Forecasting with Temporal Hierarchies

Nikolaos Kourentzes

Lancaster University

International Symposium on Forecasting 2016

22/06/2106

Lancaster University Management School | Lancaster Centre for Forecasting

Forecasting Society

Agenda

1. Forecasting issues from an organisational view
2. Forecasting issues from a modelling view
3. Temporal Aggregation
4. Multiple Aggregation Prediction Algorithm
5. A more general framework
6. Applications

Forecasting & decision making

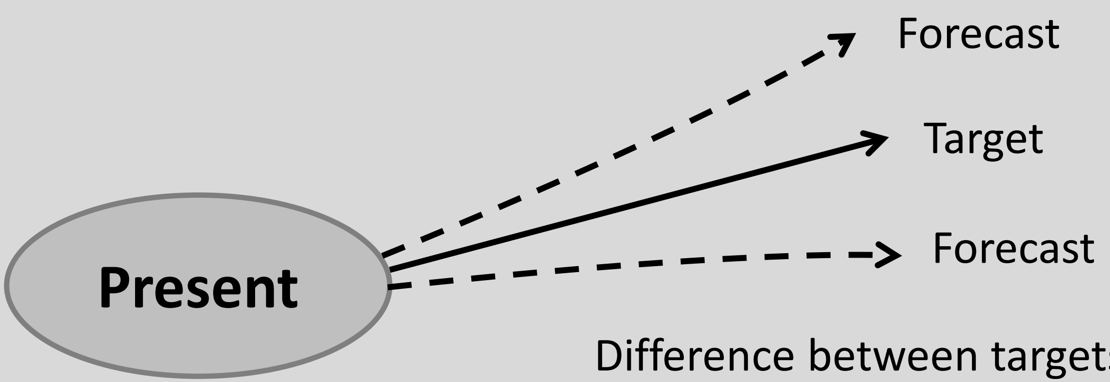
Decision making in organisations has at its core an element of forecasting

→ 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.



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

Forecasting & decision making

Decisions need to be aligned:

- Operational short-term decisions
- Tactical medium-term decisions
- Strategic long-term decisions

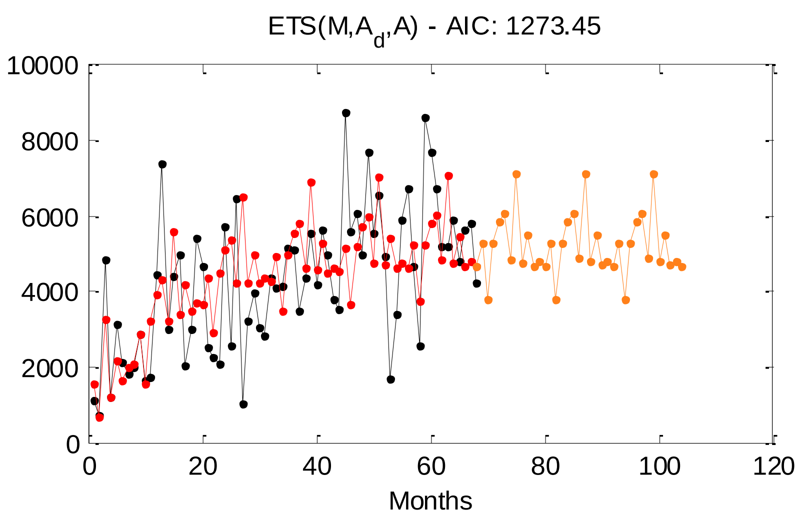
Shorter term plans are **bottom-up** and based mainly on **statistical forecasts** & expert adjustments.
Longer term plans are **top-down** and based mainly on **managerial expertise** factoring in unstructured information and organisational environment.

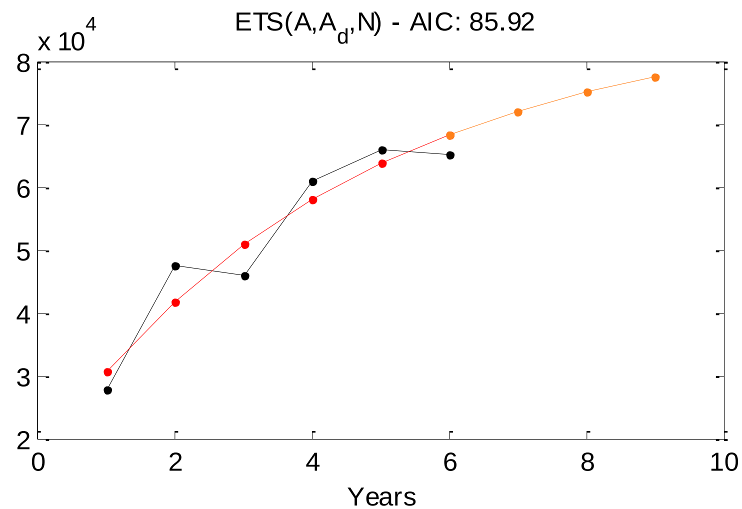
Given different sources of information (and views) forecasts will differ → plans and decisions not aligned.

Objective: construct a framework to reconcile forecasts of different levels and eventually align decisions → less waste & costs, agility to take advantage of opportunities.

Long term forecasting

- 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.





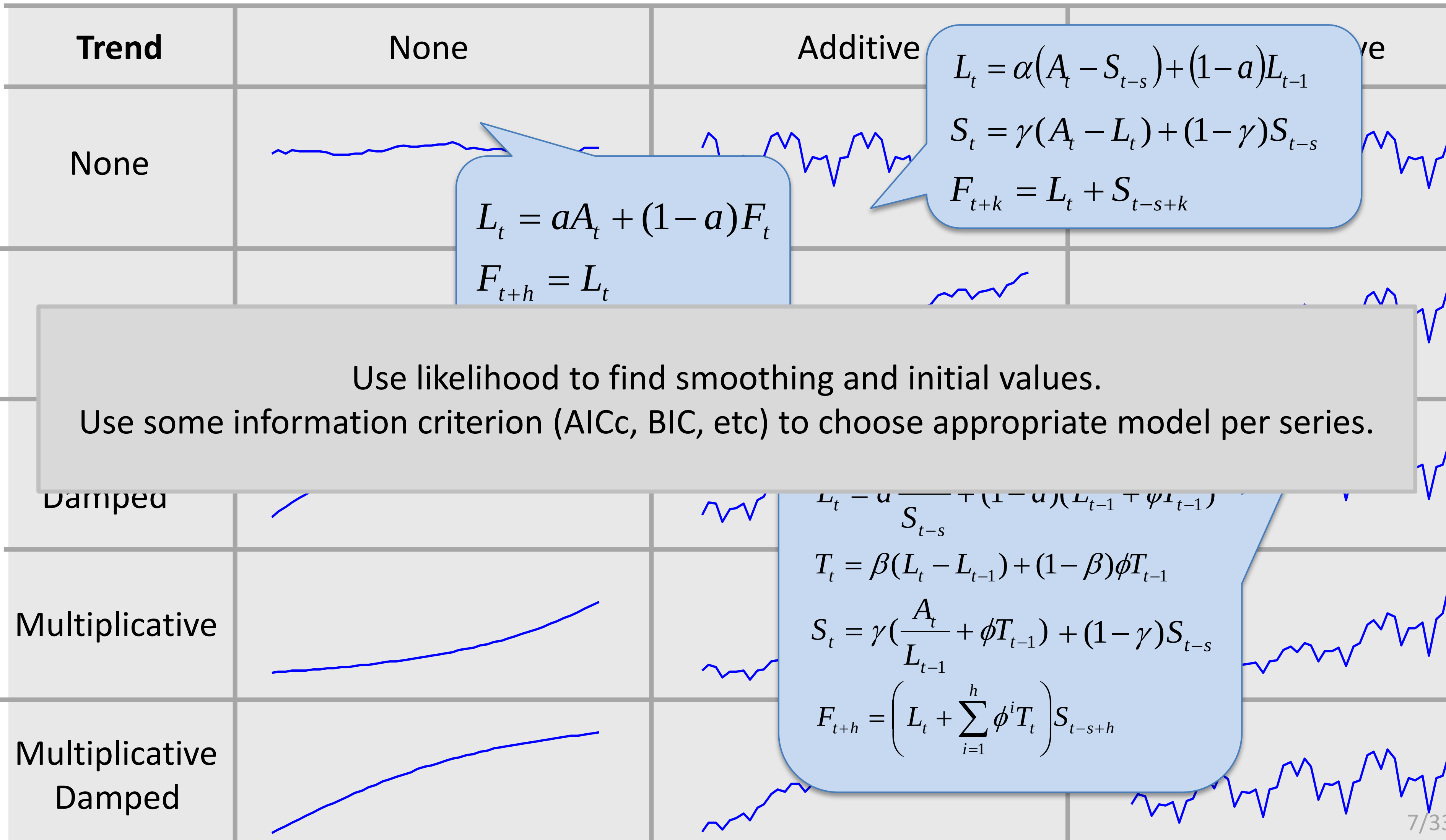
These forecasts often do not agree, which one is `correct'?

How do we build & models now?

