Hierarchical Forecasting

Name of First Author and Name of Second Author

1 Introduction

Structure of the data with an macroeconomic example

Temporal hierarchies

Forecasting hierarchical time series

- Importance of coherency
- Point forecasting
- Probabilistic forecasting

2 Hierarchical forecasting in macroeconomic data

The key macroeconomic indicators such as Gross Domestic Product (GDP), inflation and monetary policies which are used to study the behavior and performance of an economy as a whole are it self aggregates of various other components. For example, if we take the GDP growth, it is the aggregate of consumption, government expenditure, investments and net exports. These four components are again aggregates of some sub components. When we collect data for each of these individual variables over some time period, we will observe a collection of multiple time series that are bounded with some aggregation constraints. Thus the macroeconomic data are naturally forming cross sectional hierarchical time series.

———————————————

Name of First Author
Name, Address of Institute, e-mail: name@email.address

Name of Second Author
Name, Address of Institute e-mail: name@email.address

If the interest is on a single macroeconomic variable along different time granularities, then it can be considered as a temporal hierarchy. For example, suppose we have monthly consumer product index (CPI) of a particular country. The quarterly CPI is then the aggregate of corresponding monthly CPI of each quarter. Similarly the yearly CPI is the aggregate of quarterly CPI of each year. Hence it will form a temporal hierarchy.

Macroeconomic forecasts are crucial for economic and business activities of any economy. Therefore this area of study has a long history in literature. Econometricians have developed various approaches for getting reliable economic forecasts using macroeconomic data. However, the information of aggregation structure in real data is limitedly used in literature. Moreover, having coherent forecasts will help the economists and policy makers for align decision making that impact for the whole economy. Therefore, our focus in this chapter is to introduce hierarchical forecasting methods for macroeconomic forecasting particularly for cross-sectional hierarchical data structures.

Obtaining coherent forecasts are independent from the forecasting models. That means forecasters were given the freedom to use any reliable forecasting method to obtain the forecasts for individual series in the hierarchy. Getting coherent forecasts is a post-processing technique which ensures the aggregation properties are preserved in the forecasts.

briefly discuss the point forecasts as well as probabilistic forecasts in the sense of macroeconomic data

3 Notations and preliminaries

This section discusses some preliminaries related to hierarchical time series. First let us start with introducing some notations.
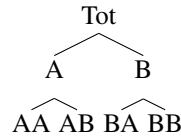

Fig. 1 Two level hierarchical diagram

Figure 1 represents a two level cross sectional hierarchical time series, where the hierarchy is defined across time series for a particular time t. In this hierarchy, *Tot* represents the most aggregate level series. We refer to this level as level 0 The observed value at a particular time t of this series is denoted by $y_{Tot,t}$. *Tot* is then disaggregated to *A* and *B* which lies in the level 1 of the hierarchy. Ob-

served values of these at time t is denoted by $y_{A,t}$ and $y_{B,t}$. Further A and B are disaggregated to AA, AB and BA, BB respectively. These series lies in the level 2 of the hierarchy and t^{th} observations are denoted by $y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}$. Then $\mathbf{y}_t = [y_{Tot,t}, y_{A,t}, y_{B,t}, y_{C,t}, y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]'$ is a vector that consists all the observations of the hierarchy at time t. Moreover, $\mathbf{b}_t = [y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]'$ is a vector consisting the observations of bottom level series at time t. n and m denote the number of total series in the hierarchy and the number of bottom level series respectively. Thus, $\mathbf{y}_t \in \mathbb{R}^n$ and $\mathbf{b}_t \in \mathbb{R}^m$.

Due to the aggregation nature of the hierarchical time series we can write,

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t, \tag{1}$$

where $\mathbf{S}$ is a $(n \times m)$ constant matrix for a given hierarchical structure. For the hierarchy given in figure 1,

$$\mathbf{S} = \begin{pmatrix} 1\ 1\ 1\ 1 \\ 1\ 1\ 0\ 0 \\ 0\ 0\ 1\ 1 \\ I_4 \end{pmatrix}. \tag{2}$$

where $\mathbf{I}_4$ is a 4-dimensional identity matrix.

Columns of $\mathbf{S}$ span the linear subspace of $\mathbb{R}^n$ for which all constraints hold. This subspace is also referred to as *coherent subspace* and denoted by $\mathfrak{s}$. Pre-multiplying a vector in $\mathbb{R}^m$ by $\mathbf{S}$ will linearly transform that vector to another vector which lies in $\mathfrak{s}$. Therefore, although it is an n-dimensional vector, it actually lies in $\mathfrak{s}$. In earlier studies, $\mathbf{S}$ is also referred to as the "summing matrix" since it aggregates the observations of bottom level to the corresponding upper levels.

