

Forecast reconciliation

A brief overview

Rob J Hyndman



MONASH University

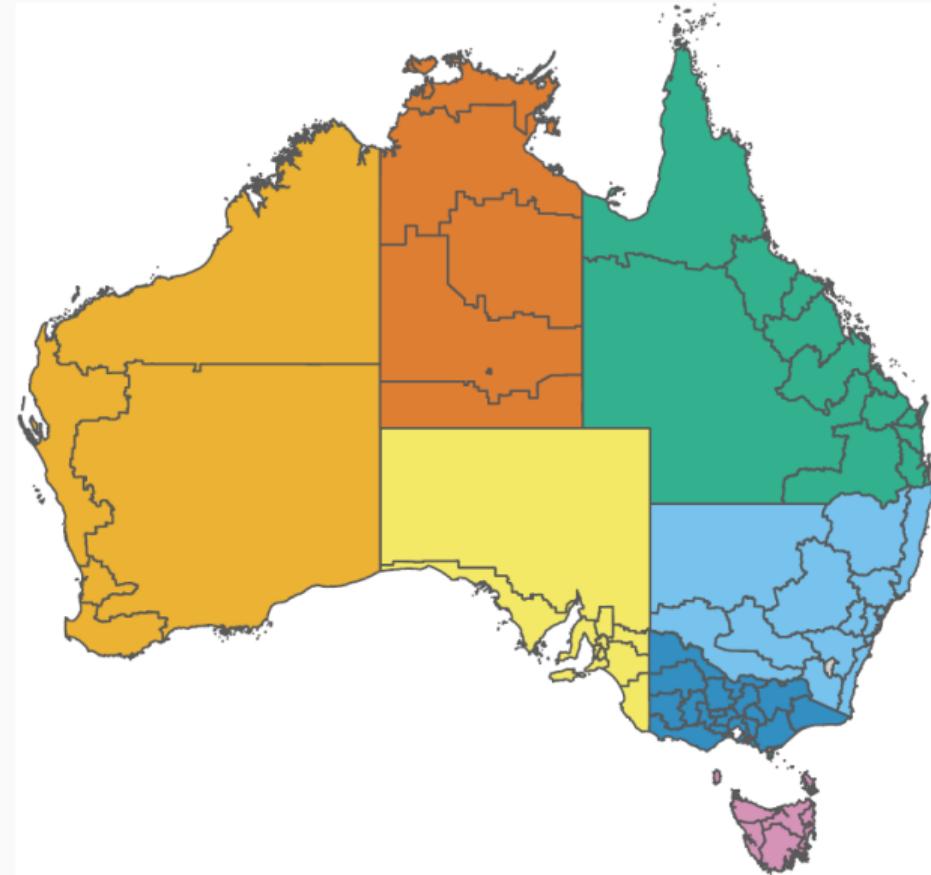
Outline

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

Outline

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

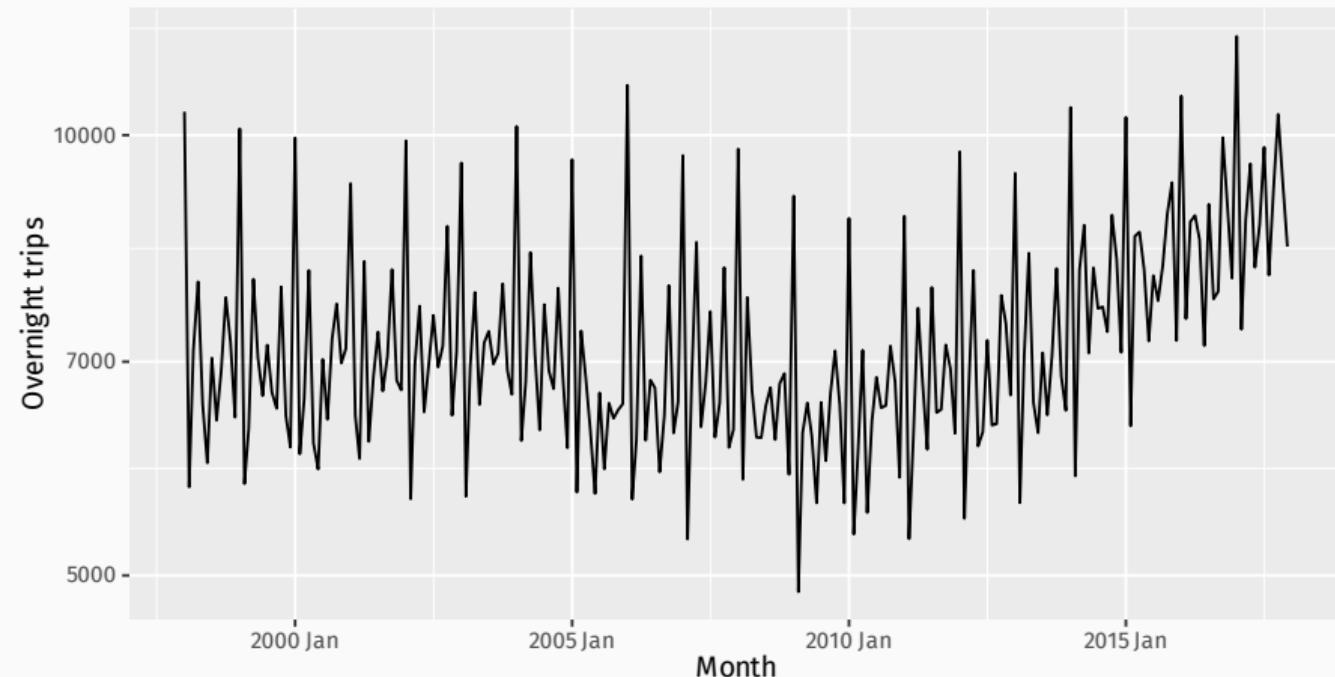
Australian tourism regions



- 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

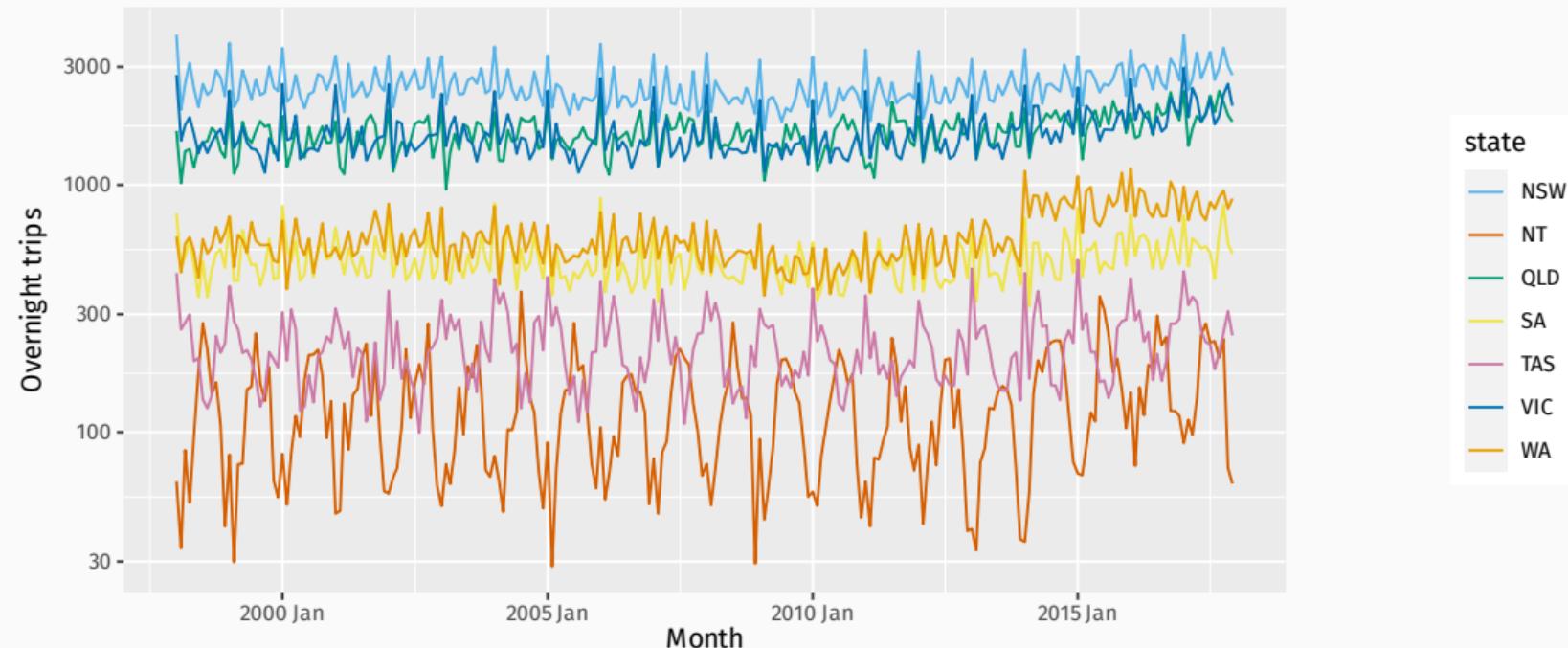
Australian tourism data