- This is by no means a resolved question, but there are some reliable approaches
- Take the example of exponential smoothing family:
  - Considered one of the most reliable and robust methods for automatic univariate forecasting .
  - It is a family of methods: **ETS (error type, trend type, seasonality type)**
    - Error: **A**dditive or **M**ultiplicative
    - Trend: **N**one or **A**dditive or **M**ultiplicative, Linear or Damped/Exponential
    - Seasonality: **N**one or **A**dditive or **M**ultiplicative
  - Adequate for a most types of time series.
  - Within the state space framework we can select and fit model parameters automatically and reliably.

LCF

How do we build & models now?



$$L_t = aA_t + (1-a)F_t$$
$$F_{t+h} = L_t$$

$$L_t = \alpha(A_t - S_{t-s}) + (1-a)L_{t-1}$$
$$S_t = \gamma(A_t - L_t) + (1-\gamma)S_{t-s}$$
$$F_{t+k} = L_t + S_{t-s+k}$$

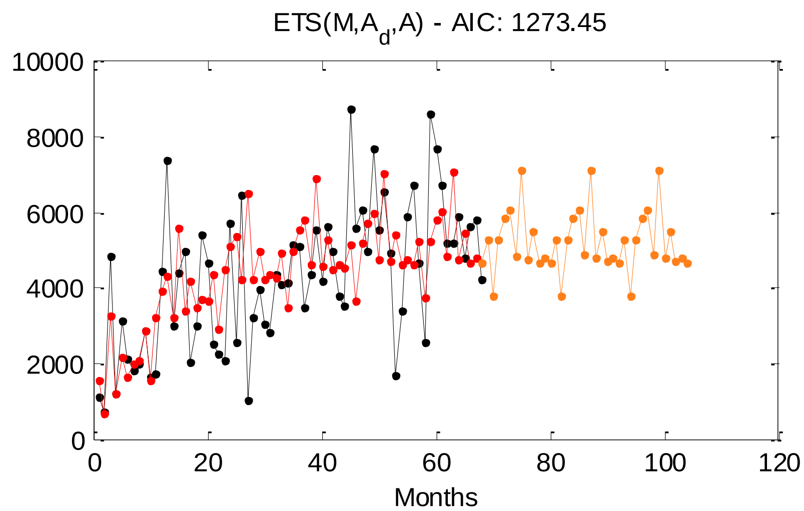
Use likelihood to find smoothing and initial values.
Use some information criterion (AICc, BIC, etc) to choose appropriate model per series.

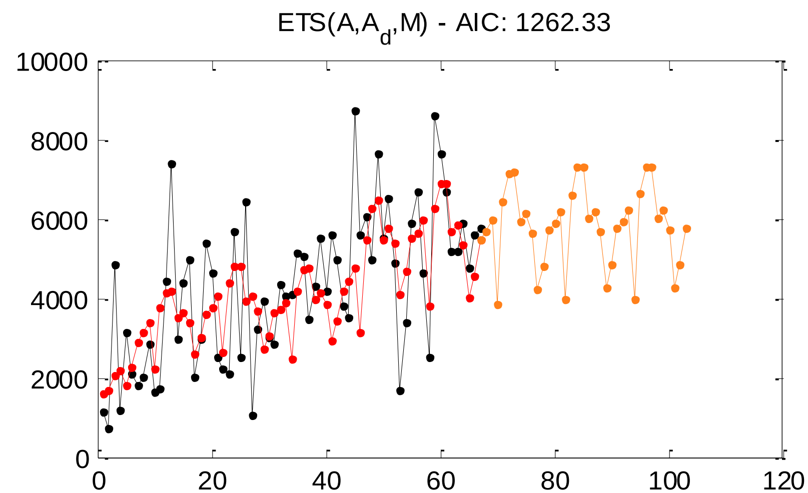
$$L_t = a\frac{\quad}{S_{t-s}} + (1-a)(L_{t-1} + \phi T_{t-1})$$
$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)\phi T_{t-1}$$
$$S_t = \gamma(\frac{A_t}{L_{t-1}} + \phi T_{t-1}) + (1-\gamma)S_{t-s}$$
$$F_{t+h} = \left(L_t + \sum_{i=1}^{h}\phi^i T_t\right)S_{t-s+h}$$

Any issues with current practice?

Issues with automatic modelling:

- 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 → Practitioners do not trust systems that change substantially

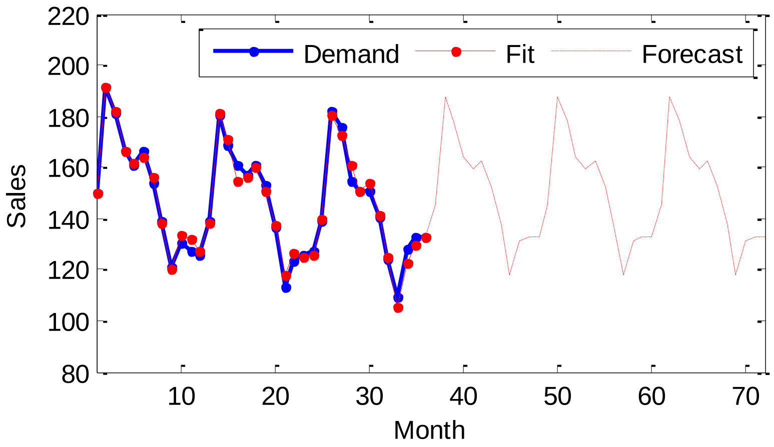




Any issues with current practice?

What can go wrong in parameter and model selection:

- Business time series are often short → Limited data
- Estimation of parameters can fail miserably (for monthly data optimise up to 18 parameters, with often no more than 36 observations)
- Model selection can fail as well (30 models → over-fitting?)
- Both optimisation and model selection are myopic → Focus on data fitting in the past, rather than '*forecastability*'
- Special cases:



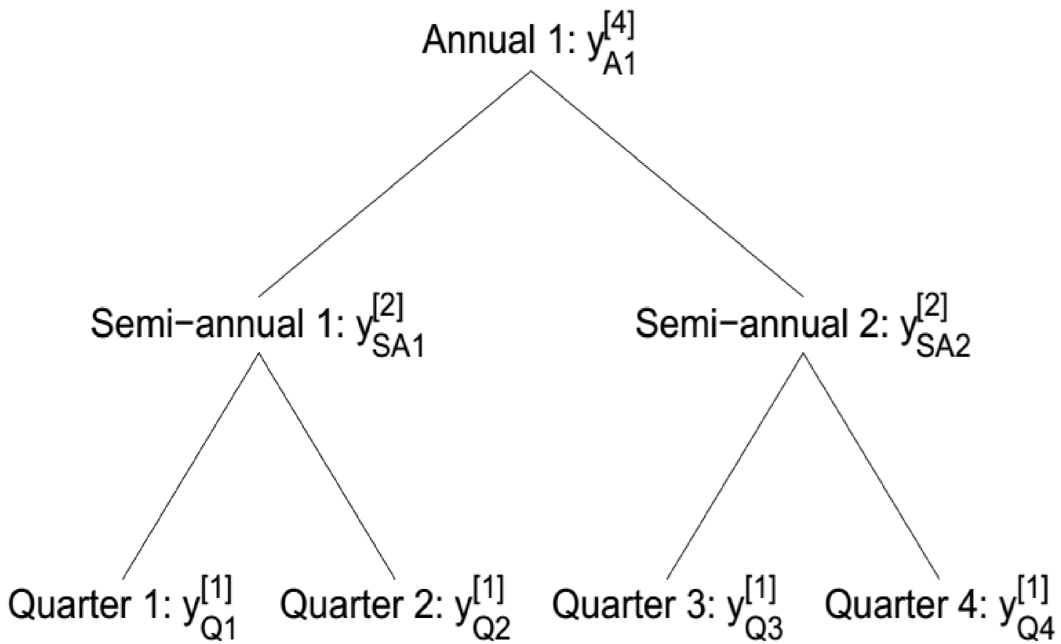
True model:
Additive trend, additive seasonality

Identified model:
No trend, additive seasonality

Why?
In-sample variance explained mostly by seasonality

A different take on modelling: temporal tricks!

