# Do we believe our forecasts?

## From simple forecasting tricks to a holistic view of the future

### Nikolaos Kourentzes

Professor of Predictive Analytics

Lancaster University Management School, UK

July 7th 2019

Marketing Analytics
and Forecasting

LUMS

# Let's start with the basics!

1.  All forecasts are wrong, it is only a matter of "what type of wrong"!
    *   **Action**: we need to understand what makes a forecast wrong.

2.  All forecasting models/methods are mere approximations of some unknown underlying demand generating process.
    *   **Action**: we need to evaluate the quality of the approximation – this cannot be untangled from the forecast objective.

3.  We do not forecast for the sake of forecasting – please take a moment to appreciate this is painful for an academic to say - **we forecast to support decisions**.
    *   **Action**: understand the decisions and their context!

*   In all of the above there is the implicit question of what is an appropriate criterion of "goodness" for forecasts.
    *   Accurate? (what does it mean, how to measure?)
    *   Profitable? (what does it mean, how to measure?)
    *   ???

# Forecasting & uncertainty

The great thing about forecasting is that you will get it wrong and that is fine → most probable outcome!
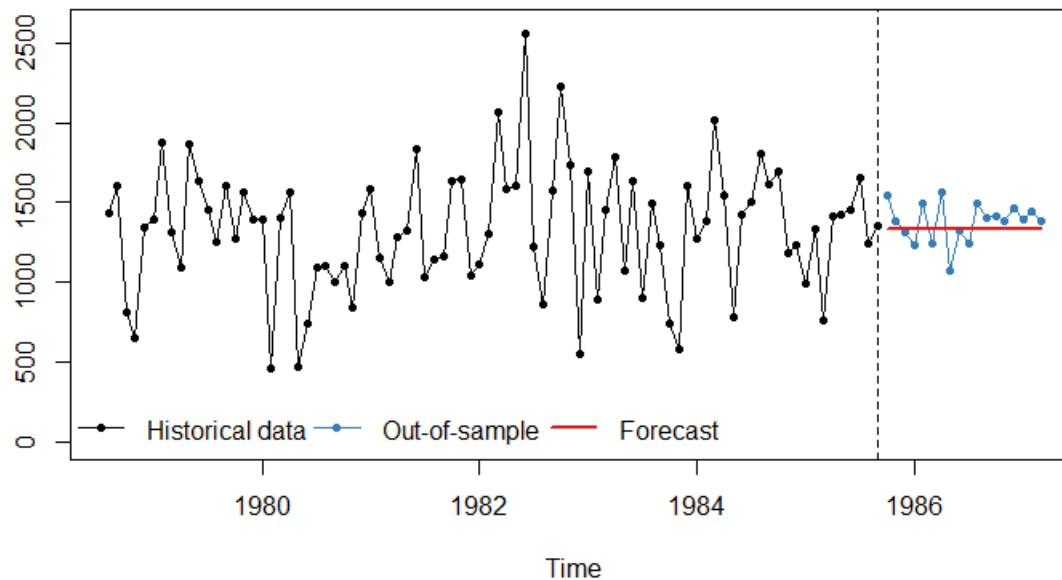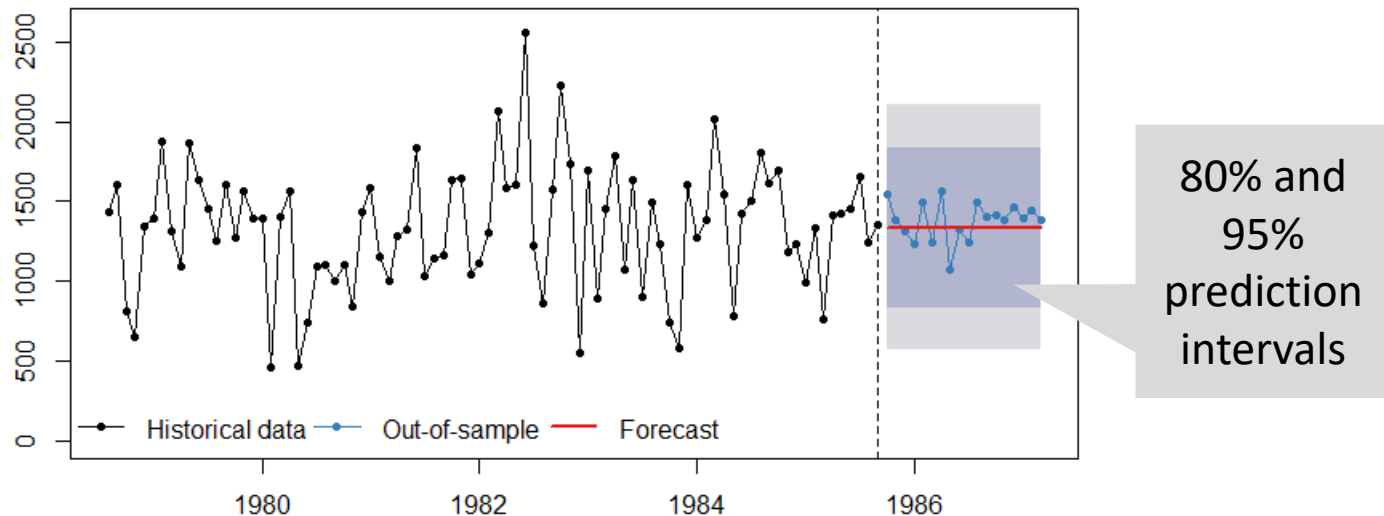


All forecasts come with uncertainties

- We try to identify and manage these uncertainties

# Forecasting & uncertainty

One of the most important jumps in forecasting as a discipline has been the move to accompany forecasts with an explicit representation of uncertainty.



80% and 95% prediction intervals

Better forecasts:

- Correspond to lower uncertainty – but this uncertainty much correspond to the observed uncertainty, not some expression based on hopeful assumptions!

- How to get better forecasts? Incorporate more information (≠ more complex models).

# Forecasting & uncertainty

Given some historical data, a forecast will attempt to capture the key patterns in the data and extrapolate these to the future.

+ Marketing actions
+ Macroeconomic indicators
+ Online behaviour of customers
+ etc.
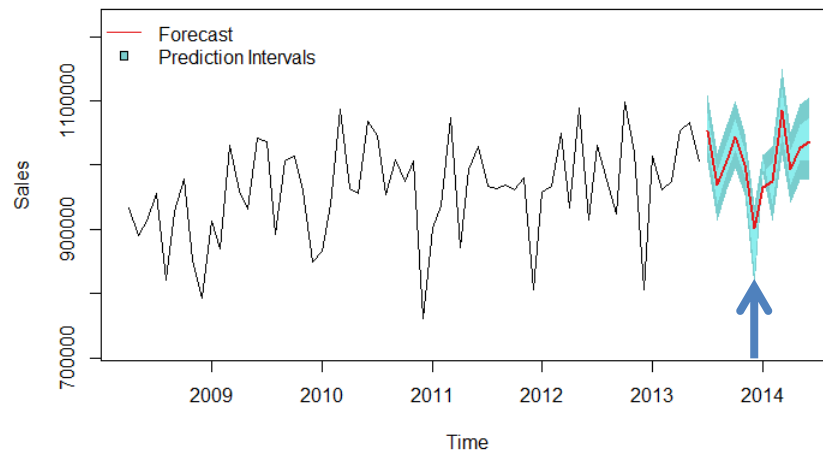


Related to the cost of the associated decision

- In fact, what we care about is the uncertainty of the forecast
  → the forecast will never be spot on, **the world is stochastic**!

