Cross-temporal coherent forecasts for Australian tourism

Nikolaos Kourentzes

Professor of Predictive Analytics

Lancaster University, UK

George Athanaspoulos

Professor of Econometrics and Business Statistics

Monash University, Australia

Tuesday, June 18th 2019

LUMS Marketing Analytics and Forecasting

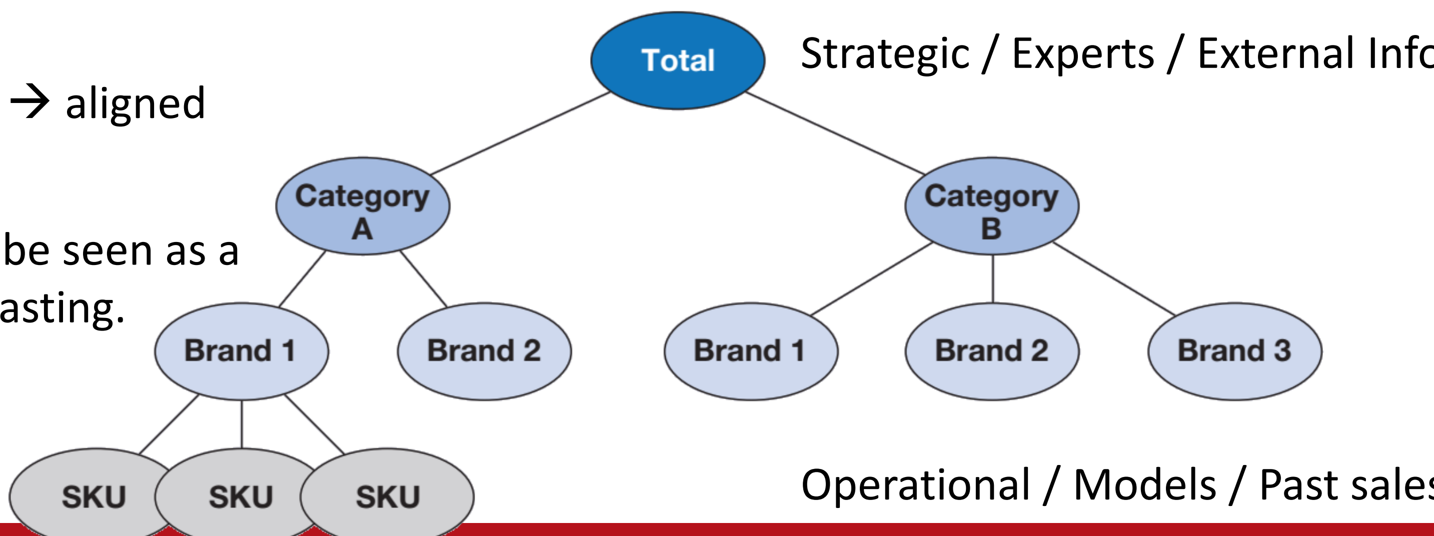
MONASH University

A classic business problem

- Companies rely on forecasts to support decision making at different levels and functions.

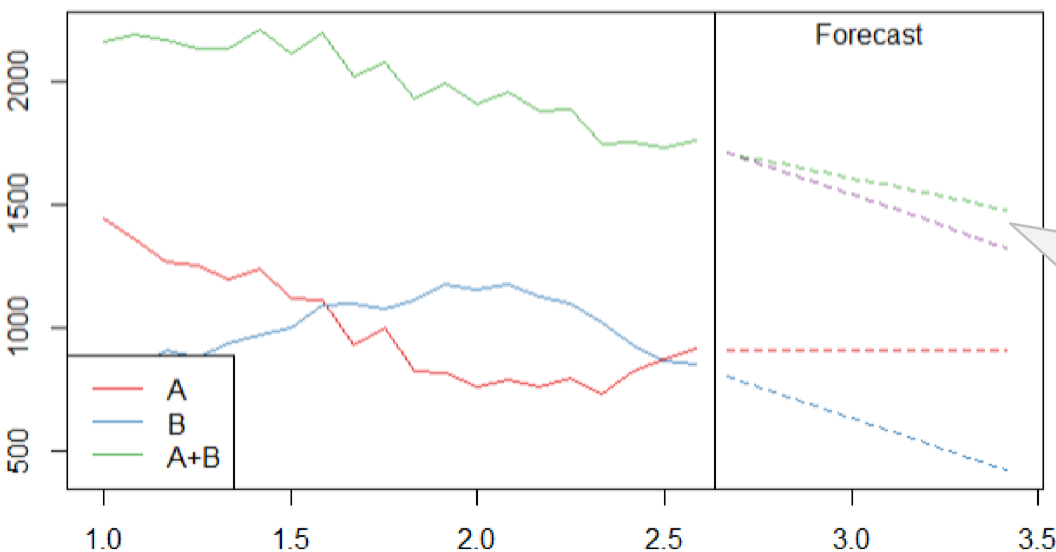
Level	Horizon	Scope	Forecasts	Methods	Information
Operational	Short	Local	Way too many	Statistical	Univariate/Hard
Tactical	Medium	Regional	↕	↕	↕
Strategic	Long	Global	Few expensive	Experts	Multivariate/Soft

- The challenge: Forecasts must be aligned.
- Aligned forecasts → aligned decisions.
- The problem can be seen as a hierarchical forecasting.



Coherent forecasts

- As we aggregate data, some structures become more prominent (trends, seasonality), while others become less obvious (promotional activity) and noise is filtered.
- Although all series are based on the same information, this does not mean that the same information is useable → different models/parameters/forecasts.
- Example: forecasting A and B separately or forecasting their sum does not lead to the same result!



F(A+B) and F(A)+F(B) will typically be different, we need to impose equality (coherency of forecasts).

F(A+B) or F(A)+F(B) is correct? Coherency avoids this question

Tourism forecasting

- For many countries the tourism sector contributes significantly to the economy (we are in Greece!), Australia is no exception (more than 3% of GDP).
  - Employees >500k people, ~5% national workforce;
  - Continuous investment to retain healthy sector, 2016-2017: $37.8bn.
- Classic hierarchical problem:
  - Split tourism flows by: geographies, type, market segments, etc.
  - Well researched for forecasting purposes (see Athanasopoulos et al., 2009).
  - Hierarchical forecasting provides accuracy gains and coherent forecasts across cross-sections.