4 Hierarchical point forecasting

Efficient and accurate coherent point forecasting methods have been well studied and established in hierarchical literature over the past few decades. This section is devoted for a comprehensive discussion on these methodologies.

4.1 Coherent point forecasts

Coherent forecasts preserve the aggregation structure of hierarchical time series. That is the forecasts of lower level series aggregates up to their corresponding upper level series of the hierarchy. In geometric language, the coherent forecasts lie in the coherent subspace $\mathfrak{s}$. Suppose $\check{\mathbf{y}}_{t+h|t}$ consists point forecasts of all series in the hierarchy at time $t+h$ conditional on the past observations upto time t. Elements of $\check{\mathbf{y}}_{t+h|t}$ are stacked in the same order as $\mathbf{y}_t$. Then, $\check{\mathbf{y}}_{t+h|t}$ is said to be coherent if $\check{\mathbf{y}}_{t+h|t} \in \mathfrak{s}$.

Let us consider the smallest possible hierarchy with two bottom level series, A and B that add up to the top level Tot. Suppose $\check{\mathbf{y}}_{t+h|t}$ of this hierarchy is given by $\check{\mathbf{y}}_{t+h|t} = [\check{y}_{Tot,t+h|t}, \check{y}_{A,t+h|t}, \check{y}_{B,t+h|t}]$. Due to the aggregation structure we have $\check{y}_{Tot,t+h|t} = \check{y}_{A,t+h|t} + \check{y}_{B,t+h|t}$. This implies that, even though $\check{y}_{Tot,t+h|t} \in \mathbb{R}^3$, the points actually lie in $\mathfrak{s} \subset \mathbb{R}^3$, which is a two dimensional subspace within $\mathbb{R}^3$ space.

The most earlier studies on hierarchical literature involves forecasting series only in one layer of the hierarchy and then aggregate or disaggregate through the corresponding structure to get coherent forecasts. Let us briefly discuss about these traditional methods.

4.1.1 Bottom-up approach

In bottom-up approach, forecasts of the lowest level series are first generated and these are aggregated to get the forecasts at upper levels of the hierarchy (Dunn et al. 1976). That is, considering the hierarchy given in Figure 1, first the forecasts for series AA, AB, BA and BB are obtained. Then the forecasts of series A is obtained by simply adding the forecasts of AA and AB. Similarly the forecasts of B is obtained by adding the forecasts of BA and BB. The forecasts of Tot series is then the aggregation of A and B which in turn the sum of forecasts of AA, AB, BA and BB.

In terms of notations, let $\hat{\mathbf{b}}_{t+h|t} \in \mathbb{R}^m$ consists h-step ahead forecasts of the bottom levels series. i.e. $\hat{\mathbf{b}}_{t+h|t} = (\hat{y}_{AA,t+h|t}, \hat{y}_{AB,t+h|t}, \hat{y}_{BA,t+h|t}, \hat{y}_{BB,t+h|t})$, where, $\hat{y}_{i,t+h|t}$ is the h-step ahead forecasts of ith series. Then, the bottom-up forecasts are given by,

$$\check{\mathbf{y}}_{t+h|t}^{BU} = \mathbf{S}\hat{\mathbf{b}}_{t+h|t}. \tag{3}$$

Even though this method looks easy to construct, it is less reliable. In fact, bottom up approach provides accurate forecasts only if the bottom level series of the hierarchy are accurately forecast. However, if the bottom level series are highly volatile or too noisy, then they are challenging to forecast. Thus the bottom-up approach would produce unreliable point forecasts.

4.1.2 Top-down approach

In contrast to the bottom-up method, the top-down approach involves forecasting the most aggregated series first and then disaggregating these forecasts down the hierarchy.

In general, all top-down forecasts can be written as,

$$\check{\mathbf{y}}_{t+h|t}^{TD} = \mathbf{S}\hat{y}_{Tot,t+h|t}\mathbf{p}, \tag{4}$$

where $\mathbf{p} = (p_1, ..., p_m)$ is a m dimensional vector of proportions consisting the disaggregation proportions of the bottom level series with respect to the top level series.

