



# **Forecast Linear Augmented Projection (FLAP): A free lunch to reduce forecast error variance**

Yangzhuoran Fin Yang  
George Athanasopoulos  
Rob J Hyndman  
Anastasios Panagiotelis

# Forecast Linear Augmented Projection (FLAP)

A model-independent post-forecast adjustment method that can reduce forecast error variance.

- Averaging indirect forecasts from linear combinations (components)
- Projecting forecasts of augmented series
- Free lunch: no additional data or information needed

# What to expect

- Intuition with data
- Literature
- Method formulation
- Properties
- Empirical applications and simulation

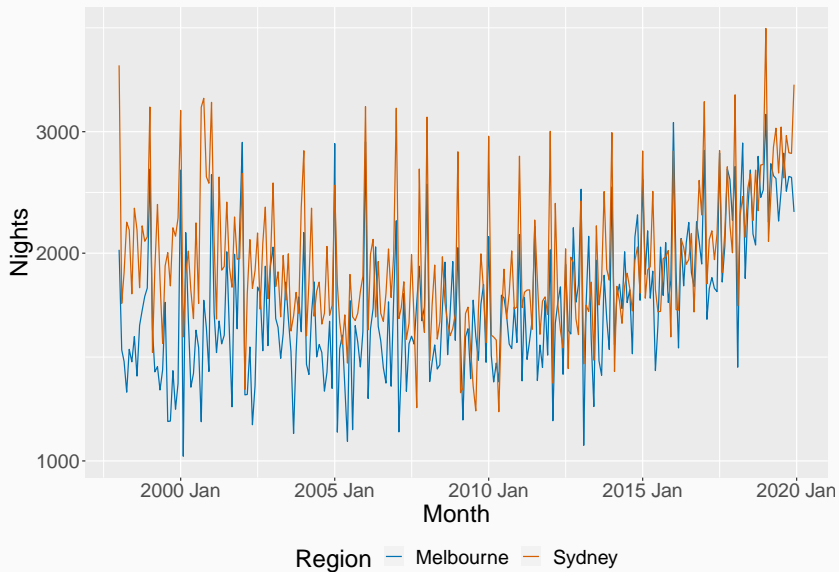
# Australian tourism data

- The data include tourism information on seven states and territories which can be divided into 77 regions
  - ▶ For example, Melbourne, Sydney, East Coast

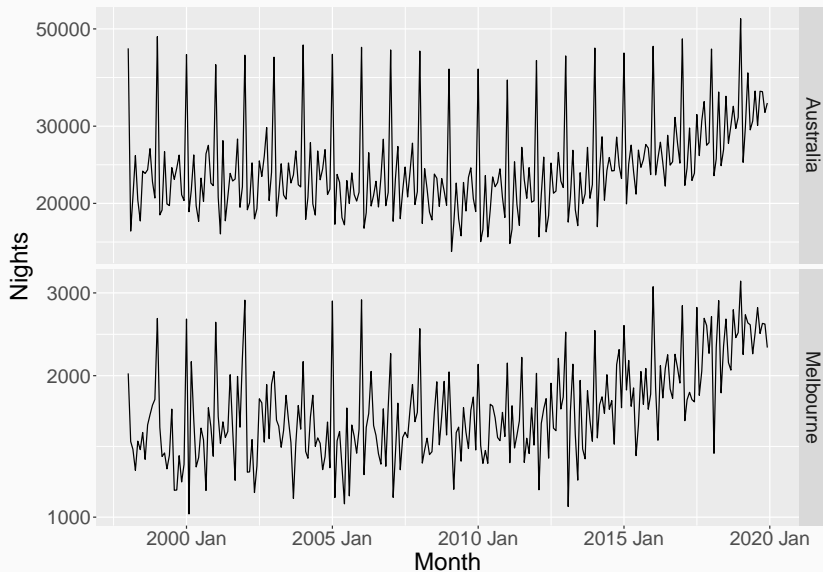
## Visitor nights

The total number of nights spent by Australians away from home recorded monthly

# Melbourne and Sydney



# Total and Region



# Intuition

## Observation

1. Similar patterns are shared by different series.
2. Better signal-noise ratio in the linear combination.

# Intuition

## Observation

1. Similar patterns are shared by different series.
2. Better signal-noise ratio in the linear combination.

## One step further

Finding components that

1. are easy to forecast;
2. can capture the common signals;
3. can improve forecast of original series.



