MONASH University

MONASH BUSINESS SCHOOL

Forecasting prison numbers: a grouped time series approach

George Athanasopoulos

with Tom Steel & Don Weatherburn

Outline

Aim

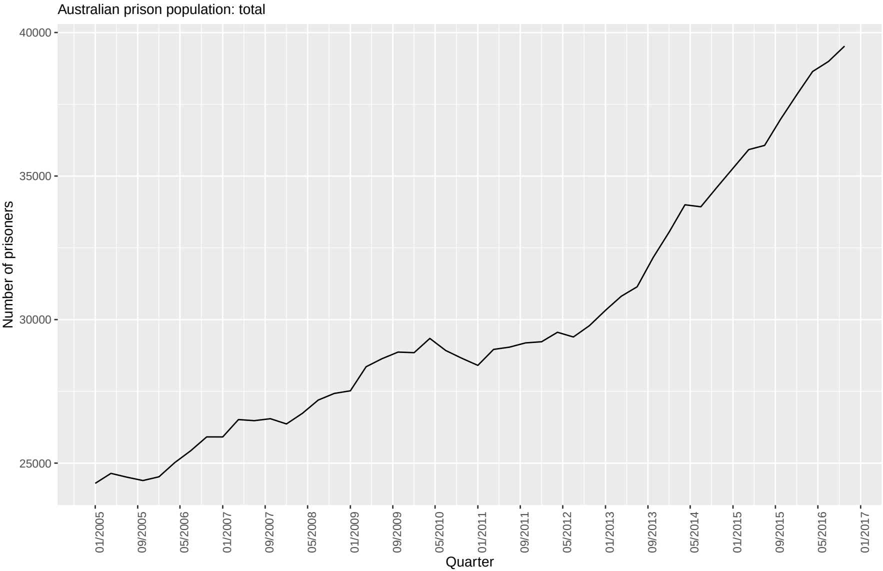
- Produce accurate but also detailed forecasts of prisoner numbers at the aggregate national level but also for multiple groupings based on attributes (and their interactions) that are of interest to a variety of policy makers and correctional administrators.

- The level of detail and the coherent nature of the forecasts enables informed and importantly aligned decision making across multiple departments and at all levels of management: strategic, tactical and operational.

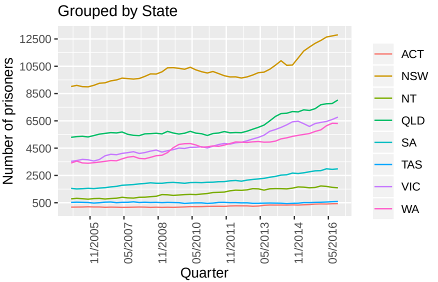
Aim

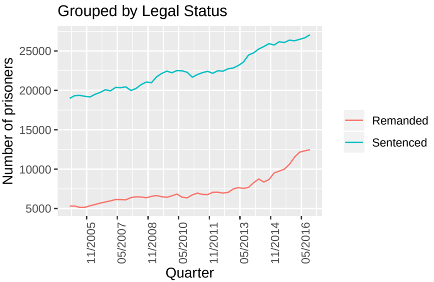
- Produce accurate but also detailed forecasts of prisoner numbers at the aggregate national level but also for multiple groupings based on attributes (and their interactions) that are of interest to a variety of policy makers and correctional administrators.
- The level of detail and the coherent nature of the forecasts enables informed and importantly aligned decision making across multiple departments and at all levels of management: strategic, tactical and operational.

