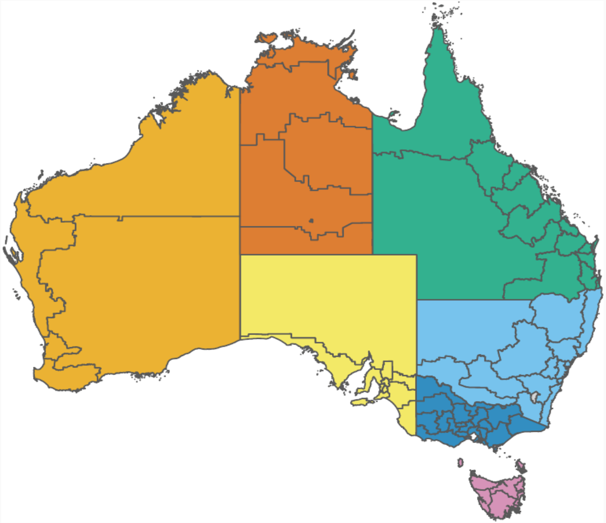
Forecast Linear Augmented Projection (FLAP)

Fin Yang, **George Athanasopoulos**, Rob J Hyndman, Anastasios Panagiotelis

MONASH University

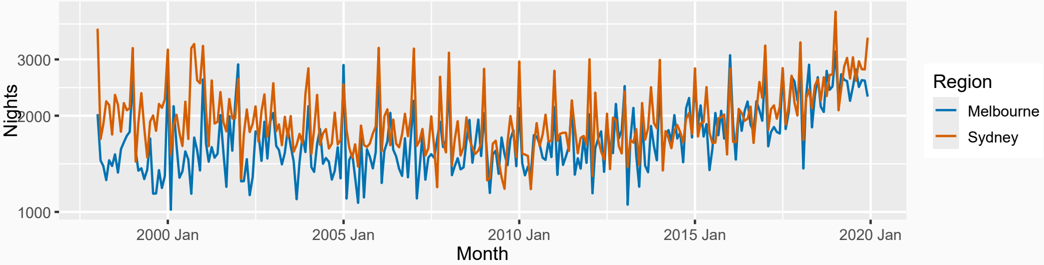
Photo by Stanley Cheung on Unsplash

Australian tourism regions

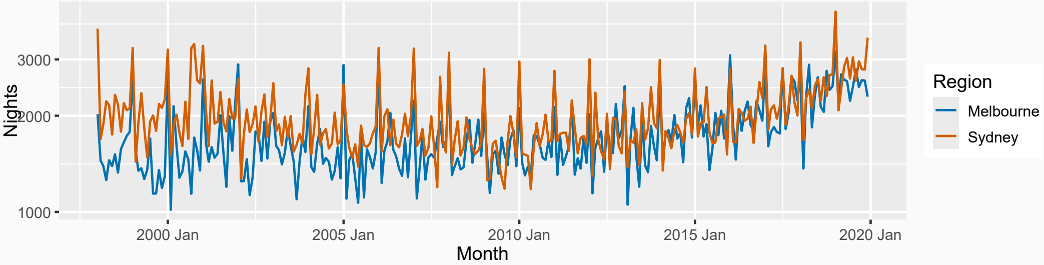


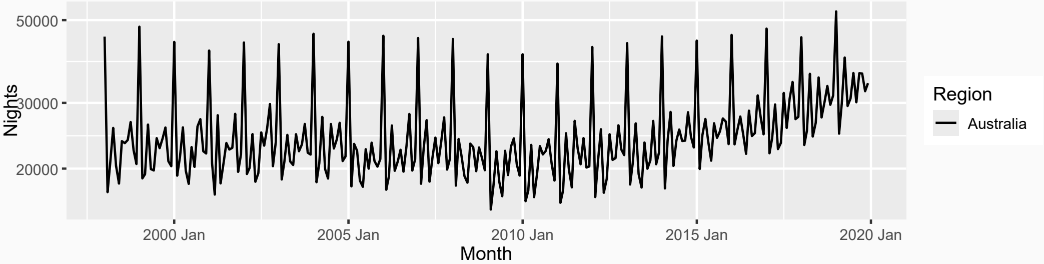
- Visitor Nights
- Monthly time series
- 1998 – 2019
- **77 regions**

Regions V Aggregate



Regions V Aggregate





FLAP Intuition

We have multivariate times series:

- which share similar patterns;
- with a better signal-noise ratio in the linear combination.