Total domestic travel: Australia



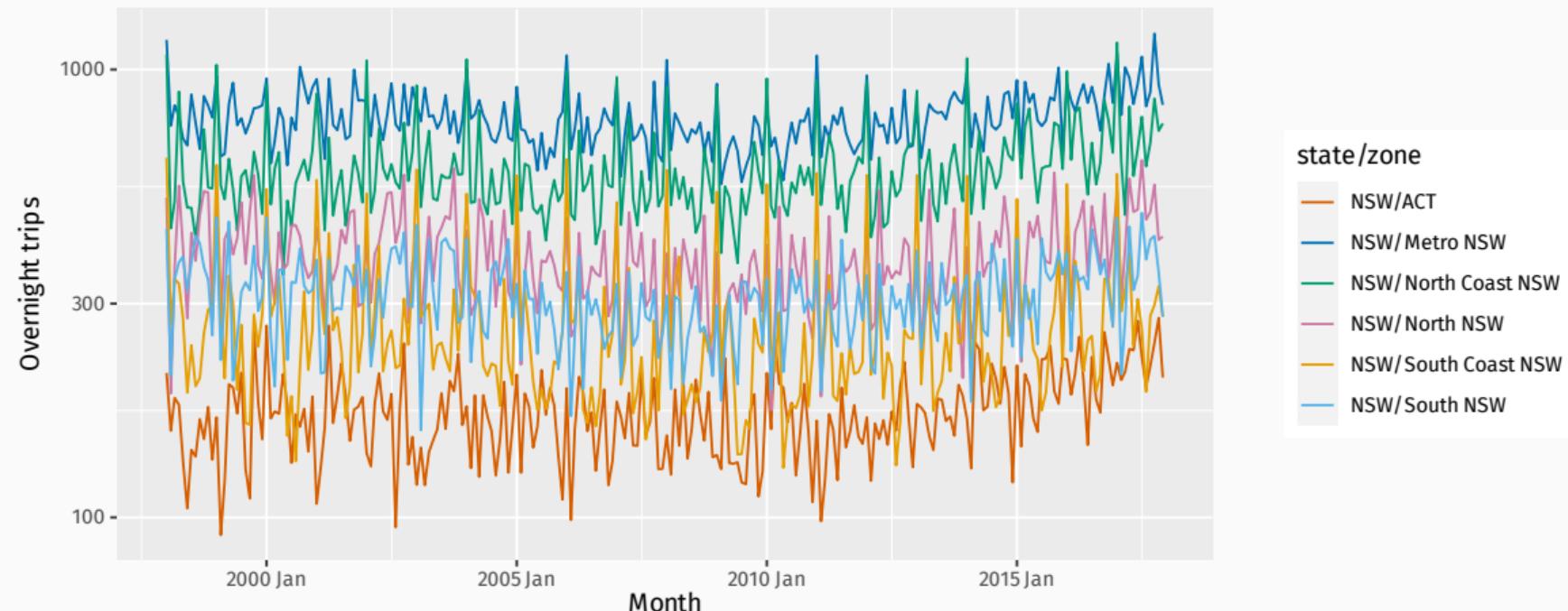
Australian tourism data

Total domestic travel: by state



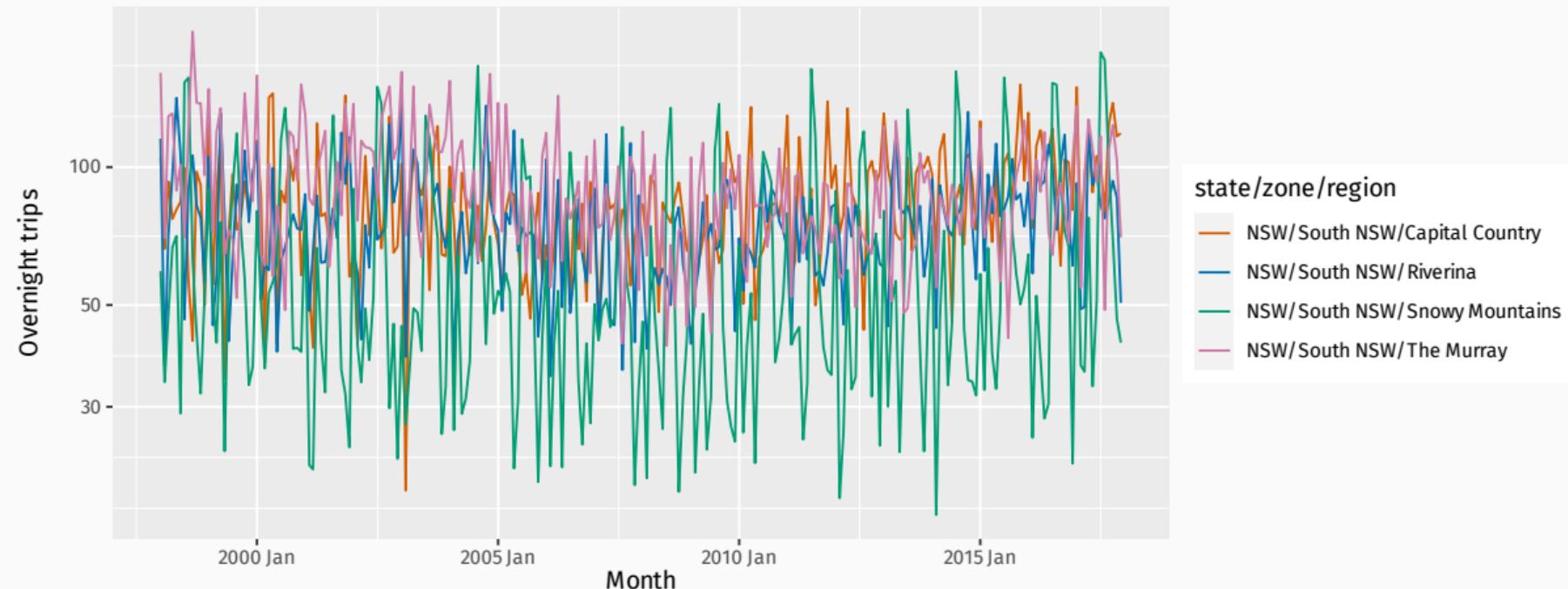
Australian tourism data

Total domestic travel: NSW by zone



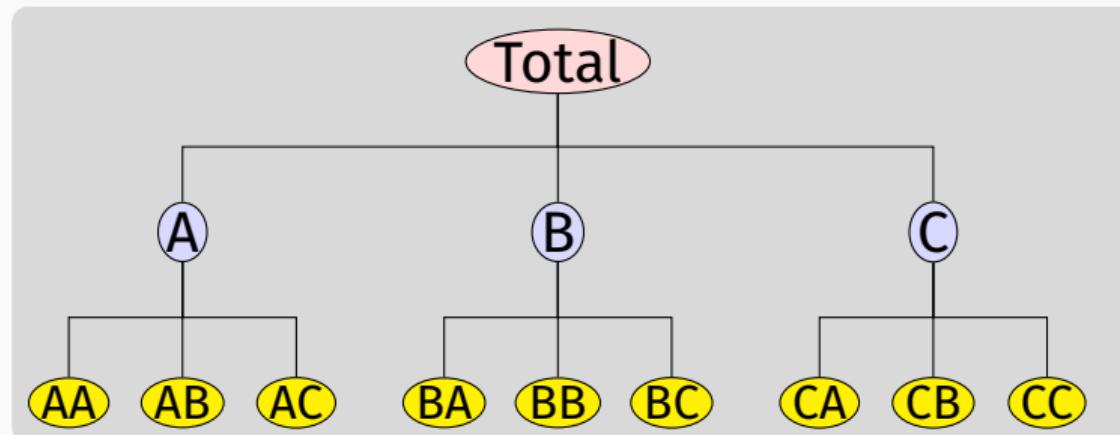
Australian tourism data

Total domestic travel: South NSW by region



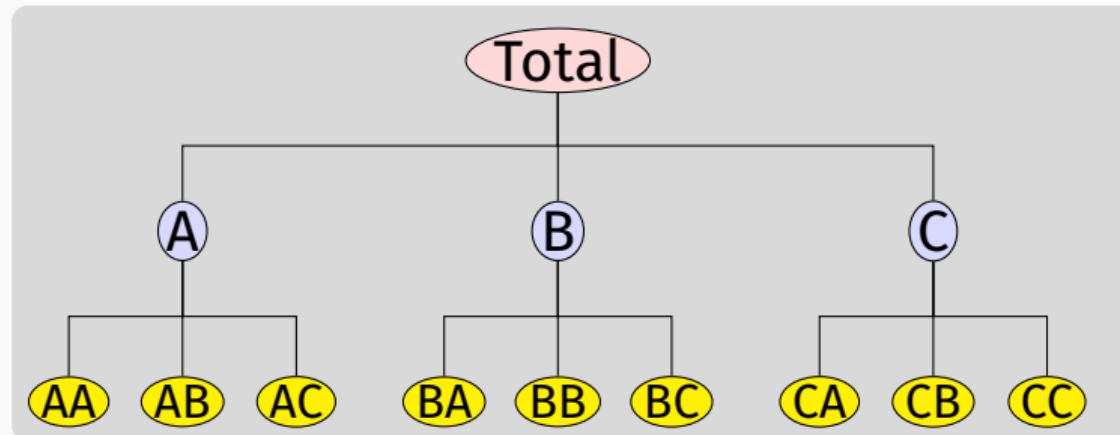
Hierarchical time series

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.