Tourism forecasting

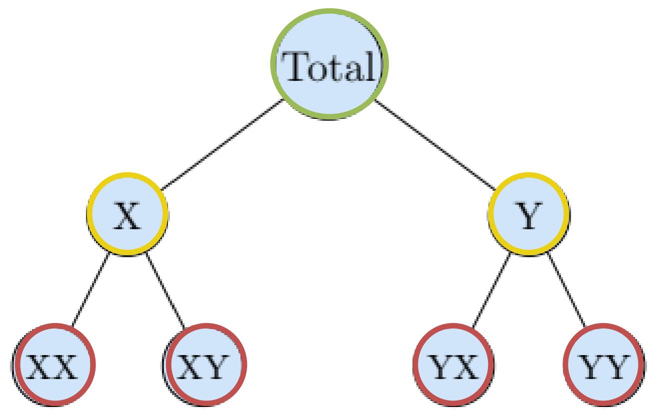
- Classic hierarchical problem. **However**:
  - Hierarchies are used as a statistical device, rather than to support decision making.
    - Consider the most disaggregate forecast: The fish tavern at the corner, does it need daily/weekly sales forecasts or 5-yearly?
    - Consider an aggregate forecast: We want to build a new hotel, as part of a major chain. Do we need daily/weekly forecasts of regional flows or 5-yearly?
    - Cross-sectional hierarchical forecasting offers **synchronous forecasts**, I.e. all levels are locked on the same forecast horizon and data frequency, although decisions for disaggregate and aggregate levels are not.
- Temporal hierarchies (Athanasopoulos et al, 2017 & Kourentzes et al., 2018) solve the complimentary: Go beyond forecast horizons and sampling frequencies, but are locked on the **unit of analysis**: still forecasting fish for the tavern at the corner.
- **To support decision making we need to be flexible on the unit and time of analysis, we need cross-temporal hierarchies.**

Hierarchical forecasting

The recently proposed **optimal combinations** recast the hierarchical problem in the following way – we recast the problem as a forecast reconciliation model (Hyndman et al., 2011).

- $\boldsymbol{b} = (y_{xx}, y_{xy}, y_{yx}, y_{yy})'$ Lower level series
- $\boldsymbol{y} = (y_{tot}, y_x, y_y, \boldsymbol{b}')'$ All series
- $\boldsymbol{y} = \boldsymbol{S}\boldsymbol{b}$ Mapping of lower to all

$$\boldsymbol{S} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ & \boldsymbol{I}_m & & \end{bmatrix} \begin{array}{l} \text{Top level} \\ \text{Middle level(s)} \\ \\ \text{Bottom level} \end{array}$$



- $\hat{\boldsymbol{y}}_h$ h-step ahead forecasts for $\boldsymbol{y}$, i.e. all series.

Then we can write: $\tilde{\boldsymbol{y}}_h = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}_h$

- where $\boldsymbol{G}$ projects (somehow!) linearly the forecasts to the lowest levels, so as to minimise $\tilde{\boldsymbol{y}}_h$- $\hat{\boldsymbol{y}}_h$, and $\tilde{\boldsymbol{y}}_h$ the coherent forecasts.

Hierarchical forecasting

The conventional **top-down** and **bottom-up** can be written as $\tilde{\boldsymbol{y}}_h = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}_h$, and in these cases $\boldsymbol{G}$ uses a single level, ignoring all other information available.

With optimal combinations:

- $\boldsymbol{G} = \left(\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}$, where $\boldsymbol{W}_h^{-1}$ is the variance-covariance matrix of h-step ahead errors.
- Given some forecasts $\hat{\boldsymbol{y}}_h$, obtained in any way, the only unknown to achieve coherent forecasts is $\boldsymbol{W}_h$.
- In principle, it should be the variance-covariance matrix of the reconciliation errors, but that poses a “chicken and the egg problem”. Wickramasuriya et al. (2018) showed that we can use the forecast errors instead (MinT reconciliation).

Hierarchical forecasting

The exact estimation of $\boldsymbol{W}_h$ is problematic:

- Obtaining h-step ahead forecast errors is computationally demanding and at times, depending on the available sample and forecasting approach, not feasible.
- The dimension of $\boldsymbol{W}_h$ can easily become very large, causing estimation problems.

We typically assume that $\boldsymbol{W}_h = k\boldsymbol{W}_1$, i.e. that it is proportional to the 1-step ahead errors of the variance-covariance matrix.

- Estimation of $\boldsymbol{W}_1$ is non-trivial, but there are several approximation methodologies that perform well (shrinkage being one of the most successful, Wickramasuriya et al., 2018).

Variance scaling

$$\begin{bmatrix} \hat{\sigma}^2_{Tot} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \hat{\sigma}^2_{X} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \hat{\sigma}^2_{Y} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \hat{\sigma}^2_{XX} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \hat{\sigma}^2_{XY} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \hat{\sigma}^2_{YX} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \hat{\sigma}^2_{YY} \end{bmatrix}$$

Structural scaling

$$\begin{bmatrix} 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

MinT shrinkage ($\rho_{i,j} \rightarrow 0$)