- The forecast can be enhanced with **additional explanatory (causal?) information**.

- Superimposed with **managerial judgment** to account for soft information.
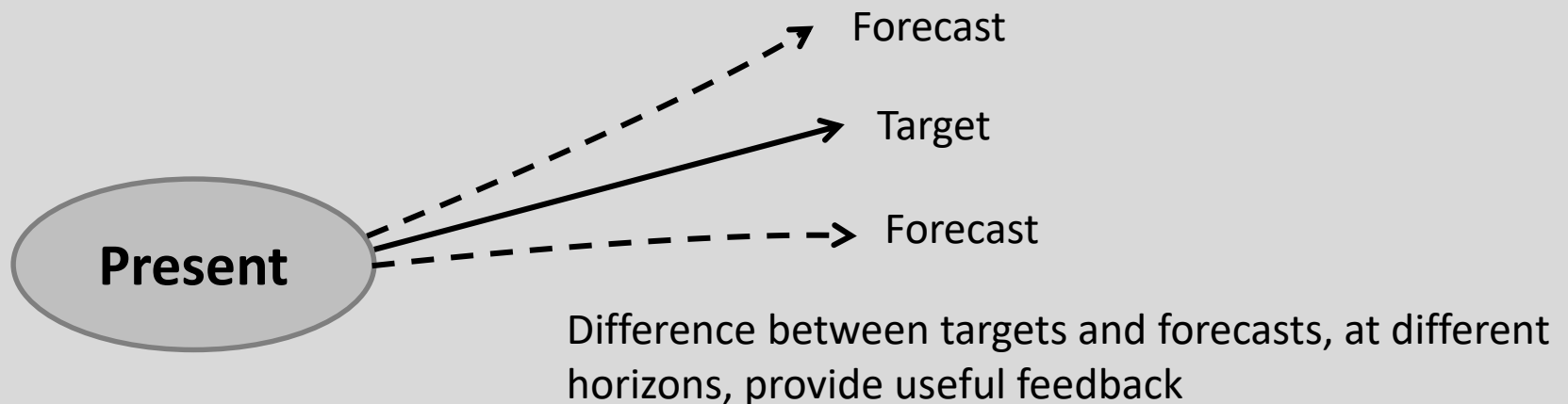
# Reduced uncertainty and decision making

Decision making in organisations has at its core an element of forecasting

  → Accurate forecasts lead to reduced uncertainty → better decisions

  → Forecasts maybe implicit or explicit

Forecasts aims to provide information about the future, conditional on historical and current knowledge

Company targets and plans aim to provide direction towards a desirable future.

**Present**

Forecast

Target

Forecast

Difference between targets and forecasts, at different horizons, provide useful feedback

# 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.



Strategic / Experts / External Info

Operational / Models / Past sales
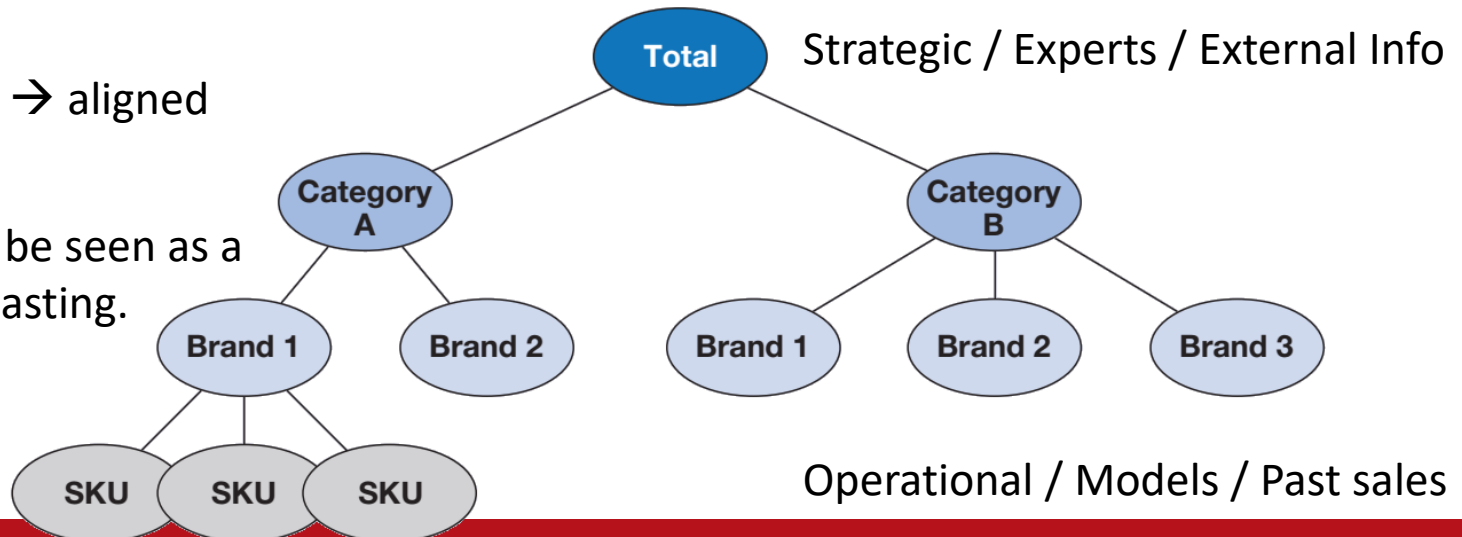
# Three curious cases: How to look at your data?

- We know that different forecasting models are better for different forecast horizons

- We also know that it helps to forecast long horizons using aggregate data
  - → Forecasting a quarter ahead using daily data is `adventurous' (90 steps ahead)
  - → Forecasting a quarter ahead using quarterly data is easier (1 step ahead)

- At different data frequencies different components of the series dominate.
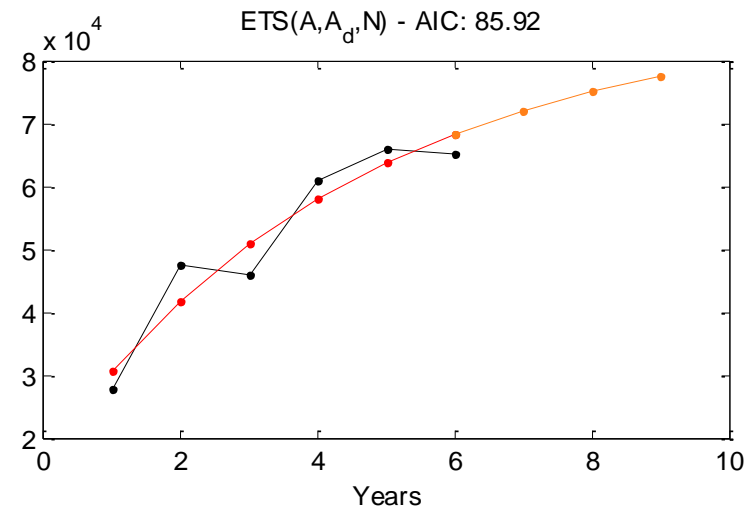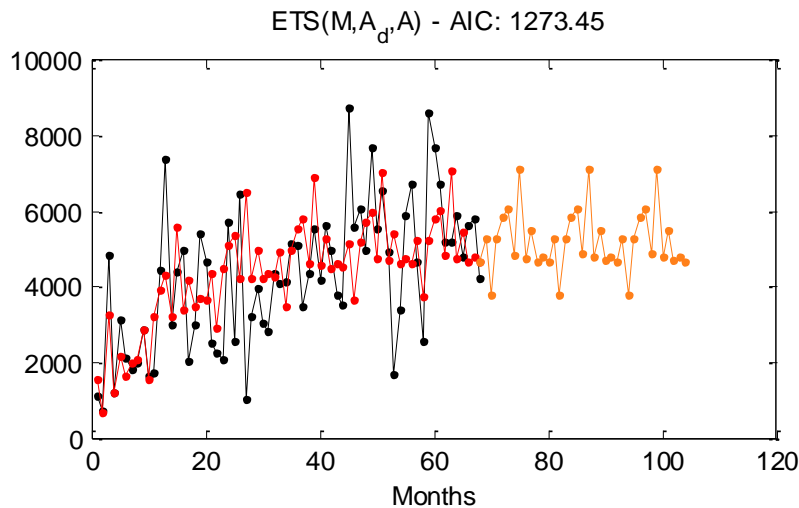