Usually, these proportions are calculated based on observed data, which is referred to as historical proportions. Gross & Sohl (1990) provide a comprehensive summary on using historical proportions in this context. Commonly there are two approaches of using historical proportions. One is to use the average of historical proportions of desired bottom level series relative to the total series. That is, the historical proportion of jth series is given by,

$$p_j = \frac{1}{T} \sum_{t=1}^{T} \frac{Y_{j,t}}{Y_{tot,t}}.$$ (5)

The second approach is to use the proportion of average of the bottom level series $y_{j,t}$ over the time $t = 1, ..., T$ relative to that of the total series $y_{Tot,t}$. i.e.

$$p_j = \frac{\frac{1}{T} \sum_{t=1}^{T} y_{j,t}}{\frac{1}{T} \sum_{t=1}^{T} y_{Tot,t}}$$ (6)

However, the largest limitation of these approaches is that its inability to reflect the characteristics of individual series such as trends, seasonality or other special events, in the forecasts of disaggregate levels. This will mainly effect the hierarchies with series having different patterns in different levels.

To overcome this limitation, Athanasopoulos et al. (2009) introduced a new top-down approach which disaggregate the top level forecasts according to the proportions of forecasts rather than historical proportions. They found that their method outperforms the conventional top-down approach through an empirical application.

Recently, Mirčetić et al. (2017) proposed a modified approach to the conventional top-down methods, which involves forecasting h-step ahead proportions of bottom level series relative to the top level series. That is, let the ratio between the jth bottom level series and the top level series is given by, $p_{j,t} = y_{j,t}/y_{Tot,t}$, for $t = 1, ..., T$. Then this series of proportions will be forecast for the h-step ahead period, which is denoted by $\hat{p}_{j,T+h|T}$. These proportions will be then used as the disaggregate proportions to construct coherent top-down forecasts. Thus giving the elements of vector **p** as,

$$p_j = \hat{p}_{j,T+h|T}.$$ (7)

This approach is much reliable, since it calculates the proportions as a function of time, rather than using the simple averages in traditional top-down methods.

As shown by Hyndman et al. (2011), top-down methods are adding a bias component to each disaggregate level. Therefore, all these top-down approaches are producing biased forecasts even if the top level base forecasts are unbiased.

4.1.3 Middle-out approach

A compromise between these two approaches is the middle-out method which entails forecasting each series of a selected middle level in the hierarchy and then

forecasting upper levels by the bottom-up method and lower levels by the top-down method.

Since this is a hybrid approach of bottom-up and top-down approaches it still remains the limitations of those two.

4.2 Point forecast reconciliation

The common limitation of traditional hierarchical forecasting methods is that the loss of structural information of individual series, as it precisely forecasting only one level of series in the hierarchy.

As an alternative to these traditional methods, Hyndman et al. (2011) proposed to utilize the information from all levels of the hierarchy to obtain coherent point forecasts in a two stage process. In the first stage, the forecasts of all series are independently obtained by fitting univariate models for individual series in the hierarchy. It is very unlikely that these forecasts are coherent. Thus in the second stage, these forecasts are optimally combined through a regression model to obtain coherent forecasts. This second step is referred to as "reconciliation" since it takes a set of incoherent forecasts and revises them to be coherent.

This was the first study to introduced the concept of forecast reconciliation in the hierarchical literature. Since this approach starts from the individual forecasts, it uses all the relevant information from the hierarchy in producing coherent forecasts. Thus improves the drawbacks from traditional approaches, in particularly the loss of information. Further reconciliation step is independent from the initial step of getting individual forecasts for all series in the hierarchy. Therefore, forecasters are given the choice to use the best possible forecasting method to get forecasts for all individual series. However the unbiasedness property of the reconciled forecasts will be preserved only if all the individual forecast are unbiased. We will discuss this property as we follows in this section.

Suppose we fit univariate time series models for all series based on data upto time t and obtained h-step ahead forecasts. These forecasts are then formed in a vector $\hat{\mathbf{y}}_{t+h|t}$ by stacking the forecasts at each node in the same order as $\mathbf{y}_t$. Since these do not satisfy the aggregate constraints of the hierarchy, they are often referred to as "incoherent" forecasts. In some texts these also referred to as "base" forecasts.

Reconciliation of $\hat{\mathbf{y}}_{t+h|t}$ involves, revising these forecasts such that they lies in the coherent subspace. We do this by projecting $\hat{\mathbf{y}}_{t+h|t}$ onto the coherent subspace through the projection matrix $\mathbf{SG}$ for any $m \times n$ matrix $\mathbf{G}$. Let $\tilde{\mathbf{y}}_{t+h|t}$ be a vector that contains reconciled forecasts. Then,

$$\tilde{\mathbf{y}}_{t+h|t} = \mathbf{SG}\hat{\mathbf{y}}_{t+h|t}. \tag{8}$$

Pre-multiplying the incoherent forecasts, $\hat{\mathbf{y}}_{t+h|t}$ by $\mathbf{G}$ will maps $\hat{\mathbf{y}}_{t+h|t}$ to the $\mathbb{R}^m$ space. This will gives the reconciled forecasts of bottom level series of the hier-

archy. Next these forecasts will be mapped to the coherent subspace by the pre-multiplication of $\mathbf{S}$ matrix.