FLAP Intuition

We have multivariate times series:

- which share similar patterns;
- with a better signal-noise ratio in the linear combination.

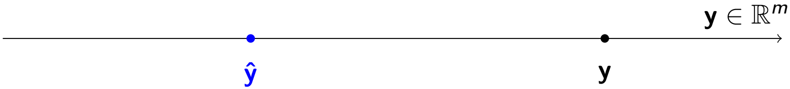
Can we find components that:

1. are easy to forecast (or easier than the original series);
2. can capture possible common signals;
3. can improve forecast of original series.

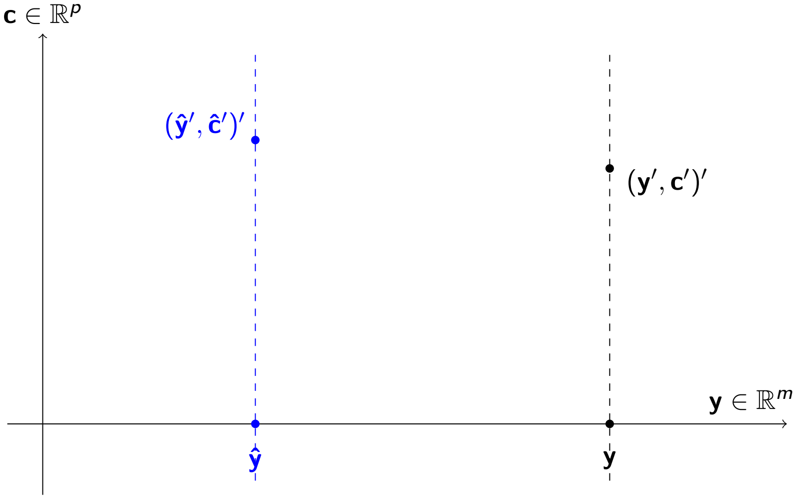
Outline of FLAP Implementation

- We want to forecast a multivariate series $\boldsymbol{y}_t \in \mathbb{R}^m$.
- Construct many linear combinations $\boldsymbol{c}_t = \Phi \boldsymbol{y}_t \in \mathbb{R}^p$ of the multivariate series.
- Produce univariate forecasts of all series $\hat{\boldsymbol{y}}_{t+h}$ and all linear combinations $\hat{\boldsymbol{c}}_{t+h}$.
- Project forecasts onto the $\mathbb{R}^m$ coherent subspace, resulting in $\tilde{\boldsymbol{y}}_{t+h}$.