## Forecast reconciliation

- Wickramasuriya, Athanasopoulos, and Hyndman (2019): Projecting forecasts to be consistent with the hierarchical structure

## Forecast combination

- Combining forecasts of the target series
- Hollyman, Petropoulos, and Tipping (2021): Combining direct and indirect forecasts
- Petropoulos and Spiliotis (2021): Combining forecasts of selections and transformations of the target series (“wisdom of data”)

# Literature

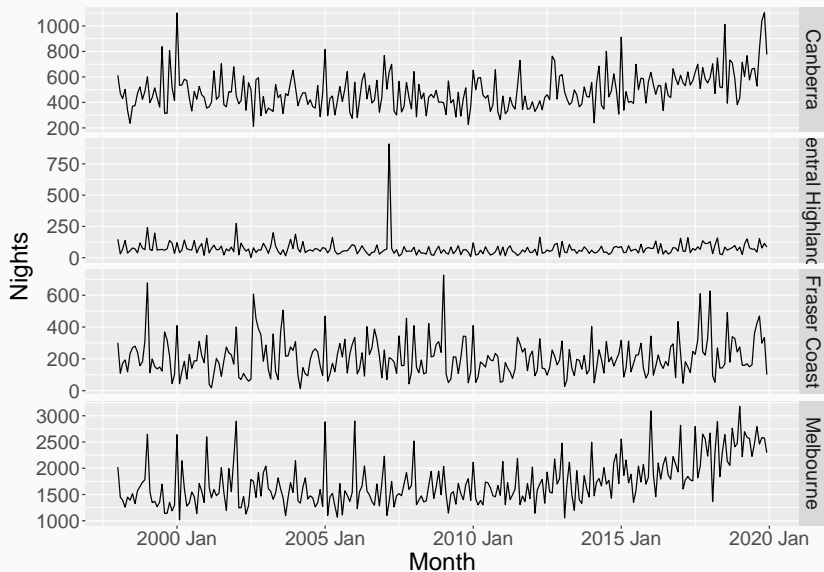
## Bagging

- Bergmeir, Hyndman, and Benítez (2016): Bagging ETS models to forecast
- Petropoulos, Hyndman, and Bergmeir (2018): The benefits of bagging originate from the model uncertainty

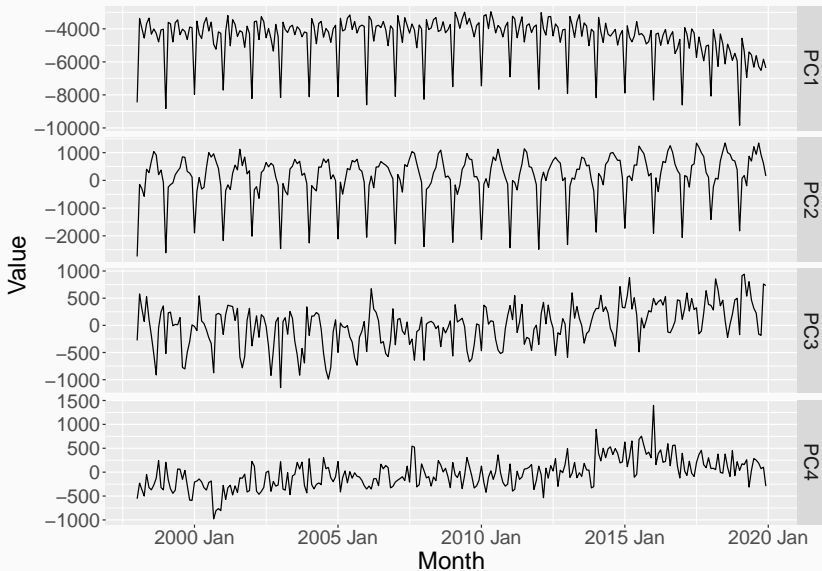
## Dynamic factor model (DFM)

- Stock and Watson (2002a), Stock and Watson (2002b)
- De Stefani et al. (2019): Machine learning extension