It is also worth discussing the unbiased property of these reconciled point forecasts. Let us assume the incoherent forecasts, $\hat{\mathbf{y}}_{t+h|t}$ are unbiased. That is, $E_{1:t}(\hat{\mathbf{y}}_{t+h|t}) = \mu_{t+h|t}$ where $\mu_{t+h|t} = E_{1:t}[\mathbf{y}_{t+h}|\mathbf{y}_1,...,\mathbf{y}_t]$ is the true mean of the hierarchy. Here the expectation is taken over the observed data upto time t. Then for any $\mathbf{G}$ such that $\mathbf{SGS} = \mathbf{S}$ produces unbiased reconciled forecasts (Hyndman et al. (2011)). Further any reconciliation method that define through a projection $\mathbf{SG}$ will always produce unbiased reconciled forecasts given that the incoherent forecasts are unbiased (Gamakumara et al. 2018).

4.3 Existing point forecast reconciliation methods

Structures of $\mathbf{G}$ define different reconciliation methods. First let us reconsider the traditional bottom up approach. Suppose we have a set of incoherent forecasts, $\hat{\mathbf{y}}_{t+h|t}$. Following equation (3), we can write, $\check{\mathbf{y}}_{t+h|t}^{BU} = \mathbf{S}\left(0_{(m \times n-m)} \; \mathbf{I}_m\right)\hat{\mathbf{y}}_{t+h|t}$. where $\mathbf{G} = \left(0_{(m \times n-m)} \; \mathbf{I}_m\right)$. Since $\mathbf{SG}$ in this case is a projection matrix, bottom-up method also produces unbiased coherent forecasts. However, it is not always a preferred approach since it ignores the information from aggregate levels of the hierarchy.

4.3.1 OLS reconciliation

Intuitively, the reconciliation starts with a set of incoherent forecasts, $\hat{\mathbf{y}}_{t+h|t}$. Hyndman et al. (2011) proposed to optimally combine these incoherent forecasts through the following regression model.

$$\hat{\mathbf{y}}_{t+h|t} = \mathbf{S}\beta_{t+h|t} + \varepsilon_{t+h|t}, \tag{9}$$

where $\beta_{t+h|t} = E[\mathbf{b}_{t+h}|\mathbf{b}_1,.....,\mathbf{b}_t]$ is the unknown conditional mean of the bottom level series at time $t+h$ and $\varepsilon_{t+h|t}$ is the reconciliation error with mean zero and variance $\mathbf{V}$. The ordinary least squares (OLS) solution for the above regression model gives,

$$\hat{\beta}_{t+h|t} = (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\hat{\mathbf{y}}_{t+h|t}, \tag{10}$$

and thus giving the reconciled forecasts,

$$\tilde{\mathbf{y}}_{t+h|t} = \mathbf{S}\hat{\beta}_{t+h|t} = \mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\hat{\mathbf{y}}_{t+h|t}. \tag{11}$$

$(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'$ is the corresponding $\mathbf{G}$ matrix in OLS reconciliation. Further in this reconciliation, the incoherent forecasts are orthogonally projected to the coherent subspace $\mathfrak{s}$. Thus it will minimises the Euclidean distance between $\hat{\mathbf{y}}_{t+h|t}$ and $\tilde{\mathbf{y}}_{t+h|t}$.

In a later study by Wickramasuriya et al. (2018) showed that the variance covariance matrix of the reconciliation error, $\varepsilon_{t+h|t}$ in the above regression model is not identifiable and thus the GLS solution that Hyndman et al. (2011) proposed for the same model is unattainable.

4.4 MinT reconciliation

As an improved work of Hyndman et al. (2011), Wickramasuriya et al. (2018) find an unique analytical solution for producing unbiased reconciled point forecasts, by minimising the sum of variances of reconciled forecast errors. Suppose the variance of h-step ahead base forecast errors is denoted by, $Var(\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h|t}|\mathbf{y}_1,...,\mathbf{y}_t) := \mathbf{W}_{t+h|t}$. Wickramasuriya et al. (2018) first showed that the variance of the reconciled forecast errors, i.e $Var(\mathbf{y}_{t+h} - \tilde{\mathbf{y}}_{t+h|t}|\mathbf{y}_1,...,\mathbf{y}_t) := \tilde{\mathbf{W}}_{t+h|t}$ is given by,

$$\tilde{\mathbf{W}}_{t+h|t} = \mathbf{SGW}_{t+h|t}\mathbf{G}'\mathbf{S}' \tag{12}$$