These forecasts often do not agree, which one is `correct'?

# Three curious cases: Do we trust past data?

Things can go badly wrong in model parametrisation and selection:

- Business time series are often short → Limited data per SKU;

- Estimation of parameters can fail miserably (more parameters → over-fitting);

- Model selection can fail as well (choose from many models → over-fitting?);

- Both optimisation and model selection are myopic → Focus on data fitting in the past, rather than '*forecastability*'.
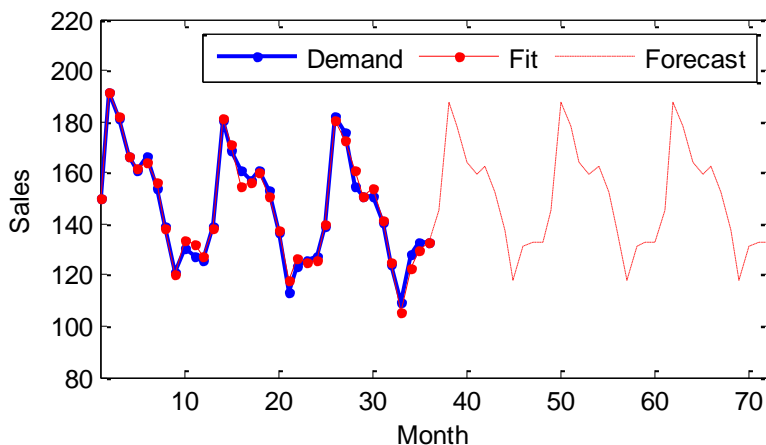
**True model**:
Additive trend, additive seasonality

**Identified model**:
No trend, additive seasonality

**Why?**
In-sample variance explained mostly by seasonality

# Three curious cases: How stable is the forecast?

Issues with automatic modelling **over time**:

- Model selection → How good is the best fit model? How reliable?
- Sampling uncertainty → Identified model/parameters stable as new data appear?
- Model uncertainty → Appropriate model structure and parameters?
- Transparency/Trust → Do we trust forecasts that change substantially?



$ETS(M,A_d,A)$ - AIC: 1273.45

$ETS(A,A_d,M)$ - AIC: 1262.33
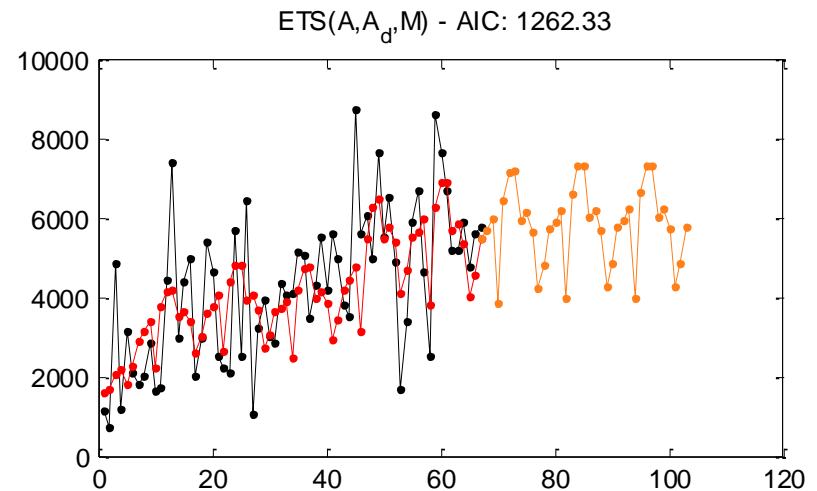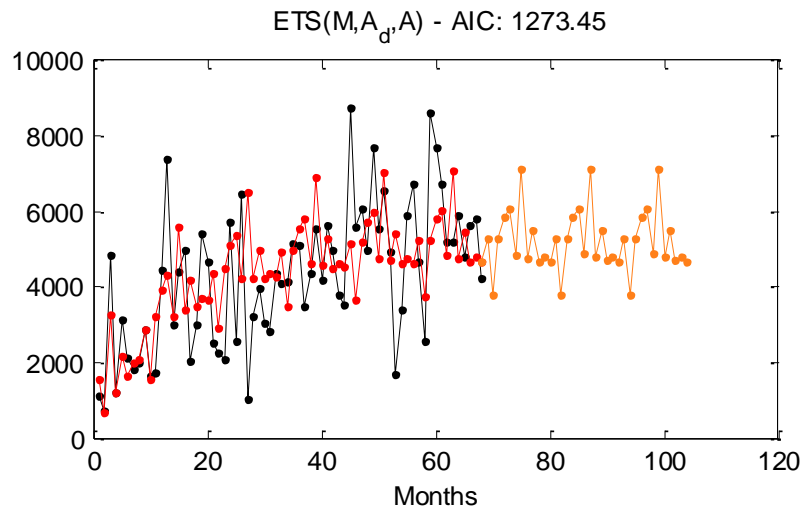
# A trick to take advantage of uncertainties

Instead of contemplating the cost of these uncertainties, let us take advantage of the forecast disagreements they introduce.

- For any time series we can artificially construct a **temporal hierarchy**.

  - Look at the series at aggregate views as well

    - But, forecasts will differ ☹

  - Great, I didn't believe my forecasts anyways We take advantage of this disagreements to reach a better consensus forecast! ☺

  - Incorporate different information at different views!