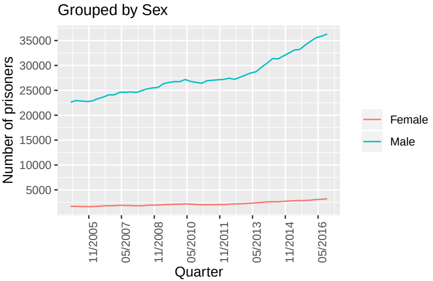
Australian prison population

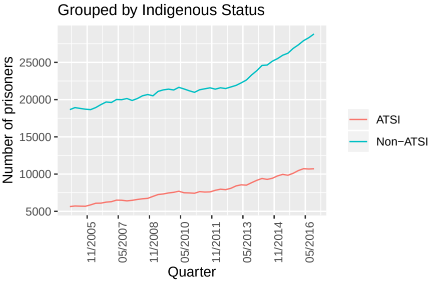


Australian prison population

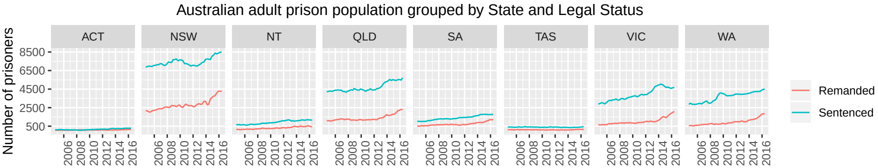


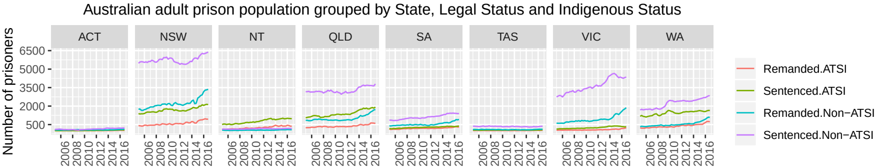






Australian prison population





Forecasting prison population

- Demographics (243 series = 1 + 16 + 60 + 104 + 64):
  - State (8)
  - Sex (2)
  - Legal Status (2)
  - Indigenous Status (2)
- ANZ Standard Offence Classification (243 series):
  - Divisions (16) (Homicide, Sexual Assault, Robbery, Illicit drugs, etc.)
  - Subdivisions (66) (Manslaughter and driving causing death, Murder, Attempted Murder, etc.)
  - Groups (160) (Manslaughter, Driving causing death, etc)

Forecasting prison population

- Demographics (243 series = 1 + 16 + 60 + 104 + 64):
  - State (8)
  - Sex (2)
  - Legal Status (2)
  - Indigenous Status (2)
- ANZ Standard Offence Classification (243 series):
  - Divisions (16) (Homicide, Sexual Assault, Robbery, Illicit drugs, etc.)
  - Subdivisions (66) (Manslaughter and driving causing death, Murder, Attempted Murder, etc.)
  - Groups (160) (Manslaughter, Driving causing death, etc)

Forecasting prison population

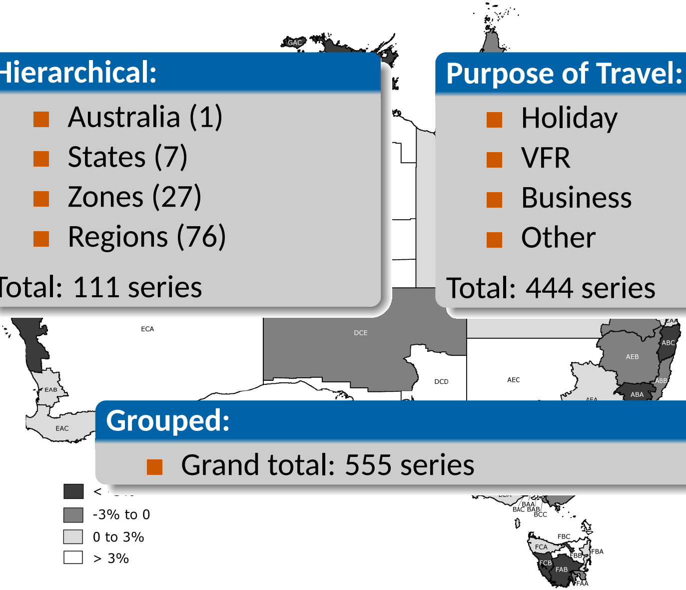
- Demographics (243 series = 1 + 16 + 60 + 104 + 64):
  - State (8)
  - Sex (2)
  - Legal Status (2)
  - Indigenous Status (2)

➥ Grouped time series.

- ANZ Standard Offence Classification (243 series):
  - Divisions (16) (Homicide, Sexual Assault, Robbery, Illicit drugs, etc.)
  - Subdivisions (66) (Manslaughter and driving causing death, Murder, Attempted Murder, etc.)
  - Groups (160) (Manslaughter, Driving causing death, etc)

➥ Hierarchical time series.

Australian domestic tourism



Hierarchical:

- Australia (1)
- States (7)
- Zones (27)
- Regions (76)

Total: 111 series

Purpose of Travel: $(\times 4)$

- Holiday
- VFR
- Business
- Other

Total: 444 series

Grouped:

- Grand total: 555 series

Forecasting student numbers

- Total number of Monash Students
  - Faculty (8)
  - Campus (2 + Other)
  - Funding source (3)
  - Course level (3)
  - Commencing/returning (2)
  - Courses (457)
  - Units (5605) (not sure we will get here).

Forecasting student numbers

- Total number of Monash Students
  - Faculty (8)
  - Campus (2 + Other)
  - Funding source (3)
  - Course level (3)
  - Commencing/returning (2)
  - Courses (457)
  - Units (5605) (not sure we will get here).

Total: 152,289 time series.

Forecasting student numbers

Challenges:

- Large number of series to forecast.
- We want a flexible forecasting process using all information available.
- We want forecasts to be coherent (add up).

- Courses (457)
- Units (5605) (not sure we will get here).

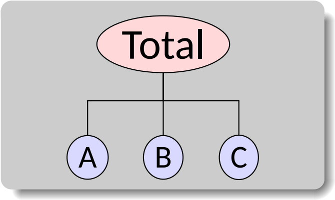
Key idea

- Forecast all series at all levels of aggregation or groupings (in contrast to typical bottom-up, top-down or middle-out approaches).
- Reconcile the forecasts so they add up correctly using least squares optimization, i.e., find closest reconciled forecasts to the original forecasts.

Key idea

- Forecast all series at all levels of aggregation or groupings (in contrast to typical bottom-up, top-down or middle-out approaches).
- Reconcile the forecasts so they add up correctly using least squares optimization, i.e., find closest reconciled forecasts to the original forecasts.