for any choice of $\mathbf{G}$. Then they minimize the trace of $\tilde{\mathbf{W}}_{t+h|t}$ with respect to $\mathbf{SGS} = \mathbf{S}$ to obtain an optimal $\mathbf{G}$. By imposing the unbiased constraint they ensure that the resulting reconciled forecasts are unbiased. The closed form solution to this minimization problem is given by,

$$\mathbf{G} = (\mathbf{S}'\mathbf{W}_{t+h|t}^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_{t+h|t}^{-1}, \tag{13}$$

and they named this as MinT approach. The reconciled point forecasts are then given by,

$$\tilde{\mathbf{y}}_{t+h|t} = \mathbf{SG}\hat{\mathbf{y}}_{t+h|t} = \mathbf{S}(\mathbf{S}'\mathbf{W}_{t+h|t}^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_{t+h|t}^{-1}\hat{\mathbf{y}}_{t+h|t}. \tag{14}$$

In terms of projections, MinT is doing an oblique projection of $\hat{\mathbf{y}}_{t+h|t}$ onto the coherent subspace $\mathfrak{s}$. Further in terms of distances, MinT minimises the Mahalonobis distance between $\hat{\mathbf{y}}_{t+h|t}$ and $\tilde{\mathbf{y}}_{t+h|t}$ as proven by Wickramasuriya et al. (2018).

Unlike OLS, MinT reconciliation imposes the correlation structure of the whole hierarchy to reconcile the point forecasts. However, estimating $\mathbf{W}_{t+h|t}$ is challenging. Wickramasuriya et al. (2018) discussed possible alternative estimates of $\mathbf{W}_{t+h|t}$ and how these lead to different $\mathbf{G}$ matrices. Let us briefly summarise these below.

Suppose we have the in-sample data up to time T. Since the variance covariance matrix of 1-step ahead incoherent forecast errors are approximately proportional to that of h-step ahead incoherent forecasts errors, we have $\mathbf{W}_{T+h|T} = \alpha\mathbf{W}_{T+1|T}$ where $\alpha > 0$. Therefore, to estimating $\mathbf{W}_{T+h|T}$ is sufficient to determine $\mathbf{G}$. If $\mathbf{W}_{T+1}$ is approximated by an identity matrix, then $\mathbf{G}$ in (14) collapses to OLS reconciliation. Following are a few other estimates for $\mathbf{W}_{T+1}$.

MinT(Sample)

The unbiased sample variance covariance matrix of 1-step ahead incoherent forecast errors can be used as the most reasonable estimator for $\mathbf{W}_{T+1|T}$. That is,

$$\hat{\mathbf{W}}_{T+1|T} = \frac{1}{T} \sum_{t=1}^{T} \hat{\mathbf{e}}_{t+1} \hat{\mathbf{e}}'_{t+1}, \tag{15}$$

where $\hat{\mathbf{e}}_{t+1} = \mathbf{y}_{t+1} - \hat{\mathbf{y}}_{t+1|t}$ for $t = 1,...,t-1$, consists the 1-step ahead insample forecast errors.

The resulting $\mathbf{G}$ matrix is referred to as MinT(Sample). This estimator is performing well in small hierarchies. However for the hierarchies with large dimensions, especially when $m > T$, this causes singularity problems.

MinT(Shrink)

Another alternative estimator for $\mathbf{W}_{T+1|T}$ is a shrinkage estimator, which is apparently useful in large dimensional hierarchies. This estimator is given by,

$$\hat{\mathbf{W}}^{shr}_{T+1|T} = \tau \hat{\mathbf{W}}^{D}_{T+1|T} + (1-\tau)\hat{\mathbf{W}}_{T+1|T}, \tag{16}$$

where $\hat{\mathbf{W}}^{D}_{T+1|T}$ is a diagonal matrix comprising diagonal entries of $\hat{\mathbf{W}}_{T+1|T}$ and

$$\tau = \frac{\sum_{i \neq j} \hat{Var}(\hat{r}_{ij})}{\sum_{i \neq j} \hat{r}^2_{ij}}$$

is a shrinkage parameter proposed by Schäfer & Strimmer (2005) where, $\hat{r}_{ij}$ is the ij-th element of sample correlation matrix. In this estimation, the off-diagonal elements of 1-step ahead sample covariance matrix will be shrunk to zero depending on the sparsity.

WLS

If $\mathbf{W}_{T+1|T}$ is approximated by a diagonal matrix with diagonal elements being the variances of incoherent forecast errors, then the resulting reconciliation is referred to as WLS reconciliation. This was first proposed by Hyndman et al. (2016) with reference to the regression model introduced in Hyndman et al. (2011). However a proper theoretical justification for the WLS reconciliation is given through MinT solution.