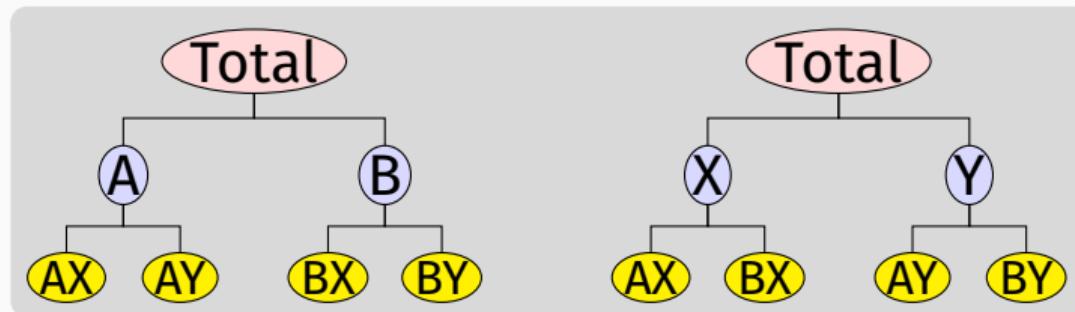


Examples

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

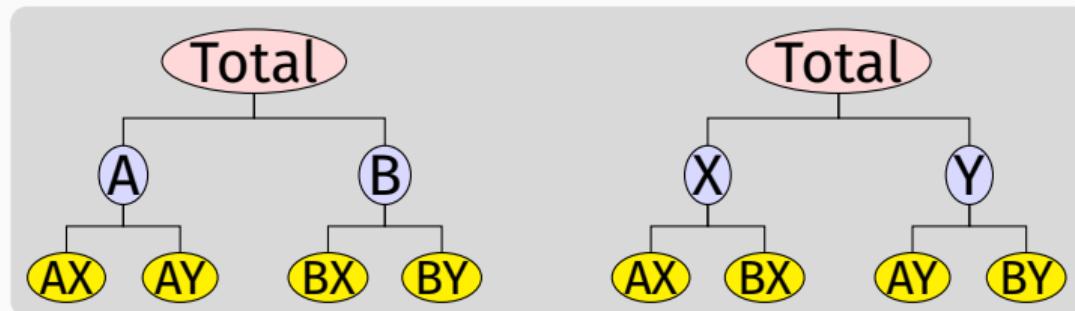
Grouped time series

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

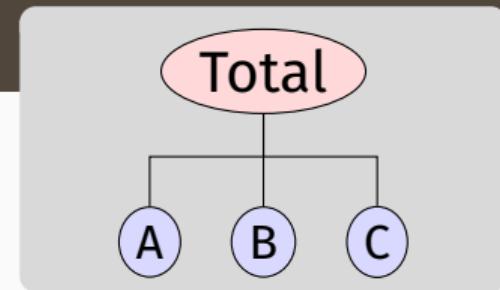
- Tourism by state and purpose of travel
- Retail sales by product groups/sub groups, and by countries/regions

Hierarchical and grouped time series

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

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

- \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} = “structural matrix” containing the linear constraints.



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

The forecast reconciliation problem

- We want “coherent” forecasts of all series – i.e., the forecasts should obey the same constraints as the data.
- We model and forecast all series independently. The resulting forecasts will almost certainly be incoherent.
- We “reconcile” the forecasts to make them coherent.

Notation

- \mathbf{y}_t = vector of data for all series at time t .
- $\hat{\mathbf{y}}_{t+h|t}$ = vector of incoherent “base” forecasts for time $t + h$ using data to time t .
- $\tilde{\mathbf{y}}_{t+h|t}$ = vector of coherent “reconciled” forecasts for time $t + h$ using data to time t .

Outline

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

The coherent subspace

Coherent subspace

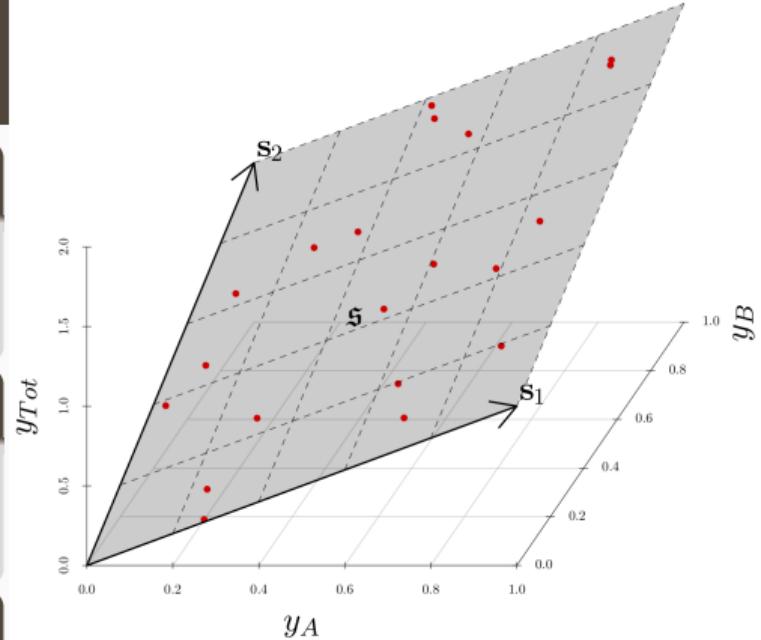
n_b -dimensional linear subspace $\mathfrak{s} \subset \chi^n$ for which linear constraints hold for all $\mathbf{y} \in \mathfrak{s}$.

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

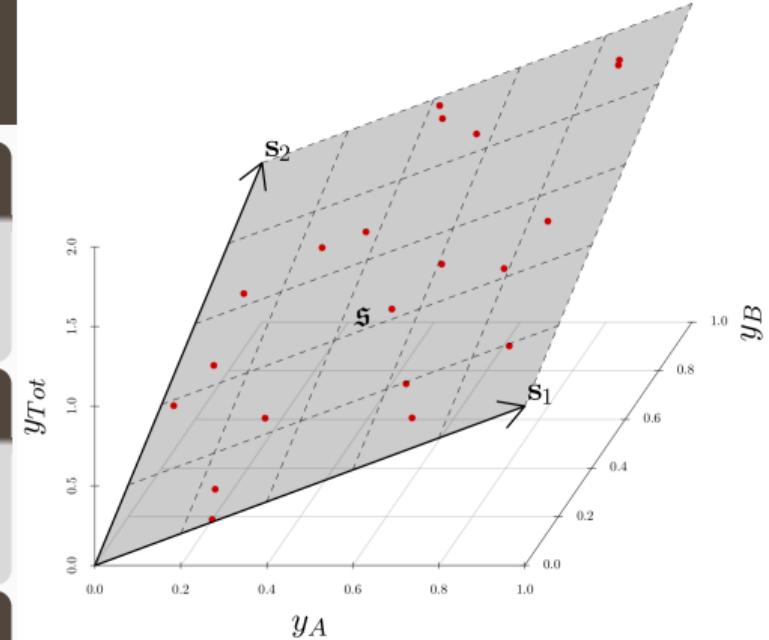
n_b -dimensional linear subspace $\mathfrak{s} \subset \chi^n$ for which linear constraints hold for all $\mathbf{y} \in \mathfrak{s}$.

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



$$\mathbf{y}_{Tot} = \mathbf{y}_A + \mathbf{y}_B$$

Base forecasts

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

The coherent subspace

Coherent subspace

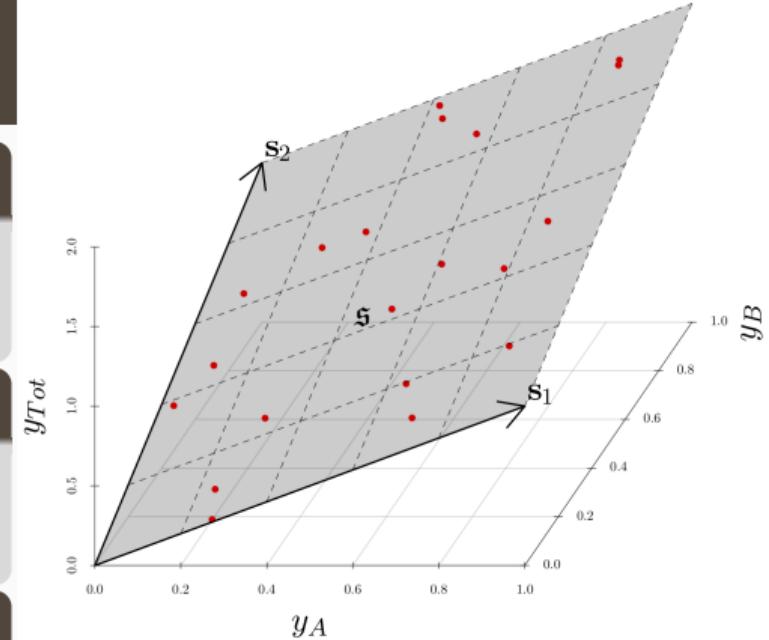
n_b -dimensional linear subspace $\mathfrak{s} \subset \chi^n$ for which linear constraints hold for all $\mathbf{y} \in \mathfrak{s}$.

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” $\hat{\mathbf{y}}_{t+h|t}$.

Linear projection reconciliation

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

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

Minimum trace (MinT) reconciliation

If \mathbf{M} is a projection, then trace of \mathbf{V}_h is minimized when $\Psi = \mathbf{W}_h$, so that

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

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

Reconciled forecasts

Base forecasts

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

Reconciliation method \mathbf{G}

OLS $(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'$

WLS(var) $(\mathbf{S}'\Lambda_v\mathbf{S})^{-1}\mathbf{S}'\Lambda_v$

WLS(struct) $(\mathbf{S}'\Lambda_s\mathbf{S})^{-1}\mathbf{S}'\Lambda_s$

MinT(sample) $(\mathbf{S}'\hat{\mathbf{W}}_{\text{sam}}^{-1}\mathbf{S})^{-1}\mathbf{S}'\hat{\mathbf{W}}_{\text{sam}}^{-1}$

MinT(shrink) $(\mathbf{S}'\hat{\mathbf{W}}_{\text{shr}}^{-1}\mathbf{S})^{-1}\mathbf{S}'\hat{\mathbf{W}}_{\text{shr}}^{-1}$