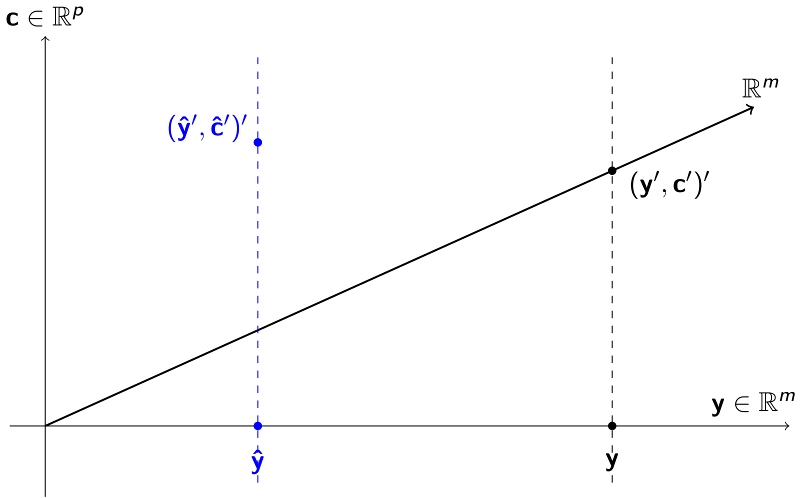
Geometry of FLAP



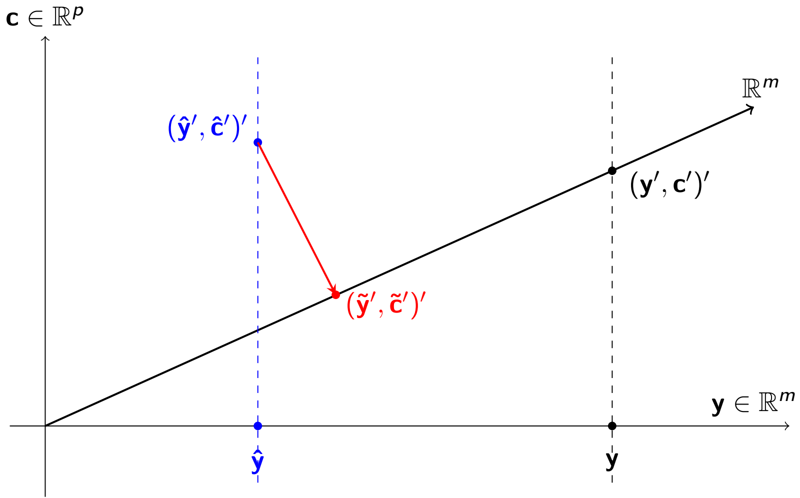
Geometry of FLAP



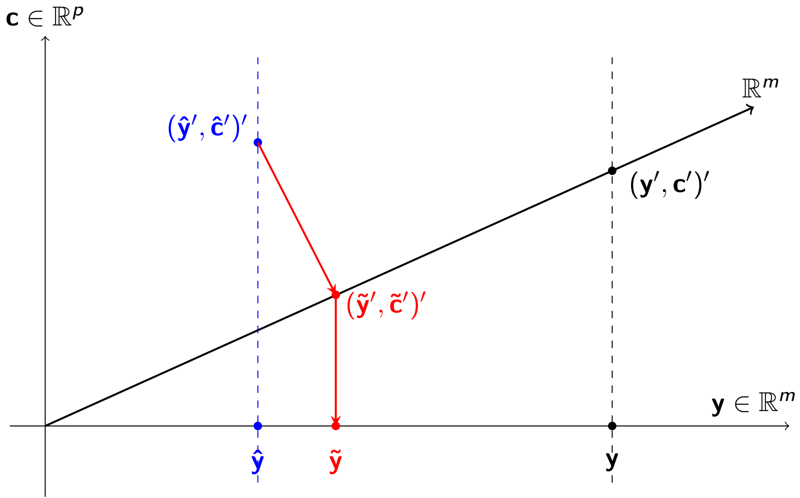
Geometry of FLAP



Geometry of FLAP



Geometry of FLAP



FLAP Projection

$$\boldsymbol{z}_t = \begin{bmatrix} \boldsymbol{y}_t \\ \boldsymbol{c}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}_t \\ \Phi \boldsymbol{y}_t \end{bmatrix}, \qquad \hat{\boldsymbol{z}}_{t+h} = \begin{bmatrix} \hat{\boldsymbol{y}}_{t+h} \\ \hat{\boldsymbol{c}}_{t+h} \end{bmatrix}, \qquad \tilde{\boldsymbol{z}}_{t+h} = \boldsymbol{M}\hat{\boldsymbol{z}}_{t+h}$$

where $\boldsymbol{M}$ is a projection matrix onto the $\mathbb{R}^m$ coherent subspace

$$\begin{aligned}
\boldsymbol{M} &= \boldsymbol{I}_{m+p} - \boldsymbol{W}_h \boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C} \\
\boldsymbol{C} &= \begin{bmatrix} -\Phi & \boldsymbol{I}_p \end{bmatrix} \\
\boldsymbol{W}_h &= \mathrm{Var}(\boldsymbol{z}_{t+h} - \hat{\boldsymbol{z}}_{t+h})
\end{aligned}$$