$$\begin{bmatrix} \hat{\sigma}^2_{Tot} & \hat{\rho}_{Tot,X} & \hat{\rho}_{Tot,Y} & \hat{\rho}_{Tot,XX} & \hat{\rho}_{Tot,XY} & \hat{\rho}_{Tot,YX} & \hat{\rho}_{Tot,YY} \\ \hat{\rho}_{X,Tot} & \hat{\sigma}^2_{X} & \hat{\rho}_{X,Y} & \hat{\rho}_{X,XX} & \hat{\rho}_{X,XY} & \hat{\rho}_{X,YX} & \hat{\rho}_{X,YY} \\ \hat{\rho}_{Y,Tot} & \hat{\rho}_{Y,X} & \hat{\sigma}^2_{Y} & \hat{\rho}_{Y,XX} & \hat{\rho}_{Y,XY} & \hat{\rho}_{Y,YX} & \hat{\rho}_{Y,YY} \\ \hat{\rho}_{XX,Tot} & \hat{\rho}_{XX,X} & \hat{\rho}_{XX,Y} & \hat{\sigma}^2_{XX} & \hat{\rho}_{XX,XY} & \hat{\rho}_{XX,YX} & \hat{\rho}_{XX,YY} \\ \hat{\rho}_{XY,Tot} & \hat{\rho}_{XY,X} & \hat{\rho}_{XY,Y} & \hat{\rho}_{XY,XX} & \hat{\sigma}^2_{XY} & \hat{\rho}_{XY,YX} & \hat{\rho}_{XY,YY} \\ \hat{\rho}_{YX,Tot} & \hat{\rho}_{YX,X} & \hat{\rho}_{YX,Y} & \hat{\rho}_{YX,XX} & \hat{\rho}_{YX,XY} & \hat{\sigma}^2_{YX} & \hat{\rho}_{YX,YY} \\ \hat{\rho}_{YY,Tot} & \hat{\rho}_{YY,X} & \hat{\rho}_{YY,Y} & \hat{\rho}_{YY,XX} & \hat{\rho}_{YY,XY} & \hat{\rho}_{YY,YX} & \hat{\sigma}^2_{YY} \end{bmatrix}$$

We may have made our lives more difficult…

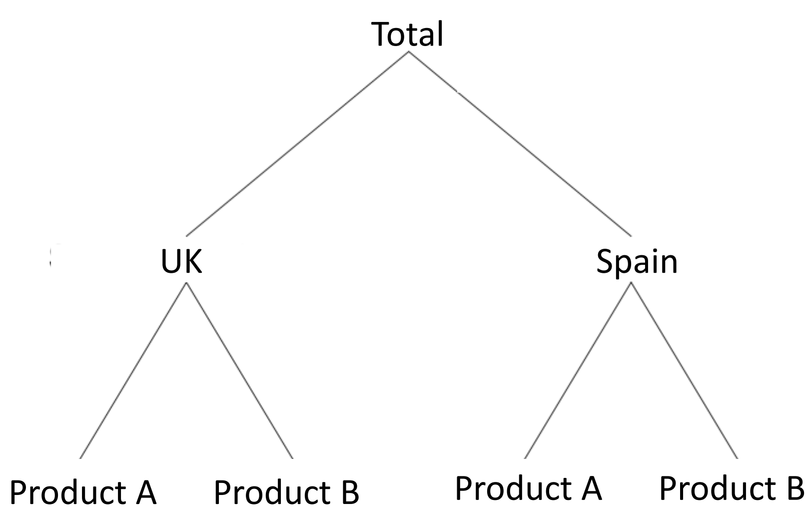
Sources of uncertainty in this problem?

- The classic ones:
  - Data – I don’t believe my observations, the world is inherently stochastic.
  - Model – I believe neither my forecasting method nor its parameters, impossible to get the true DGP.
- And a new one:
  - Model selection for forecast reconciliation, i.e. what is the best estimate of $\boldsymbol{W}_1$?
- On the upside, forecast reconciliation forces:
  - Forecast combination – this is typically helping accuracy and mitigating the risk of model selection (Kourentzes et al., 2019).
  - Makes explicit that we do not believe any of the forecasts, due to the classic sources of uncertainty. So, all and all, it’s a trade-off between the classic uncertainties and a “safer” one.

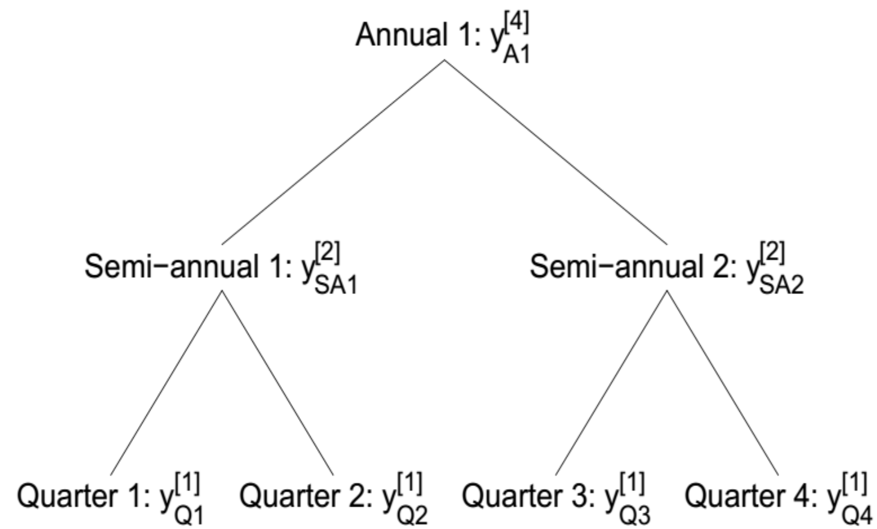
Temporal Hierarchies

Kourentzes et al. (2014) and Athanasopoulos et al. (2017) proposed the temporal analogue to hierarchical forecasting. The objective now is to join short-term and long-term forecasting.