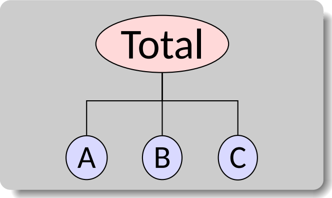
Hierarchical time series



$y_{Tot,t}$: observed aggregate of all series at time t.

$y_{X,t}$: observation on series X at time t. .

Hierarchical time series



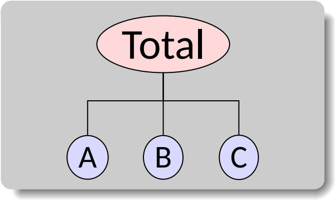
$y_{Tot,t}$: observed aggregate of all series at time t.

$y_{X,t}$: observation on series X at time t. .

Key concept:

➥ I can construct all time series in my collection if I know the aggregation structure and the bottom-level series.

Hierarchical time series

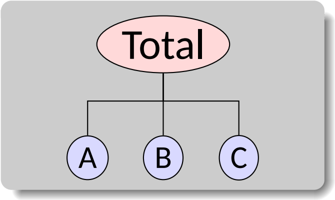


$$\boldsymbol{y}_t = \begin{pmatrix} y_{Tot,t} \\ y_{A,t} \\ y_{B,t} \\ y_{C,t} \end{pmatrix}$$

$y_{Tot,t}$: observed aggregate of all series at time t.

$y_{X,t}$: observation on series X at time t. .

Hierarchical time series

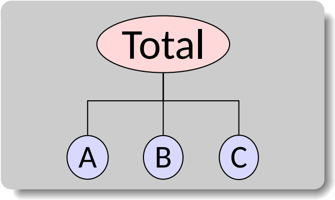


$y_{Tot,t}$: observed aggregate of all series at time t.

$y_{X,t}$: observation on series X at time t. .

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

Hierarchical time series



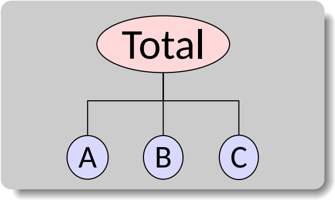
$y_{Tot,t}$: observed aggregate of all series at time t.

$y_{X,t}$: observation on series X at time t.

$\boldsymbol{b}_t$: vector of all bottom-level series at time t.

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

Hierarchical time series



$y_{Tot,t}$: observed aggregate of all series at time t.

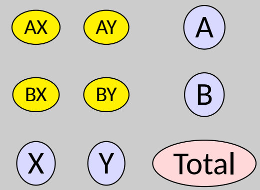
$y_{X,t}$: observation on series X at time t.

$\boldsymbol{b}_t$: vector of all bottom-level series at time t.

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

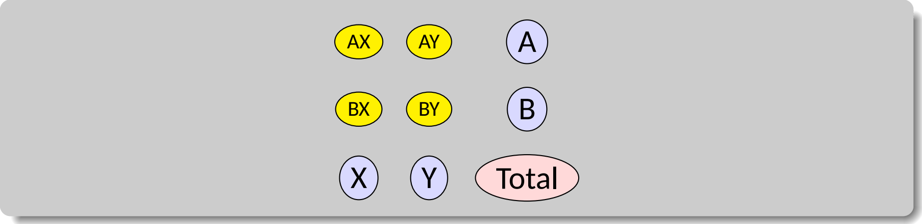
$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t$$

Grouped time series



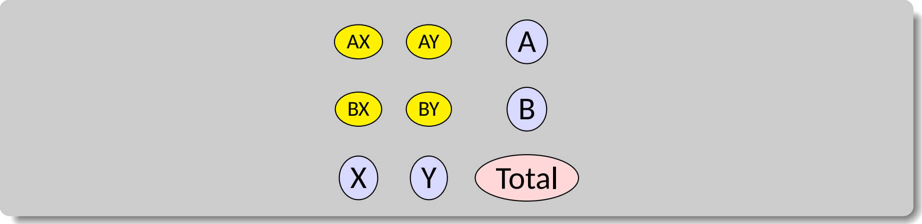
$$\boldsymbol{y}_t = \begin{pmatrix} y_{Tot,t} \\ y_{A,t} \\ y_{B,t} \\ y_{X,t} \\ y_{Y,t} \\ y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{S}} \underbrace{\begin{pmatrix} y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{pmatrix}}_{\boldsymbol{b}_t}$$

Grouped time series



$$\boldsymbol{y}_t = \begin{pmatrix} y_{Tot,t} \\ y_{A,t} \\ y_{B,t} \\ y_{X,t} \\ y_{Y,t} \\ y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{S}} \underbrace{\begin{pmatrix} y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{pmatrix}}_{\boldsymbol{b}_t}$$

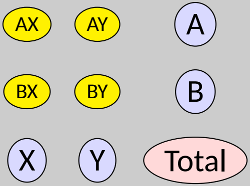
Grouped time series



$$\boldsymbol{y}_t = \begin{pmatrix} y_{Tot,t} \\ y_{A,t} \\ y_{B,t} \\ y_{X,t} \\ y_{Y,t} \\ y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{S}} \underbrace{\begin{pmatrix} y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{pmatrix}}_{\boldsymbol{b}_t}$$