Aggregate external information: e.g. macroeconomic

Annual 1: $y_{A1}^{[4]}$

Semi-annual 1: $y_{SA1}^{[2]}$

Semi-annual 2: $y_{SA2}^{[2]}$
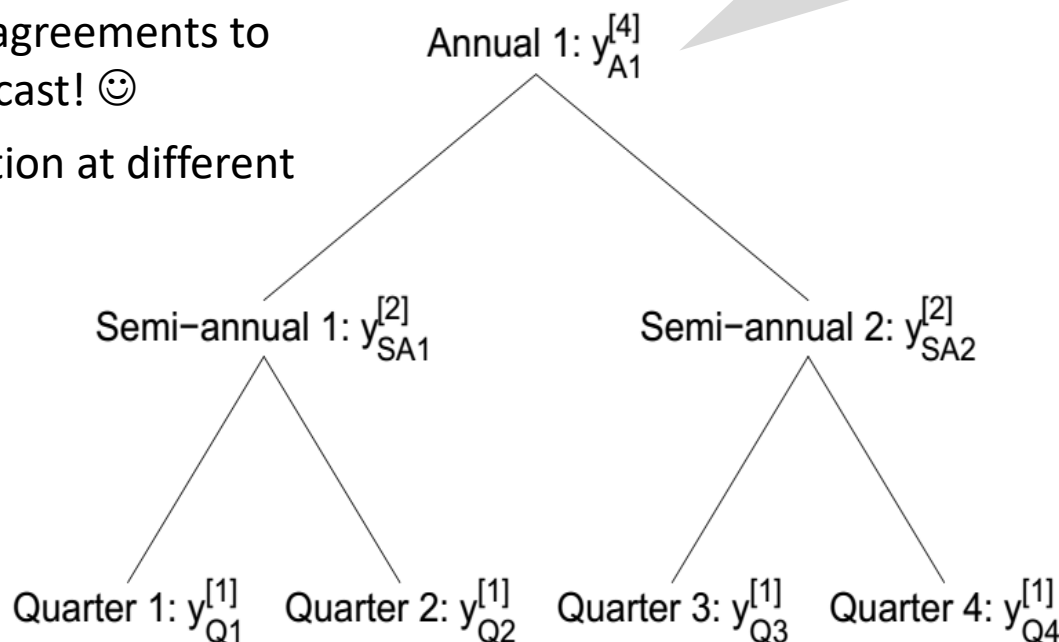
Disaggregate internal information: e.g. promotions

Quarter 1: $y_{Q1}^{[1]}$    Quarter 2: $y_{Q2}^{[1]}$    Quarter 3: $y_{Q3}^{[1]}$    Quarter 4: $y_{Q4}^{[1]}$

# Temporal Aggregation

- Temporal aggregation filters high frequency components (e.g. seasonality), strengthening low frequency ones (e.g. trend)

- Reduces sample size, harming estimation efficiency.
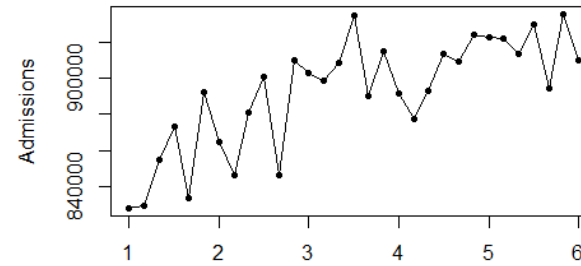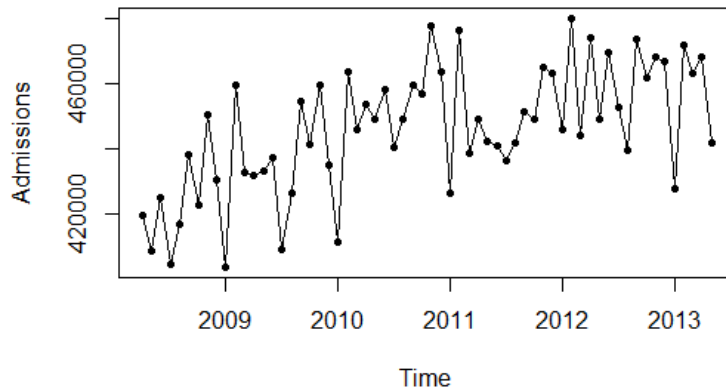
**Bi-monthly**
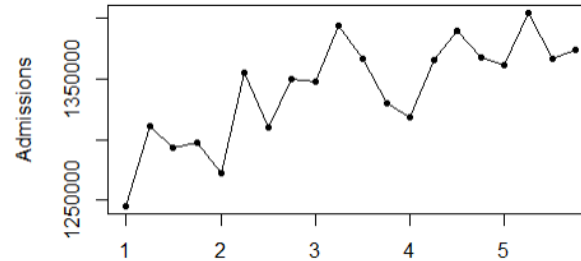
**Quarterly**

**Monthly**
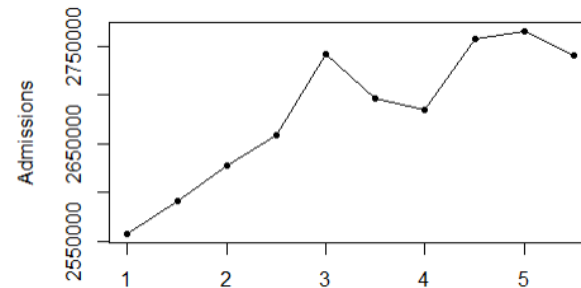
**Half-annually**

**Annually**



NHS A&E admissions
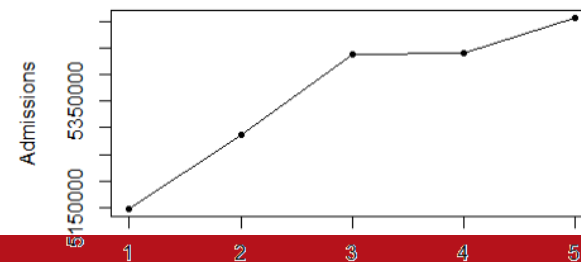
# Temporal Hierarchies

As it happens we know the maths how to solve Temporal Hierarchies. They are the same* as the ones for the well known cross-sectional hierarchical forecasting problem.

Cross-sectional hierarchy
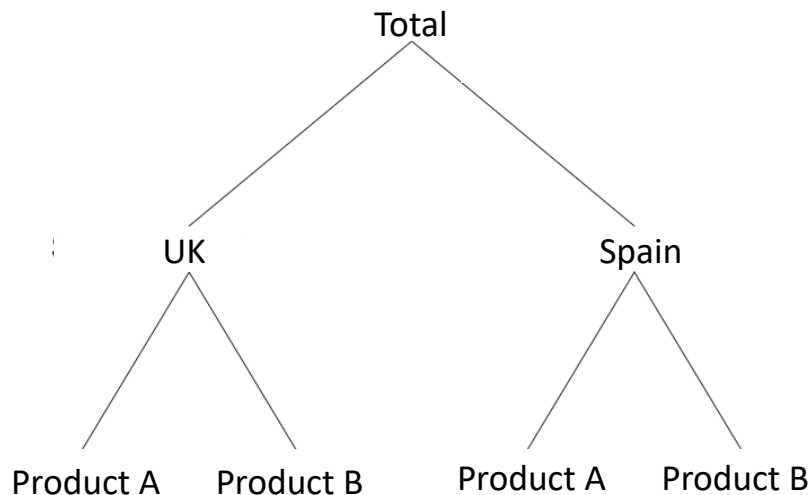
Temporal hierarchy
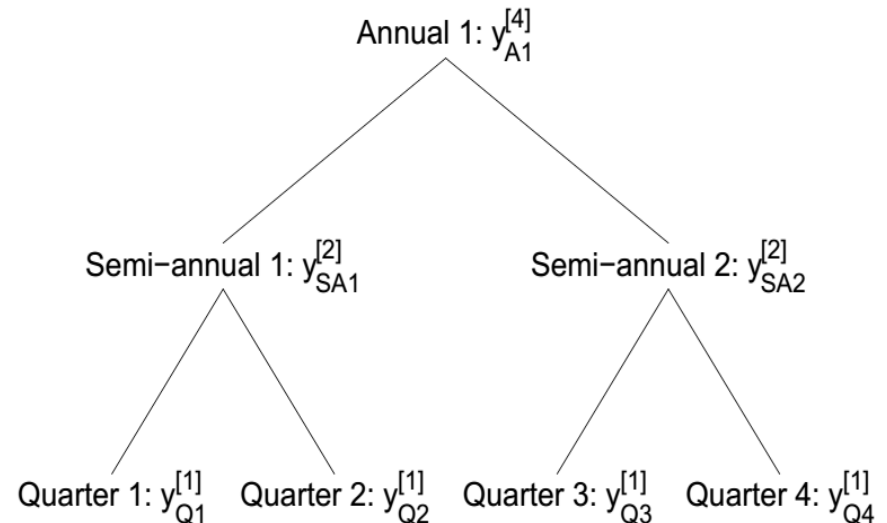


* Terms & conditions apply! Issues with sample size and definition/estimation of covariance matrices, ask me for the details.

