Forecasting with Temporal Hierarchies

George Athanasopoulosa, Rob J. Hyndmana, Nikolaos Kourentzesb,,
Fotios Petropoulosc

Zhangzhihong
2019/11/22

Contents

1 Introduction

Problem:

- strategic decisions require **long-run forecasts** at an aggregate level, while decisions at the highly dynamic operational level require **short-term** very detailed forecasts.

- these forecasts are produced by dierent approaches .

1 Introduction

This paper:

- introduce a novel approach for time series modelling and forecasting: **Temporal Hierarchies**.This lead to:

(a) reconciled forecasts, supporting better decisions across planning horizons;

(b) Increased forecast accuracy;

(c) mitigating modelling risks.

2 Methodology

Temporal hierarchies

More generally, we are interested in a time series $\{y_t;\ t = 1, \ldots, T\}$ observed at the highest available sampling frequency per year, m, and in the k-aggregates that can be constructed where k is a factor of m. The various aggregated series can be written as

$$y_j^{[k]} = \sum_{t=t^\star+(j-1)k}^{t^\star+jk-1} y_t, \qquad (1)$$

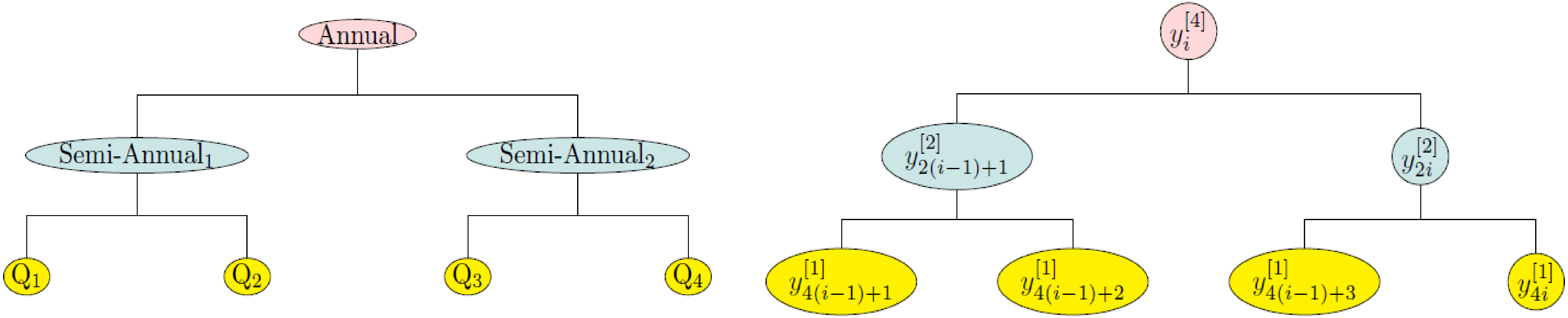


Figure 1: *Temporal hierarchy for quarterly series.*

2 Methodology

Temporal hierarchies

We stack the observations for each aggregation level below the annual level in column vectors such that

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

Collecting these in one column vector, $\boldsymbol{y}_i = \left(y_i^{[m]}, \dots, \boldsymbol{y}_i^{[k_3]'}, \boldsymbol{y}_i^{[k_2]'}, \boldsymbol{y}_i^{[1]'}\right)'$, we can write

$$\boldsymbol{y}_i = \boldsymbol{S}\boldsymbol{y}_i^{[1]} \tag{3}$$

where for quarterly data

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

is referred to as the “summing” matrix (drawing from the work of Hyndman et al. 2011) and $\boldsymbol{y}_i^{[1]}$ is the vector of observations of the time series observed at the highest available frequency.

2 Methodology

Forecasting framework

conditional on $\lfloor T/k \rfloor$ observations. We refer to these as base forecasts. For each forecast horizon h we stack the forecasts the same way as the data, i.e.,

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

where each $\hat{\boldsymbol{y}}_h^{[k]} = (\hat{y}^{[k]}_{M_k(h-1)+1}, \hat{y}^{[k]}_{M_k(h-1)+2}, \dots, \hat{y}^{[k]}_{M_k h})'$ is of dimension M_k and therefore $\hat{\boldsymbol{y}}_h$ is of dimension $\sum_{\ell=1}^{p} k_\ell$.