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

Grouped time series



$y_{Tot,t}$ $\quad$ $1 \quad 1 \quad 1 \quad 1$

Key contribution:

➥ We can now deal effectively with both hierarchical and grouped aggregation structures.

$$\boldsymbol{y}_t = \begin{pmatrix} \vdots \\ y_{Y,t} \\ y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{pmatrix} = \underbrace{\begin{pmatrix} \vdots \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{S}} \underbrace{\begin{pmatrix} \vdots \\ y_{BX,t} \\ y_{BY,t} \end{pmatrix}}_{\boldsymbol{b}_t}$$

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

Hierarchical and grouped time series

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

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

where

- $\boldsymbol{y}_t$ is a vector of all series at time t.
- $\boldsymbol{S}$ is a "summing matrix" containing the aggregation constraints.
- $\boldsymbol{b}_t$ is a vector of the most disaggregated series at time t.

Outline

Forecasting framework

Let $\hat{\boldsymbol{y}}_T(h)$ be a vector of base (initial) h-step forecasts made at time T, stacked in same order as $\boldsymbol{y}_t$.
(These will almost certainly never add up.)

Reconciled (coherent) forecasts must be of the form:

$$\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_T(h)$$

for some matrix $\boldsymbol{P}$.

- $\boldsymbol{P}$ extracts and combines base forecasts $\hat{\boldsymbol{y}}_T(h)$ to get bottom-level forecasts, $\boldsymbol{P}\hat{\boldsymbol{y}}_T(h) = \hat{\boldsymbol{b}}_T(h)$. E.g., $\boldsymbol{P} = [\boldsymbol{0}|\boldsymbol{I}_m]$ for bottom-up, $\boldsymbol{P} = [\boldsymbol{p}|\boldsymbol{0}_{n-1}]$ for top-down.
- $\boldsymbol{S}$ adds them up, $\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}\hat{\boldsymbol{b}}_T(h)$.

Forecasting framework

Let $\hat{\boldsymbol{y}}_T(h)$ be a vector of base (initial) h-step forecasts made at time T, stacked in same order as $\boldsymbol{y}_t$.
(These will almost certainly never add up.)

Reconciled (coherent) forecasts must be of the form:

$$\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{SP}\hat{\boldsymbol{y}}_T(h)$$

for some matrix $\boldsymbol{P}$.

- $\boldsymbol{P}$ extracts and combines base forecasts $\hat{\boldsymbol{y}}_T(h)$ to get bottom-level forecasts, $\boldsymbol{P}\hat{\boldsymbol{y}}_T(h) = \hat{\boldsymbol{b}}_T(h)$. E.g., $\boldsymbol{P} = [\boldsymbol{0}|\boldsymbol{I}_m]$ for bottom-up, $\boldsymbol{P} = [\boldsymbol{p}|\boldsymbol{0}_{n-1}]$ for top-down.
- $\boldsymbol{S}$ adds them up, $\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}\hat{\boldsymbol{b}}_T(h)$.

Forecasting framework

Let $\hat{\boldsymbol{y}}_T(h)$ be a vector of base (initial) h-step forecasts made at time T, stacked in same order as $\boldsymbol{y}_t$.
(These will almost certainly never add up.)

Reconciled (coherent) forecasts must be of the form:

$$\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_T(h)$$

for some matrix $\boldsymbol{P}$.

- $\boldsymbol{P}$ extracts and combines base forecasts $\hat{\boldsymbol{y}}_T(h)$ to get bottom-level forecasts, $\boldsymbol{P}\hat{\boldsymbol{y}}_T(h) = \hat{\boldsymbol{b}}_T(h)$. E.g., $\boldsymbol{P} = [\boldsymbol{0}|\boldsymbol{I}_m]$ for bottom-up, $\boldsymbol{P} = [\boldsymbol{p}|\boldsymbol{0}_{n-1}]$ for top-down.
- $\boldsymbol{S}$ adds them up, $\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}\hat{\boldsymbol{b}}_T(h)$.

Forecasting framework

Let $\hat{\boldsymbol{y}}_T(h)$ be a vector of base (initial) h-step forecasts made at time T, stacked in same order as $\boldsymbol{y}_t$.
(These will almost certainly never add up.)

Reconciled (coherent) forecasts must be of the form:

$$\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_T(h)$$

for some matrix $\boldsymbol{P}$.

- $\boldsymbol{P}$ extracts and combines base forecasts $\hat{\boldsymbol{y}}_T(h)$ to get bottom-level forecasts, $\boldsymbol{P}\hat{\boldsymbol{y}}_T(h) = \hat{\boldsymbol{b}}_T(h)$. E.g., $\boldsymbol{P} = [\boldsymbol{0}|\boldsymbol{I}_m]$ for bottom-up, $\boldsymbol{P} = [\boldsymbol{p}|\boldsymbol{0}_{n-1}]$ for top-down.
- $\boldsymbol{S}$ adds them up, $\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}\hat{\boldsymbol{b}}_T(h)$.

Forecasting framework

Let $\hat{\boldsymbol{y}}_T(h)$ be a vector of base (initial) h-step forecasts made at time T, stacked in same order as $\boldsymbol{y}_t$.
(These will almost certainly never add up.)