4.5 GTOP reconciliation

Van Erven & Cugliari (2014) proposed to use a game theoretical approach to produce reconciled forecasts that are at least as good as incoherent forecasts. Initially starting with the best possible incoherent forecasts, they show that the sum of weighted squared loss of these incoherent forecasts will never be less than that for the reconciled forecasts.

This approach does not require unbiasedness assumption on incoherent forecasts and it allows forecasters to choose the best possible incoherent forecasts without depending on any assumptions. However this method doesn't guarantee to produce unbiased reconciled forecasts. Further, there is no closed form solution to this approach and thus would result a high computational cost in producing reconciled forecasts for large scale hierarchies.

Any reconciliation approach via projections such as MinT are preferred as they improve these limitations in the GTOP method. Nevertheless, they guarantee to produce coherent forecasts that are at least as good as incoherent forecasts (Wickramasuriya et al. (2018), Gamakumara et al. (2018)).

Another important feature of hierarchical time series is that they often consist of thousands or millions of individual series and thus imposes computational challenges in implementing any forecasting solutions. The MinT reconciliation method was further generalized to handle these constraints and to scale to large hierarchies by Wickramasuriya et al. (2018).

5 Hierarchical probabilistic forecasting

Point forecasts are limited because they provide no indication of forecast uncertainty. Providing prediction intervals helps (Shang 2017), but a richer description of forecast uncertainty is obtained by estimating the entire forecast distribution. These are often called "probabilistic forecasts" (Gneiting & Katzfuss 2014). For example, McSharry et al. (2005) produced probabilistic forecasts for electricity demand, Ben Taieb et al. (2017) for smart meter data, Pinson et al. (2009) for wind power generation, and Gel et al. (2004), Gneiting et al. (2005) and Gneiting & Raftery (2005) for various weather variables.

Even though the point forecasts in hierarchical time series is well studied and having an established literature, the hierarchical probabilistic forecasting is still an emerging area of interest. There's is only a few studies in published literature. The uncertainly around $\mathbf{y}_{t+h|t}$ is referred to as the probabilistic forecasts in hierarchical time series. Due to the aggregation nature of the data, these should also lie in the coherent subspace $\mathfrak{s}$. If so, we call them as coherent probabilistic forecasts.

Taieb et al. (2017) defined the coherent probabilistic forecasts in terms of convolution. That is, the predictive distributions of a hierarchy is said to be coherent, if the convolution of forecast distributions of disaggregate series is equal to the forecast

distribution of the corresponding aggregate series. Based on this, they proposed an algorithm to produce coherent probabilistic forecasts and applied it to UK electricity smart meter data.

Analogous to point forecast reconciliation, we can reconcile the forecast distributions to get coherent probabilistic forecasts. This will start with an incoherent forecast distribution of the whole hierarchy which was obtained by gathering all information available, and then projecting that to the coherent subspace similar in the point forecasting case. Geometric intuition behind the probabilistic forecast reconciliation is discussed in Gamakumara et al. (2018).

5.1 Probabilistic forecast reconciliation in the Gaussian framework

Suppose we have a set of hierarchical time series where each realisation follows a multivariate Guassian distribution. i.e., $\mathbf{y}_T \sim \mathcal{N}(\mu_T, \Sigma_T)$ where both μ_T and Σ_T lives in the coherent subspace $\mathfrak{s}$ due to the aggregation structure of the hierarchy. We are interested in estimating the predictive Gaussian distribution of $\mathbf{Y}_{T+h}|\mathscr{I}_T$ where $\mathscr{I}_T = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T\}$, which should also lives in $\mathfrak{s}$. Since Gaussian distributions are uniquely characterised by the first two moments, it is sufficient to have the mean and variance forecasts to get the Gaussian predictive distributions of the hierarchy.

Assume we have fit the time series models for each series of the hierarchy by considering all the available information. Using these fitted models we can estimate the means and variance forecasts of the hierarchy which are denoted by $\hat{\mu}_{T+h}$ and $\hat{\Sigma}_{T+h}$ respectively. Each element in $\hat{\mu}_{T+h}$ corresponds to the mean forecast of each series in the hierarchy and these elements are stacked in the same order as $\mathbf{y}_t$. Similarly $\hat{\Sigma}_{T+h}$ contains variances and covariances of all series in the hierarchy. One can either fit univariate models for each individual series of the hierarchy or fit multivariate models by considering the correlation structure of the series for getting these forecasts. However as long as the aggregation structure is not imposed, it is very unlikely that these forecasts will be coherent. Thus it comes to the point of reconciliation.