We can write the $\boxed{\text{base forecasts}}$ as

$$\hat{\boldsymbol{y}}_h = \boldsymbol{S}\boldsymbol{\beta}(h) + \boldsymbol{\varepsilon}_h \tag{5}$$

where $\boldsymbol{\beta}(h) = \mathrm{E}[\boldsymbol{y}^{[1]}_{\lfloor T/m \rfloor + h} \mid y_1, \dots, y_T]$ is the unknown conditional mean of the future values of the most disaggregated observed series, and $\boldsymbol{\varepsilon}_h$ represents the "reconciliation error", the difference between the base forecasts $\hat{\boldsymbol{y}}_h$ and their expected value if they were reconciled. We assume that $\boldsymbol{\varepsilon}_h$ has zero mean and covariance matrix $\boldsymbol{\Sigma}$. We refer to (5) as the temporal reconciliation regression model. It is analogous to the cross-sectional hierarchical reconciliation

2 Methodology

Forecasting framework

If $\boldsymbol{\Sigma}$ was known, the generalised least squares (GLS) estimator of $\boldsymbol{\beta}(h)$ would lead to reconciled forecasts given by

$$\tilde{\boldsymbol{y}}_h = \boldsymbol{S}\hat{\boldsymbol{\beta}}(h) = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{\Sigma}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{y}}_h = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_h, \tag{6}$$

where $\boldsymbol{P} = (\boldsymbol{S}'\boldsymbol{\Sigma}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{\Sigma}^{-1}$. The reconciled forecasts would be optimal in that the base forecasts are adjusted by the least amount (in the sense of least squares) so that these become reconciled. In general, $\boldsymbol{\Sigma}$ is not known and needs to be estimated. Hyndman et al. (2011) and Athanasopoulos et al. (2009) avoid estimating $\boldsymbol{\Sigma}$ by using ordinary least squares (OLS), replacing $\boldsymbol{\Sigma}$ by $\sigma^2\boldsymbol{I}$ in (6). Recently, Wickramasuriya et al. (2015) show that $\boldsymbol{\Sigma}$ is not identifiable.

2 Methodology

Forecasting framework

ing $\boldsymbol{\Sigma}$ by $\sigma^2\boldsymbol{I}$ in (6). Recently, Wickramasuriya et al. (2015) show that $\boldsymbol{\Sigma}$ is not identifiable. Assuming that the base forecasts are unbiased they show that for the forecast errors of the reconciled forecasts,

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

where $\boldsymbol{W} = \mathrm{Var}(\boldsymbol{y}_{\lfloor T/m \rfloor+h} - \hat{\boldsymbol{y}}_h)$ is the covariance matrix of the base forecast errors. By minimizing the variances of the reconciled forecast errors, they propose an estimator which results in unbiased reconciled forecasts given by

$$\tilde{\boldsymbol{y}}_h = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{W}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}^{-1}\hat{\boldsymbol{y}}_h. \tag{7}$$

Note that this coincides with the GLS estimator in (6) but with a different covariance matrix. Defining the in-sample one-step-ahead base forecast errors as $\boldsymbol{e}_i = \left(e_i^{[m]}, \ldots, \boldsymbol{e}_i^{[k_3]'}, \boldsymbol{e}_i^{[k_2]'}, \boldsymbol{e}_i^{[1]'}\right)'$, for $i = 1, \ldots, \lfloor T/m \rfloor$, the sample covariance estimator of $\boldsymbol{W}$ is given by,

$$\boldsymbol{\Lambda} = \frac{1}{\lfloor T/m \rfloor} \sum_{i=1}^{\lfloor T/m \rfloor} \boldsymbol{e}_i\boldsymbol{e}_i', \tag{8}$$

2 Methodology

Forecasting framework

To better illustrate the differences between the three proposed scaling methods we show the different matrices for quarterly data:

$$
\begin{aligned}
\mathbf{\Lambda}_H &= \mathrm{diag}\big(\hat{\sigma}_A^{[4]}, \hat{\sigma}_{SA_1}^{[2]}, \hat{\sigma}_{SA_2}^{[2]}, \hat{\sigma}_{Q_1}^{[1]}, \hat{\sigma}_{Q_2}^{[1]}, \hat{\sigma}_{Q_3}^{[1]}, \hat{\sigma}_{Q_4}^{[1]}\big)^2, \\
\mathbf{\Lambda}_V &= \mathrm{diag}\big(\hat{\sigma}^{[4]}, \hat{\sigma}^{[2]}, \hat{\sigma}^{[2]}, \hat{\sigma}^{[1]}, \hat{\sigma}^{[1]}, \hat{\sigma}^{[1]}, \hat{\sigma}^{[1]}\big)^2, \\
\mathbf{\Lambda}_S &= \mathrm{diag}\big(4, 2, 2, 1, 1, 1, 1\big),
\end{aligned}
$$