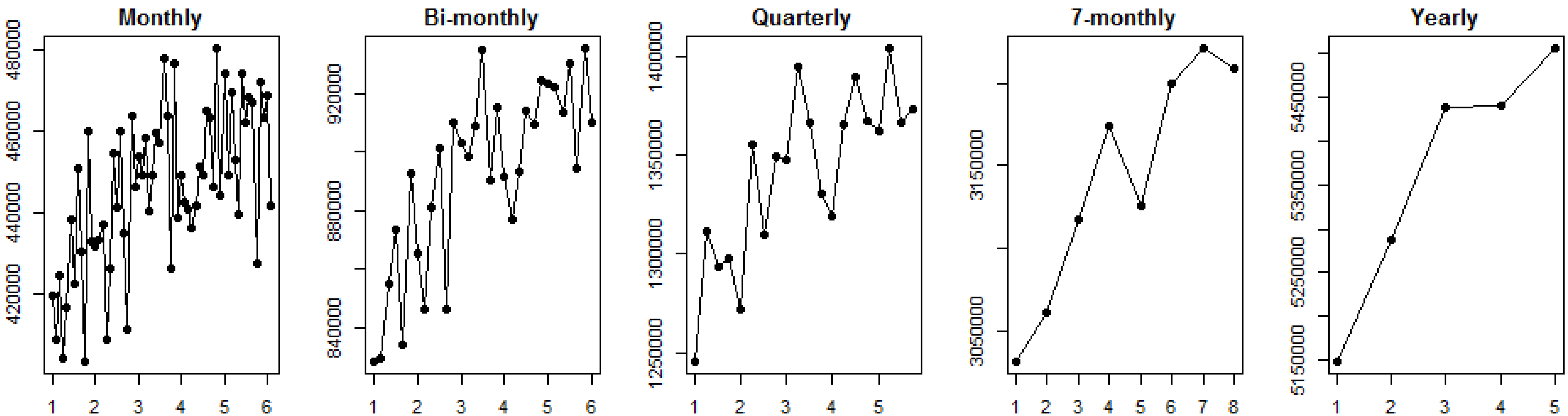
Traditionally we model time series at the frequency that we sampled them or take decisions. However, a time series can be view in many different ways, adapting the notion of product hierarchies to **temporal hierarchies**:



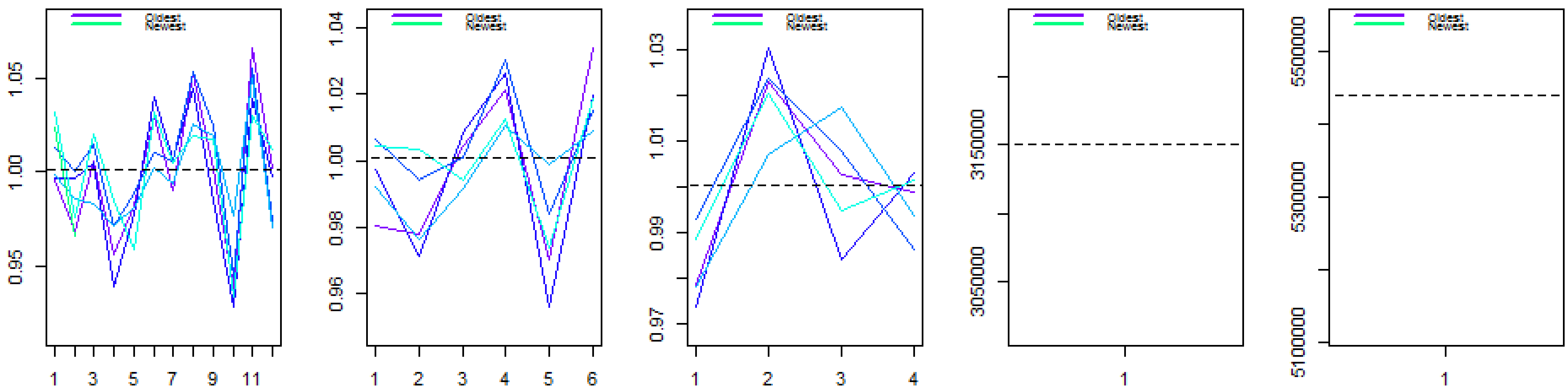
The advantages of temporal hierarchies can be highlighted by examining the data at **both time and frequency domains**.

LCF

How temporal aggregation changes the series



Seasonal diagrams



The Idea

- Temporal aggregation strengthens and attenuates different elements of the series:

  → at an aggregate level trend/cycle is easy to distinguish
  → at a disaggregate level high frequency elements like seasonality typically dominate.

- Modelling a time series at a very disaggregate level (e.g. weekly) → short-term forecast. The opposite is true for aggregate levels (e.g. annual)

- **Propose Temporal Hierarchies that** provide a framework to optimally combine information from various levels (**irrespective of forecasting method**) to:

  - **reconcile forecasts**
  - **avoid over-reliance on a single planning level**
  - **avoid over-reliance on a single forecasting method/model**

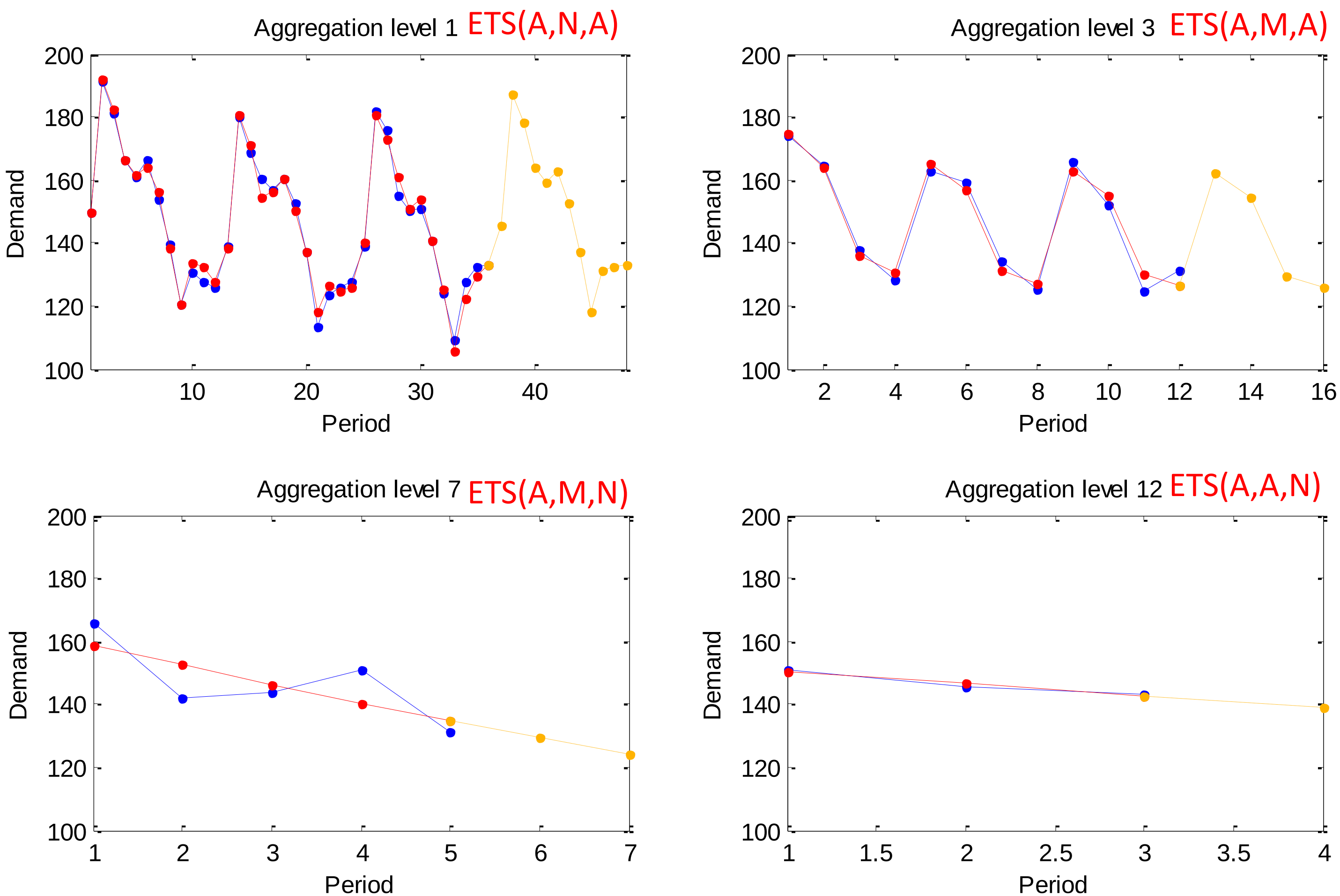
LCF

Temporal aggregation and forecasting

- It is not new, but the question has been at **which single level to model the time series**. Econometrics have investigate the question for decades → inconclusive
- Supply chain applications: ADIDA → beneficial to slow and fast moving items forecast accuracy (like everything… not always!):
  - **Step 1**: Temporally aggregate time series to the appropriate level
  - **Step 2**: Forecast
  - **Step 3**: Disaggregate forecast and use
  - Selection of aggregation level → No theoretical grounding for general case, but good understanding for AR(1)/MA(1) cases.