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

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

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))
```

```
# 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

```
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 |>  
  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.

Outline

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

The coherent subspace

Coherent subspace

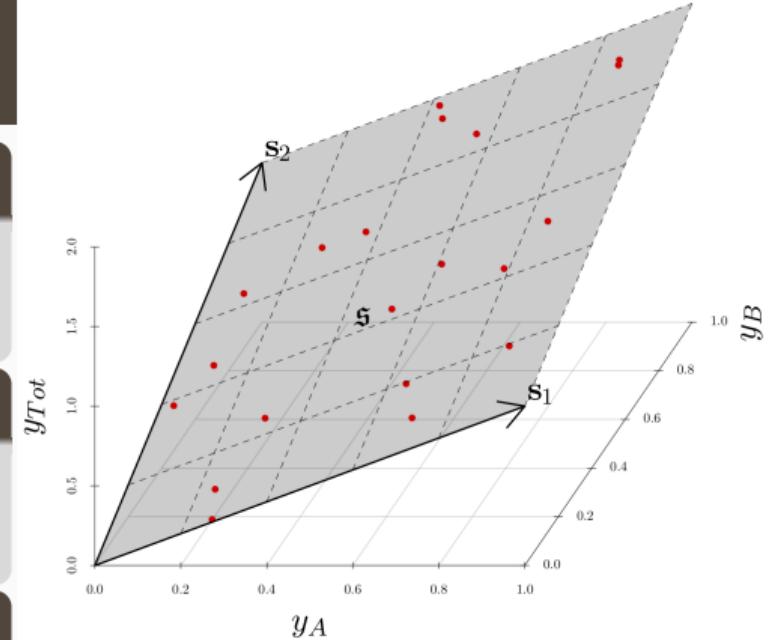
m -dimensional linear subspace $\mathfrak{s} \subset \chi^n$ for which linear constraints hold for all $\mathbf{y} \in \mathfrak{s}$.

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” $\hat{\mathbf{y}}_{t+h|t}$.

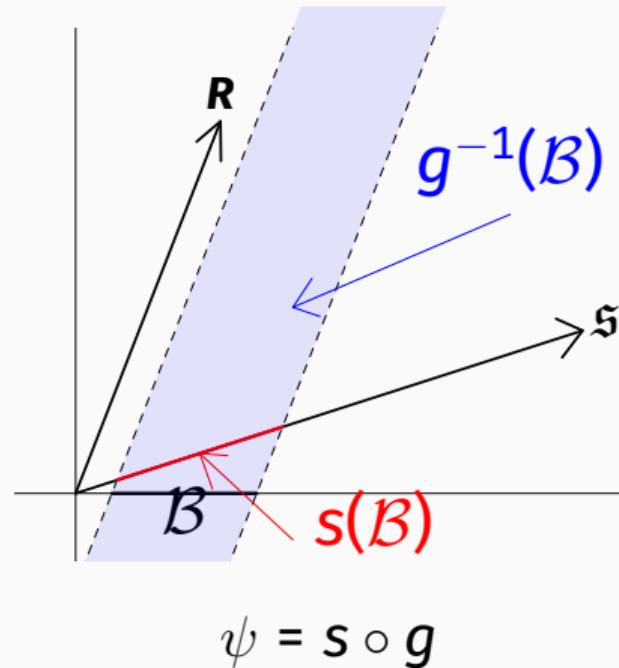
Coherent probabilistic forecasts

Coherent probabilistic forecasts

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

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

- 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} .