Reconciled (coherent) forecasts must be of the form:

$$\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_T(h)$$

for some matrix $\boldsymbol{P}$.

- $\boldsymbol{P}$ extracts and combines base forecasts $\hat{\boldsymbol{y}}_T(h)$ to get bottom-level forecasts, $\boldsymbol{P}\hat{\boldsymbol{y}}_T(h) = \hat{\boldsymbol{b}}_T(h)$. E.g., $\boldsymbol{P} = [\boldsymbol{0}|\boldsymbol{I}_m]$ for bottom-up, $\boldsymbol{P} = [\boldsymbol{p}|\boldsymbol{0}_{n-1}]$ for top-down.
- $\boldsymbol{S}$ adds them up, $\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}\hat{\boldsymbol{b}}_T(h)$.

Forecasting framework

Let $\hat{\boldsymbol{y}}_T(h)$ be a vector of base (initial) h-step forecasts made at time T, stacked in same order as $\boldsymbol{y}_t$.
(These will almost certainly never add up.)

Reconciled (coherent) forecasts must be of the form:

$$\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_T(h)$$

for some matrix $\boldsymbol{P}$.

- $\boldsymbol{P}$ extracts and combines base forecasts $\hat{\boldsymbol{y}}_T(h)$ to get bottom-level forecasts, $\boldsymbol{P}\hat{\boldsymbol{y}}_T(h) = \hat{\boldsymbol{b}}_T(h)$. E.g., $\boldsymbol{P} = [\boldsymbol{0}|\boldsymbol{I}_m]$ for bottom-up, $\boldsymbol{P} = [\boldsymbol{p}|\boldsymbol{0}_{n-1}]$ for top-down.
- $\boldsymbol{S}$ adds them up, $\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}\hat{\boldsymbol{b}}_T(h)$.

Forecasting framework

Let $\hat{\boldsymbol{y}}_T(h)$ be a vector of base (initial) h-step forecasts made at time T, stacked in same order as $\boldsymbol{y}_t$.
(These will almost certainly never add up.)

Reconciled (coherent) forecasts must be of the form:

$$\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_T(h)$$

for some matrix $\boldsymbol{P}$.

- $\boldsymbol{P}$ extracts and combines base forecasts $\hat{\boldsymbol{y}}_T(h)$ to get bottom-level forecasts, $\boldsymbol{P}\hat{\boldsymbol{y}}_T(h) = \hat{\boldsymbol{b}}_T(h)$. E.g., $\boldsymbol{P} = [\boldsymbol{0}|\boldsymbol{I}_m]$ for bottom-up, $\boldsymbol{P} = [\boldsymbol{p}|\boldsymbol{0}_{n-1}]$ for top-down.
- $\boldsymbol{S}$ adds them up, $\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}\hat{\boldsymbol{b}}_T(h)$.

Forecasting framework

Let $\hat{\boldsymbol{y}}_T(h)$ be a vector of base (initial) h-step forecasts made at time T, stacked in same order as $\boldsymbol{y}_t$.
(These will almost certainly never add up.)

Reconciled (coherent) forecasts must be of the form:

$$\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_T(h)$$

Key limitation:

- Traditional approaches use information only from a single level.
- Can we do better?

bottom-level forecasts, $\boldsymbol{P}\hat{\boldsymbol{y}}_T(h) = \boldsymbol{b}_T(h)$. E.g., $\boldsymbol{P} = [\boldsymbol{0}|\boldsymbol{I}_m]$ for bottom-up, $\boldsymbol{P} = [\boldsymbol{p}|\boldsymbol{0}_{n-1}]$ for top-down.

- $\boldsymbol{S}$ adds them up, $\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}\hat{\boldsymbol{b}}_T(h)$.

Optimal reconciliation approach

$$\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_T(h)$$

The error variance of the reconciled forecasts is

$$\text{Var}(\boldsymbol{y}_{T+h} - \tilde{\boldsymbol{y}}_T(h)) = \boldsymbol{S}\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'\boldsymbol{S}'$$

where $\boldsymbol{W}_h = \text{Var}(\boldsymbol{y}_{T+h} - \hat{\boldsymbol{y}}_T(h))$, error variance of base forecasts.

Theorem: BLUF via trace minimisation (MinT)

For any $\boldsymbol{P}$ satisfying $\boldsymbol{S}\boldsymbol{P}\boldsymbol{S} = \boldsymbol{S}$

$$\min_{\boldsymbol{P}} \ \text{tr}[\boldsymbol{S}\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'\boldsymbol{S}']$$

has unique solution at $\boldsymbol{P} = (\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}$.

- Estimating $\boldsymbol{W}_h$ is challenging especially for $h > 1$.

Optimal reconciliation approach

$$\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_T(h)$$

The error variance of the reconciled forecasts is

$$\text{Var}(\boldsymbol{y}_{T+h} - \tilde{\boldsymbol{y}}_T(h)) = \boldsymbol{S}\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'\boldsymbol{S}'$$

where $\boldsymbol{W}_h = \text{Var}(\boldsymbol{y}_{T+h} - \hat{\boldsymbol{y}}_T(h))$, error variance of base forecasts.