Multiple temporal aggregation

What if we do not select an aggregation level? → use multiple



Issues:

- Different model
- Different length
- Combination

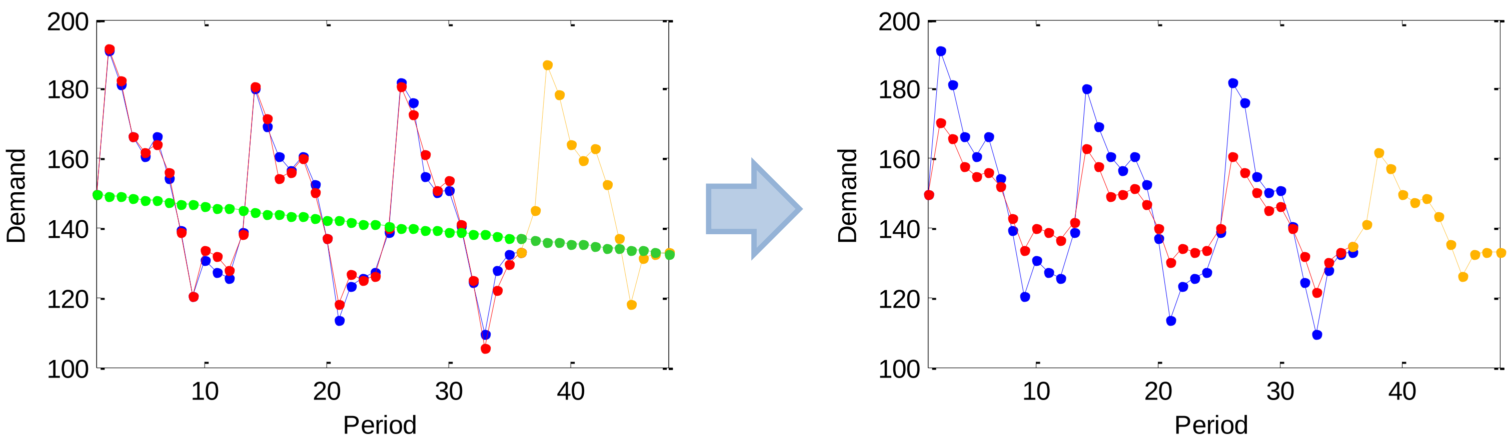
Multiple temporal aggregation

Forecast combination:

- Forecast combination is widely considered as beneficial for forecast accuracy
- Simple combination methods (average, median) considered robust, relatively accurate to more complex methods

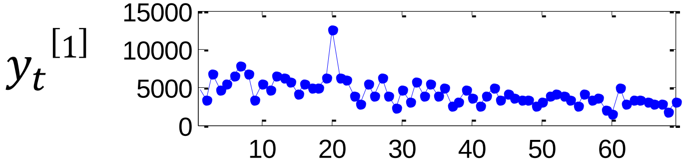
Issue:

- If there are different model types to be combined then the resulting forecast does not fit well at any component!

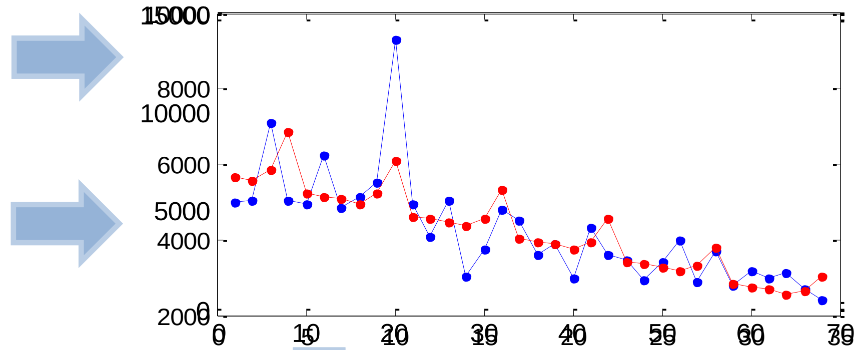


LCF

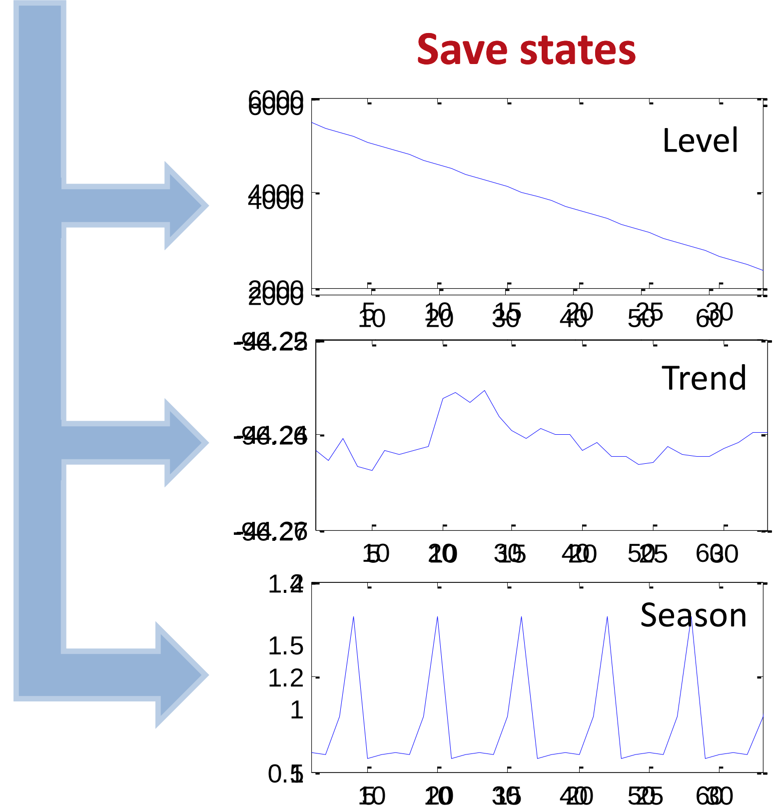
Aggregate



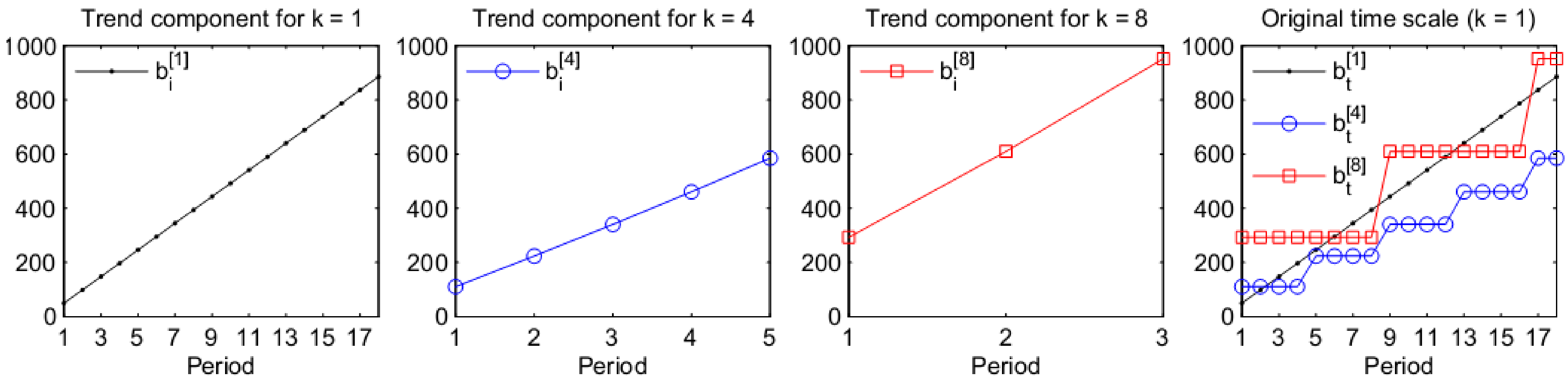
Fit state space ETS



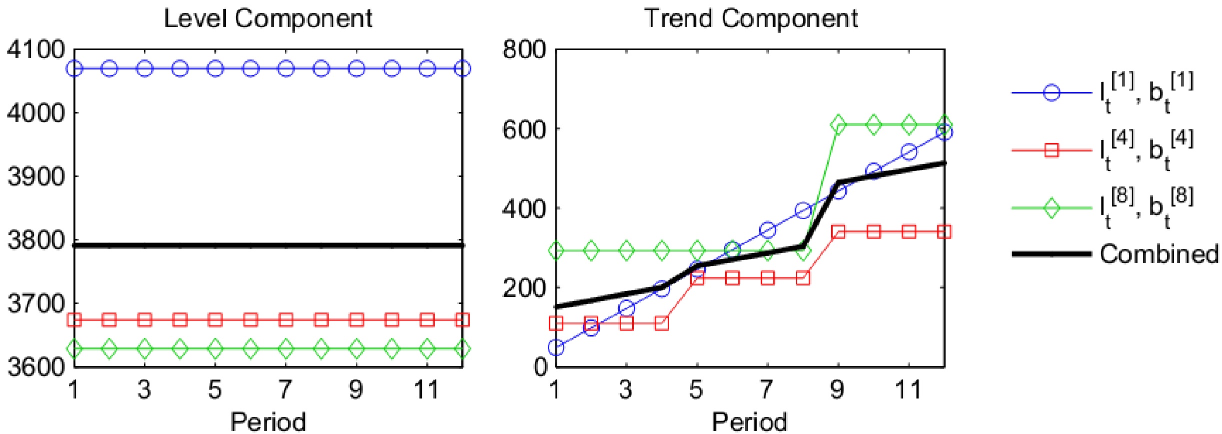
Save states



Transform states to additive and to original sampling frequency



Combine states (components)