Series  $y_t \in \mathbb{R}^m$



# Components $\mathbf{c}_t = \Phi \mathbf{y}_t \in \mathbb{R}^p$



$$\mathbf{z}_t = \begin{bmatrix} \mathbf{y}_t \\ \mathbf{c}_t \end{bmatrix} \quad \tilde{\mathbf{z}}_{t+h} = \mathbf{M}\hat{\mathbf{z}}_{t+h}$$

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

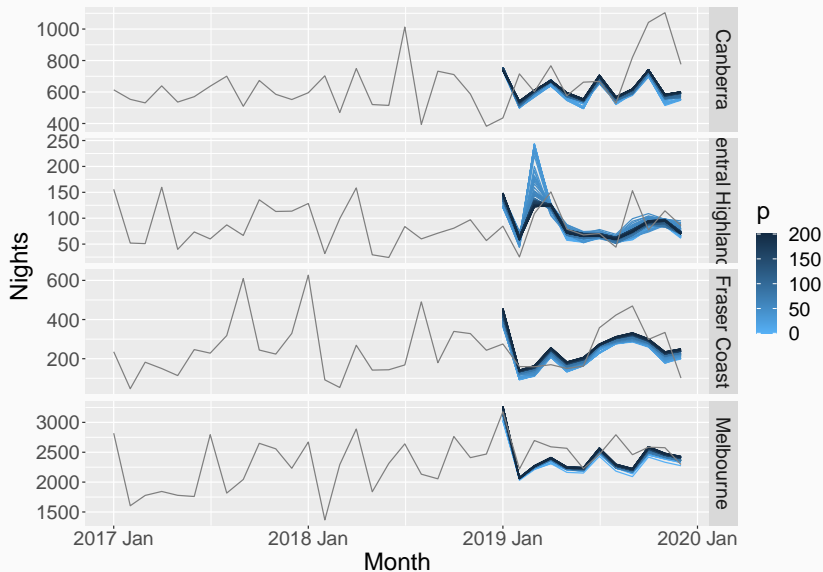
$$\mathbf{M} = \mathbf{I}_{m+p} - \mathbf{W}_h \mathbf{C}' (\mathbf{C} \mathbf{W}_h \mathbf{C}')^{-1} \mathbf{C}$$

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

$$\mathbf{C} = \begin{bmatrix} -\Phi & \mathbf{I}_p \end{bmatrix}$$

$$\mathbf{W}_h = \text{Var}(\mathbf{z}_{t+h} - \hat{\mathbf{z}}_{t+h})$$

# Forecasts and FLAP of series



# Unbiasedness

## Unbiasedness

If the base forecasts are unbiased, then the FLAP forecasts are also unbiased.

## Projection matrix

- 1 The matrix  $\mathbf{M}$  is a projection onto the space where the constraint  $\mathbf{C}\mathbf{z}_t = \mathbf{0}$  is satisfied.
- 2 The projected forecast  $\tilde{\mathbf{z}}_{t+h}$  satisfies the constraint  $\mathbf{C}\tilde{\mathbf{z}}_{t+h} = \mathbf{0}$ .
- 3 For  $\mathbf{z}_{t+h}$  that already satisfies the constraint, the projection does not change its value:  
$$\mathbf{M}\mathbf{z}_{t+h} = \mathbf{z}_{t+h}$$

# Nonnegative variance reduction

Under unbiasedness, the variance reduction is **positive semi-definite**:

$$\begin{aligned}\text{Var}(\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h}) - \text{Var}(\mathbf{y}_{t+h} - \tilde{\mathbf{y}}_{t+h}) \\ = \mathbf{J}\mathbf{W}_h\mathbf{C}'(\mathbf{C}\mathbf{W}_h\mathbf{C}')^{-1}\mathbf{C}\mathbf{W}_h\mathbf{J}'\end{aligned}$$



## Example $W_h = I_{m+p}$

$$\begin{aligned}\text{Var}(\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h}) - \text{Var}(\mathbf{y}_{t+h} - \tilde{\mathbf{y}}_{t+h}) \\&= \mathbf{J}\mathbf{C}'(\mathbf{C}\mathbf{C}')^{-1}\mathbf{C}\mathbf{J}' \\&= \Phi'(\Phi\Phi' + \mathbf{I})^{-1}\Phi\end{aligned}$$