Theorem: BLUF via trace minimisation (MinT)

For any $\boldsymbol{P}$ satisfying $\boldsymbol{S}\boldsymbol{P}\boldsymbol{S} = \boldsymbol{S}$

$$\min_{\boldsymbol{P}} \ \text{tr}[\boldsymbol{S}\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'\boldsymbol{S}']$$

has unique solution at $\boldsymbol{P} = (\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}$.

- Estimating $\boldsymbol{W}_h$ is challenging especially for $h > 1$.

Optimal reconciliation forecasts

$$\tilde{\mathbf{y}}_T(h) = \mathbf{S}(\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h^{-1}\hat{\mathbf{y}}_T(h)$$

Reconciled forecasts

Base forecasts

Optimal reconciliation forecasts

$$\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}\hat{\boldsymbol{y}}_T(h)$$

Reconciled forecasts

Base forecasts

WLS Solution

- We assume that $\boldsymbol{W}_h = k_h \boldsymbol{W}_1$ and approximate $\boldsymbol{W}_1$ by its diagonal using in-sample one-step ahead forecast errors.
- Easy to estimate, and places weight where we have best forecasts.

Shanika L Wickramasuriya, George Athanasopoulos, and Rob J Hyndman (2019). "Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization". *Journal of the American Statistical Association*, 1–45.

Outline

Australian Prison Population

Location:

- NSW
- QLD
- SA
- TAS
- VIC
- WA
- ACT
- NT

Sex:

- Male
- Female

Legal status:

- Remanded
- Sentenced

Indigenous status:

- ATSI
- Non-ATSI

Australian Prison Population

Location:

- NSW
- QLD
- SA
- TAS
- VIC
- WA
- ACT
- NT

Sex:

- Male
- Female

Legal status:

- Remanded
- Sentenced

Indigenous status:

- ATSI
- Non-ATSI

Grouped Total:

243 series

Australian Prison Population

Location:

- NSW
- QLD
- SA
- TAS
- VIC
- WA
- ACT
- NT

Sex:

- Male
- Female

Legal status:

- Remanded
- Sentenced

Indigenous status:

- ATSI
- Non-ATSI

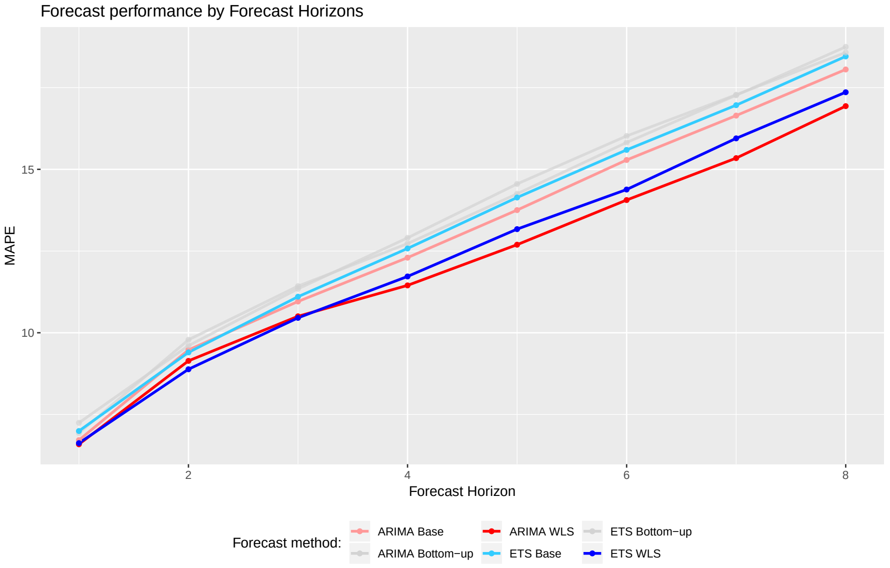
Grouped Total:

243 series

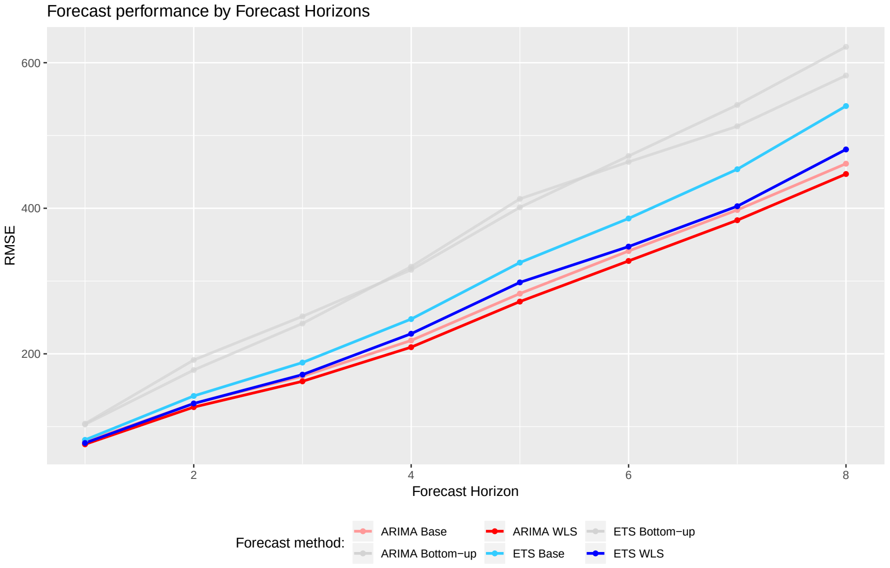
Forecast evaluation setup

- All adult prisoners in Australia: 2005Q1-2016Q4. (ABS corrective services database).
- 36 obs as training set and generate base forecasts with `auto.arima()` and `ets()` for $h = 1$ to 8-steps ahead.
- Obtain coherent forecasts using optimal reconciliation (WLS), and bottom-up.
- Use a rolling window: 12 1-step, 11 2-steps,...,4 8-steps ahead forecasts for evaluation.

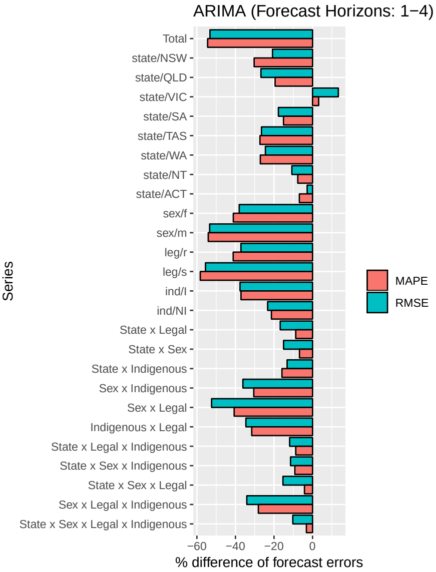
Forecast evaluation - MAPE

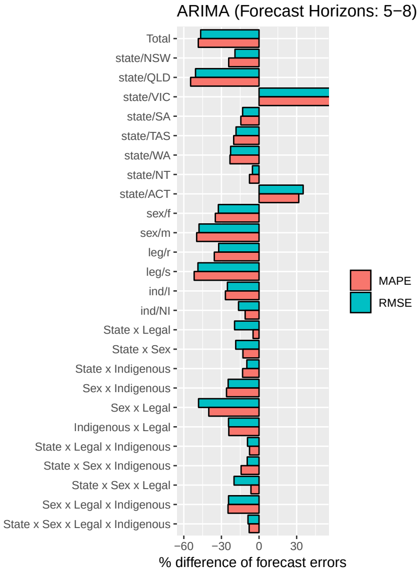


Forecast evaluation - RMSE

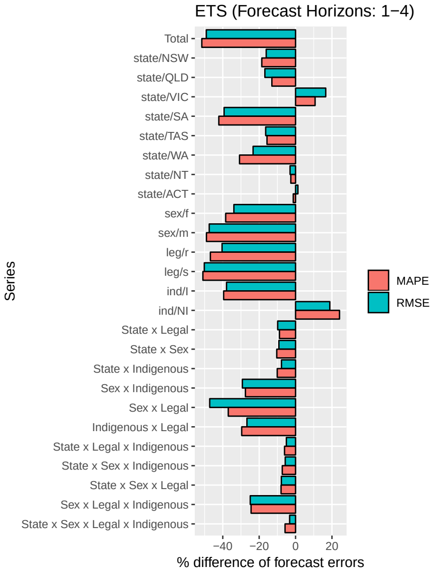


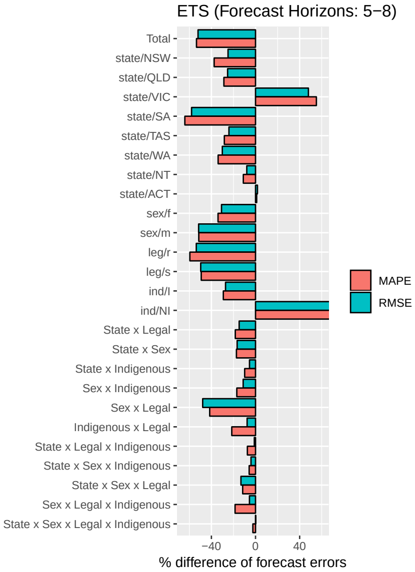
Forecast evaluation - Levels





Forecast evaluation - Levels





Outline

Temporal reconciliation

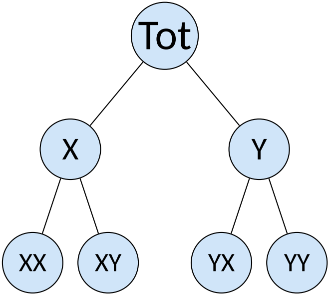


Figure: A simple two-level cross-sectional hierarchy.

Temporal reconciliation

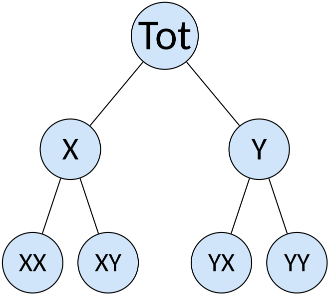


Figure: A simple two-level cross-sectional hierarchy.