Reconciling Gaussian forecast distribution is all about reconciling the incoherent means and variance of forecasts. That is, the reconciled Gaussian distribution is given by $\mathcal{N}(\tilde{\mu}_{T+h}, \tilde{\Sigma}_{T+h})$, where,

$$\tilde{\mu}_{T+h} = \mathbf{SG}\hat{\mu}_{T+h}, \tag{17}$$

and

$$\tilde{\Sigma}_{T+h} = \mathbf{SG}\hat{\Sigma}_{T+h}\mathbf{G}'\mathbf{S}'. \tag{18}$$

A more rigorous explanation about deriving these equations can be found in Gamakumara et al. (2018). Depending on the choice of $\mathbf{G}$ we will get different Gaussian forecast distributions. The same choices of $\mathbf{G}$ matrices that has been used in the point forecasting case can be used in reconciling Gaussian forecast distributions. It is also worth mentioning that, due to the aggregation constraints, the distributions

lies in $\mathfrak{s}$, which is a lower dimensional subspace of $\mathbb{R}^n$. Therefore, these reconciled Gaussian distributions are degenerate.

5.2 Probabilistic forecast reconciliation in the non-parametric framework

In many applications it may not reasonable to assume parametric forecasting distributions. Therefore, non-parametric approaches has been widely used for probabilistic forecasts in different disciplines. For example, ensemble forecasting in weather applications (Gneiting & Raftery (2005), Gneiting & Katzfuss (2014), Gneiting et al. (2008)), bootstrap based approaches (Manzan & Zerom (2008), Vilar & Vilar (2013)).

Non-parametric approaches are also important in hierarchical forecasting as in most applications they have millions of time series which are often difficult to assume parametric distributions. Further, for data with heavy tails, it is often misleading to assume Gaussianity. The algorithm introduced by Taieb et al. (2017) is also a non-parametric approach as it does not make any distributional assumptions. They first generate, a sample from the bottom level predictive distribution, and then aggregate to obtain coherent probabilistic forecasts of the upper levels of the hierarchy. Initially they use MinT algorithm to reconcile the means of the bottom level forecast distributions, and then a copula-based approach is employed to model the dependency structure of the hierarchy. Resulting multi-dimensional distribution is used to generate the empirical forecast distributions for all bottom-level series which are then aggregated to obtain the empirical forecast distribution for the entire hierarchy. Although the means of forecast distributions are reconciled, the predictive distributions were obtained through a bottom-up based approach. Therefore this approach is not a reconciliation method as it does not use all the information from the hierarchy when producing coherent probabilistic forecasts.

Jeon et al. (2018) is the only existing study that does reconciliation in probabilistic hierarchical forecasts. This method is based on cross-validation and it also does not assume any parametric distributions for predictive densities. However they applied this method particularly for temporal hierarchies.

Gamakumara et al. (2018) discussed that reconciling the draws from incoherent forecast distribution will give draws from a reconciled forecast distribution. Employing this fact, we introduce a novel method based on a bootstrap approach. This involves a two steps process. In the first step we obtain possible sample paths from the incoherent forecast distributions. This follows by the reconciliation step which projects each sample path to the coherent subspace. These steps will be discussed in detail below.

Step 1: Obtaining incoherent probabilistic forecasts

Suppose we fit univariate time series models for the series at each node in the hierarchy, by using past observations up to time T. Using these fitted models, we can generate h step ahead future sample paths for each node, by conditioning on past observations. In order to capture the dependency structure of the hierarchy, we incorporate bootstrapped training errors from the fitted models. That is, let $\Gamma_{(T \times n)} = (\mathbf{e}_1, \mathbf{e}_2, \ldots\ldots, \mathbf{e}_T)'$ denote the in-sample residual matrix where, $\mathbf{e}_t = \mathbf{y}_t - \hat{\mathbf{y}}_t$ is a vector that consists of residuals in each node at time t and stacked in the same order as $\mathbf{y}_t$. Γ is also referred to as the matrix of in-sample errors. Then we block bootstrap a sample of size h from Γ, and these bootstrapped errors will be incorporated as the error series for simulating future paths. Taking block bootstrapped in-sample errors in generating future paths will implicitly model the dependency structure of the hierarchy. These simulated future sample paths will be then formed in a vector $\hat{\mathbf{y}}^b_{T+h|T}$ by stacking the sample paths of each node in the same order as $\mathbf{y}_t$. As such we generate a sample of N future paths for the hierarchy. We can denote this sample by a $(N \times n)$ matrix $\hat{\mathbf{Y}}^b_{T+h|T}$ where,

$$
\hat{\mathbf{Y}}^b_{T+h|T} = \begin{pmatrix} \hat{\mathbf{y}}^b_{1,T+h|T} \\ \hat{\mathbf{y}}^b_{2,T+h|T} \\ \vdots \\ \hat{\mathbf{y}}^b_{N,T+h|T} \end{pmatrix}.
\tag{19}
$$

Although the dependency structure of the hierarchy is captured through the bootstrap in-sample errors, the aggregation structure will not be reflected through these sample paths. Therefore, these future paths are incoherent and thus considered as a possible sample from the incoherent probabilistic forecast distribution of the hierarchy.