Let  $\Phi$  consist of orthogonal unit vectors:

$$\Phi\Phi' = \mathbf{I}_p \text{ when } p \leq m$$

$$\Phi'\Phi = \mathbf{I}_m \text{ when } p = m.$$

$$\begin{aligned}\text{tr}(\text{Var}(\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h}) - \text{Var}(\mathbf{y}_{t+h} - \tilde{\mathbf{y}}_{t+h})) \\&= \frac{1}{2}\text{tr}(\Phi'\Phi) = \frac{1}{2}p\end{aligned}$$

# Positive condition

For the first component to have a guaranteed reduction of forecast error variance, the following condition must be satisfied:

$$\phi_1 \mathbf{W}_{y,h} \neq \mathbf{w}_{c_1 y, h},$$

- $\phi_1$  is the weight vector of the first component
- $\mathbf{W}_{y,h} = \text{Var}(\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h})$
- $\mathbf{w}_{c_1 y, h}$  is the forecast covariance between the first component and the original series.

# Monotonicity

The forecast error variance reductions, i.e. the diagonal elements of

$$\begin{aligned} \text{Var}(\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h}) - \text{Var}(\mathbf{y}_{t+h} - \tilde{\mathbf{y}}_{t+h}) \\ = \mathbf{J}\mathbf{W}_h\mathbf{C}'(\mathbf{C}\mathbf{W}_h\mathbf{C}')^{-1}\mathbf{C}\mathbf{W}_h\mathbf{J}' \end{aligned}$$

is non-decreasing as  $p$  increases.

# Minimum variance of individual series

The projection is equivalent to the mapping

$$\tilde{\mathbf{y}}_{t+h} = \mathbf{G}\hat{\mathbf{z}}_{t+h},$$

where  $\mathbf{G} = [\mathbf{g}_1 \ \mathbf{g}_2 \ \dots \ \mathbf{g}_m]' \in \mathbb{R}^{m \times (m+p)}$  is the solution to

$$\arg \min_{\mathbf{G}} \mathbf{G}\mathbf{W}_h\mathbf{G}' \quad \text{s.t. } \mathbf{G}\mathbf{S} = \mathbf{I}$$

or

$$\arg \min_{\mathbf{g}_i} \mathbf{g}_i' \mathbf{W}_h \mathbf{g}_i \quad \text{s.t. } \mathbf{g}_i' \mathbf{s}_j = \mathbf{1}(i = j),$$

where  $\mathbf{S} = \begin{bmatrix} \mathbf{I}_m \\ \Phi \end{bmatrix} = [\mathbf{s}_1 \ \dots \ \mathbf{s}_m]$ .

# Key results

- 1 The forecast error variance is **reduced** with FLAP
- 2 The forecast error variance **monotonically** decreases with increasing number of components
- 3 The forecast projection is **optimal** to achieve minimum forecast error variance of each series

# Estimation

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

$$\mathbf{M} = \mathbf{I}_{m+p} - \mathbf{W}_h \mathbf{C}' (\mathbf{C} \mathbf{W}_h \mathbf{C}')^{-1} \mathbf{C}$$

$$\mathbf{W}_h = \text{Var}(\mathbf{z}_{t+h} - \hat{\mathbf{z}}_{t+h})$$

$$\mathbf{C} = \begin{bmatrix} -\Phi & \mathbf{I}_p \end{bmatrix}$$

- Estimation of  $\mathbf{W}_h$
- Construction of  $\Phi$

# Estimation of $\mathbf{W}_h$

**Shrinking variance** towards their median (Opge-Rhein and Strimmer 2007) and **shrinking covariance** towards zero (Schäfer and Strimmer 2005).

The shrinkage estimator is

- Positive definite, and
- Numerically stable.