3 Application

We perform an extensive empirical evaluation of forecasting with temporal hierarchies using the 1,428 monthly and 756 quarterly time series from the M3 competition (Makridakis and Hibon, 2000). In order to ensure the comparability of our results with the original competition, we withhold as test samples the last 18 observations of each monthly series and the last 8 observations of each quarterly series.

The forecasts are evaluated using the Relative Mean Absolute Error (RMAE) (see Davydenko and Fildes, 2013) and the Mean Absolute Scaled Error (MASE) (see Hyndman and Koehler, 2006). Both these measures permit calculating forecasting accuracy across time series of different scales. For h-step-ahead forecasts:

$$\text{RMAE} = \frac{\text{MAE}^a}{\text{MAE}^{\text{Base}}} \tag{10}$$

where $\text{MAE}^a = \frac{1}{h}\sum_{j=1}^{h}|y_j - \hat{y}_j|$ is the mean absolute error for forecasts of method a, MAE^{Base} is the mean absolute error of the base forecasts, y_j and $\hat{y}_j$ are the actual and forecast values at period j respectively; and

$$\text{MASE} = \frac{\text{MAE}^a}{Q} \tag{11}$$

$Q = \frac{1}{T-m}\sum_{t=1}^{T}|y_t - y_{t-m}|$ is the scaling factor where m is the sampling frequency per year.

3 Application

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	1.0	-19.6	-22.0	-22.0	**-25.1**	1.0	-28.6	-33.1	-32.8	**-33.4**
Semi-annual	3	1.0	0.6	-4.0	-3.6	**-5.4**	1.0	-3.4	-8.2	-8.3	**-9.9**
Four-monthly	4	1.0	2.0	-2.4	-2.2	**-3.0**	1.0	-1.7	-5.5	-5.9	**-6.7**
Quarterly	6	1.0	2.4	-1.6	-1.7	**-2.8**	1.0	-3.6	-7.2	-8.1	**-9.1**
Bi-monthly	9	1.0	0.7	-2.9	-3.3	**-4.3**	1.0	-1.5	-4.4	-5.3	**-6.3**
Monthly	18	1.0	0.0	-2.2	-3.2	**-3.9**	1.0	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	1.1	-12.1	-17.9	-17.8	**-18.5**	1.3	-25.4	-29.9	-29.9	**-30.2**
Semi-annual	3	1.0	0.0	-6.3	-6.0	**-6.9**	1.1	-2.9	-8.1	-8.2	**-9.4**
Four-monthly	4	0.9	3.1	-3.2	-3.0	**-3.4**	0.9	-1.8	-6.2	-6.5	**-7.1**
Quarterly	6	0.9	3.2	-2.8	-2.7	**-3.4**	1.0	-2.6	-6.9	-7.4	**-8.1**
Bi-monthly	9	0.9	2.7	-2.9	-3.0	**-3.7**	0.9	-1.3	-5.0	-5.5	**-6.3**
Monthly	18	0.9	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***