FLAP Projection

$$\boldsymbol{z}_t = \begin{bmatrix} \boldsymbol{y}_t \\ \boldsymbol{c}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}_t \\ \Phi \boldsymbol{y}_t \end{bmatrix}, \qquad \hat{\boldsymbol{z}}_{t+h} = \begin{bmatrix} \hat{\boldsymbol{y}}_{t+h} \\ \hat{\boldsymbol{c}}_{t+h} \end{bmatrix}, \qquad \tilde{\boldsymbol{z}}_{t+h} = \boldsymbol{M}\hat{\boldsymbol{z}}_{t+h}$$

where $\boldsymbol{M}$ is a projection matrix onto the $\mathbb{R}^m$ coherent subspace

$$\begin{aligned} \boldsymbol{M} &= \boldsymbol{I}_{m+p} - \boldsymbol{W}_h \boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C} \\ \boldsymbol{C} &= \begin{bmatrix} -\Phi & \boldsymbol{I}_p \end{bmatrix} \\ \boldsymbol{W}_h &= \mathrm{Var}(\boldsymbol{z}_{t+h} - \hat{\boldsymbol{z}}_{t+h}) \end{aligned}$$

$$\tilde{\boldsymbol{y}}_{t+h} = \boldsymbol{G}\hat{\boldsymbol{z}}_{t+h} = \boldsymbol{J}\boldsymbol{M}\hat{\boldsymbol{z}}_{t+h}$$

$$\boldsymbol{J} = \begin{bmatrix} \boldsymbol{I}_m & \boldsymbol{O}_{m\times p} \end{bmatrix}$$

Minimum variance of individual series

The projection is equivalent to the mapping

$$\tilde{\boldsymbol{y}}_{t+h} = \boldsymbol{G}\hat{\boldsymbol{z}}_{t+h} \quad \text{and} \quad \text{Var}(\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_{t+h}) = \boldsymbol{G}\boldsymbol{W}_h\boldsymbol{G}',$$

where $\boldsymbol{G} = \begin{bmatrix}\boldsymbol{g}_1 & \boldsymbol{g}_2 & \dots & \boldsymbol{g}_m\end{bmatrix}' \in \mathbb{R}^{m\times(m+p)}$ is the solution to

$$\underset{\boldsymbol{G}}{\arg\min}\ \text{tr}(\boldsymbol{G}\boldsymbol{W}_h\boldsymbol{G}') \qquad \text{s.t. } \boldsymbol{G}\boldsymbol{S} = \boldsymbol{I}$$

or

$$\underset{\boldsymbol{g}_i}{\arg\min}\ \boldsymbol{g}_i'\boldsymbol{W}_h\boldsymbol{g}_i \qquad \text{s.t. } \boldsymbol{g}_i'\boldsymbol{s}_j = \boldsymbol{1}(i=j),$$

where $\boldsymbol{S} = \begin{bmatrix}\boldsymbol{I}_m \\ \boldsymbol{\Phi}\end{bmatrix} = \begin{bmatrix}\boldsymbol{s}_1 \cdots \boldsymbol{s}_m\end{bmatrix}.$

Key results

1. The forecast error variance is **reduced** with FLAP
   - $\text{Var}(\boldsymbol{y}_{t+h} - \hat{\boldsymbol{y}}_{t+h}) - \text{Var}(\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_{t+h})$ is **positive semi-definite**.

Key results

1. The forecast error variance is **reduced** with FLAP
   - $\text{Var}(\boldsymbol{y}_{t+h} - \hat{\boldsymbol{y}}_{t+h}) - \text{Var}(\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_{t+h})$ is **positive semi-definite**.

2. The forecast error variance **monotonically** decreases with increasing number of components
   - the diagonal elements of $\text{Var}(\boldsymbol{y}_{t+h} - \hat{\boldsymbol{y}}_{t+h}) - \text{Var}(\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_{t+h})$ are non-decreasing as the number of components increases.

Key results

1. The forecast error variance is **reduced** with FLAP
   - $\text{Var}(\boldsymbol{y}_{t+h} - \hat{\boldsymbol{y}}_{t+h}) - \text{Var}(\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_{t+h})$ is **positive semi-definite**.
2. The forecast error variance **monotonically** decreases with increasing number of components
   - the diagonal elements of $\text{Var}(\boldsymbol{y}_{t+h} - \hat{\boldsymbol{y}}_{t+h}) - \text{Var}(\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_{t+h})$ are non-decreasing as the number of components increases.
3. The forecast projection is **optimal** to achieve minimum forecast error variance of each series.

In practice, we need to:

- Estimate $\boldsymbol{W}_h = \text{Var}(\boldsymbol{z}_{t+h} - \hat{\boldsymbol{z}}_{t+h})$.
  - Use in-sample residuals, shrink variances to their median, covariances to zero.