In empirical applications, we assume

$$\widehat{\mathbf{W}}_h^{shr} = \eta_h \widehat{\mathbf{W}}_1^{shr}.$$

# Construction of $\Phi$

## Principal component analysis (PCA)

Finding the weights matrix so that the resulting components **maximise variance**

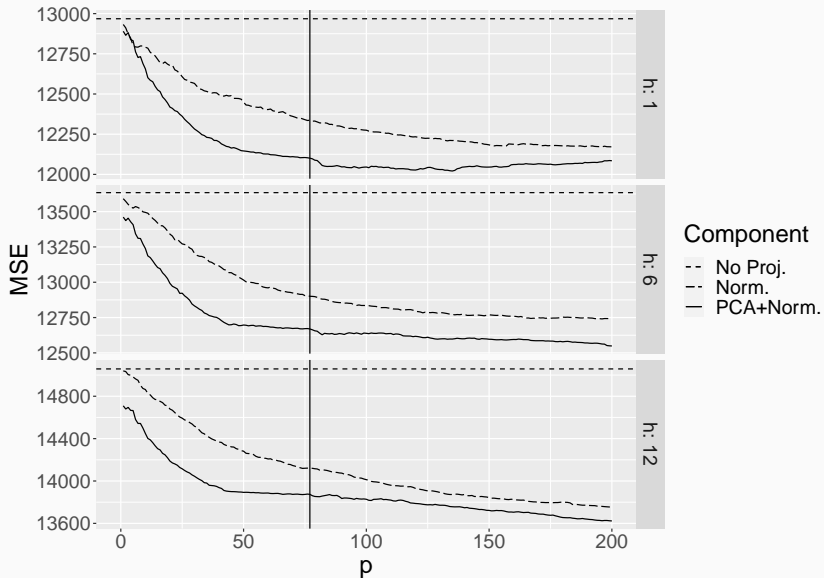
## Simulation

Generating values from a random distribution and normalising them to unit vectors

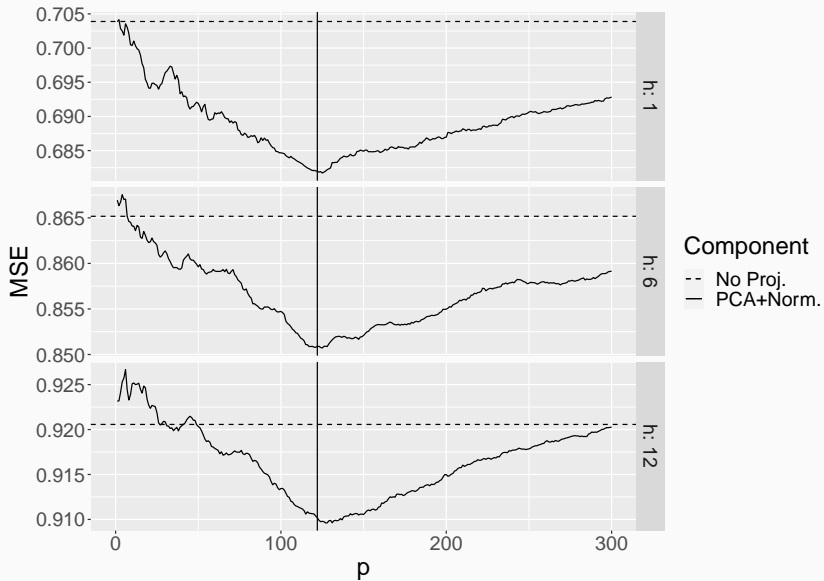
- Normal distribution
- Uniform distribution
- Orthonormal matrix (Borchers 2023)