Cross-sectional hierarchy



Temporal hierarchy



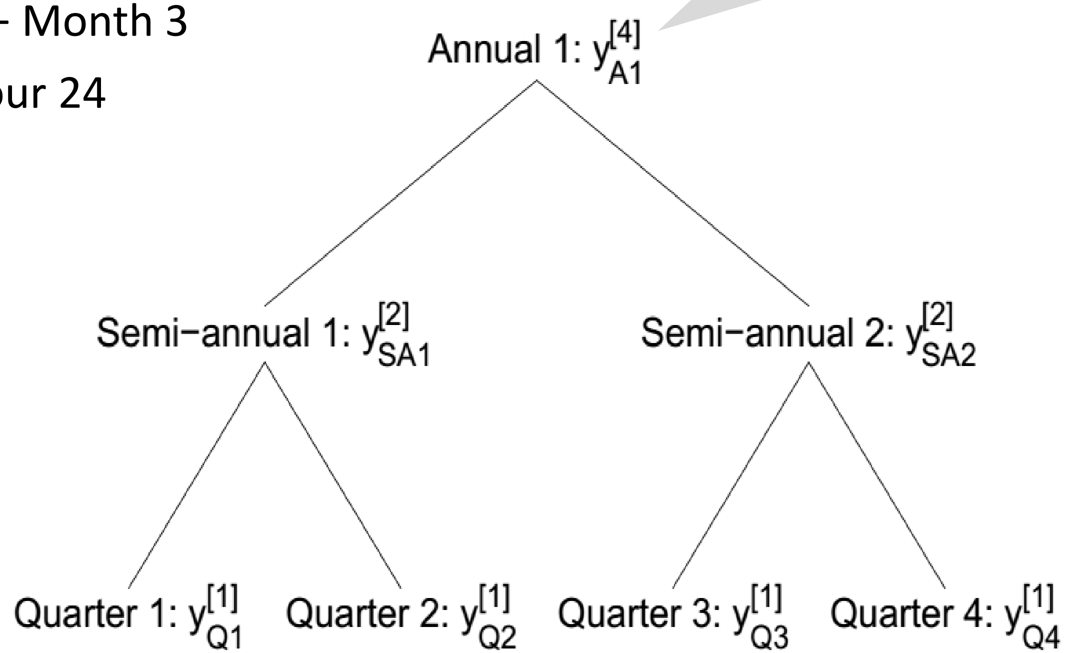
Temporal Hierarchies

Operational planning is done at detailed daily/weekly/monthly series, while tactical planning is done at monthly/quarterly/yearly series.

These series are naturally connected, for example:

- Year 1 = Quarter 1 + Quarter 2 + Quarter 3 + Quarter 4
- Quarter 1 = Month 1 + Month 2 + Month 3
- Day 1 = Hour 1 + Hour 2 + … + Hour 24
- etc.

Aggregate external information: e.g. macroeconomic

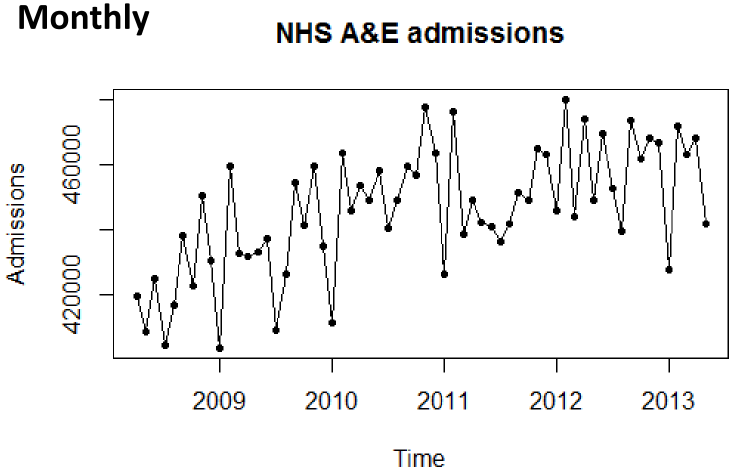


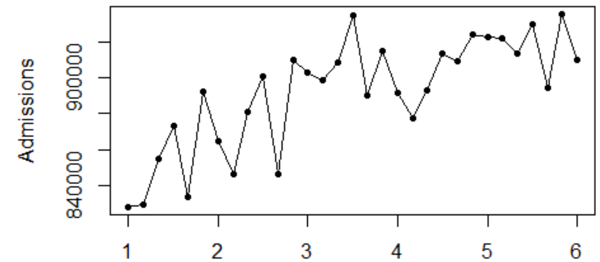
Disaggregate internal information: e.g. promotions

Temporal Aggregation

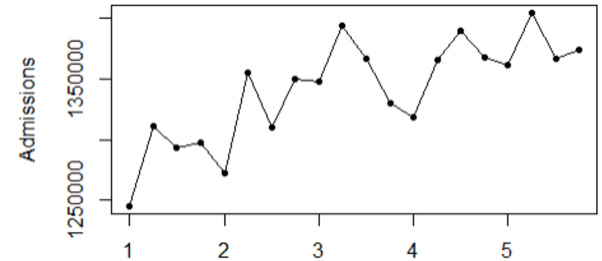
- Temporal aggregation filters high frequency components (e.g. seasonality), strengthening low frequency ones (e.g. trend)
- Reduces sample size, harming estimation efficiency.

