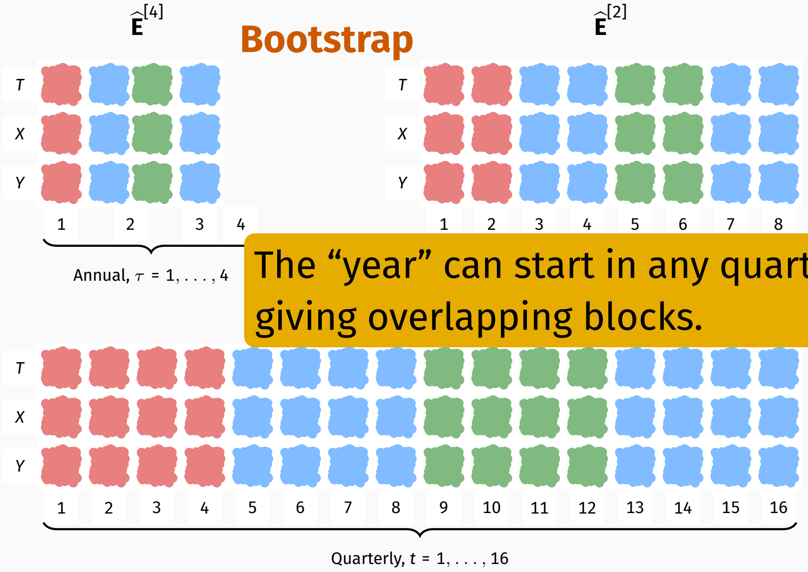




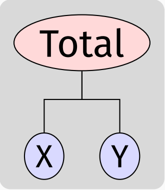
Year 1
Year 2
Year 3
Year 4

Cross-temporal probabilistic forecast reconciliation

Bootstrap



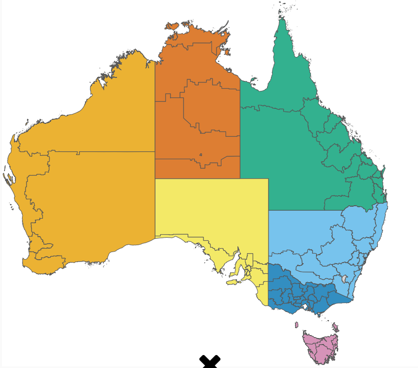
The "year" can start in any quarter, giving overlapping blocks.



Year 1
Year 2
Year 3
Year 4

Monthly Australian Tourism Demand

Geographical division



×

Purpose of travel

Holiday, Visiting friends & relatives, Business, Other

- **Cross-sectional aggregations** (geographical divisions × purpose of travel)

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

- **Temporal aggregations**, 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 data – CRPS skill scores

Reconciliation using different covariance matrix ($\boldsymbol{W}_h$) estimates

	Worse than benchmark $\forall k \in \{12, 6, 4, 3, 2, 1\}$	**Best** $k = 1$
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

Forecast reconciliation 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

More information

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

References

- Athanasopoulos, G, RJ Hyndman, N Kourentzes, and A Panagiotelis (2024). "Forecast reconciliation: a review". forthcoming. http://robjhyndman.com/publications/hfreview.html.
- 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 (2024). Probabilistic reconciliation of count time series. *International Journal of Forecasting*. forthcoming.
- Di Fonzo, T and D Girolimetto (2023). Cross-temporal forecast reconciliation: Optimal combination method and heuristic alternatives. *International Journal of Forecasting* **39**(1), 39–57.
- Girolimetto, D, G Athanasopoulos, T Di Fonzo, and RJ Hyndman (2024). Cross-temporal probabilistic forecast reconciliation. *International J Forecasting*. forthcoming.

References

- Hyndman, RJ, RA Ahmed, G Athanasopoulos, and HL Shang (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis* **55**(9), 2579–2589.
- Hyndman, RJ, A Lee, and E Wang (2016). Fast computation of reconciled forecasts for hierarchical and grouped time series. *Computational Statistics & Data Analysis* **97**, 16–32.
- 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.
- 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.