# Example: Predicting A&E admissions
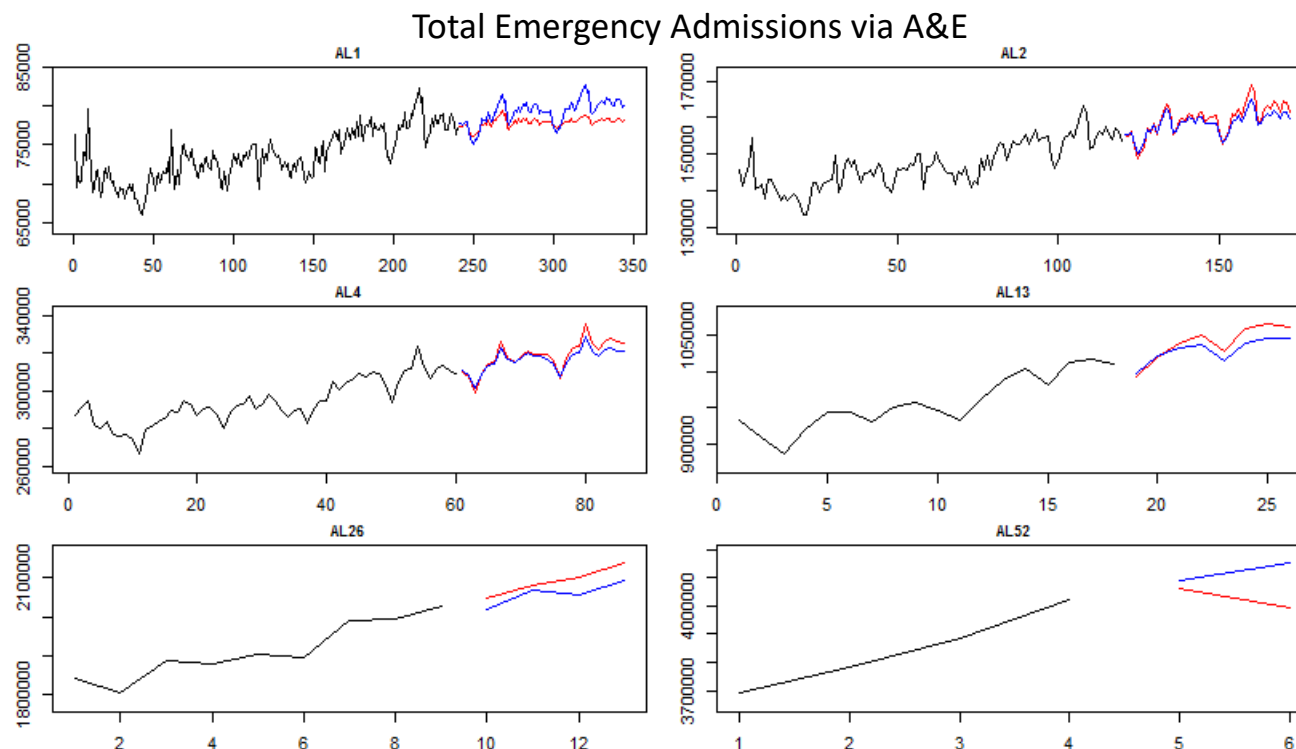


Total Emergency Admissions via A&E

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

# Example: Predicting A&E admissions

| Aggr. Level | h | Base | Reconciled | Change |
|---|---|---|---|---|
| Weekly | 1 | 1.6 | 1.3 | -17.2% |
| Weekly | 4 | 1.9 | 1.5 | -18.6% |
| Weekly | 13 | 2.3 | 1.9 | -16.2% |
| Weekly | 1-52 | 2.0 | 1.9 | -5.0% |
| Annual | 1 | 3.4 | 1.9 | -42.9% |

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

- ARIMA forecasts; MASE accuracy metric; Rolling evaluation over 52 weeks.

- Accuracy gains at all planning horizons

- Crucially, forecasts are reconciled leading to aligned plans

Athanasopoulos, et al. Forecasting with temporal hierarchies. EJOR, 2017

# Postulate: Always better than base forecast

- Simulations of known ARIMA: 4 sample sizes x 1000 repetitions each.
  - Scenario 1: No uncertainty;
  - Scenario 2: Parameter uncertainty;
  - Scenario 3: Model uncertainty;
  - Scenario 4: Forced misspecification.
- Negative entries = percentage gain over base.

Sample size: specified at the annual aggregation level
*(Forecast horizon: specified at the annual aggregation level)*

| | 4 (1) | 12 (3) | 20 (5) | 40 (10) | 4 (1) | 12 (3) | 20 (5) | 40 (10) | 4 (1) | 12 (3) | 20 (5) | 40 (10) | 4 (1) | 12 (3) | 20 (5) | 40 (10) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Scenario 1 | | | | Scenario 2 | | | | Scenario 3 | | | | Scenario 4 | | | |
| | WLS combination forecasts using variance scaling | | | | | | | | | | | | | | | |
| Annual | −0.3 | 0.0 | 0.0 | 0.0 | −4.3 | −7.9 | −6.1 | −3.3 | −66.2 | −5.1 | −2.6 | −0.4 | −24.7 | 1.6 | 0.5 | −1.8 |
| Semi-annual | −0.1 | −0.1 | 0.0 | 0.0 | −5.2 | −3.5 | −1.6 | −0.2 | −50.6 | −4.9 | −2.6 | −1.2 | −42.6 | −5.5 | −2.7 | −1.1 |
| Four-monthly | −0.1 | 0.0 | 0.0 | 0.0 | −3.8 | −1.5 | −0.4 | −0.1 | −10.1 | −6.2 | −2.0 | −1.2 | −9.4 | −6.7 | −2.7 | −4.3 |
| Quarterly | −0.1 | 0.0 | 0.0 | 0.0 | −3.9 | −0.6 | −0.2 | −0.1 | −16.4 | −4.1 | −1.9 | −0.8 | −1.2 | −8.3 | −5.5 | −6.0 |
| Bi-monthly | 0.0 | 0.0 | 0.0 | 0.0 | −1.1 | 0.0 | 0.1 | 0.0 | −7.5 | −3.3 | −0.7 | −0.9 | −1.0 | −8.3 | −9.3 | −8.6 |
| Monthly | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 | 0.5 | 0.1 | 0.0 | −0.9 | −0.5 | −0.8 | −1.9 | −1.4 | −7.3 | −11.3 | −17.0 |
| | Bottom-up | | | | | | | | | | | | | | | |
| Annual | −0.7 | −0.1 | 0.2 | 0.1 | −5.3 | −9.5 | −7.1 | −3.4 | −64.2 | −1.2 | 5.9 | 27.9 | −20.9 | 69.1 | 101.6 | 150.4 |
| Semi-annual | −0.5 | −0.1 | 0.1 | 0.0 | −7.6 | −4.8 | −2.4 | −0.2 | −48.5 | −2.8 | 2.3 | 13.8 | −40.0 | 35.5 | 63.8 | 105.3 |
| Four-monthly | −0.2 | −0.1 | 0.1 | −0.1 | −5.5 | −2.7 | −1.0 | −0.2 | −7.1 | −5.1 | 1.4 | 8.7 | −5.8 | 23.4 | 47.8 | 73.1 |
| Quarterly | −0.2 | 0.0 | 0.0 | 0.0 | −6.1 | −1.8 | −0.7 | −0.2 | −14.0 | −3.0 | 0.4 | 6.5 | 2.3 | 15.5 | 33.4 | 54.9 |
| Bi-monthly | −0.1 | −0.1 | 0.0 | 0.0 | −2.8 | −0.9 | −0.2 | −0.1 | −5.8 | −2.4 | 1.2 | 3.8 | 1.9 | 8.2 | 16.1 | 32.7 |
| Monthly | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