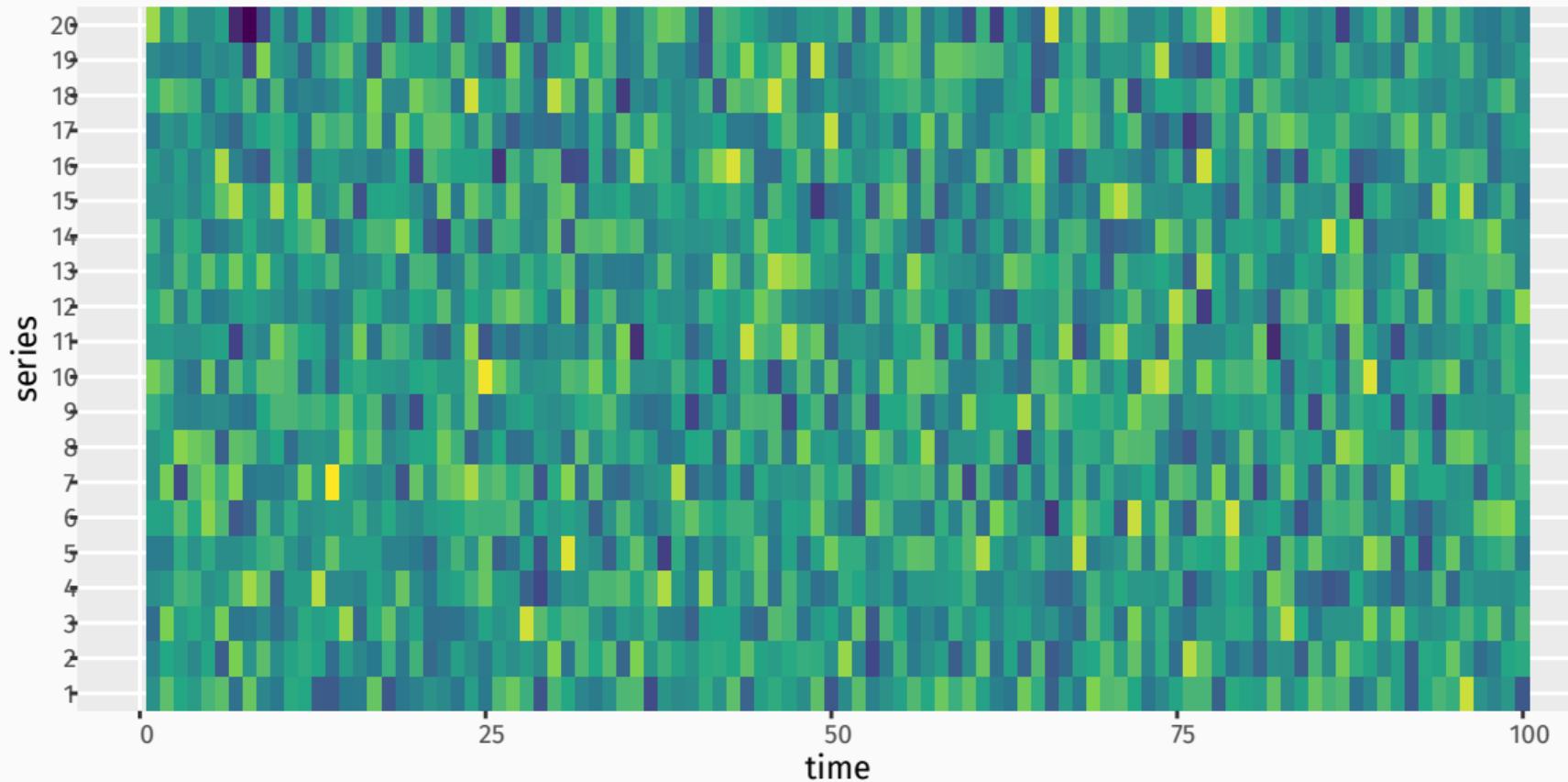


Simulation from a reconciled distribution

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

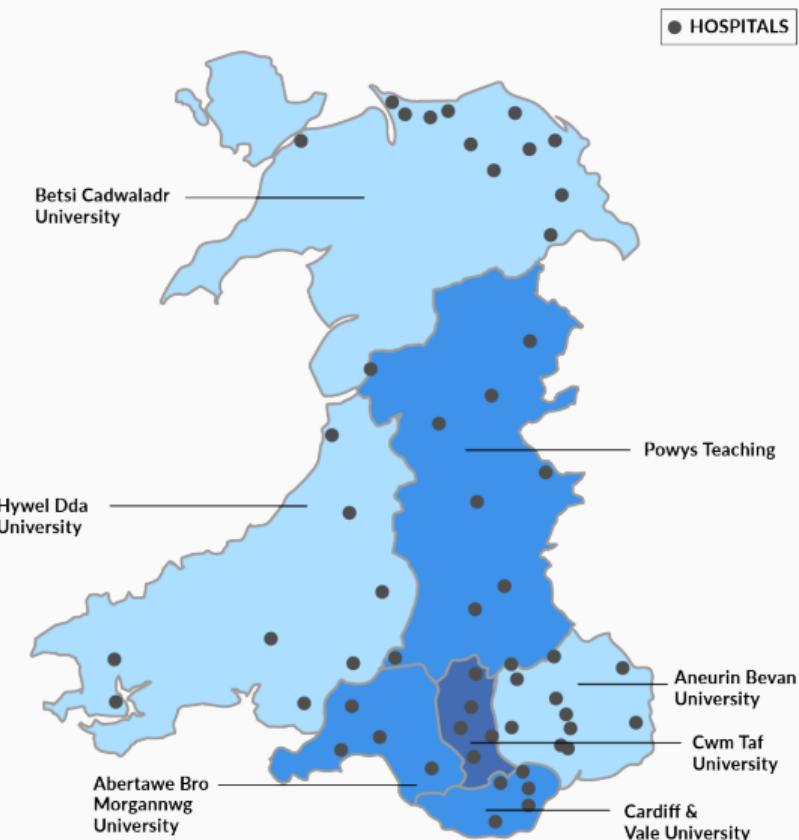
- Simulate future sample paths for each series, by simulating from each model using a multivariate bootstrap of the residuals (to preserve cross-correlations).
- Reconcile the sample paths.
- The reconciled sample paths are a sample from the reconciled distribution.

Simulation from a reconciled distribution



Simulation from a reconciled distribution

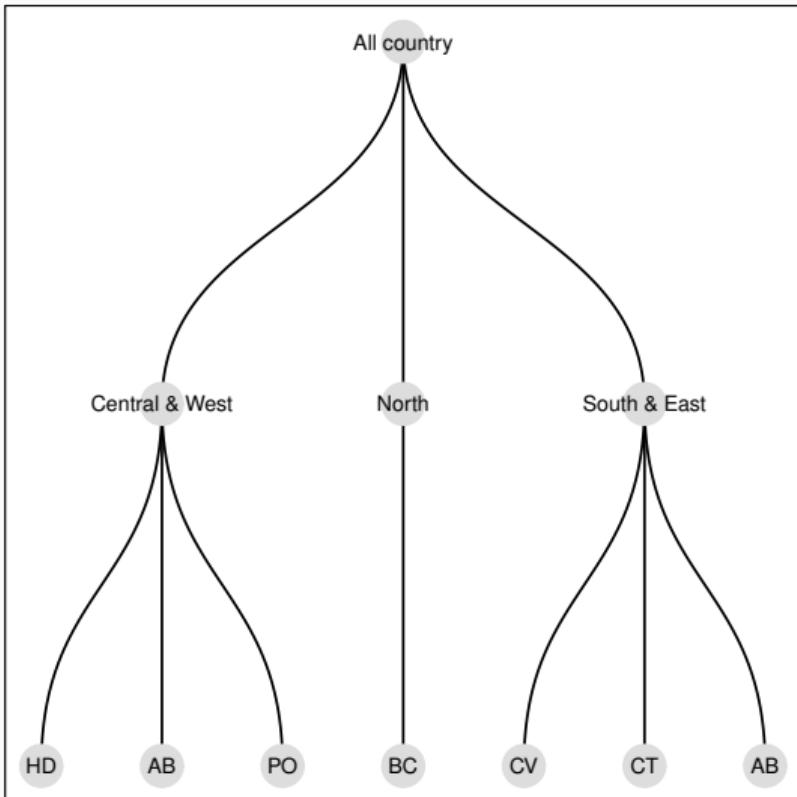
Wales Health Board Areas



Data

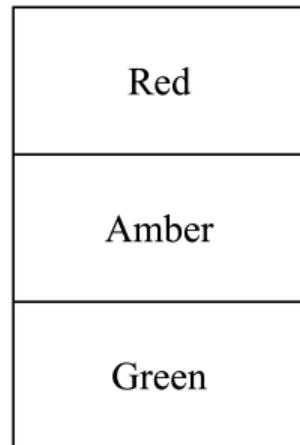
- Daily number of attended incidents:
1 October 2015 – 31 July 2019
- Disaggregated by:
 - ▶ control area (x3)
 - ▶ health board (x7)
 - ▶ priority (x3)
 - ▶ nature of incidents (x35)
- 2,142,000 rows observations from 1,530 time series.

Data structure



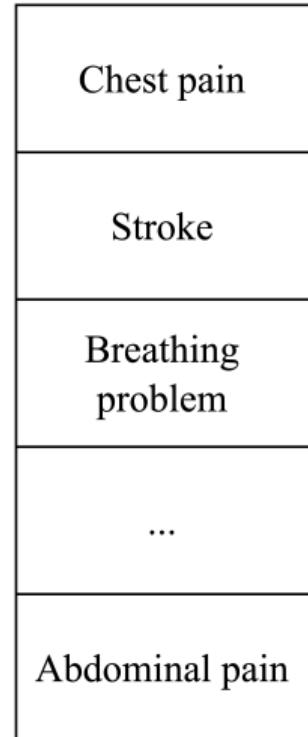
*

Priority



*

Nature of incident



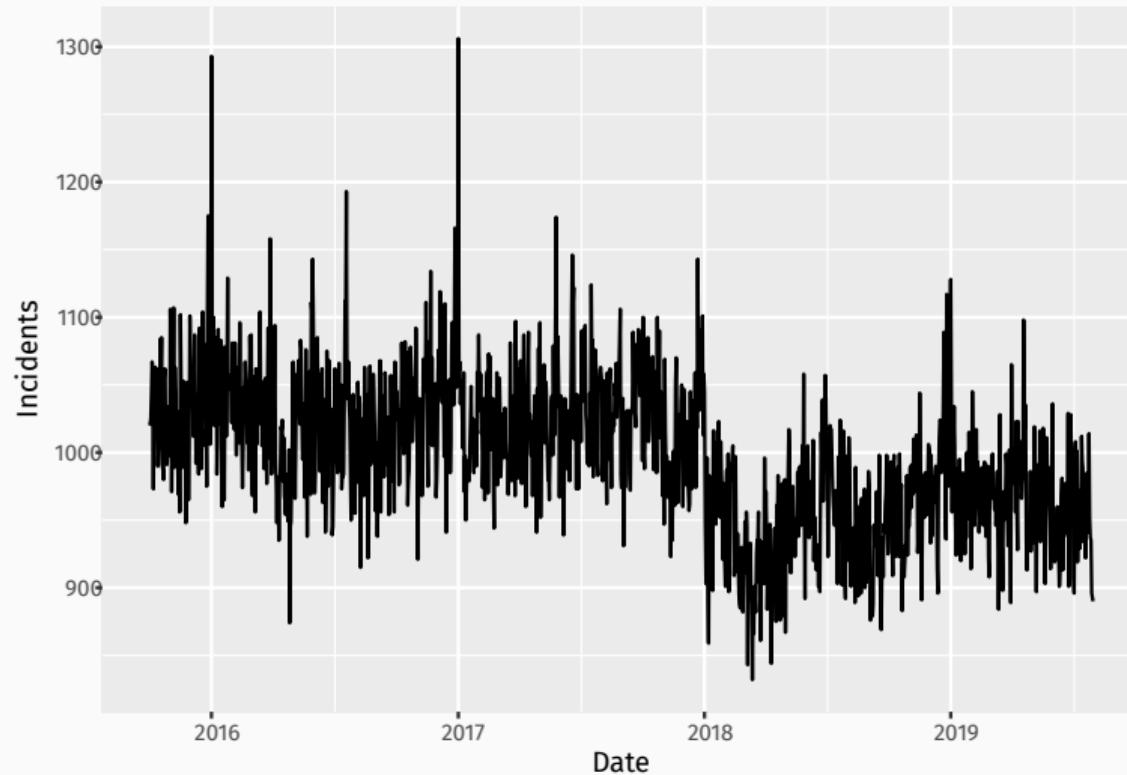
Data

