MONASH University

MONASH
BUSINESS
SCHOOL

Free Lunch Multivariate Forecast Projection: reducing forecast variance using linear combinations

Yangzhuoran Fin Yang
Rob J Hyndman
George Athanasopoulos
Anastasios Panagiotelis

Free lunch forecast projection

A model-independent post-forecast adjustment method that can reduce forecast 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

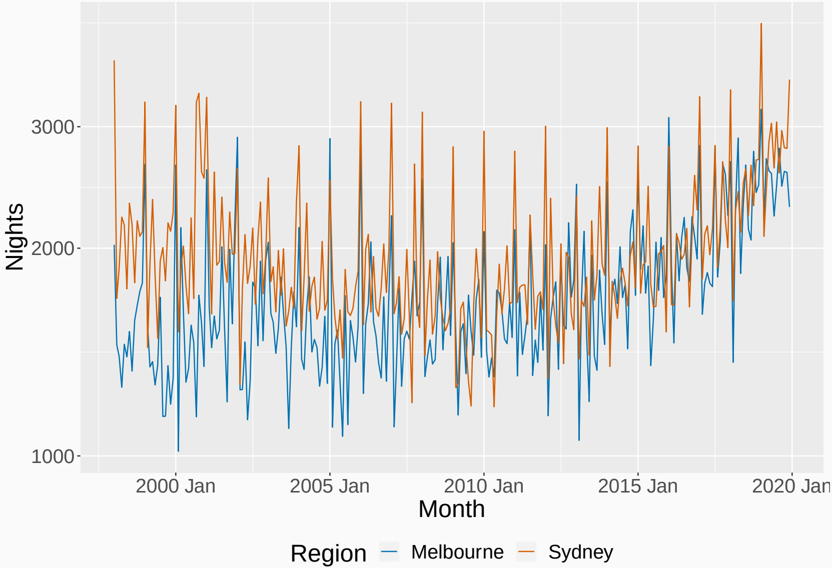
Australian tourism data

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

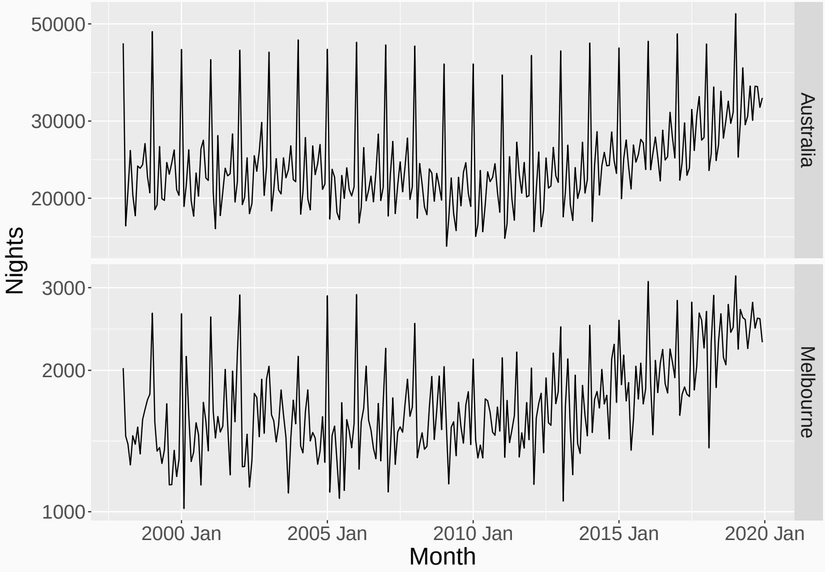
Visitor nights

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.