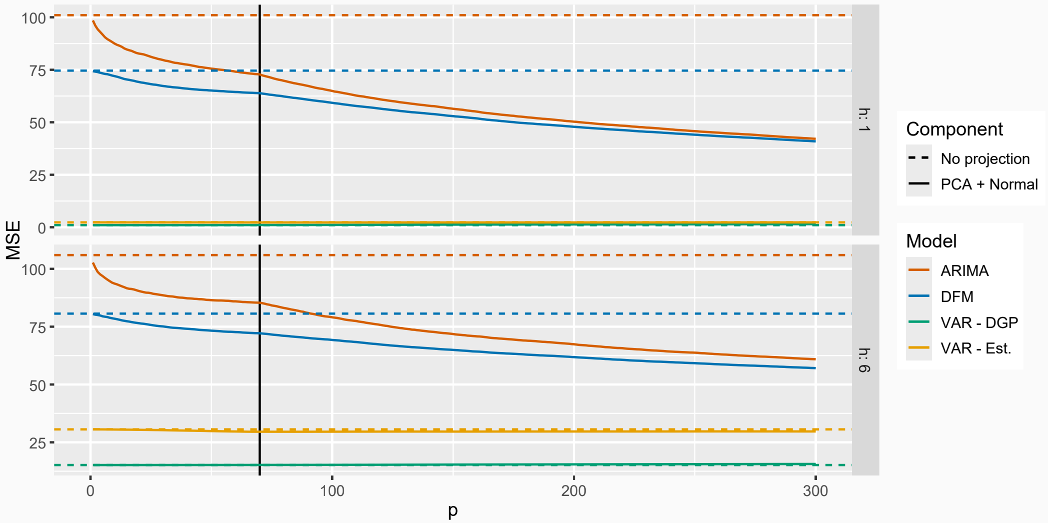
In practice, we need to:

- Estimate $\boldsymbol{W}_h = \text{Var}(\boldsymbol{z}_{t+h} - \hat{\boldsymbol{z}}_{t+h})$.
  - Use in-sample residuals, shrink variances to their median, covariances to zero.
- Construct the components, Φ.
  - Principal component analysis (PCA): find the weights matrix Φ so that the resulting components **maximise variance**.
  - Simulation: generate values of Φ from a random distribution and normalising them to unit vectors.
    - Normal distribution
    - Uniform distribution
    - Orthonormal matrix

Simulation

- Data generating process: VAR(3) with $m = 70$ variables
- Innovations $\sim N(0, \boldsymbol{I}_m)$
- Sample size: $T = 400$
- Number of repeated samples: 220
- Base forecasts:
  - ARIMA models using AICc (`auto.arima()` in `forecast` package).
  - DFM structure using BIC (different model for each horizon).

Simulation

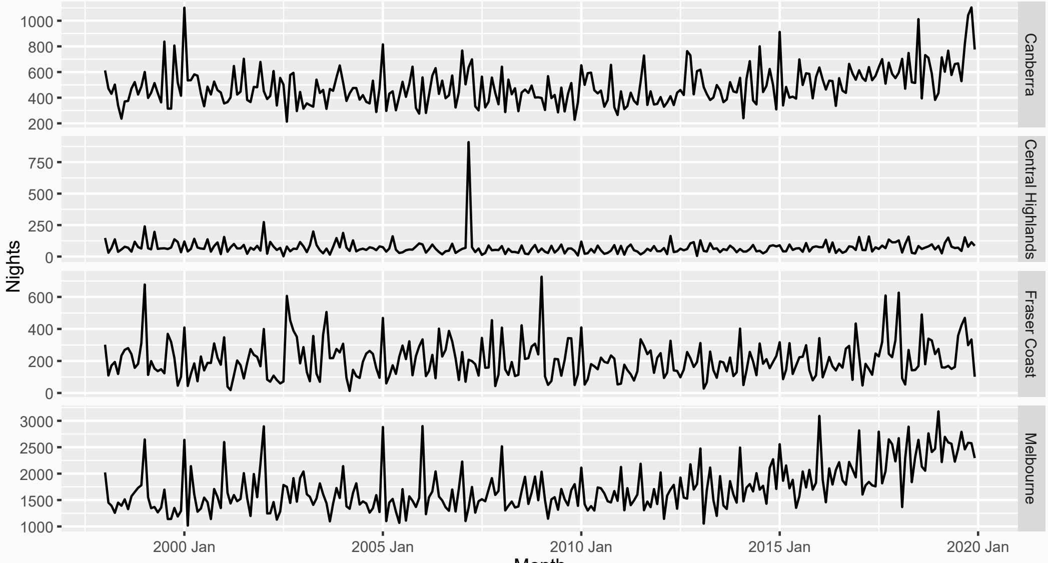


MSE
100
75
50
25
0
h: 1
100
75
50
25
h: 6
0
100
200
300
p
Component
No projection
PCA + Normal
Model
ARIMA
DFM
VAR - DGP
VAR - Est.