# What is the intuition?

- 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!
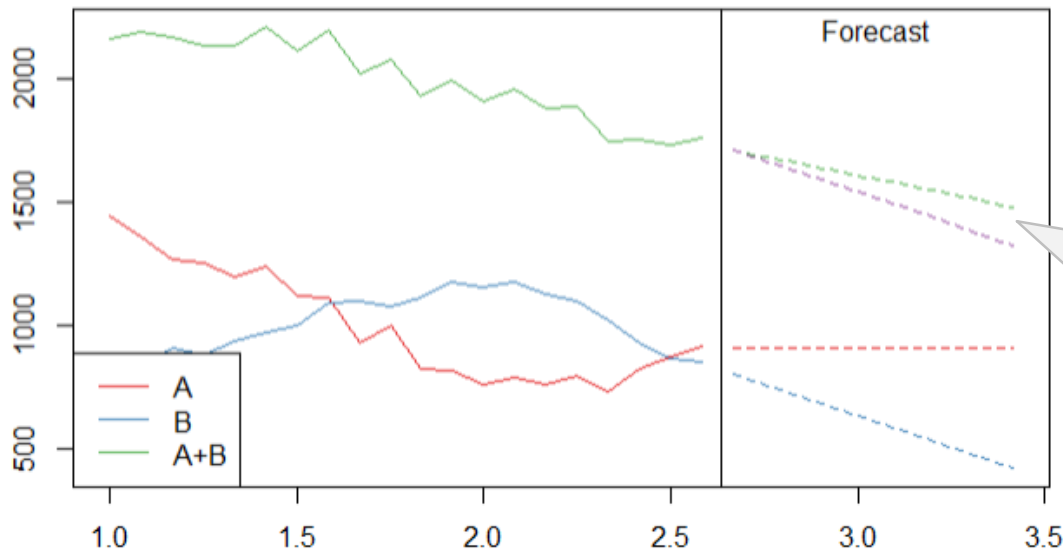


F(A+B) and F(A)+F(B) will typically be different, we need to impose equality (coherency of forecasts).

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

# What is the intuition?

- We produce forecasts at different aggregation levels:

  - We end up with multiple predictions, based on **different information for the same quantity**.

  - We **combine** all these together so as to ensure coherency of forecasts.

  - Forecast combination on average increases accuracy, **particularly when the combined forecasts consider different information**.

  - So instead of hoping that a single well calibrated and selected model approximates the underlying demand process well enough, we:

    - Thrive in the uncertainty! If all forecasts agree then there is no uncertainty. If forecasts disagree a lot, we take advantage of that to improve quality of combined forecast!

    - We mitigate the risk from identifying the "correct" model.

- Same logic explains the gains from cross-sectional forecasting.

# What we got so far?

- Temporal hierarchies is a device to:
  - Mitigate modelling uncertainty, by looking at the data from different views;
  - Results in more accurate forecasts, due to the explicit handling of modelling uncertainty;
  - Results in more reliable forecasts (accurate over time) even when competing with favorable conditions base forecasts (e.g. knowledge of the process form);

- A main benefit of using temporal hierarchies is that is allows merging information from different levels of planning
  - Operational short-term vs. Strategic long-term;
  - Operational univariate (+ features) vs. Tactical/Strategic multivariate/scenario based.
  - **Reconciles across forecast planning horizons**.

- Provides a machinery to balance information flows:
  - From *strategising operations* (i.e. top-down information flows) to…
  - *Operationalising strategies*, that is a bottom-up flow as operations as close to the customer.

# 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.
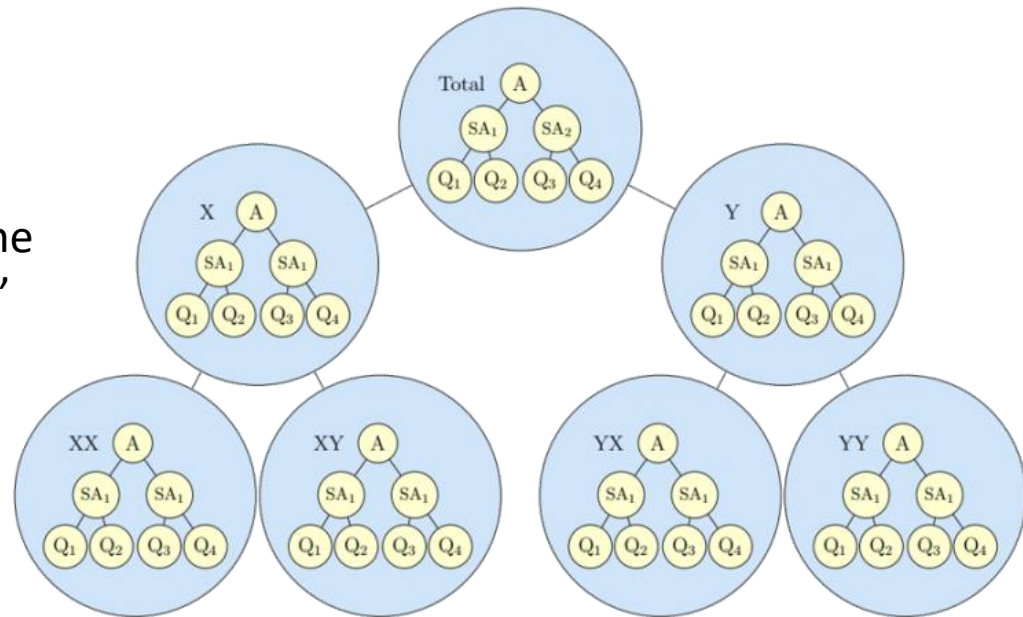


* Terms & conditions apply! We split the hierarchical problem to aid estimation.