Monthly

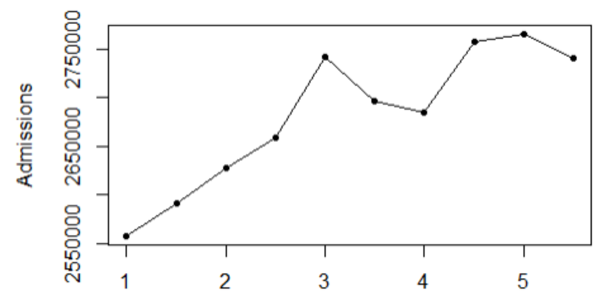




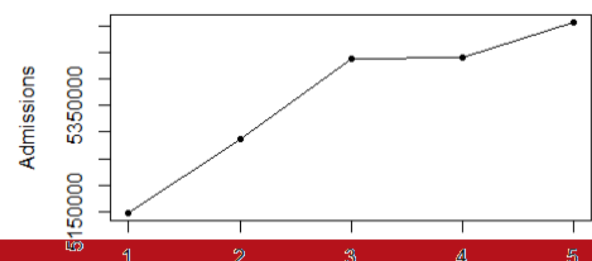
Bi-monthly



Quarterly

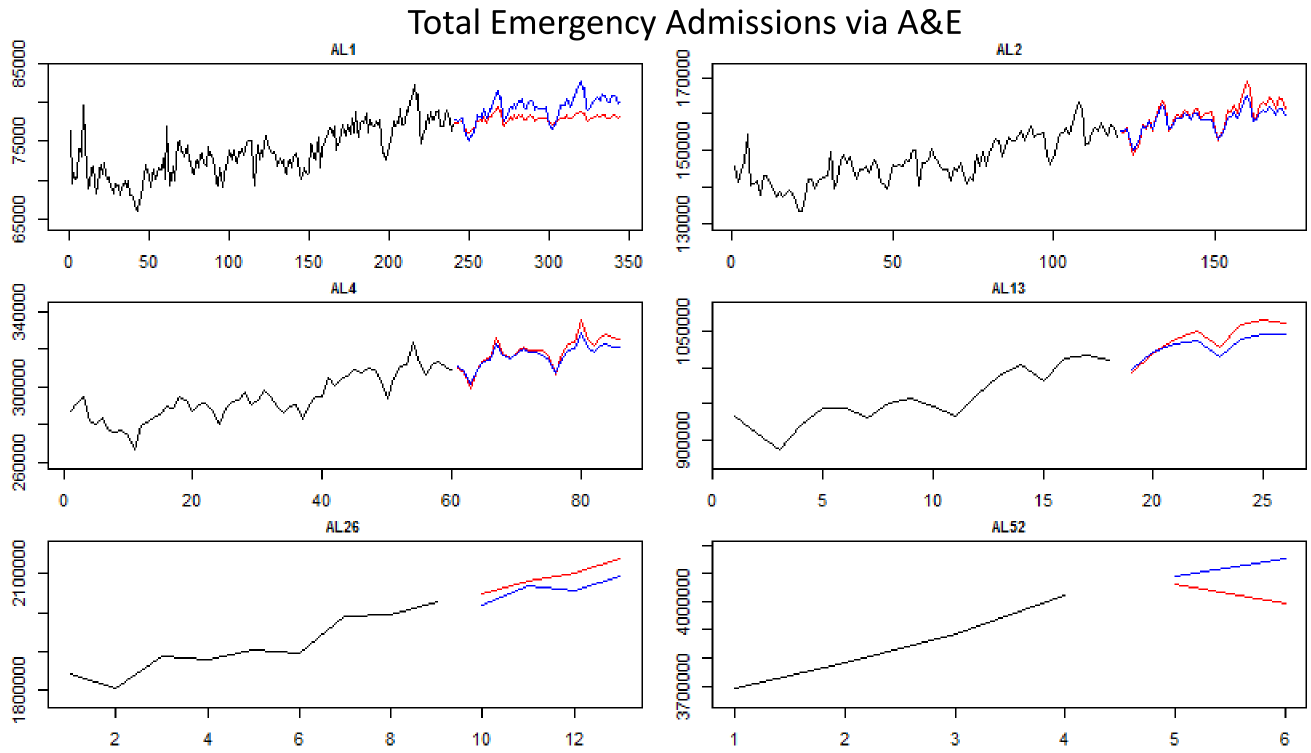


Half-annually



Annually

Example: Predicting A&E admissions



Red is the prediction of the base model – at each level separately
Blue is the temporal hierarchy forecasts

Observe how information is `borrowed' between temporal levels. Base models for instance provide very poor weekly and annual forecasts

Cross-Temporal Hierarchies

The two sides of hierarchical forecasting have limitations:

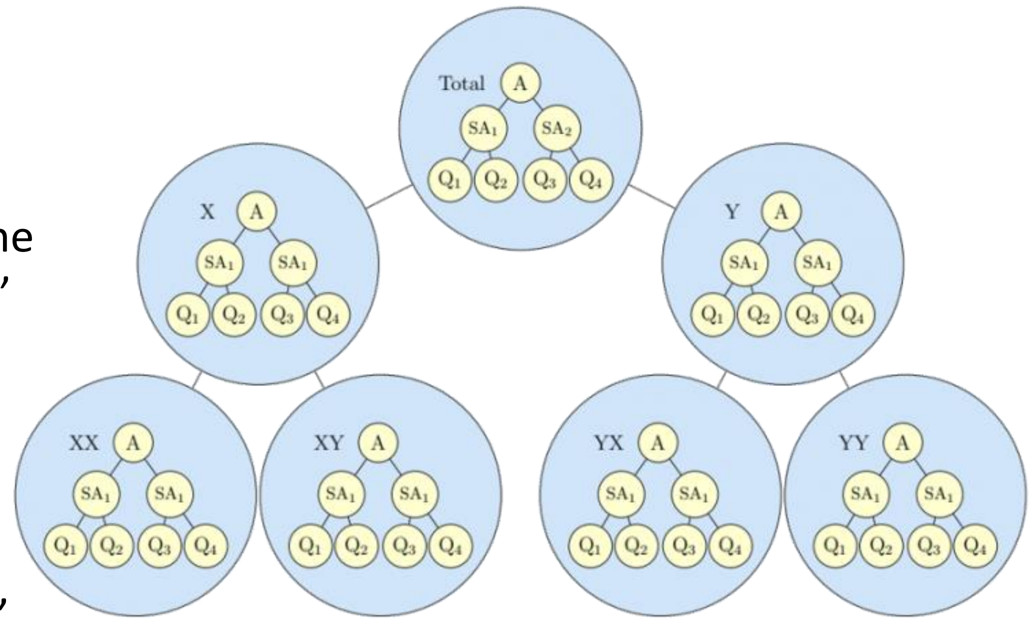
- Cross-sectional: is locked to the **time** of analysis;
- Temporal: is locked to the **unit** of analysis;
- So both are statistical devices to improve the forecasts, but are somewhat disjoint from decision making at different levels.

What we need is to combine both using **cross-temporal hierarchies**.

- Achieve coherency across units and time of analysis, the so called “one-number” forecast exists!

The same formulation applies.