# Tourism (ETS)



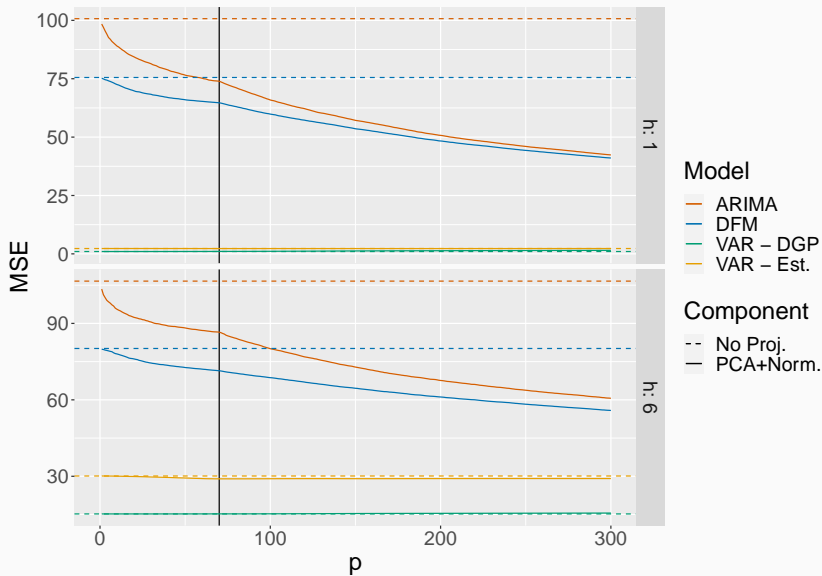
# FRED-MD (DFM)



# Simulation

- Data generating process (DGP): VAR(3) with  $m = 70$  variables
- Sample size:  $T = 400$
- Number of repeated samples: 220
- Base model: ARIMA and DFM

# Simulation



# 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  $\Phi$
- Use forecast projection and forecast reconciliation together

# R Package flap

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")
```

# Contact

- Email: [Fin.Yang@monash.edu](mailto:Fin.Yang@monash.edu)
- <https://yangzhuoranyang.com>



# References i

- Bergmeir, Christoph, Rob J Hyndman, and José M Benítez. 2016. "Bagging Exponential Smoothing Methods Using STL Decomposition and Box-Cox Transformation." *International Journal of Forecasting* 32 (April): 303–12.  
<https://doi.org/10.1016/j.ijforecast.2015.07.002>.
- Borchers, Hans W. 2023. *pracma: Practical Numerical Math Functions*.  
<https://CRAN.R-project.org/package=pracma>.
- De Stefani, Jacopo, Yann-Aël Le Borgne, Olivier Caelen, Dalila Hattab, and Gianluca Bontempi. 2019. "Batch and Incremental Dynamic Factor Machine Learning for Multivariate and Multi-Step-Ahead Forecasting." *International Journal of Data Science and Analytics* 7 (June): 311–29.  
<https://doi.org/10.1007/s41060-018-0150-x>.
- Hollyman, Ross, Fotios Petropoulos, and Michael E Tipping. 2021. "Understanding Forecast Reconciliation." *European Journal of Operational Research* 294 (October): 149–60. <https://doi.org/10.1016/j.ejor.2021.01.017>.
- Opgen-Rhein, Rainer, and Korbinian Strimmer. 2007. "Accurate Ranking of Differentially Expressed Genes by a Distribution-Free Shrinkage Approach." *Statistical Applications in Genetics and Molecular Biology* 6 (February).  
<https://doi.org/10.2202/1544-6115.1252>.



# References ii

- Petropoulos, Fotios, Rob J Hyndman, and Christoph Bergmeir. 2018. "Exploring the Sources of Uncertainty: Why Does Bagging for Time Series Forecasting Work?" *European Journal of Operational Research* 268 (July): 545–54. <https://doi.org/10.1016/j.ejor.2018.01.045>.
- Petropoulos, Fotios, and Evangelos Spiliotis. 2021. "The Wisdom of the Data: Getting the Most Out of Univariate Time Series Forecasting." *Forecasting* 3 (June): 478–97. <https://doi.org/10.3390/forecast3030029>.
- Schäfer, Juliane, and Korbinian Strimmer. 2005. "A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics." *Statistical Applications in Genetics and Molecular Biology* 4 (November). <https://doi.org/10.2202/1544-6115.1175>.
- Stock, James H, and Mark W Watson. 2002a. "Macroeconomic Forecasting Using Diffusion Indexes." *Journal of Business & Economic Statistics* 20 (April): 147–62. <https://doi.org/10.1198/073500102317351921>.
- . 2002b. "Forecasting Using Principal Components from a Large Number of Predictors." *Journal of the American Statistical Association* 97 (December): 1167–79. <https://doi.org/10.1198/016214502388618960>.

# References iii

Wickramasuriya, Shanika L, 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* 114 (April): 804–19. <https://doi.org/10.1080/01621459.2018.1448825>.