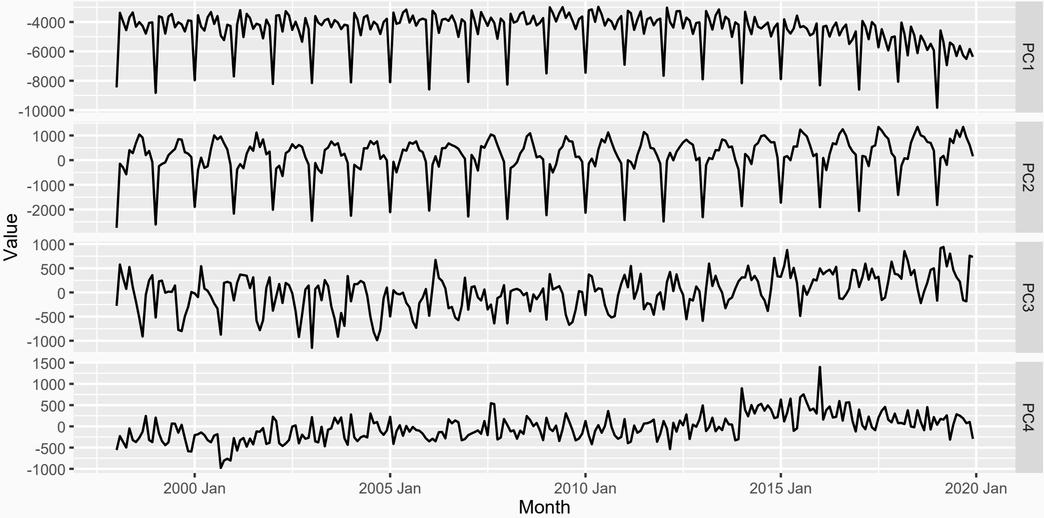
Monthly Australian regional tourism

- Monthly Australian tourism data by region giving 77 series, from Jan 1998 to Dec 2019
- Use expanding window time series cross-validation with $T = 84$ observations in first training set, and forecast horizons $h = 1, 2, \dots, 12$.
- Estimate `ets()` models using the `forecast` package.