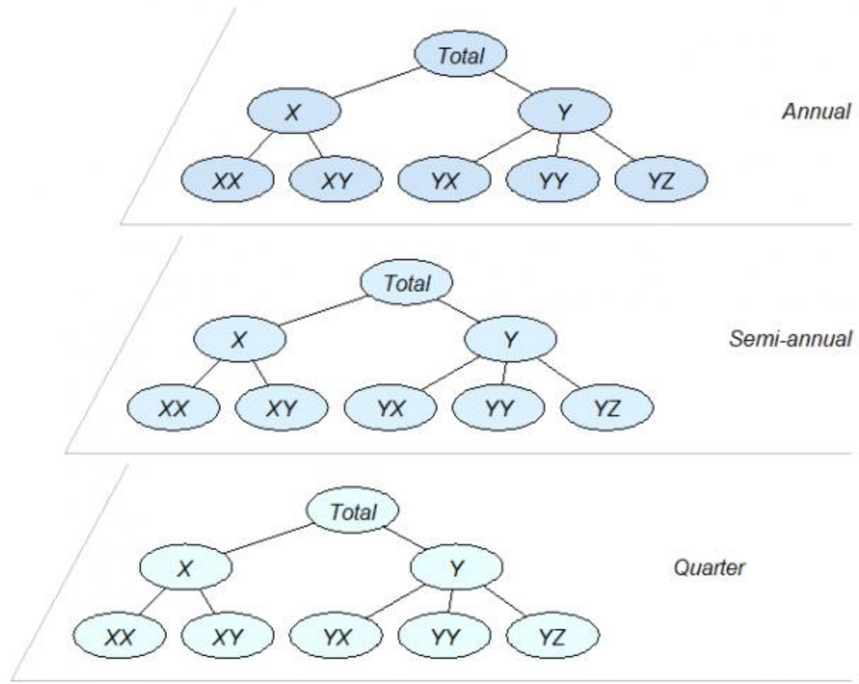
- But now constructing $\boldsymbol{S}$ and $\boldsymbol{G}$ is non-trivial and computationally demanding, because of their **dimensionality** and **non-unique mapping**.



Cross-Temporal Hierarchies

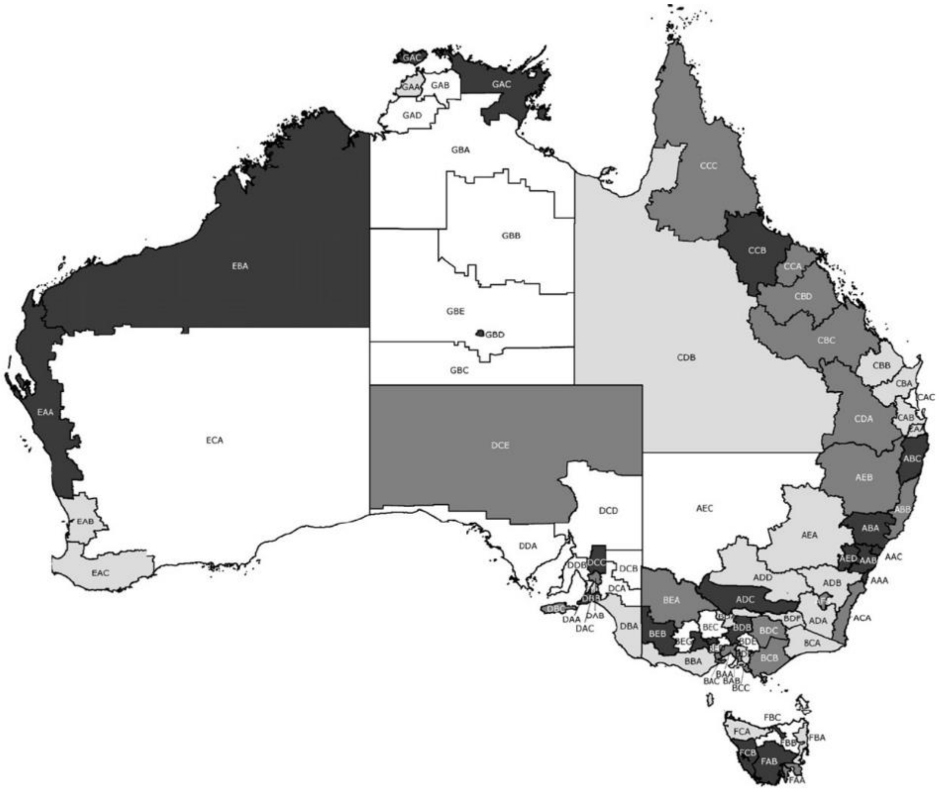
Kourentzes & Athanasopoulos (2019) proposes a methodology to split the estimation, reducing the size of the problem and still achieving cross-temporally coherent forecasts.

1. Produce temporal hierarchy forecasts for all time series in the cross-sectional hierarchy → **coherent in the time dimension**.
2. Estimate $\boldsymbol{G}$ at all temporal levels (as in figure!)
3. Calculate common $\overline{\boldsymbol{G}} = k^{-1}\sum_{=1}^{k} \boldsymbol{G}_k$, where k is the number of temporal levels.
4. Reconcile using $\widetilde{\boldsymbol{y}}_h = \boldsymbol{S}\overline{\boldsymbol{G}}\widehat{\boldsymbol{y}}_h$.



Empirical evaluation

- Total to regional monthly tourism flows for Australia. 111 series, spanning 10 years.
- Test set 6 years, with rolling origin evaluation. Relative RMSE (**<1 better**) to base forecast.
- Forecast using exponential smoothing. Results with ARIMA similar.