Literature

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

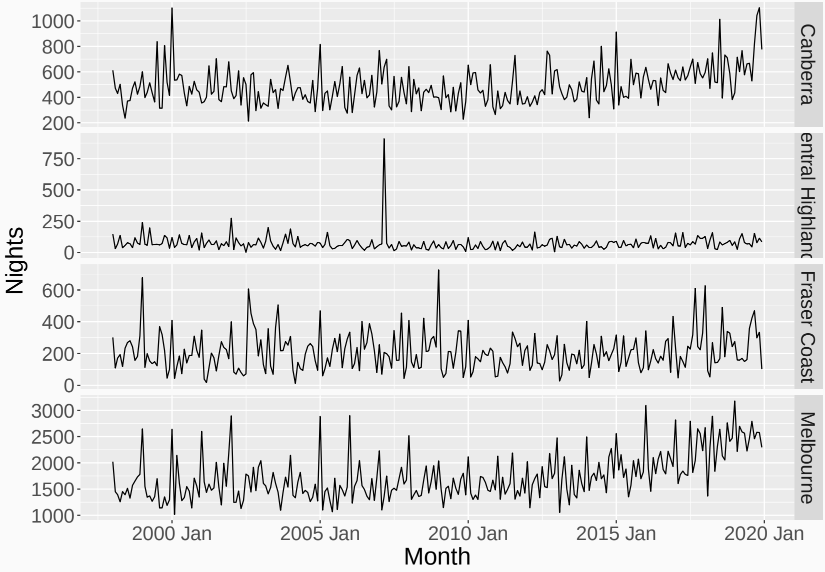
Bagging

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

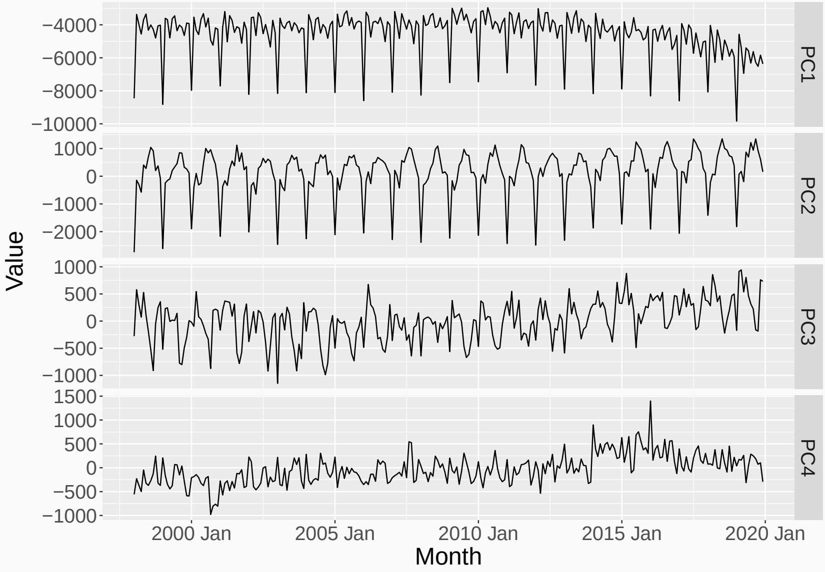
Dynamic factor model (DFM)

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

Series $\boldsymbol{z}_t \in \mathbb{R}^m$



Components $\boldsymbol{c}_t = \Phi \boldsymbol{z}_t \in \mathbb{R}^p$



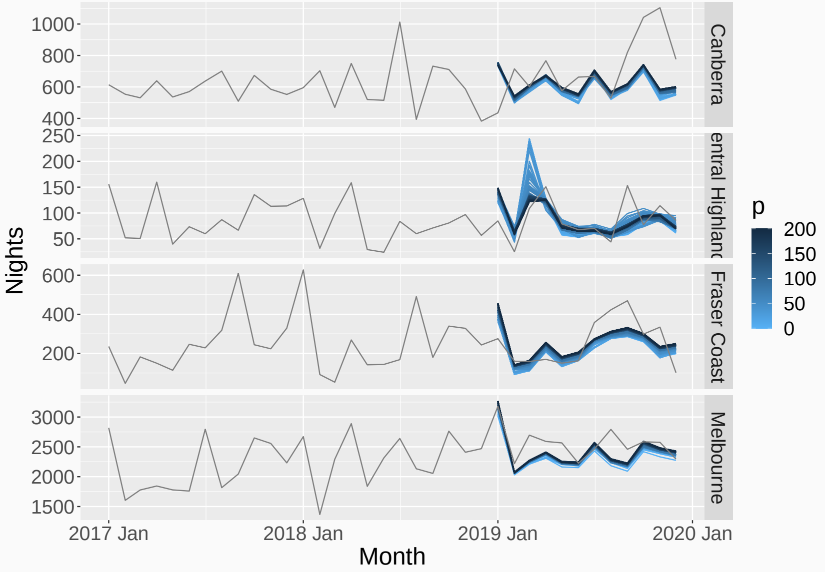
Free lunch forecast projection

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

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

$$\begin{aligned} \boldsymbol{M} &= \boldsymbol{I}_{m+p} - \boldsymbol{W}_h\boldsymbol{C}'\left(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}'\right)^{-1}\boldsymbol{C} \\ \boldsymbol{J} = \boldsymbol{J}_{m,p} &= \begin{bmatrix} \boldsymbol{I}_m & \boldsymbol{O}_{m\times p} \end{bmatrix} \\ \boldsymbol{C} &= \begin{bmatrix} -\Phi & \boldsymbol{I}_p \end{bmatrix} \\ \boldsymbol{W}_h &= \mathrm{Var}(\hat{\boldsymbol{y}}_{t+h}) \end{aligned}$$