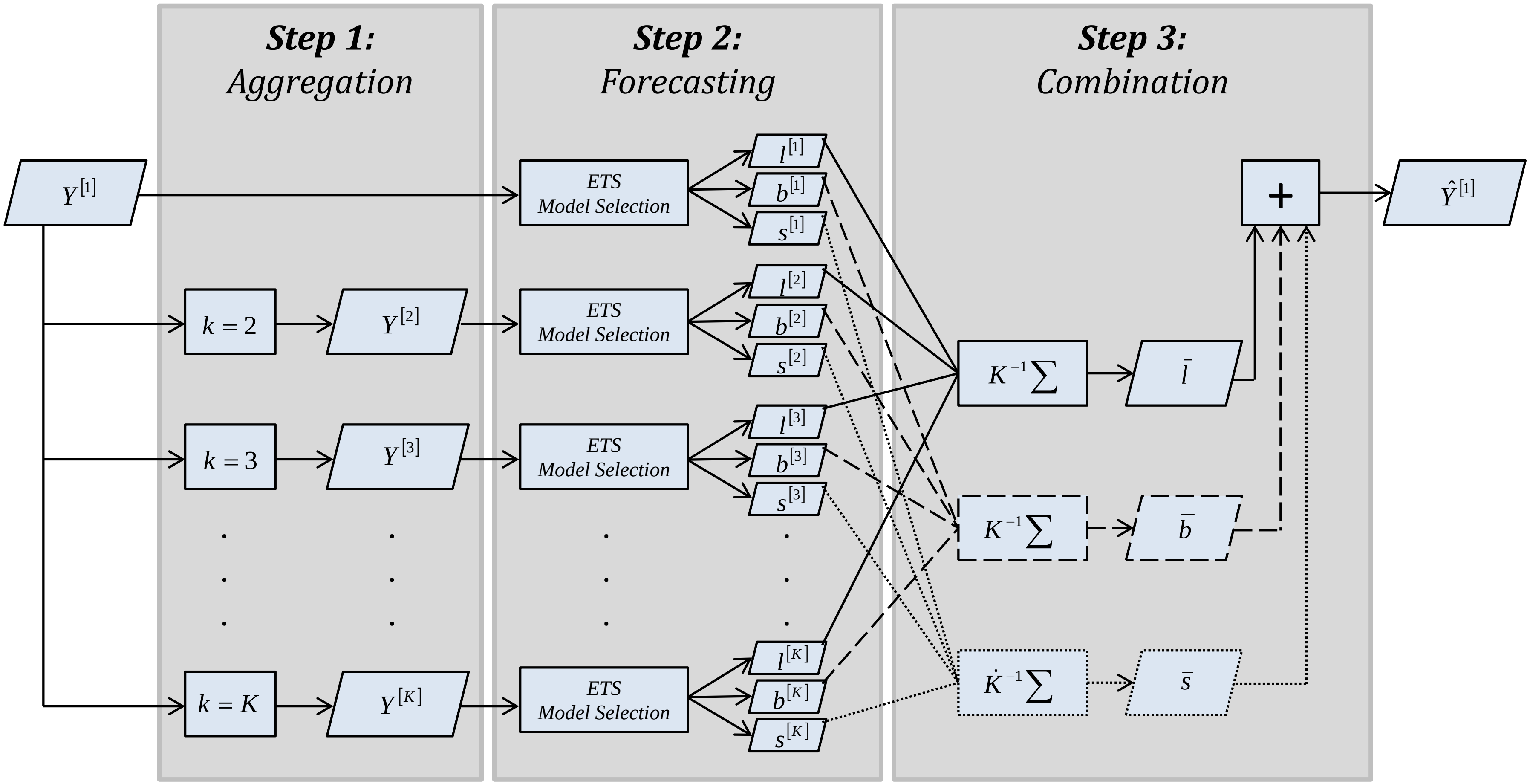


Produce forecasts

$$\hat{y}^{[1]}_{t+h^{[1]}} = \bar{l}_{t+h^{[1]}} + \bar{b}_{t+h^{[1]}} + \bar{s}_{t-S+h^{[1]}}$$

LCF

Multiple Aggregation Prediction Algorithm (MAPA)

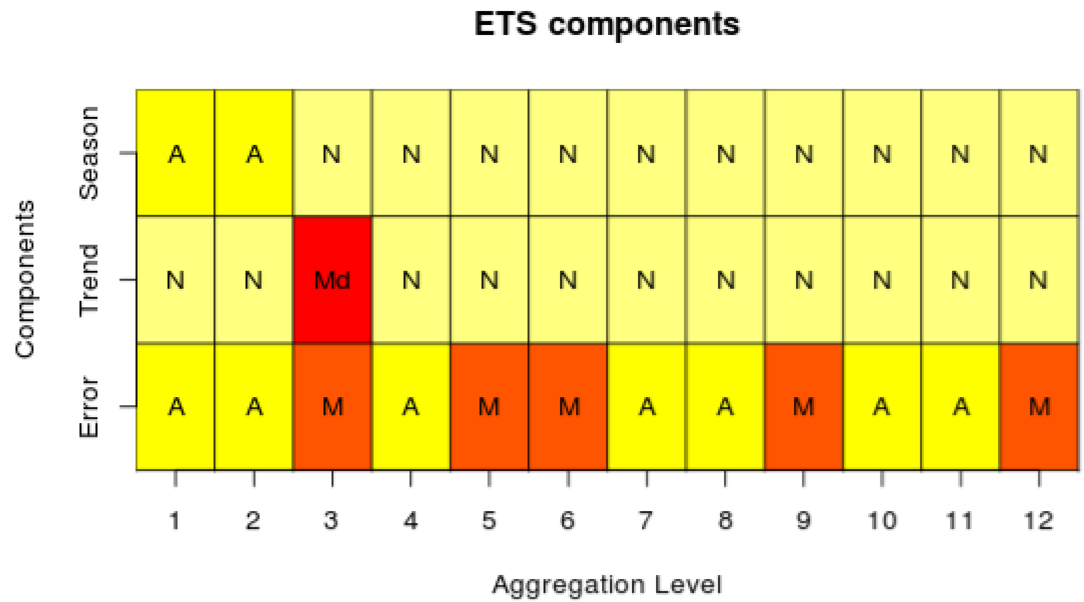


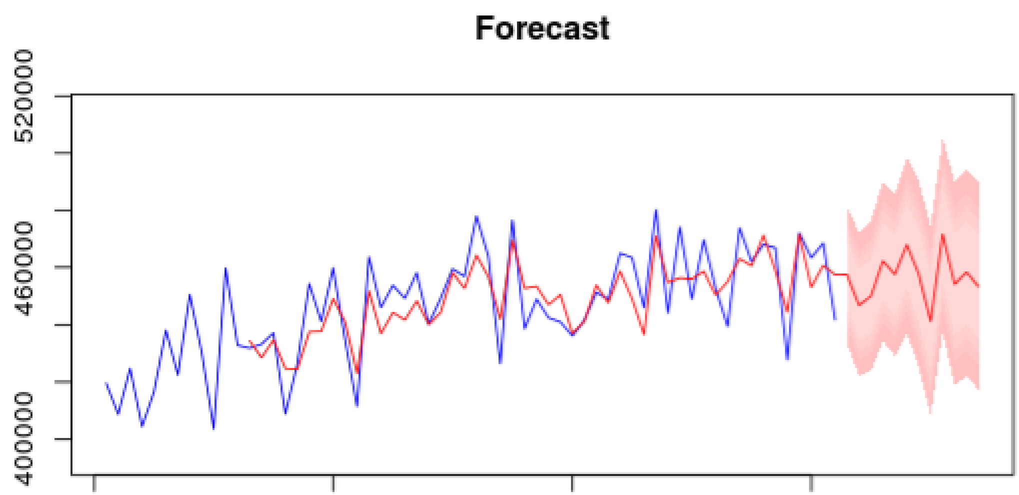
Strengthens and attenuates components

Estimation of parameters at multiple levels

Robustness on model selection and parameterisation

Multiple Aggregation Prediction Algorithm (MAPA)





Multiple Aggregation Prediction Algorithm (MAPA)

MAPA was developed to take advantage of temporal aggregation and hierarchies:

- MAPA provides a framework to better identify and estimate the different time series components → better forecasts
- On average outperforms ETS, one of the most widely used, robust and accurate univariate forecasting methods
- It provides reconciled forecasts across planning levels and forecast horizons
- Robust against model selection and parameterisation issues
- Shown to be useful for fast moving items, promotional modelling and intermittent time series forecasting.