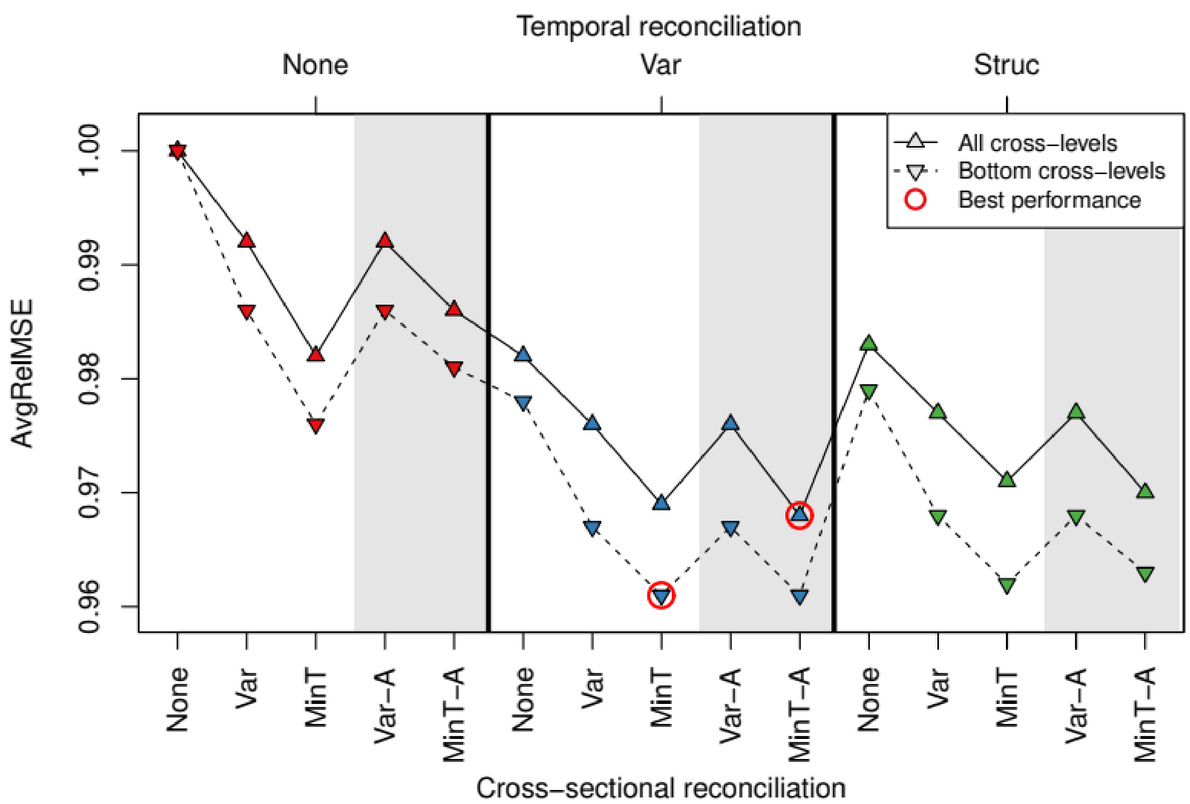


$$RelRMSE = \frac{RMSE_{Hierarchical}}{RMSE_{Base}}$$

$$AvgRelRMSE = \sqrt[n]{\sum_{i=1}^{n} RelRMSE}$$

n is the number of time series

Empirical evaluation

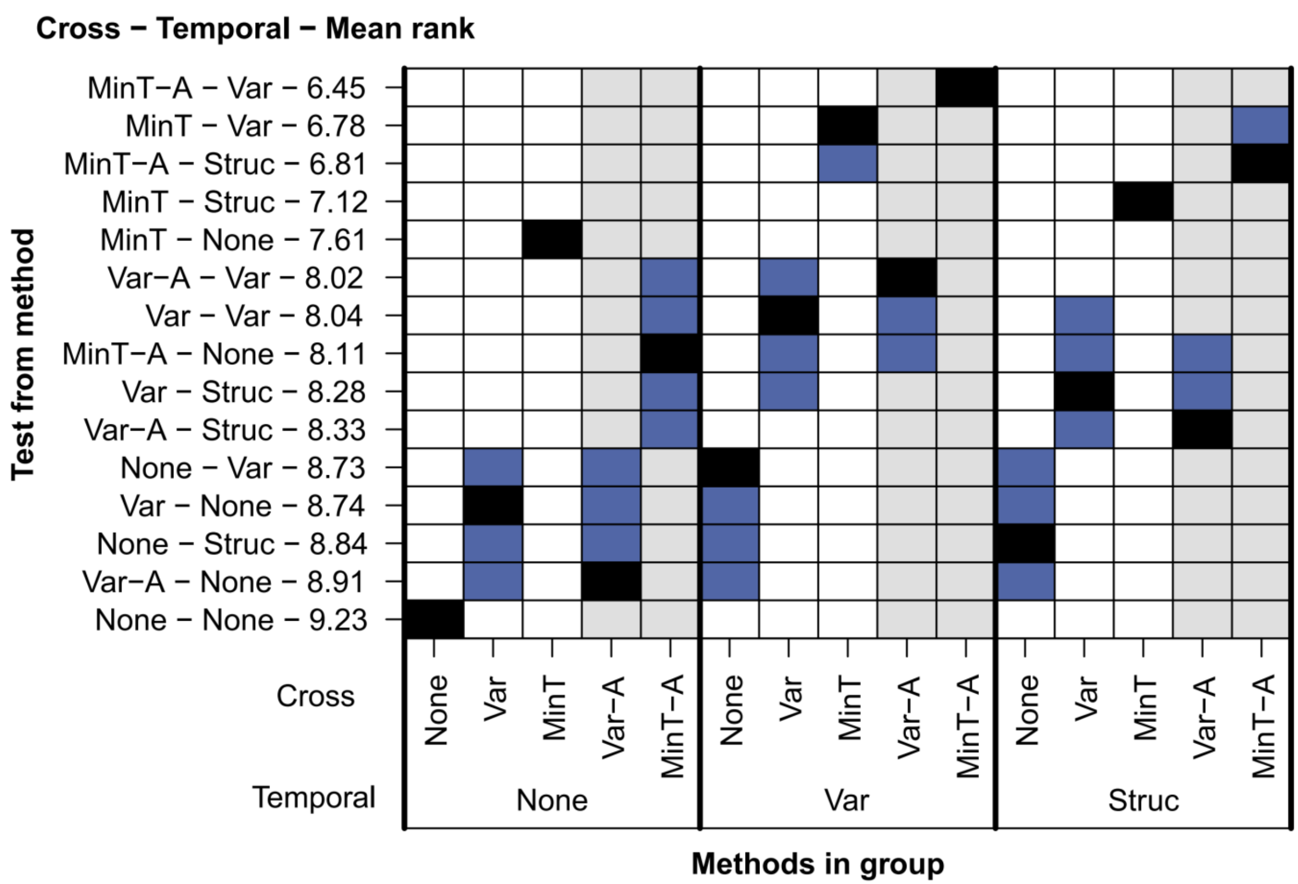


Figures in grey are cross-temporally coherent

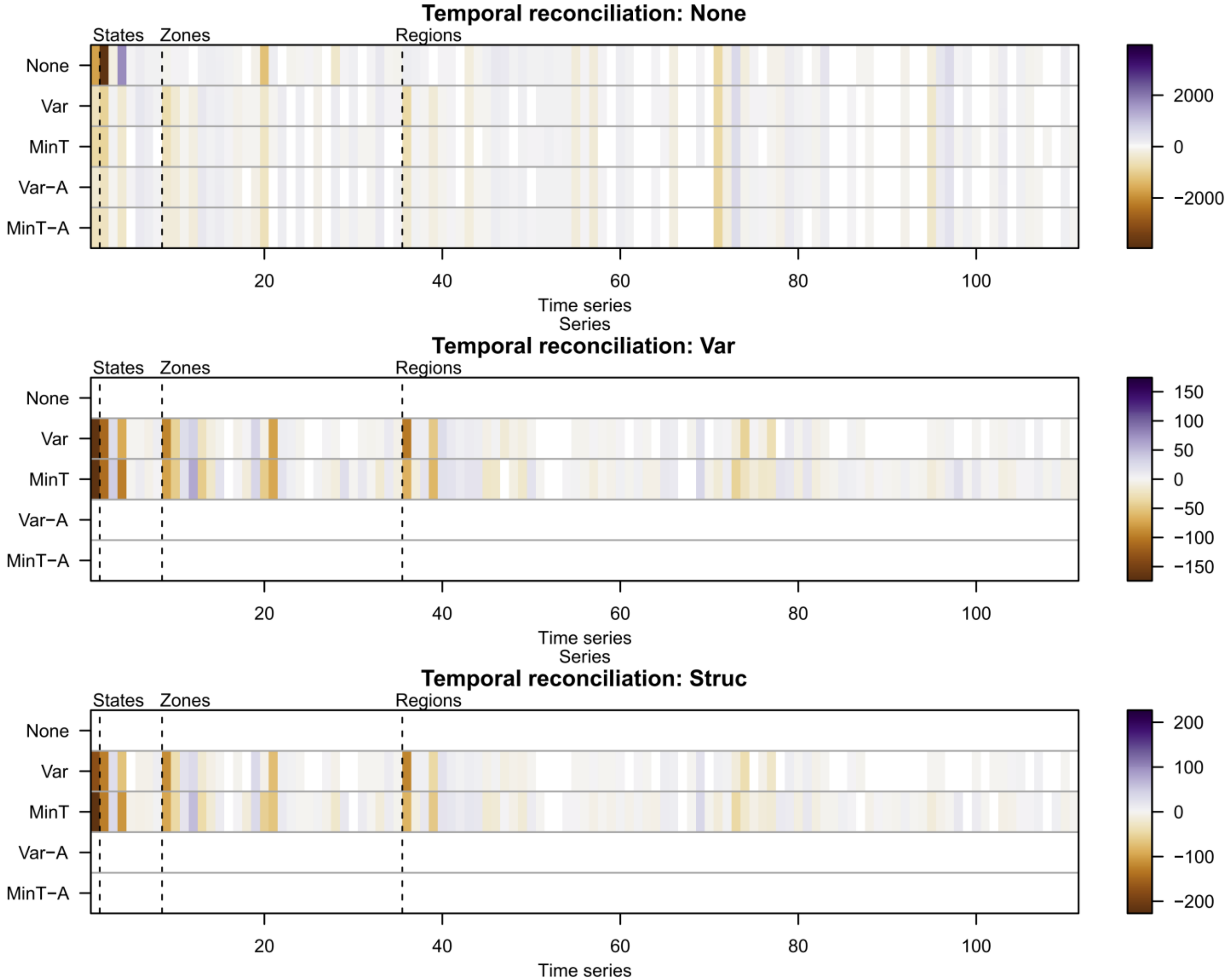
- Biggest accuracy gains from temporal hierarchies.
- Hierarchical forecasts always **more accurate** & and uni- or multi-dimensionally **coherent**.

Empirical evaluation

- Testing of statistical significance of accuracy differences using Friedman and Nemenyi non-parametric tests.



Reconciliation errors – single forecast origin



Conclusions

- Cross-temporal hierarchy forecasts provide a single view of the future across market demarcations and planning horizons → “one number forecast”.
- Cross-temporally coherent forecast come with accuracy gains. Demanding coherence constrains forecasts – forecasts cannot be wildly off and still coherent.
- Blending information from all levels of the organisation (or across organisations)
  - Breaking information silos between functions/organisations the “analytics way”.
  - From **operationalising strategies** to **informed strategies**: there is valuable information in operations, close to the customer, for top-management.
  - Collaboration: different companies can have common view of the future.
- Exciting applications!

Resources


Annals of Tourism Research 75 (2019) 393–409

Contents lists available at ScienceDirect

Annals of Tourism Research

journal homepage: www.elsevier.com/locate/annals

ELSEVIER

Cross-temporal coherent forecasts for Australian tourism