Forecasts and projection of series



Unbiasedness

Unbiasedness

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

Projection matrix

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

Nonnegative variance reduction

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

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

Example $\boldsymbol{W}_h = \boldsymbol{I}_{m+p}$

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

Let Φ consist of orthogonal unit vectors:

$$\Phi\Phi' = \boldsymbol{I}_p \text{ when } p \leq m$$
$$\Phi'\Phi = \boldsymbol{I}_m \text{ when } p = m.$$

$$\begin{aligned}\mathrm{tr}(\mathrm{Var}(\hat{\boldsymbol{z}}_{t+h} - \boldsymbol{z}_{t+h}) - \mathrm{Var}(\tilde{\boldsymbol{z}}_{t+h} - \boldsymbol{z}_{t+h})) \\ = \frac{1}{2}tr(\Phi'\Phi) = \frac{1}{2}p\end{aligned}$$

Positive condition

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

$$\phi_1 \boldsymbol{W}_{z,h} \neq \boldsymbol{w}_{c_1 z,h},$$

- ϕ_1 is the weight vector of the first component
- $\boldsymbol{W}_{z,h} = \mathrm{Var}(\hat{\boldsymbol{z}}_{t+h})$
- $\boldsymbol{w}_{c_1 z,h}$ is the forecast covariance between the first component and the original series.

Monotonicity

The sum of forecast variance reductions

$$\begin{aligned}\operatorname{tr}\big(\operatorname{Var}(\hat{\boldsymbol{z}}_{t+h} - \boldsymbol{z}_{t+h}) - \operatorname{Var}(\tilde{\boldsymbol{z}}_{t+h} - \boldsymbol{z}_{t+h})\big) \\ = \operatorname{tr}\big(\boldsymbol{J}\boldsymbol{W}_h\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{J}'\big)\end{aligned}$$

is non-decreasing as p increases.

Minimum variance of individual series

The projection is equivalent to the mapping

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

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

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

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

Estimation

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

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

$$\boldsymbol{W}_h = \mathrm{Var}(\hat{\boldsymbol{y}}_{t+h})$$
$$\boldsymbol{C} = \begin{bmatrix} -\Phi & \boldsymbol{I}_p \end{bmatrix}$$

- Estimation of $\boldsymbol{W}_h$
- Construction of Φ

Estimation of $\boldsymbol{W}_h$

Shrinking variance towards their median (Opgen-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

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

Construction of Φ

Principal component analysis (PCA)

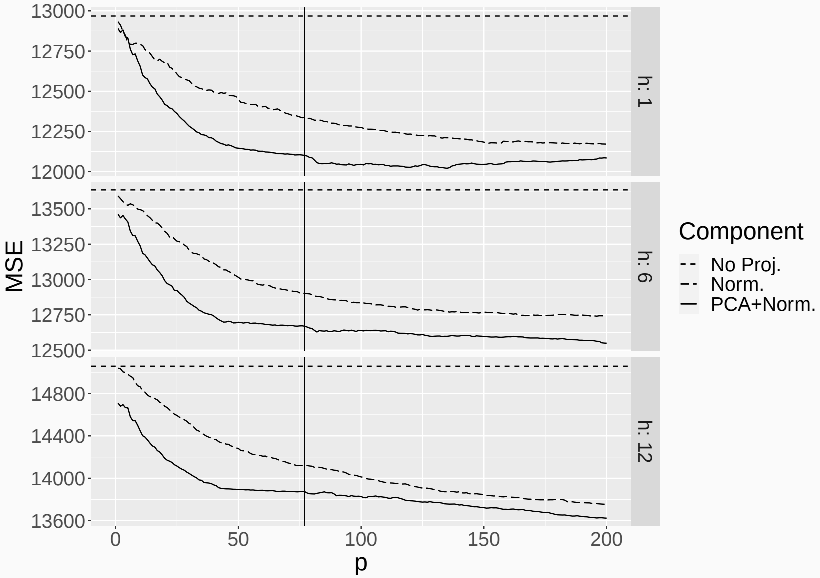
Finding the weights matrix so that the resulting components **maximise variance**