MAPA is available for R, in the **MAPA** package:
http://cran.r-project.org/web/packages/MAPA/index.html
Its intermittent demand counterpart is available in the **tsintermittent** package:
http://cran.r-project.org/web/packages/tsintermittent/index.html
Examples and interactive demos for both are available at my blog:
http://nikolaos.kourentzes.com

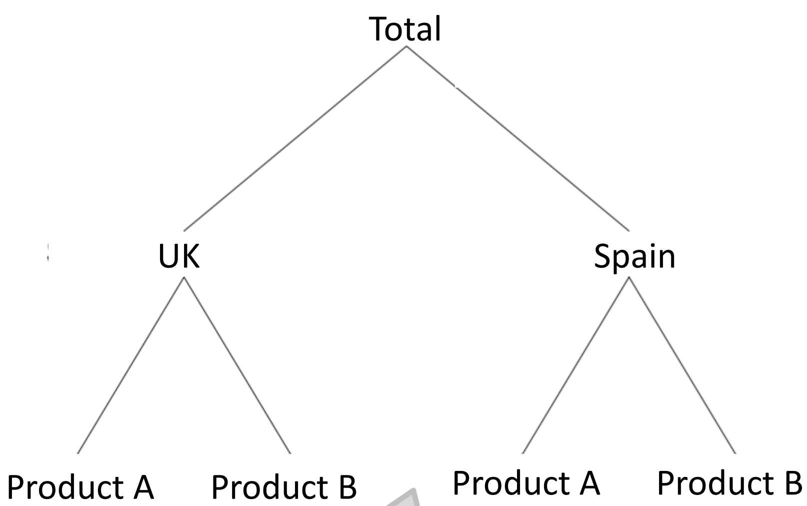
Temporal Hierarchies: A modelling framework

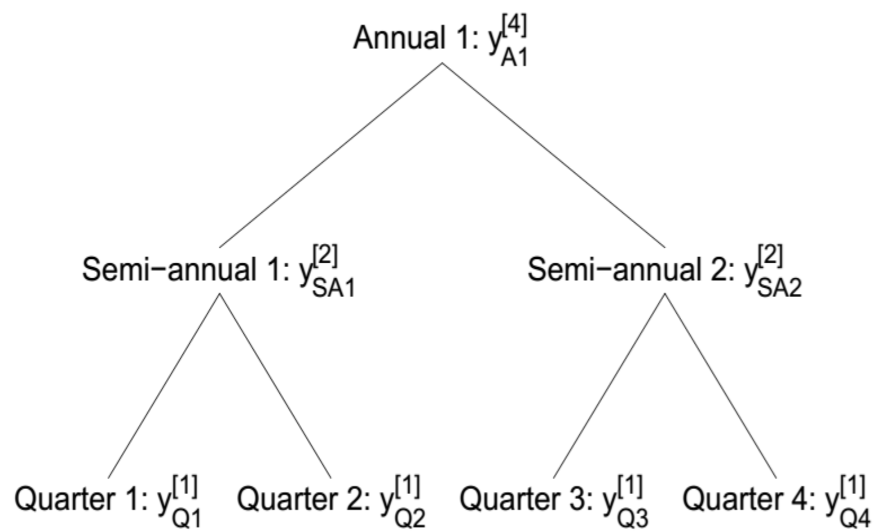
- MAPA demonstrated the strength of the approach, but it is not general:
  - How to incorporate forecasts from any model/method?
  - How to incorporate judgement?
- We can introduce a general framework for **temporal hierarchies** that borrows many elements from **cross-sectional hierarchies**
- Eventually we will get to **cross-temporal hierarchies**, also touted as the **'one-number' forecast**, i.e. a reconciled forecast across planning horizons and product/customer/location groups.

LCF

Cross-sectional and Temporal Hierarchies

- We know how to do cross-sectional hierarchies
  - ~~Top-down, bottom-up, middle-out~~
  - Optimal combinations





We know how to do this!

… then we know how to do this as well, with some small-print!

but we have to correct for the different scales at each aggregation level, which is not that difficult due to the imposed temporal structure. For that we need to calculate the covariance matrix between the forecast errors at different aggregation levels. There are easy and not so easy ways to do this.

Some evidence that it actually works!

Comparison with other M3 results (symmetric Mean Absolute Percentage Error):

- **Monthly dataset**
  - **Temporal** (ETS based): 13.61%
  - **ETS**: 14.45% [Hyndman et al., 2002]
  - **MAPA**: 13.69% [Kourentzes et al., 2014]
  - **Theta**: 13.85% (best original performance) [Makridakis & Hibon, 2000]

- **Quarterly dataset**
  - **Temporal** (ETS based): 9.70%
  - **ETS**: 9.94% [Hyndman et al., 2002]
  - **MAPA**: 9.58% [Kourentzes et al., 2014]
  - **Theta**: 8.96% (best original performance) [Makridakis & Hibon, 2000]

Detailed results available if you are interested at the end of the presentation!

Application: Predicting A&E admissions

Collect weekly data for UK A&E wards.
13 time series: covering different types of emergencies and different severities (measured as time to treatment)
Span from week 45 2010 (7th Nov 2010) to week 24 2015 (7th June 2015)
Series are at England level (not local authorities).

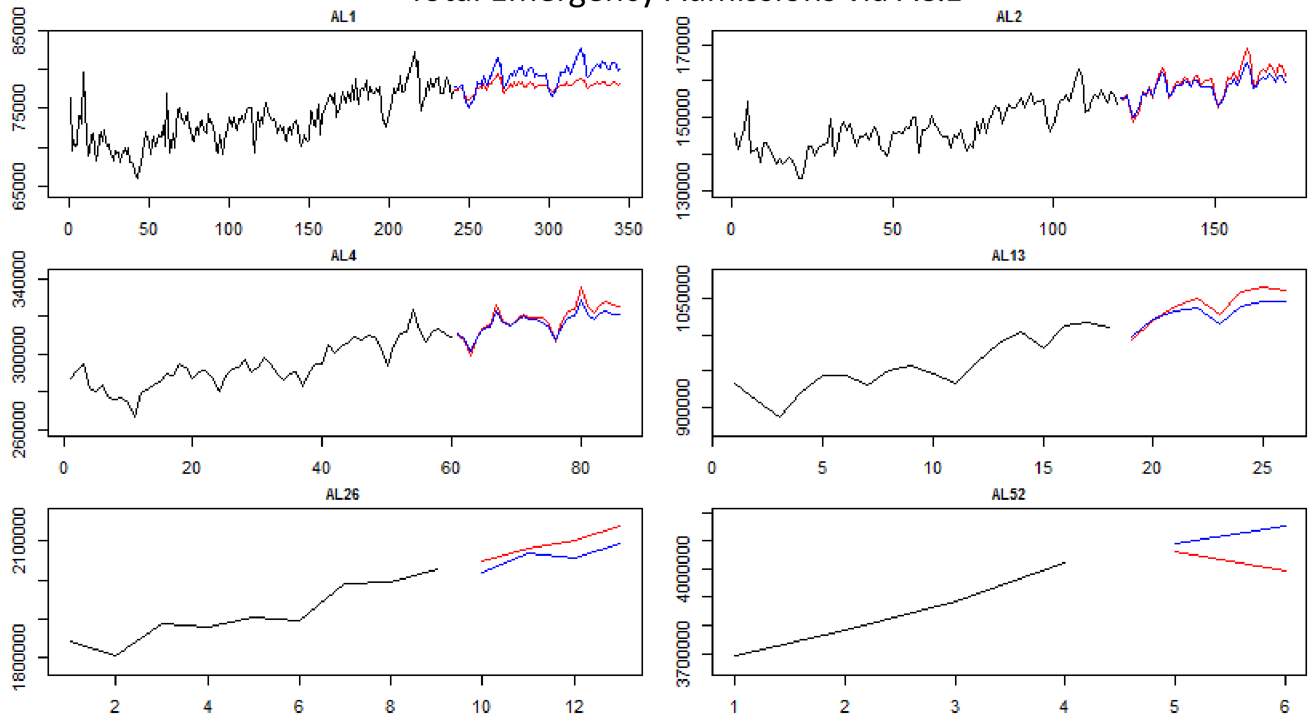
Accurately predict to support staffing and training decisions.
Aligning the short and long term forecasts is important for consistency of planning and budgeting.

Test set: 52 weeks.
Rolling origin evaluation.
Forecast horizons of interest: t+1, t+4, t+52 (1 week, 1 month, 1 year).
Evaluation MASE (relative to base model)
As a base model auto.arima (forecast package R) is used.

LCF

Application: Predicting A&E admissions

Total Emergency Admissions via A&E



Red is the prediction of the base model (ARIMA)
Blue is the temporal hierarchy reconciled forecasts (based on ARIMA)

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

Application: 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%

- Accuracy gains at all planning horizons
- Crucially, forecasts are reconciled leading to aligned plans

LCF

Hierarchical forecasting & decision making

Hierarchical (or grouped) forecasting can improve accuracy, but their true strength lies in the reconciliation of the forecasts → aligning forecasts is crucial for decision making.

Is the reconciliation achieved useful for decision making?

Cross-sectional

- Reconcile across different items.
- Units may change at different levels of hierarchy.
- Suppose an electricity demand hierarchy: lower and higher levels have same units. All levels relevant for decision making.
- Suppose a supply chain hierarchy. Weekly sales of SKU are useful. Weekly sales of organisation are not! Needed at different time scale.

Temporal

- Reconcile across time units/horizons.
- Units of items do not change.
- Consider our application. NHS admissions short and long term are useful for decision making.
- Suppose a supply chain hierarchy. Weekly sales of SKU is useful for operations. Yearly sales of a single SKU may be useful, but often not!
- Operational → Tactical → Strategic forecasts.