# 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.
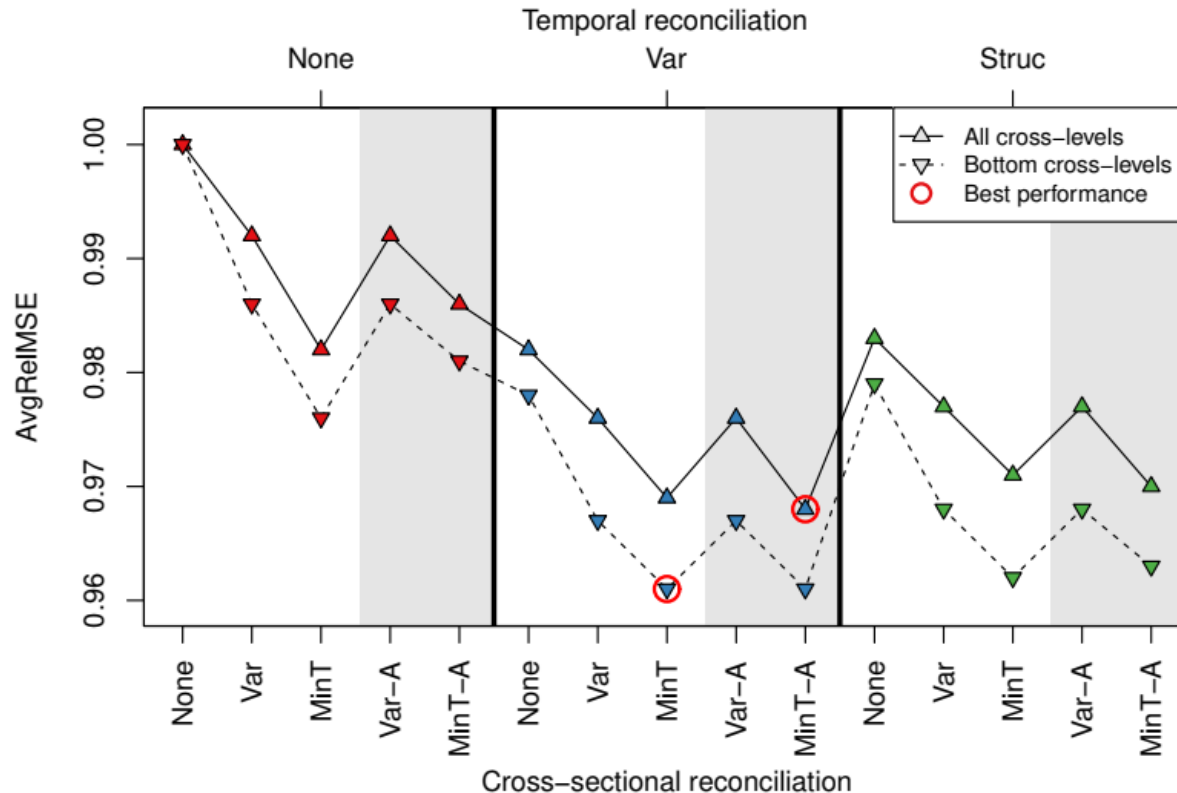


Figures in grey are cross-temporally coherent

Kourentzes and Athanasopoulos, Cross-temporal coherent forecasts. ATR, 2019

# Where next? Top-level information to inventory

| Type of indicator | Percent selected | Indicator units |
|---|---|---|
| Labor | 18.5 | Persons, USD/week, % growth, hours |
| Consumer price index | 12.4 | Index, growth rate, % growth |
| Government services | 9.6 | Persons, % growth, index |
| Retail trade | 9.3 | Persons, index, hours |
| Financial activities | 8.5 | Persons, index, % growth, USD |
| Private services | 8.3 | Persons, % growth |
| Transport and automotive | 6.4 | Persons, USD, index, % growth |
| Manufacturing | 5.8 | Hours, USD, net % growth, index |
| Food manufacturing | 5.0 | Persons, rate, % growth |
| Education and health services | 4.4 | Persons, % growth |
| Wholesale trade | 3.6 | Persons, USD, % growth, USD/week |
| Tourism | 2.6 | Persons |
| Recession indicator | 1.5 | Index |
| Construction | 1.1 | Hours, net % growth, index, USD/week |
| Mining | 1.0 | Persons, hours, % growth |
| International trade | 1.0 | Growth rate, USD, national currency, index |
| Exchange rate | 0.6 | Index, % growth |

- Top-level macroeconomic leading indicators for a manufacturer.

- Can be tied to country/across country levels

- Use hierarchies to bring this information down to operational decision unit/SKU
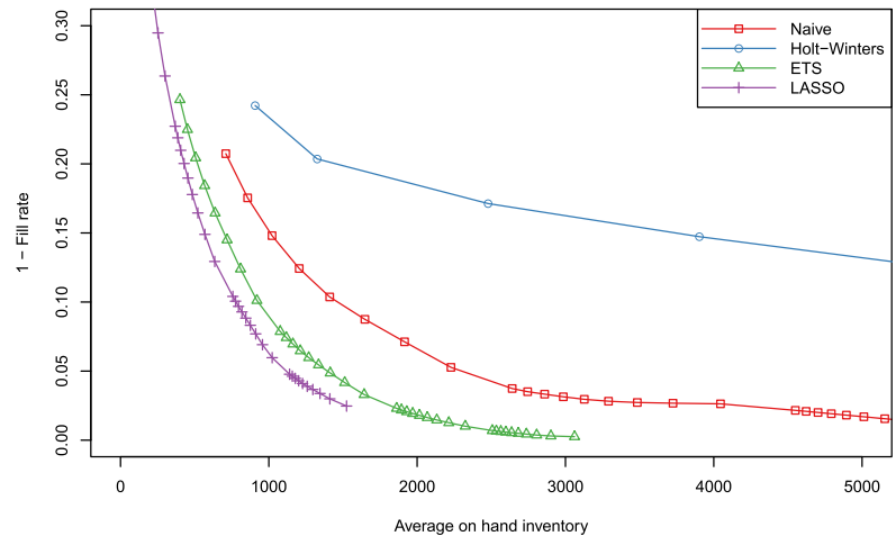
# Where next? Supply chain collaboration

- Realistic supply chains are messy to forecast as a system due to their complexity; currently most work done with simulations, and over-simplistic.



Upstream

Consumer
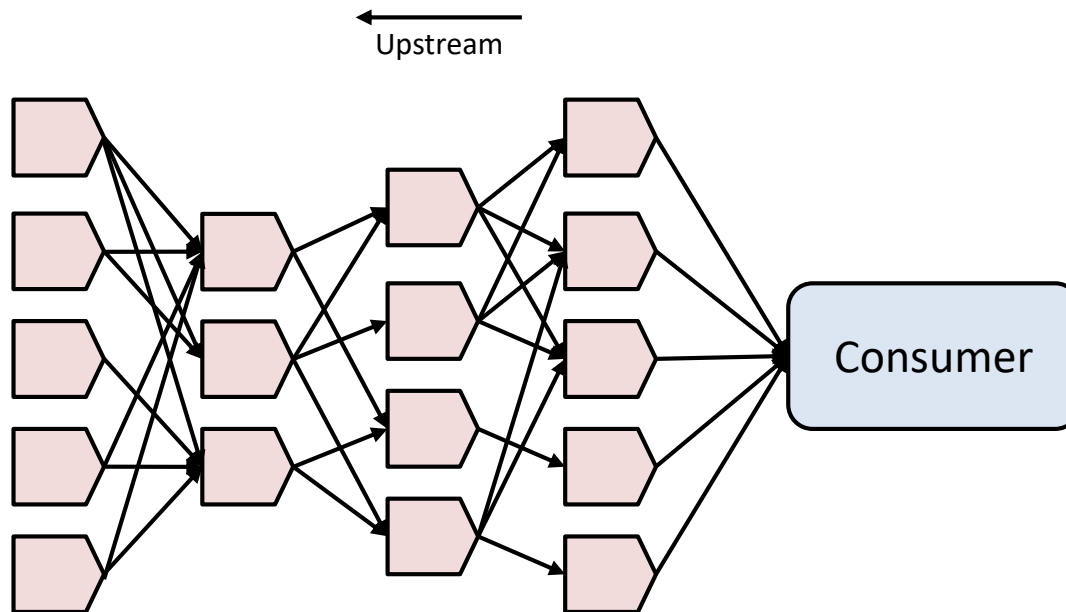
- Not fully connected.
- Multiple layers and multiple actors in each layer.
- Different demand patterns at each level and decision frequency.