Simulation

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

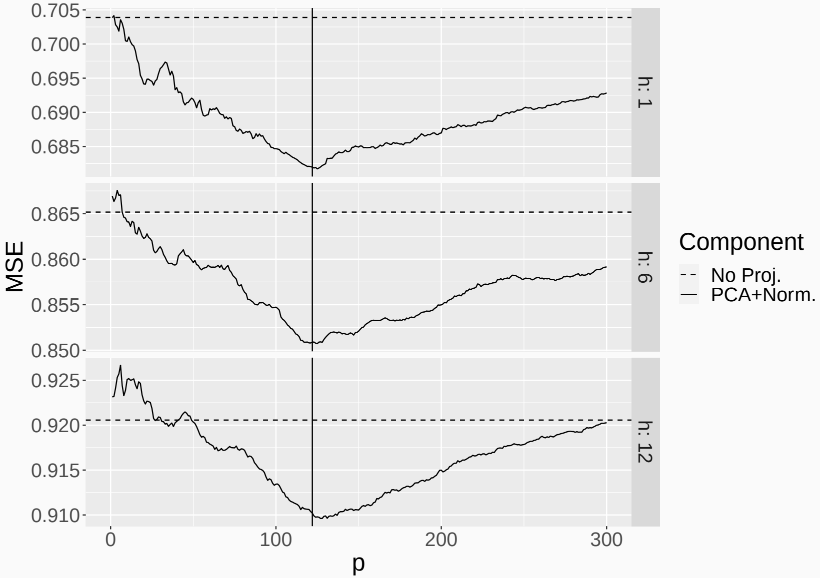
- Normal distribution
- Uniform distribution
- Orthonormal matrix (Borchers 2023)

Tourism (ETS)



MSE
13000
12750
12500
12250
12000
13500
13250
13000
12750
12500
14800
14400
14000
13600
0
50
100
150
200
p
h: 1
h: 6
h: 12
Component
No Proj.
Norm.
PCA+Norm.

FRED-MD (DFM)

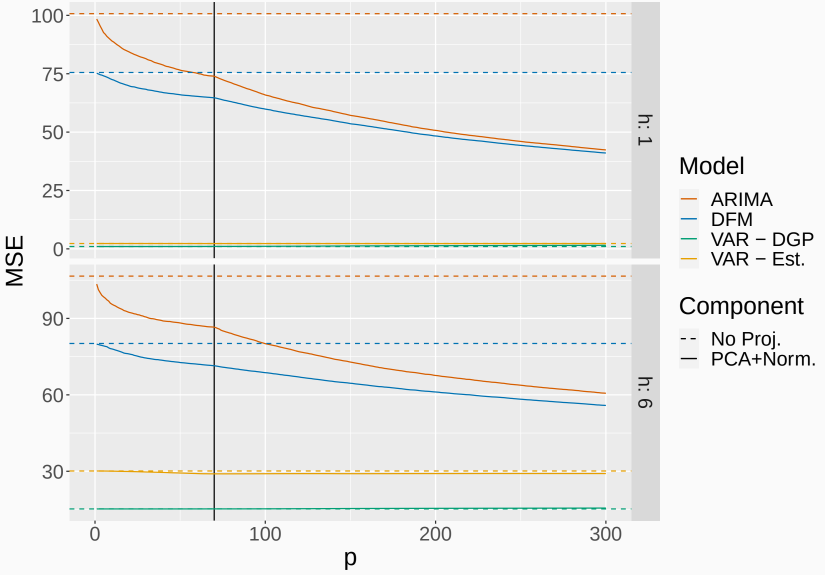


0.705
0.700
0.695
0.690
0.685
h: 1
0.865
0.860
0.855
0.850
h: 6
0.925
0.920
0.915
0.910
h: 12
MSE
0
100
200
300
p
Component
No Proj.
PCA+Norm.

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



100
75
50
25
0
90
60
30
MSE
0
100
200
300
p
h: 1
h: 6
Model
ARIMA
DFM
VAR − DGP
VAR − Est.
Component
No Proj.
PCA+Norm.

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 variance with respect to Φ
- Use forecast projection and forecast reconciliation together

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.