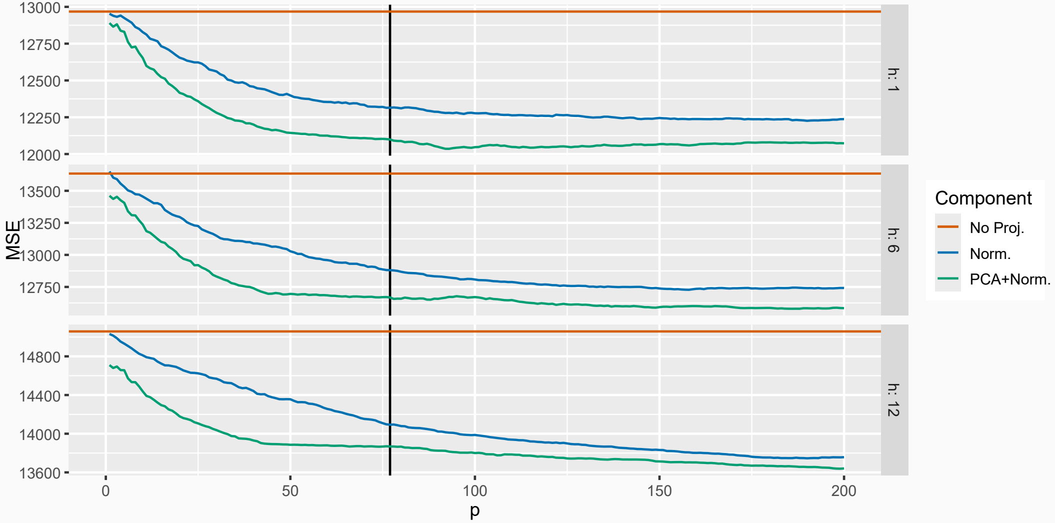
Monthly Australian regional tourism



Monthly Australian regional tourism



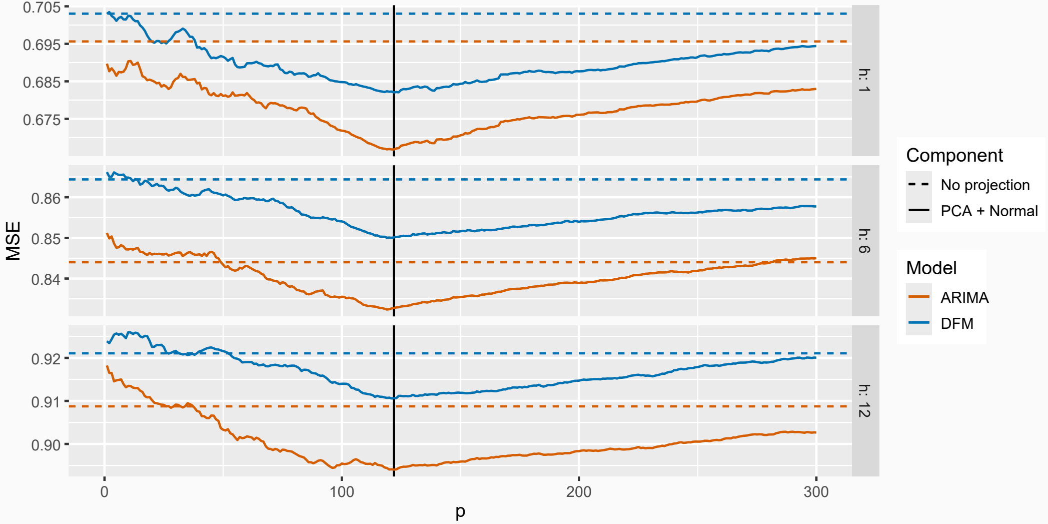
Monthly Australian regional tourism – `ets()`



FRED-MD

- Monthly data of macroeconomic variables (McCracken and Ng, 2016).
- Data from Jan 1959 – Sep 2023. 777 observations on 122 series.
- Same cleaning process as per McCracken and Ng (2016).
- All series scaled to have mean 0 and variance 1.
- Expanding time series cross-validation with initial size of 25 years and forecast horizon 12 months.

FRED-MD


0.705
0.695
0.685
0.675
h: 1
0.86
0.85
0.84
h: 6
0.92
0.91
0.90
h: 12
MSE
0
100
200
300
p
Component
No projection
PCA + Normal
Model
ARIMA
DFM

Future research directions

- Investigate why PCA performs better than random weights
- Find other components that are better than PCA
- Find optimal components by minimising forecast error variance with respect to Φ
- Use forecast projection and forecast reconciliation together

Working Paper and R Package

YF Yang, G Athanasopoulos, RJ Hyndman, and A Panagiotelis (2024). **"Forecast Linear Augmented Projection (FLAP): A free lunch to reduce forecast error variance".** Department of Econometrics and Business Statistics, Monash University, Working Paper Series 13/24. URL: https://www.monash.edu/business/ebs/research/publications/ebs/2024/wp13-2024.pdf

You can install the stable version from CRAN

```
## CRAN.R-project.org/package=flap
install.packages("flap")
```

or the development version from Github

```
## github.com/FinYang/flap
# install.packages("remotes")
remotes::install_github("FinYang/flap")
```

Slides and other information

Slides:

https://github.com/GeorgeAthana/FLAP-presentation

Other information:

https://research.monash.edu/en/persons/george-athanasopoulos

Thank you!

Yang, YF, G Athanasopoulos, RJ Hyndman, and A Panagiotelis (2024). **“Forecast Linear Augmented Projection (FLAP): A free lunch to reduce forecast error variance”.** Department of Econometrics and Business Statistics, Monash University, Working Paper Series 13/24. URL: https://www.monash.edu/business/ebs/research/publications/ebs/2024/wp13-2024.pdf.