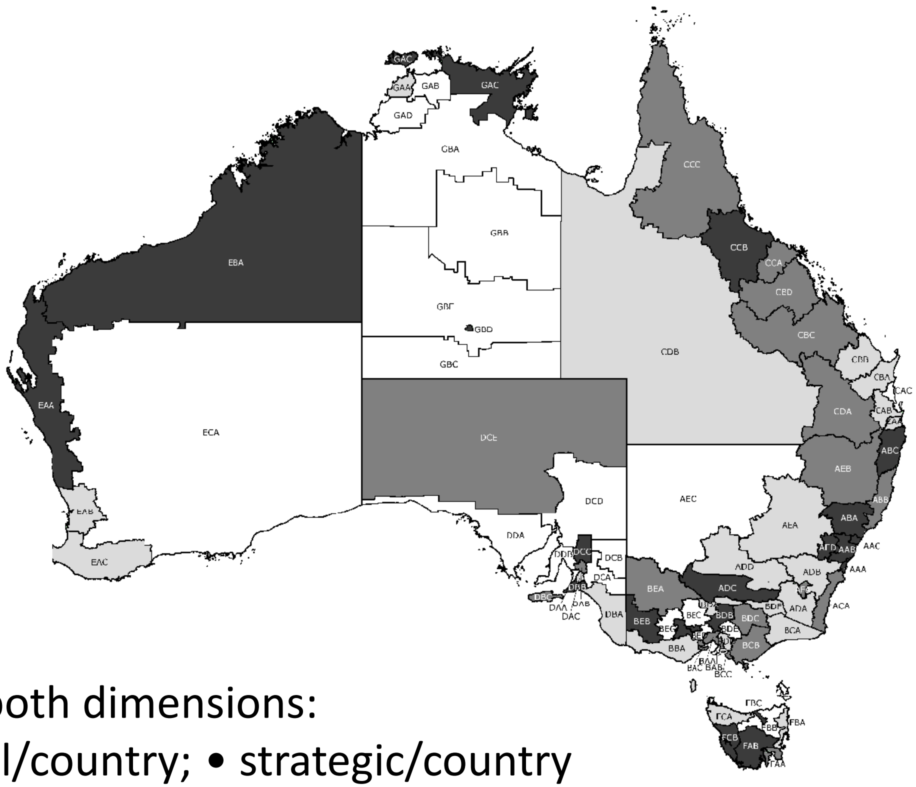
Cross-temporal hierarchies

Temporal hierarchies permit aligning operational, tactical and strategic planning, while offering accuracy gains → useful for decision making

BUT there can be cases that strategic level forecasts are not required for each item, but at an aggregate level.

Let us consider tourism demand for Australia as an example. Local authorities can make use of detailed forecasts (temporal/spatial) but at a country level weekly forecasts are of limited use.

- Temporal: tactical → strategic
- Cross-sectional: local → country



Cross temporal can support decisions at both dimensions:
• Tactical/local; • strategic/local; • tactical/country; • strategic/country

56 (bottom level) quarterly tourism demand series

- 6 years in-sample
- 3 years out-of-sample horizon: up to 2 years
- rolling origin evaluation

Cross-temporal hierarchical forecasts:

- Most accurate
- Most complete reconciliation (one number forecast)
- Flexible decision making support

MAPE %

Level	No. of series	ETS	Theta
		Base forecasts per series	
Overall	89	32.26	28.74
Top	1	5.61	5.96
Level 1	4	9.08	9.05
Level 2	28	28.39	24.68
Bottom	56	36.32	32.58
		Temporally reconciled	
Overall	89	30.46	28.19
Top	1	5.75	6.13
Level 1	4	9.29	9.04
Level 2	28	27.18	24.21
Bottom	56	34.06	31.95
		Cross-temporally reconciled	
Overall	89	30.26	**28.04**
Top	1	6.02	**5.88**
Level 1	4	9.11	**8.70**
Level 2	28	25.91	**23.87**
Bottom	56	34.39	**31.90**

Production ready?

- **Multiple Aggregation Prediction Algorithm (MAPA)**
  - Kourentzes, N.; Petropoulos, F. & Trapero, J. R. Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting*, **2014**, *30*, 291-302 (Details)
  - Petropoulos, F. & Kourentzes, N. Improving forecasting via multiple temporal aggregation. *Foresight: The International Journal of Applied Forecasting*, **2014**, *2014*, 12-17 (Easier introduction!)
  - Petropoulos, F & Kourentzes, N. Forecast combinations for intermittent demand. *Journal of the Operational Research Society* 66.6 (2014): 914-924. (Intermittent)
  - Kourentzes, N. & Petropoulos, F. Forecasting with multivariate temporal aggregation: The case of promotional modelling. *International Journal of Production Economics* (2015). (Promotional modelling)
  - R package on CRAN: **MAPA** (and **tsintermittent** for slow movers)
  - All papers, code and examples available on my website (http://nikolaos.kourentzes.com)

- **Temporal Hierarchies** → Working paper at my blog! (R code out soon)

LCF

Conclusions

- Temporal hierarchies provide a new class of hierarchical forecasts that can be produced for any time series.
- Applicable to forecasts produced by any means → theoretically elegant hierarchical combination of forecasts.
- Joins operational, tactical and strategic decision making by reconciling forecasts → satisfies a business need that has remained unmet
- Potential to increase forecasting accuracy and mitigate modelling uncertainty
- Combining cross-sectional and temporal hierarchies: forecasts reconciled across conventional hierarchy and forecast horizons → `one-number' forecast → superior decision making.

LCF

Thank you for your attention!
Questions?

Published, working papers and code available at my blog!

Nikolaos Kourentzes

email: n.kourentzes@lancaster.ac.uk

blog: http://nikolaos.kourentzes.com

Lancaster University Management School | Lancaster Centre for Forecasting

Appendix

Detailed M3 results for temporal hierarchies

Lancaster University Management School | Lancaster Centre for Forecasting

Some evidence that it actually works!

M3 quarterly dataset

% error change over base (ETS: BU, WLS$_H$, WLS$_V$, WLS$_S$) — % error change over base (ARIMA: BU, WLS$_H$, WLS$_V$, WLS$_S$)

Aggregation level	h	ETS Base	ETS BU	ETS WLS$_H$	ETS WLS$_V$	ETS WLS$_S$	ARIMA Base	ARIMA BU	ARIMA WLS$_H$	ARIMA WLS$_V$	ARIMA WLS$_S$
RMAE											
Annual	1	-	-20.9	-22.7	**-22.8**	-22.7	-	-27.7	-27.8	**-28.0**	-22.8
Semi-annual	3	-	-4.5	-6.0	**-6.2**	-4.8	-	-3.3	-3.9	**-4.4**	2.5
Quarterly	6	-	0.0	-0.2	**-1.1**	-0.3	-	0.0	-0.3	**-1.1**	5.5
Average			*-8.5*	*-9.6*	***-10.0***	*-9.3*		*-10.3*	*-10.7*	***-11.1***	*-4.9*
MASE											
Annual	2	1.5	-14.6	-15.8	-15.9	**-17.2**	1.6	-20.6	**-22.1**	-22.1	-19.7
Semi-annual	4	1.3	-6.8	-7.8	-7.9	**-9.1**	1.2	-2.9	**-4.7**	-4.5	-1.6
Quarterly	8	1.2	0.0	-0.6	-1.1	**-2.6**	1.2	0.0	**-1.6**	-1.4	1.5
Average			*-7.1*	*-8.1*	*-8.3*	***-9.6***		*-7.8*	***-9.5***	*-9.3*	*-6.6*

BU: Bottom-Up; **WLS$_H$**: Hierarchy scaling; **WLS$_V$**: Variance scaling; **WLS$_S$**: Structural scaling

756 series, forecast t+1 - t+8 quarters ahead

Some evidence that it actually works!