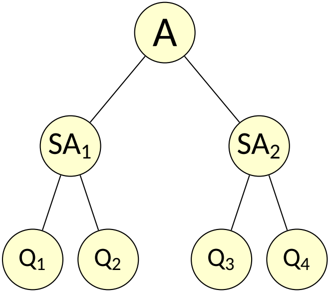
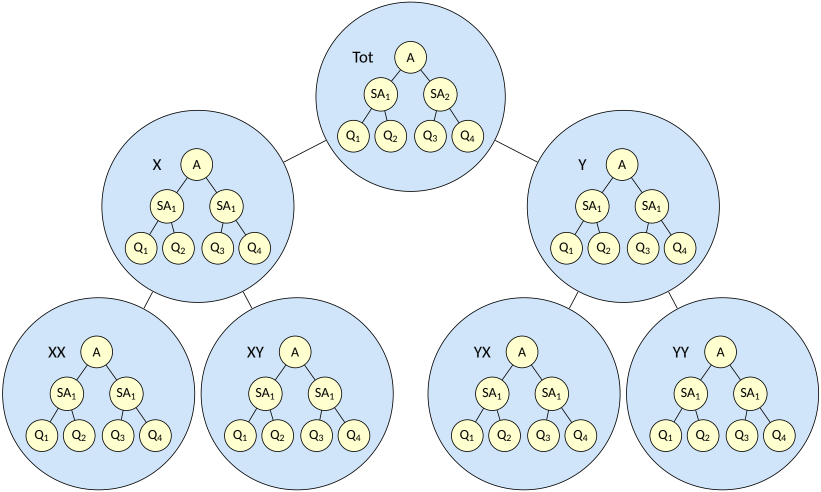


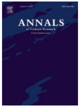
Figure: A temporal hierarchy for quarterly data.

George Athanasopoulos, Rob J Hyndman, Nikolaos Kourentzes, and Fotios Petropoulos (2017). "Forecasting with Temporal Hierarchies". *European Journal of Operational Research* **262**, 60–74.

Cross-temporal reconciliation

Cross-temporal reconciliation





Nikolaos Kourentzes and George Athanasopoulos (2019). "Cross-temporal coherent forecasts for Australian tourism". *Annals of Tourism Research* **forthcoming**.

Summary

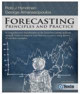
- Reconciliation (especially cross-temporal) offers a single/aligned view of the future to all decision makers, removing any organisational friction from misaligned decisions.
- More crucially, it offers a data driven way to break within and between organisations information silos.

Summary

- Reconciliation (especially cross-temporal) offers a single/aligned view of the future to all decision makers, removing any organisational friction from misaligned decisions.
- More crucially, it offers a data driven way to break within and between organisations information silos.

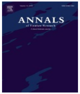
Outline

References



Rob J. Hyndman and George Athanasopoulos (2018). *Forecasting: principles and practice*. 2nd Edn. OTexts. `OTexts.org/fpp2/`. Chapter 10.

Shanika L Wickramasuriya, George Athanasopoulos, and Rob J Hyndman (2019). "Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization". *Journal of the American Statistical Association*, 1–45.



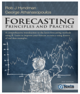
Nikolaos Kourentzes and George Athanasopoulos (2019). "Cross-temporal coherent forecasts for Australian tourism". *Annals of Tourism Research* **forthcomin**

George Athanasopoulos, Rob J Hyndman, Nikolaos Kourentzes, and Fotios Petropoulos (2017). "Forecasting with Temporal Hierarchies". *European Journal of Operational Research* **262**, 60–74

Rob J Hyndman, Alan J Lee, Earo Wang, and Shanika Wickramasuriya (2016). *hts: Hierarchical and Grouped Time Series*. R package v5.0 on CRAN.

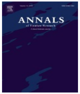
Rob J Hyndman and Nikolaos Kourentzes (2016). *thief: Temporal Hierarchical Forecasting*. R package v0.2 on CRAN.

References



Rob J. Hyndman and George Athanasopoulos (2018). *Forecasting: principles and practice*. 2nd Edn. OTexts. `OTexts.org/fpp2/`. Chapter 10.

Shanika L Wickramasuriya, George Athanasopoulos, and Rob J Hyndman (2019). "Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization". *Journal of the American Statistical Association*, 1–45.



Nikolaos Kourentzes and George Athanasopoulos (2019). "Cross-temporal coherent forecasts for Australian tourism". *Annals of Tourism Research* **forthcomin**

George Athanasopoulos, Rob J Hyndman, Nikolaos Kourentzes, and Fotios Petropoulos (2017). "Forecasting with Temporal Hierarchies". *European Journal of Operational Research* **262**, 60–74

Rob J Hyndman, Alan J Lee, Earo Wang, and Shanika Wickramasuriya (2016). *hts: Hierarchical and Grouped Time Series*. R package v5.0 on CRAN.

Rob J Hyndman and Nikolaos Kourentzes (2016). *thief: Temporal Hierarchical Forecasting*. R package v0.2 on CRAN.

Thank you!

Total Emergency Admissions via A&E