Fits to the framework!

# Where next? Supply chain collaboration
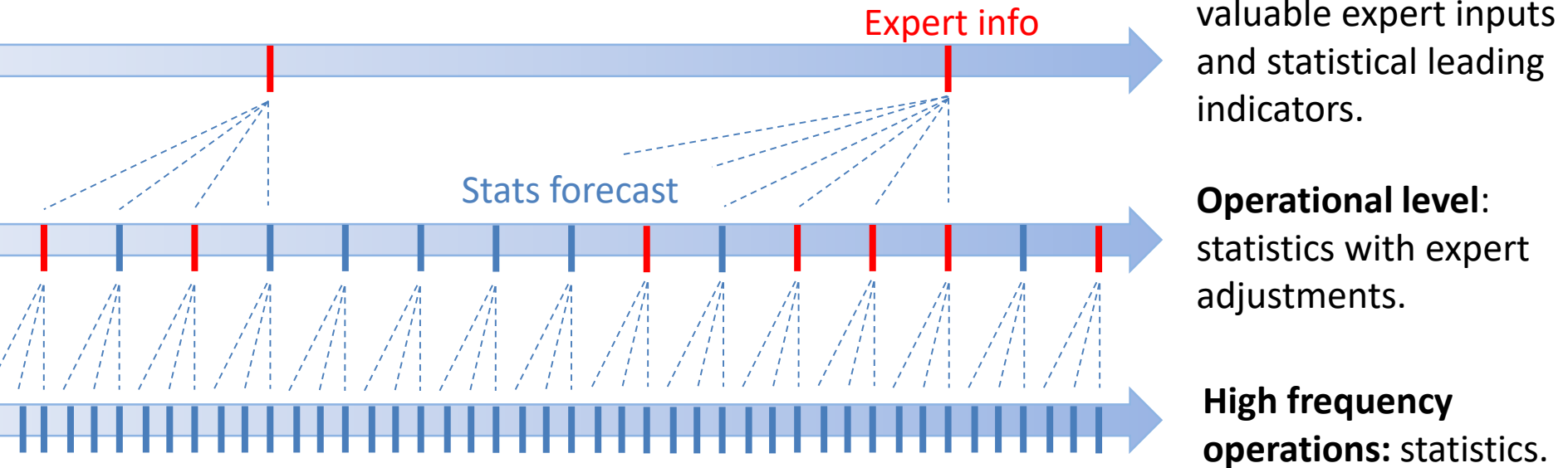
- This is more insidious than it seems at first sight
  - Total = information you know + information you don't know;
    - i.e. Total = you + competition
    - i.e. Total = you + friends + neighbour + stuff you really shouldn't know!
  - Raises issues about privacy, encryption data sharing, distributed computing, etc.
  - Who holds the negotiating power of forecasting beyond an organization, who holds the responsibility?

- Nonetheless, if done properly, the potential is tremendous:
  - Synchronised supply chains;
  - Aligned objectives;
  - Reduction of waste;
  - Sustainability;
  - …

# Where next? Ultra-high frequency decision making

- Real time decision making can lead to new interactions with customers:

  - Recommendation systems to shape consumer basket, interacting online, or via mobile phone in store.

  - Dynamic individualised promotions/pricing to maximise loyalty and consumption so as to:
    - Profit;
    - Balance inventory;
    - Shape market demand;
    - Mitigate bullwhip;
    - etc.

  - Seamless shopping experience: till-less and managed shopping trajectories (guide your customer through the "required" routes of your physical/online store).

# Where next? Ultra-high frequency decision making

- **Humans add value** to the forecasting process (but **inconsistently**).
- However, they **do not scale-up** and cannot handle very high frequency information (get **lost in randomness**).
- Temporal hierarchies to the rescue!



**Strategic** (slow) **level:** valuable expert inputs and statistical leading indicators.

**Operational level:** statistics with expert adjustments.

**High frequency operations:** statistics.

Humans can aid with low frequency adjustments and decisions. Statistics can do ultra-high frequency decisions. Meld forecast/planning levels to have human aided ultra high decision making. How: temporal hierarchies. Across functions: cross-temporal hierarchies.

# Conclusions

- Cross-temporal hierarchy forecasts provide a single view of the future across market demarcations and planning horizons → "one number forecast".

- Allows to seemingly join plans across functions within the organization. That budget forecast informs inventory decisions and that promotional forecast informs budget → without needing people to talk to each other.

- Cross-temporal forecast come with accuracy gains. Temporal hierarchy causes the biggest gains → handles modelling uncertainty explicitly.

- 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.
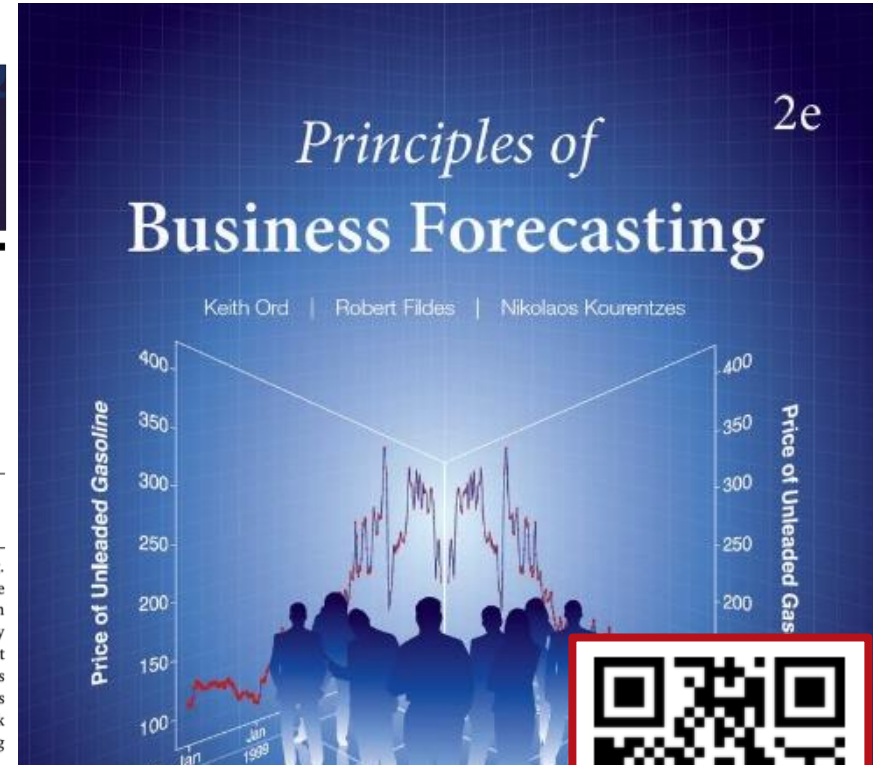
- Exciting applications!

# 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

### 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

**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 information from different sources.

## Principles of Business Forecasting 2e

Keith Ord | Robert Fildes | Nikolaos Kourentzes

- 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

Marketing Analytics and Forecasting

LUMS