```
# 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
```

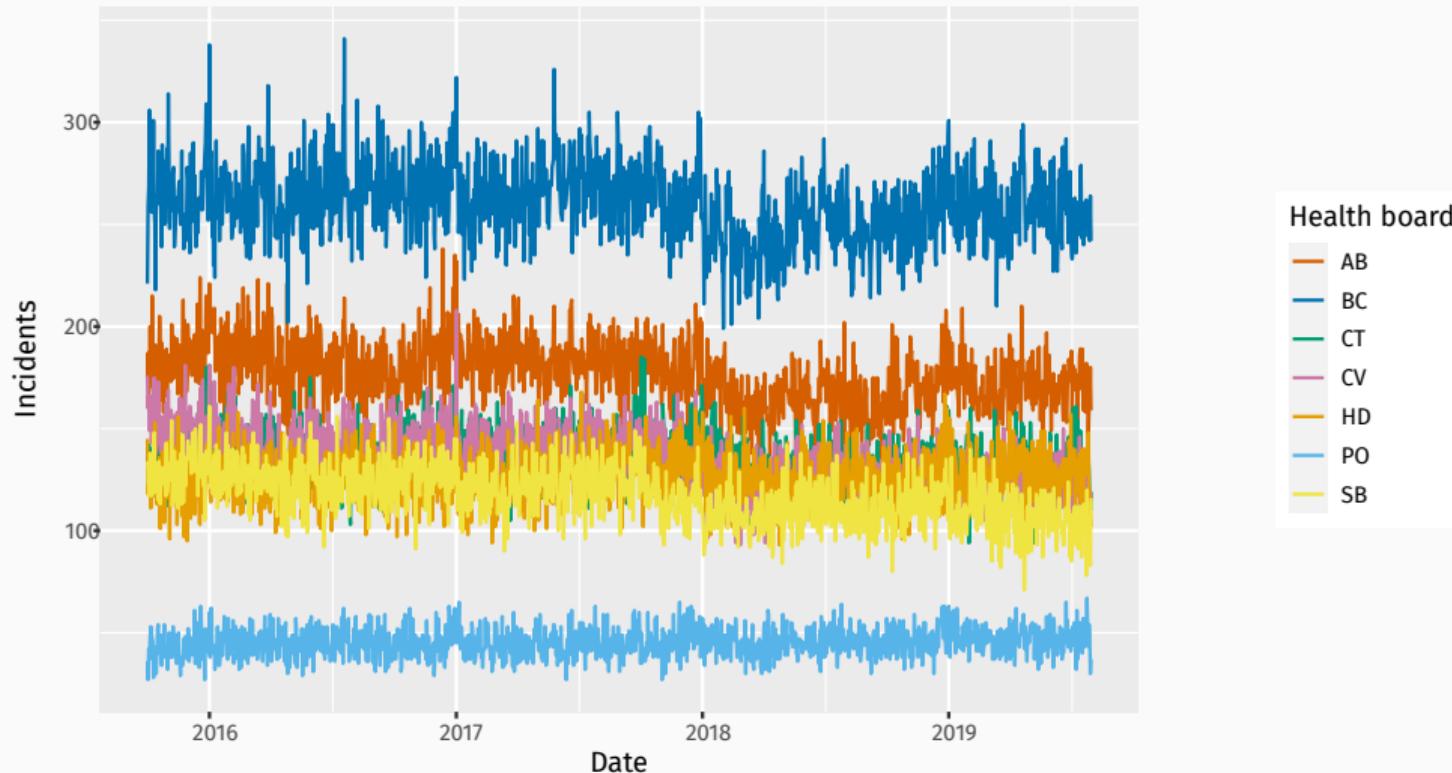
Aggregated daily incidents



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

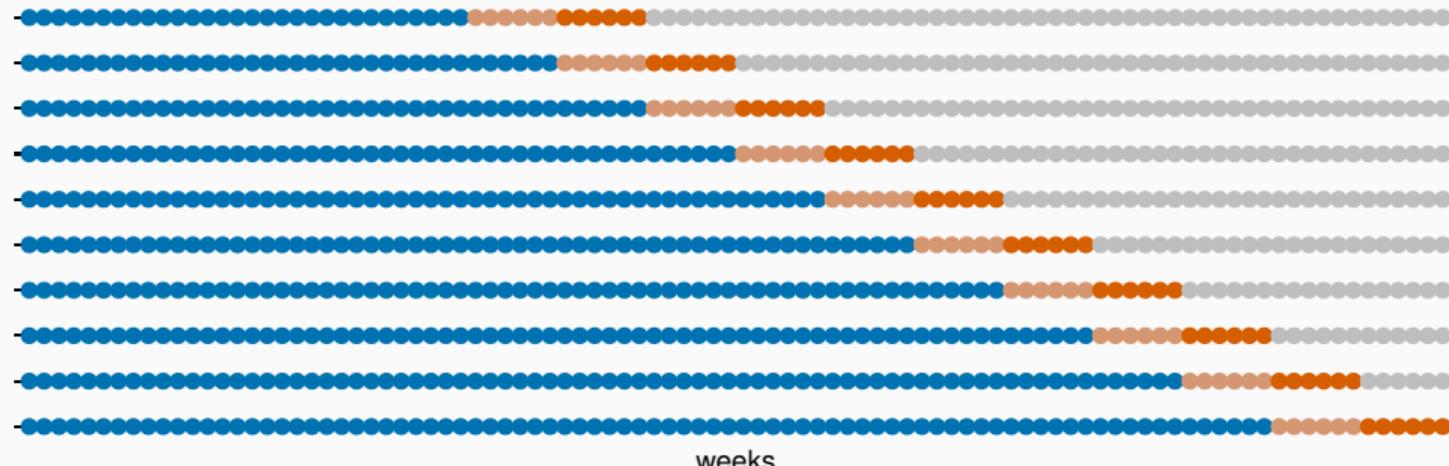


Forecasting methods

- 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: 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

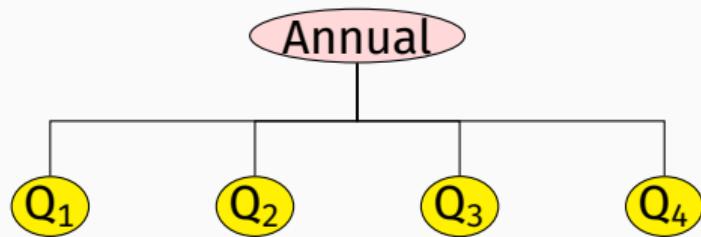
Forecast accuracy

- 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.

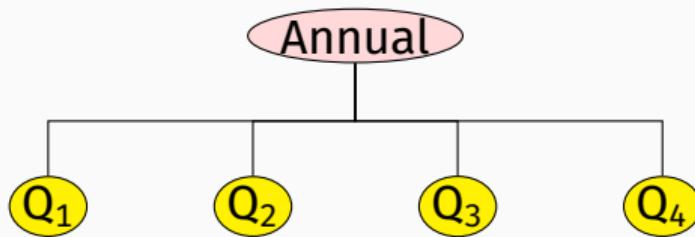
Outline

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

Temporal reconciliation: quarterly data

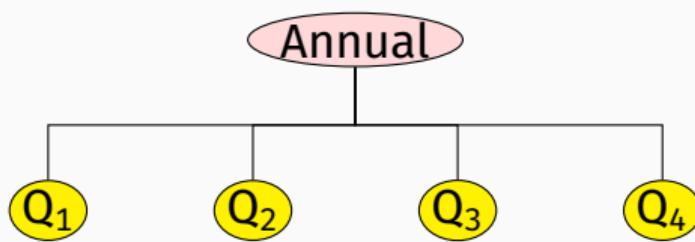


Temporal reconciliation: quarterly data



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

Temporal reconciliation: quarterly data

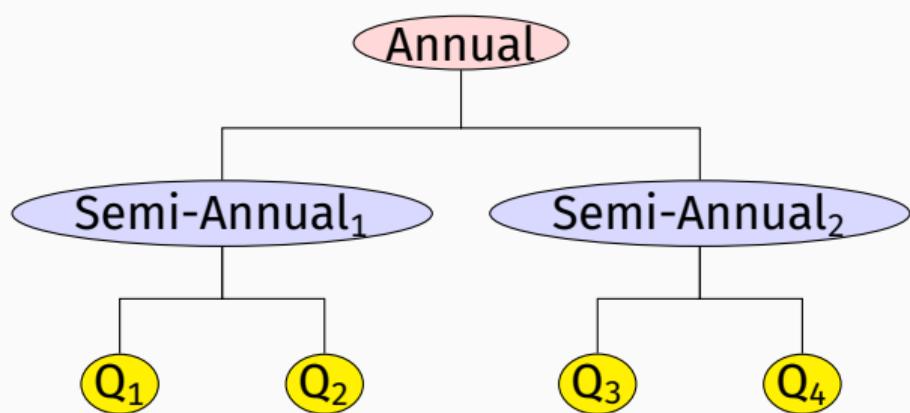


$$\mathbf{y}_\tau = \begin{bmatrix} x_\tau^{[4]} \\ x_{\tau,1}^{[1]} \\ x_{\tau,2}^{[1]} \\ x_{\tau,3}^{[1]} \\ x_{\tau,4}^{[1]} \end{bmatrix} \quad \mathbf{S} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

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

τ = index of largest temporal aggregation level.

Temporal reconciliation: quarterly data

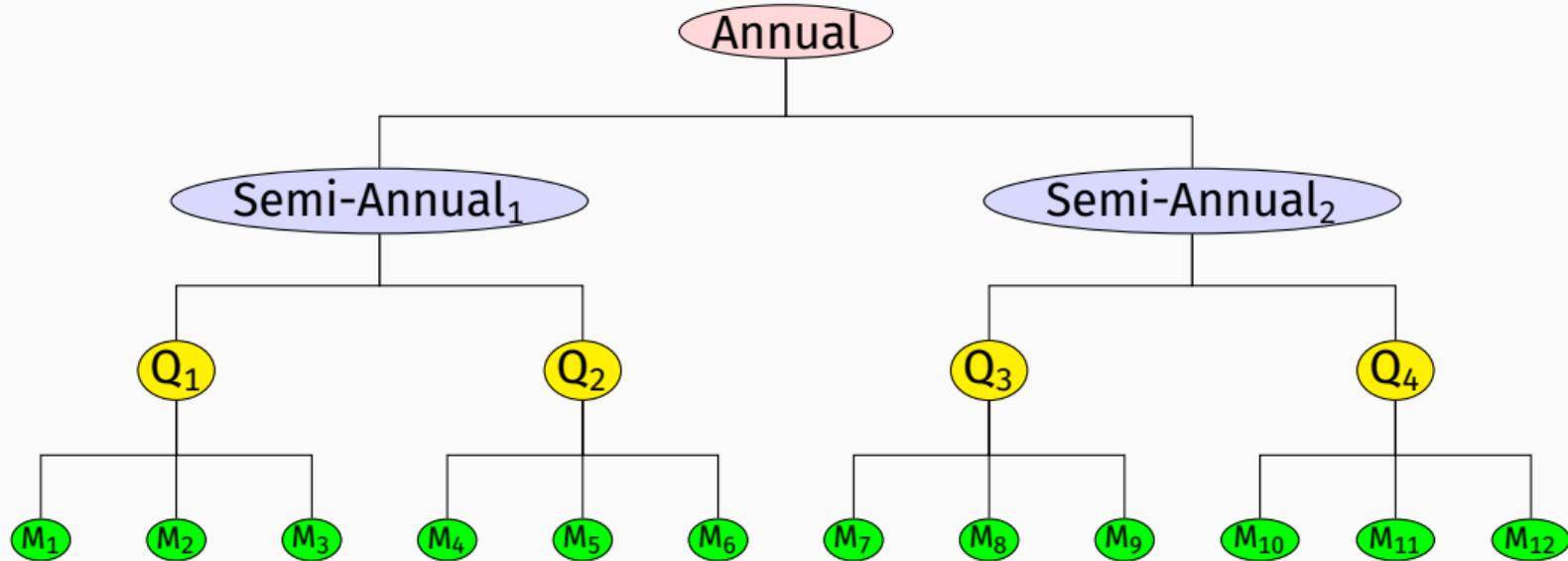


- Forecast series at each available frequency.
- Optimally combine forecasts within the same year.

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

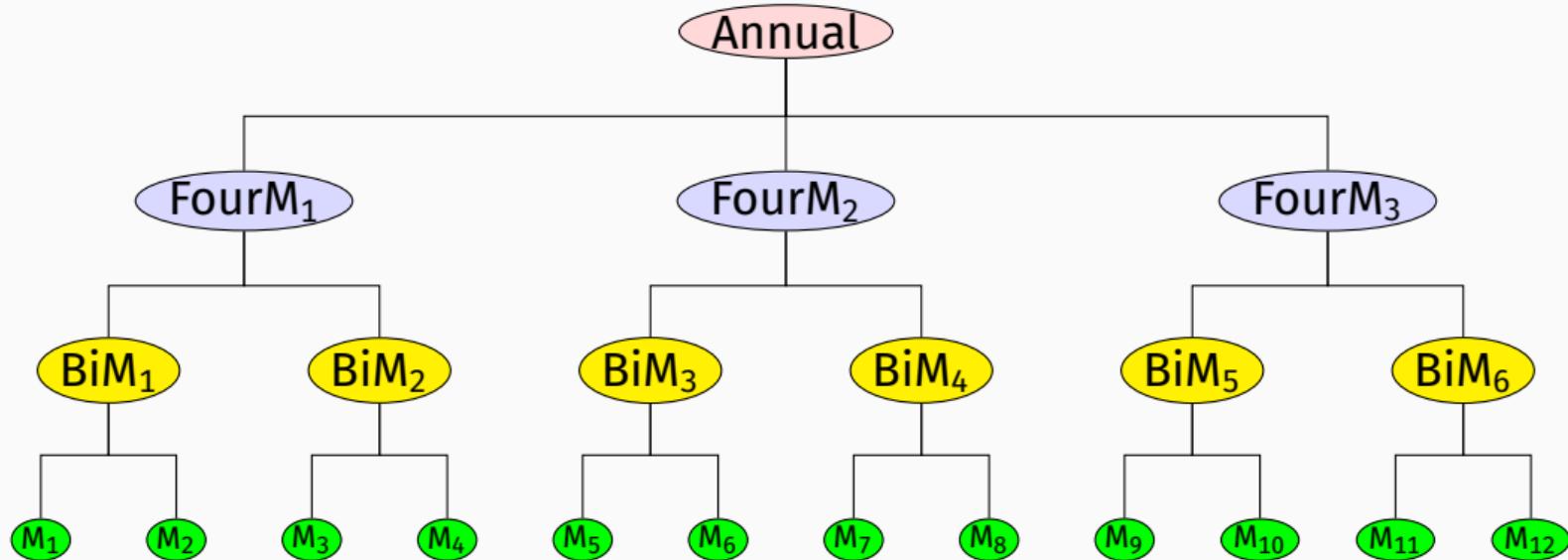
τ = index of largest temporal aggregation level.

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