M3 monthly dataset

Aggregation level	h	ETS (% error change over base) Base	BU	WLS_H	WLS_V	WLS_S	ARIMA (% error change over base) Base	BU	WLS_H	WLS_V	WLS_S
RMAE											
Annual	1	-	-19.6	-22.0	-22.0	**-25.1**	-	-28.6	-33.1	-32.8	**-33.4**
Semi-annual	3	-	0.6	-4.0	-3.6	**-5.4**	-	-3.4	-8.2	-8.3	**-9.9**
Four-monthly	4	-	2.0	-2.4	-2.2	**-3.0**	-	-1.7	-5.5	-5.9	**-6.7**
Quarterly	6	-	2.4	-1.6	-1.7	**-2.8**	-	-3.6	-7.2	-8.1	**-9.1**
Bi-monthly	9	-	0.7	-2.9	-3.3	**-4.3**	-	-1.5	-4.4	-5.3	**-6.3**
Monthly	18	-	0.0	-2.2	-3.2	**-3.9**	-	0.0	-0.9	-2.9	**-3.4**
Average			*-2.3*	*-5.9*	*-6.0*	***-7.4***		*-6.5*	*-9.9*	*-10.5*	***-11.5***
MASE											
Annual	1	1.11	-12.1	-17.9	-17.8	**-18.5**	1.3	-25.4	-29.9	-29.9	**-30.2**
Semi-annual	3	1.03	0.0	-6.3	-6.0	**-6.9**	1.1	-2.9	-8.1	-8.2	**-9.4**
Four-monthly	4	0.90	3.1	-3.2	-3.0	**-3.4**	0.9	-1.8	-6.2	-6.5	**-7.1**
Quarterly	6	0.93	3.2	-2.8	-2.7	**-3.4**	1.0	-2.6	-6.9	-7.4	**-8.1**
Bi-monthly	9	0.90	2.7	-2.9	-3.0	**-3.7**	0.9	-1.3	-5.0	-5.5	**-6.3**
Monthly	18	0.89	0.0	-3.7	-4.6	**-5.0**	0.9	0.0	-1.9	-3.2	**-3.7**
Average			*-0.5*	*-6.1*	*-6.2*	***-6.8***		*-5.7*	*-9.7*	*-10.1*	***-10.8***

BU: Bottom-Up; **WLS_H**: Hierarchy scaling; **WLS_V**: Variance scaling; **WLS_S**: Structural scaling

1453 series, forecast t+1 - t+18 months ahead

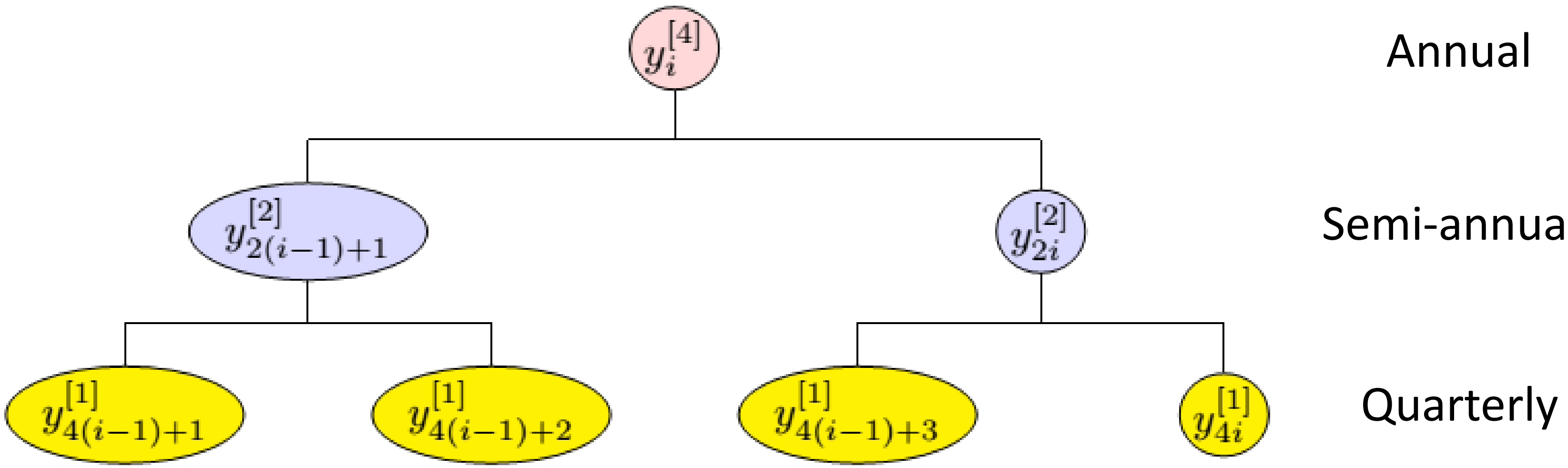
Appendix

Calculation details for temporal hierarchies

Lancaster University Management School | Lancaster Centre for Forecasting

Temporal Hierarchies - Notation

Non-overlapping temporal aggregation to k[th] level: $$y_j^{[k]} = \sum_{t=t^\star+(j-1)k}^{jk} y_t,$$



$$\boldsymbol{y}_i^{[k]} = (y_{M_k(i-1)+1}^{[k]}, y_{M_k(i-1)+2}^{[k]}, \ldots, y_{M_k i}^{[k]})'.$$

Observations at each aggregation level

LCF

Temporal Hierarchies - Notation

Collecting the observations from the different levels in a column:

$$\boldsymbol{y}_i = \left(y_i^{[m]}, \ldots, \boldsymbol{y}_i^{[k_3]'}, \boldsymbol{y}_i^{[k_2]'}, \boldsymbol{y}_i^{[1]'}\right)'$$

We can define a “summing” matrix **S** so that:

$$\boldsymbol{y}_i = \boldsymbol{S}\boldsymbol{y}_i^{[1]}$$

Lowest level observations

$$\boldsymbol{S} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Annual

Semi-annual

Quarter

Example: Monthly

$$\boldsymbol{S} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & & & & \vdots & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & & & & \vdots & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ & & & & & & \boldsymbol{I}_{12} & & & & & \end{bmatrix}$$

$$\boldsymbol{y}_i = \boldsymbol{S}\boldsymbol{y}_i^{[1]}$$

$$\boldsymbol{y}_i = \left(y_i^{[12]}, \boldsymbol{y}_i^{[6]}, \boldsymbol{y}_i^{[4]}, \boldsymbol{y}_i^{[3]}, \boldsymbol{y}_i^{[2]}, \boldsymbol{y}_i^{[1]}\right)'$$

Aggregation levels k are selected so that we do not get fractional seasonalities

Temporal Hierarchies - Forecasting

We can arrange the forecasts from each level in a similar fashion:

$$\hat{\boldsymbol{y}}_h = (\hat{y}_h^{[m]}, \dots, \hat{\boldsymbol{y}}_h^{[k_3]\prime}, \hat{\boldsymbol{y}}_h^{[k_2]\prime}, \hat{\boldsymbol{y}}_h^{[1]\prime})'$$

The reconciliation model is:

$$\hat{\boldsymbol{y}}_h = \boldsymbol{S}\boldsymbol{\beta}(h) + \boldsymbol{\varepsilon}_h$$

Reconciled forecasts

Summing matrix

Unknown conditional means of the future values at lowest level

$$\boldsymbol{\beta}(h) = \mathrm{E}[\boldsymbol{y}^{[1]}_{\lfloor T/m \rfloor + h} \mid y_1, \dots, y_T]$$

Reconciliation errors how much the forecasts across levels do not agree

The reconciliation error has zero mean and covariance matrix $\boldsymbol{\Sigma}$

Temporal Hierarchies - Forecasting

If $\boldsymbol{\Sigma}$ was known then we can write (GLS estimator):

$$\tilde{\boldsymbol{y}}_h = \boldsymbol{S}\hat{\boldsymbol{\beta}}(h) = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{\Sigma}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{y}}_h = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_h$$

But in general it is not know, so we need to estimate it.

It can be shown that $\boldsymbol{\Sigma}$ is not identifiable (you need to know the reconciled forecasts, before you reconcile them), however:

$$\mathrm{Var}(\boldsymbol{y}_{T+h} - \tilde{\boldsymbol{y}}_h) = \boldsymbol{S}\boldsymbol{P}\boldsymbol{W}\boldsymbol{P}'\boldsymbol{S}'$$

Reconciliation errors

Covariance of forecast errors

So our problem becomes:

$$\tilde{\boldsymbol{y}}_h = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{W}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}^{-1}\hat{\boldsymbol{y}}_h$$

LCF

Temporal Hierarchies - Forecasting

All we need now is an estimation of **W**

$$\Lambda = \frac{1}{\lfloor T/m \rfloor} \sum_{i=1}^{\lfloor T/m \rfloor} e_i e_i'$$

Sample covariance of in-sample errors

In principle this is fine, but its sample size is controlled by the number of top-level (annual) observations. For example 104 observations at weekly level, results in just 2 sample points (2 years).

So the estimation of Λ is typically weak in practice.

LCF

Temporal Hierarchies - Forecasting

We propose three ways to estimate it, with increasing simplifying assumptions.

Using as example quarterly data the approximations are:

Hierarchy variance scaling

$$\mathbf{\Lambda}_H = diag\left(\hat{\sigma}_A^{[4]}, \hat{\sigma}_{SA_1}^{[2]}, \hat{\sigma}_{SA_2}^{[2]}, \hat{\sigma}_{Q_1}^{[1]}, \hat{\sigma}_{Q_2}^{[1]}, \hat{\sigma}_{Q_3}^{[1]}, \hat{\sigma}_{Q_4}^{[1]}\right)^2$$

Diagonal of covariance matrix → less elements to estimate

Series variance scaling

$$\mathbf{\Lambda}_V = diag\left(\hat{\sigma}^{[4]}, \hat{\sigma}^{[2]}, \hat{\sigma}^{[2]}, \hat{\sigma}^{[1]}, \hat{\sigma}^{[1]}, \hat{\sigma}^{[1]}, \hat{\sigma}^{[1]}\right)^2$$

Assume within level equal variances. This is what conventional forecasting does. Increases sample size.

Structural scaling

$$\mathbf{\Lambda}_S = diag\left(4, 2, 2, 1, 1, 1, 1\right)$$

Assume proportional error variances. No need for estimates → can be used when unknown (e.g. expert forecasts).