Nikolaos Kourentzes[a,*], George Athanasopoulos[b]

[a] Lancaster University Management School, Department of Management Science, Lancaster LA1 4YX, UK
[b] Department of Econometrics and Business Statistics, Monash University, Australia

ARTICLE INFO

Associate editor: Haiyan Song

Keywords:
Cross-sectional aggregation
Temporal aggregation
Forecast combinations
Spatial correlations

ABSTRACT

Key to ensuring a successful tourism sector is timely policy making and detailed planning. National policy formulation and strategic planning requires long-term forecasts at an aggregate level, while regional operational decisions require short-term forecasts, relevant to local tourism operators. For aligned decisions at all levels, supporting forecasts must be 'coherent', that is they should add up appropriately, across relevant demarcations (e.g., geographical divisions or market segments) and also across time. We propose an approach for generating coherent forecasts across both cross-sections and planning horizons for Australia. This results in significant improvements in forecast accuracy with substantial decision making benefits. Coherent forecasts help break intra- and inter-organisational information and planning silos, in a data driven fashion, blending

- References within the published paper.
- Useful R packages for cross-temporally coherent forecasts
  - thief – Temporal hierarchies;
  - hts – Cross-sectional hierarchies;
  - MAPA - alternative for temporally coherent forecasts.

Thank you for your attention!
Questions?

Nikolaos Kourentzes
email: nikolaos@kourentzes.com
twitter @nkourentz
Blog: http://nikolaos.kourentzes.com