$$\mathbf{y}_\tau = \begin{bmatrix} \mathbf{x}_\tau^{[12]} \\ \mathbf{x}_\tau^{[6]} \\ \mathbf{x}_\tau^{[4]} \\ \mathbf{x}_\tau^{[3]} \\ \mathbf{x}_\tau^{[2]} \\ \mathbf{x}_\tau^{[1]} \end{bmatrix}$$

$$\mathbf{S} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & & & & \vdots & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & & & & \vdots & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \mathbf{I}_{12}$$

Temporal reconciliation

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

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

- $k \in \mathcal{K} = \{k_1, \dots, k_p\}$ 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

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

where

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

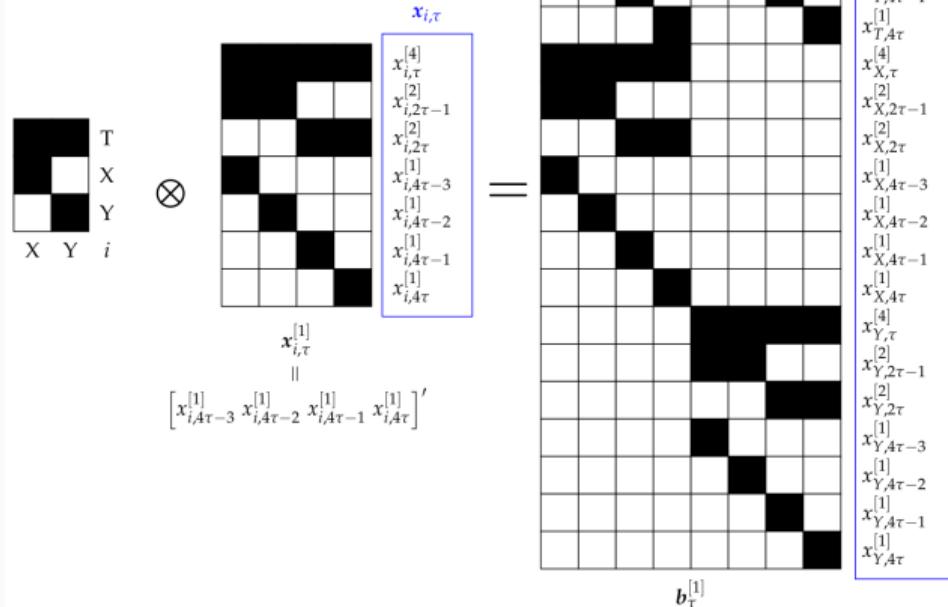
τ is time index for most aggregated series,

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

Cross-temporal forecast reconciliation

- \mathbf{S}_{cs} = structural cross-sectional matrix
- \mathbf{S}_{te} = structural temporal matrix
- $\mathbf{S}_{ct} = \mathbf{S}_{cs} \otimes \mathbf{S}_{te}$
- $\mathbf{x}_\tau = \mathbf{S}_{ct} \mathbf{b}_\tau$

$$\mathbf{b}_\tau = \begin{bmatrix} \mathbf{x}_{1,\tau}^{[1]} \\ \vdots \\ \mathbf{x}_{n,\tau}^{[1]} \end{bmatrix}.$$

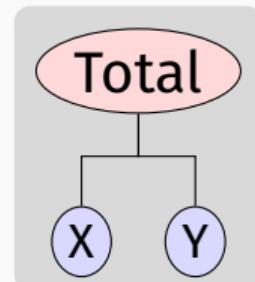
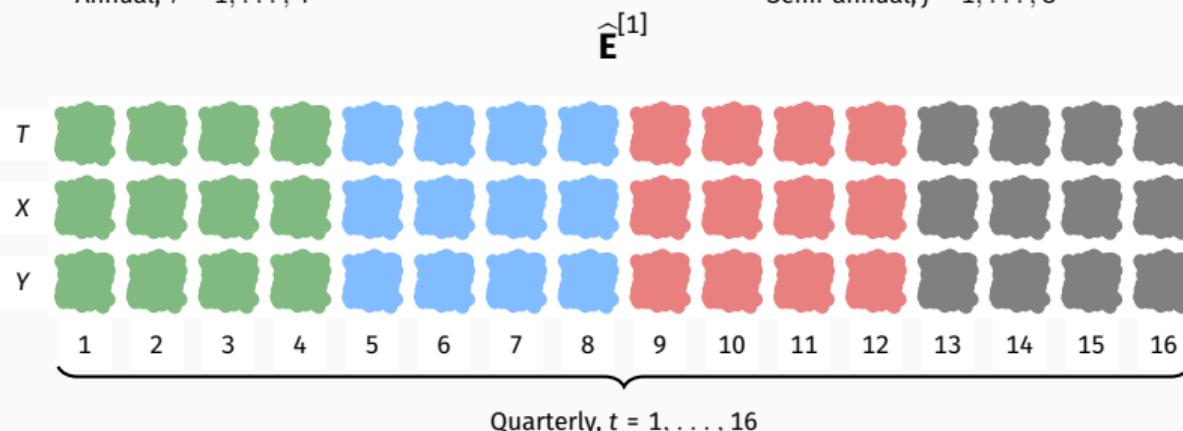
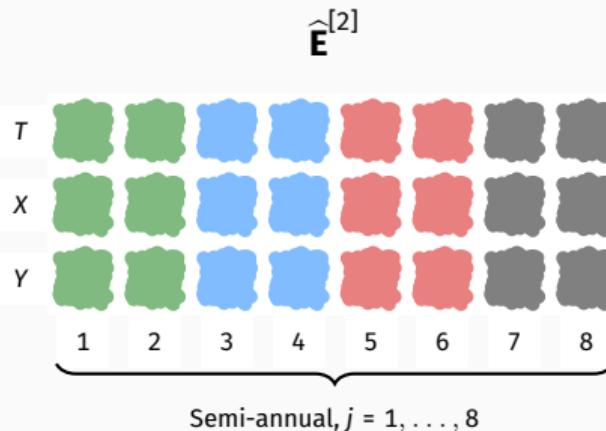
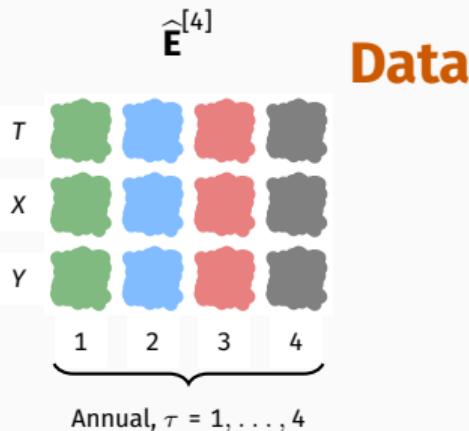


Cross-temporal probabilistic forecast reconciliation

Nonparametric bootstrap

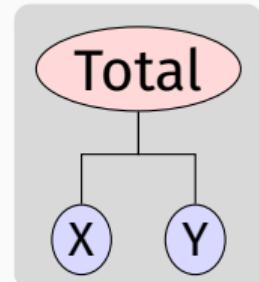
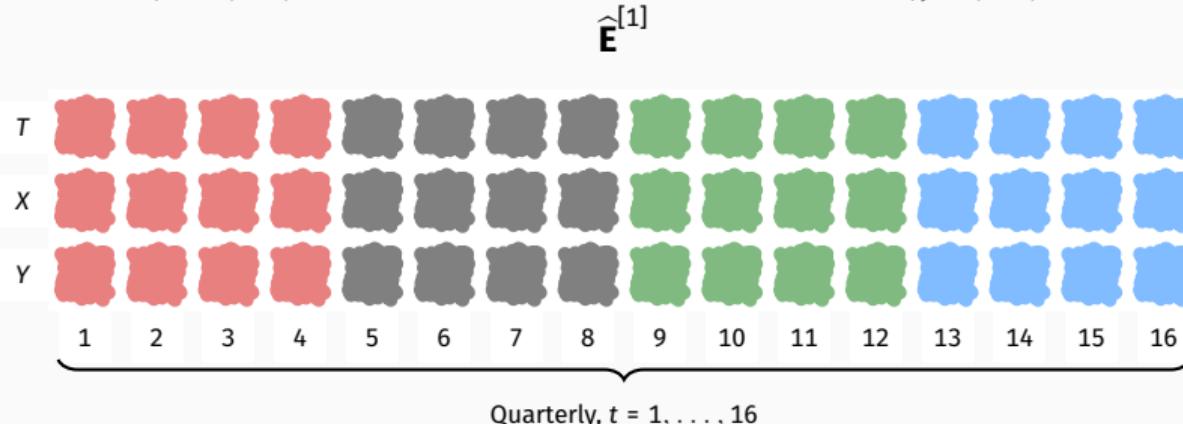
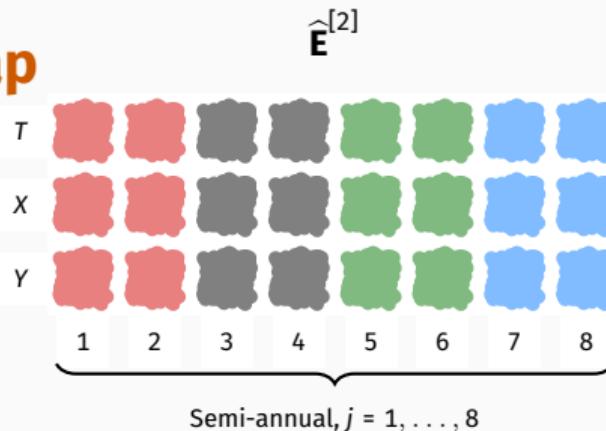