3 Application

7. Case study: predicting accident and emergency service demand

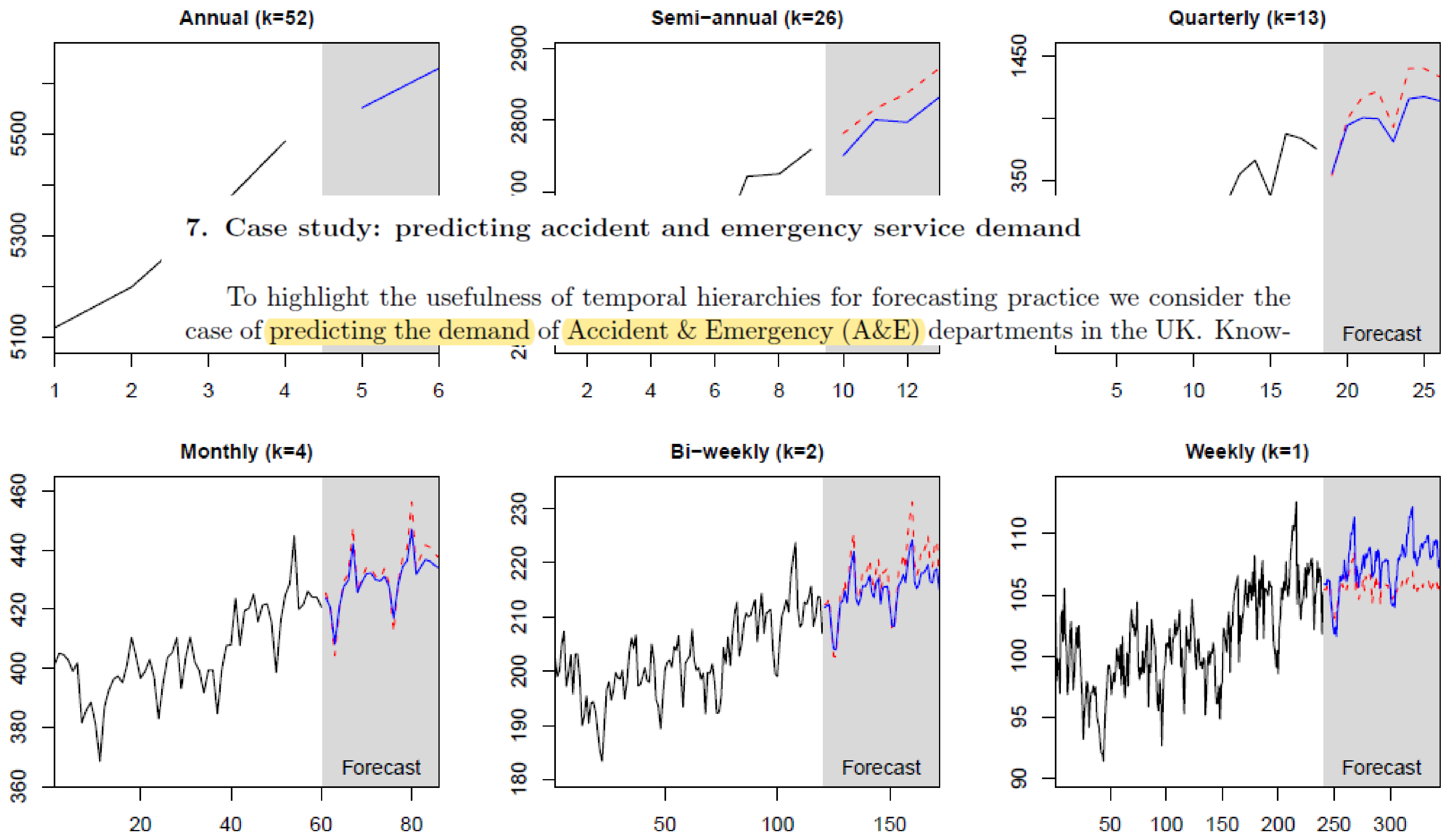
To highlight the usefulness of temporal hierarchies for forecasting practice we consider the case of predicting the demand of Accident & Emergency (A&E) departments in the UK. Know-

Each time series is split into two subsets, a training set and an out-of-sample evaluation set. The latter spans the last 52 weeks of the time series, while all the remaining observations are used for fitting appropriate forecasting models. Each series is forecasted using ARIMA, which is specified as detailed in Section 5. Forecasts are first generated conventionally to give the base forecasts, and then temporal hierarchies are used with series variance scaling (WLS_V). The alternative scaling methods result in similar performance and therefore are not reported here. We use MASE to track the accuracy in predicting 1, 4 and 13 weeks ahead, matching

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

3 Application

Figure 6: *Predictions for the* Total Emergency Admissions *time series (in 000s). Forecasts are plotted in the greyed area. Dashed line (--) is used for the base forecasts and solid line (—) for the temporal hierarchy forecasts.*



7. Case study: predicting accident and emergency service demand

To highlight the usefulness of temporal hierarchies for forecasting practice we consider the case of predicting the demand of Accident & Emergency (A&E) departments in the UK. Know-

THANKS