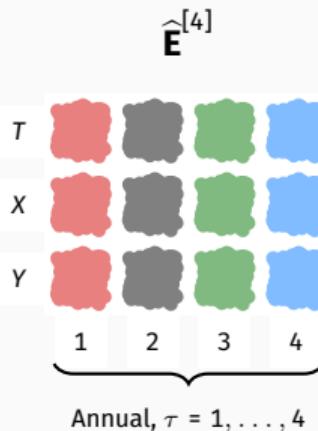
- Simulate future sample paths from all models using bootstrapped residuals, then reconcile them to obtain coherent sample paths.
- Need to generate samples that preserve cross-temporal relationships.
- Draw residual samples of all series at same time from most temporally aggregated level.
- Residuals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

Cross-temporal probabilistic forecast reconciliation



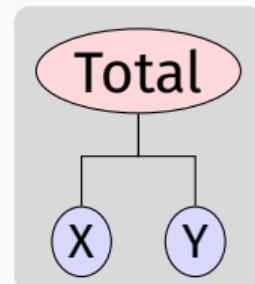
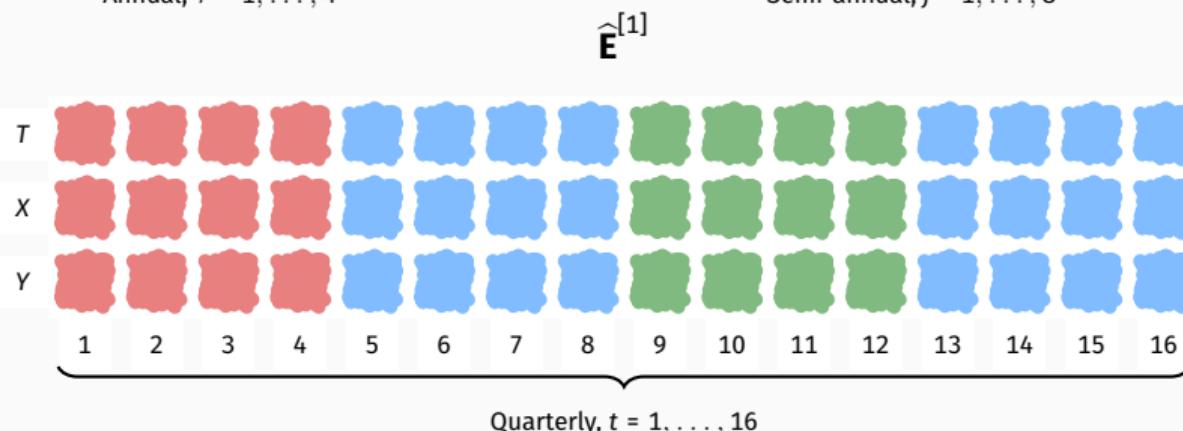
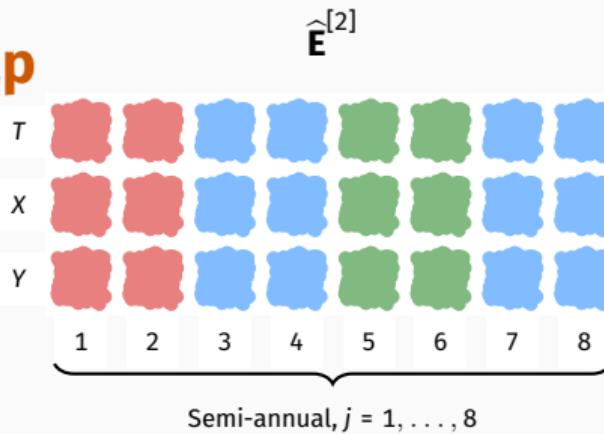
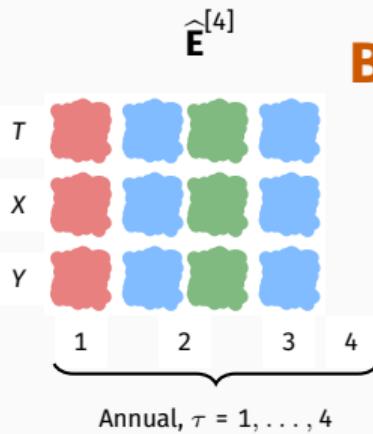
Year 1
Year 2
Year 3
Year 4

Cross-temporal probabilistic forecast reconciliation



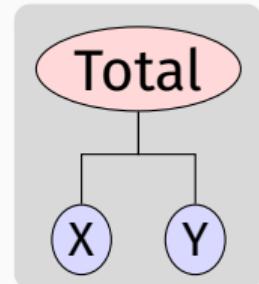
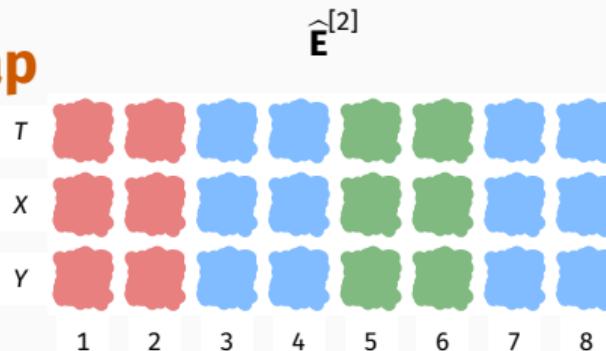
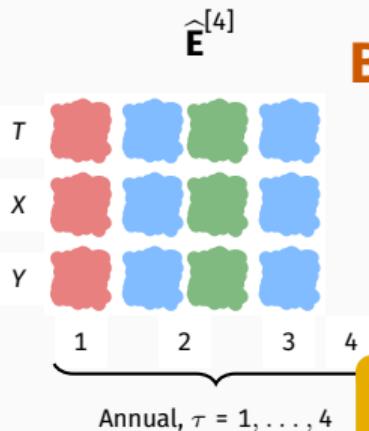
Year 1
Year 2
Year 3
Year 4

Cross-temporal probabilistic forecast reconciliation

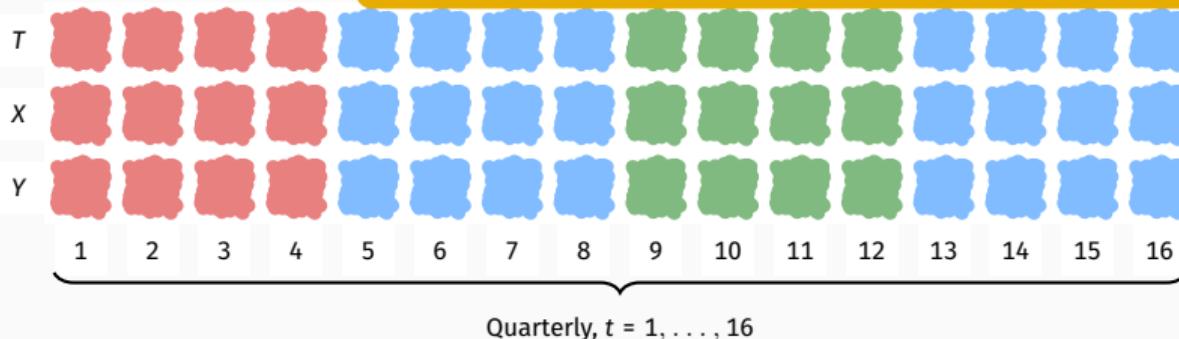


Year 1
Year 2
Year 3
Year 4

Cross-temporal probabilistic forecast reconciliation



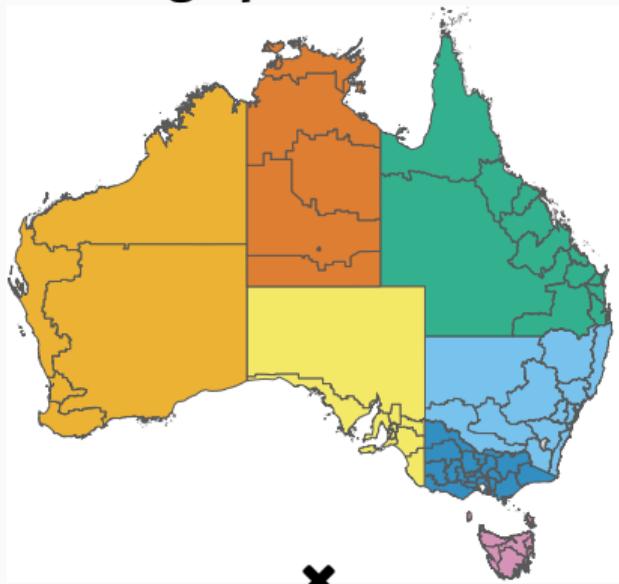
The “year” can start in any quarter, giving overlapping blocks.



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
total	5	35	105	380	525

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

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

Outline

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

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

 [@robjhyndman](https://aus.social/@robjhyndman)

 [@robjhyndman](https://github.com/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.