Step 2: Reconciling future sample paths

The second step is to project these sample paths to the coherent subspace. Then we get a set of reconciled future paths which form a possible sample from the reconciled probabilistic forecast distribution. Similar to the projection used in the point forecast reconciliation, we project each sample path through the projection $\mathbf{SG}$. Then we get,

$$
\tilde{\mathbf{y}}^b_{i,T+h} = \mathbf{SG}\hat{\mathbf{y}}^b_{i,T+h}, \quad i = 1, \ldots, N
\tag{20}
$$

and let

$$\tilde{\mathbf{Y}}^b_{T+h|T} = \begin{pmatrix} \tilde{\mathbf{y}}^b_{1,T+h|T} \\ \tilde{\mathbf{y}}^b_{2,T+h|T} \\ \vdots \\ \tilde{\mathbf{y}}^b_{N,T+h|T} \end{pmatrix} \tag{21}$$

where, $\tilde{\mathbf{y}}^b_{i,T+h} \in \mathfrak{s}$ denote a h-step-ahead reconciled future paths for $i = 1,...,N$. $\tilde{\mathbf{Y}}^b_{T+h|T}$ form an empirical coherent forecast distribution of the hierarchy that lies in the coherent subspace $\mathfrak{s}$. As we discussed in point forecast reconciliation methods, different estimates of $\mathbf{G}$ provide alternative estimates of reconciled future paths. We can also obtain bottom-up based future paths by simply aggregating the future paths of bottom level series for their respective upper levels. This is referred to as bottom-up future paths. Even though the bottom-up approach generates coherent probabilistic forecasts, it cannot be considered as a reconciliation method since it use only half of the information.

6 Evaluation

In this section we briefly discuss the methods that we use to measure the forecast accuracy of hierarchical forecasts.

6.1 Point forecast evaluation

Any existing point forecast evaluation method can be used to evaluate the accuracy of hierarchical point forecast. We are mainly using Mean Squared Error (MSE) and Mean Absolute Scaled Error (MASE) in our empirical application. Mathematical expression for these measures are given below.

$$MSE = mean(e^2_{t+h}), \tag{22}$$

and

$$MASE = mean(q_{t+h}), \tag{23}$$

where $q_{t+h} = \frac{e_{t+h}}{\frac{1}{T-m}\sum_{j=m+1}^{T}|y_j-y_{j-m}|}$. Further e_{t+h} is the forecast error, which is the difference between forecast and realisation.

MSE is a scaled dependent measure. Therefore to compare series in different scales or with different units, MASE is preferred. For more details on different point forecast accuracy measures refer to Hyndman & Athanasopoulos (2018).

6.2 Probabilistic forecast evaluation

Forecast accuracy of probabilistic forecasts can be evaluated using scoring rules Gneiting & Katzfuss (2014). Scoring rules provides a summary measurement based on the relationship between realisation and the forecast distribution. However, not all scoring rules are applicable for hierarchical probabilistic forecast evaluation. An comprehensive discussion on this can be found in Gamakumara et al. (2018).

Since the whole forecast distribution of the hierarchy is a multivariate distribution, the forecast accuracy should also be assessed in the multivariate framework. Therefore using multivariate scoring rules is helpful. Energy score (Gneiting et al. 2008) and variogram scores (Scheuerer & Hamill 2015) are the mainly using multivariate scoring rules in hierarchical framework. Further, multivariate log score (Gneiting & Raftery 2007) can also be used for evaluating parametric forecast distributions. The mathematical expressions for these are given table 1.

Log score is easy to use if the parametric forecast distributions are available. However, Gamakumara et al. (2018) showed that the multivariate log scores are inappropriate for the comparisons between incoherent and coherent forecast distributions. This is due to the degeneracy of multivariate coherent forecast distributions.

Energy score and variogram score can be used in any comparisons between multivariate hierarchical forecast distributions. The expectations in these scoring rules can be approximated by the sample means of simulated samples.

It would be also interested to see the performance of forecast distributions of individual series in the hierarchy. To measure these we use univariate scoring rules such as Continuous Rank Probability Score (CRPS) and univariate log score (Gneiting et al. 2008). The mathematical expression for CRPS is given by,

$$\text{CRPS}(\breve{F}_i, y_{T+h,i}) = E_{\breve{F}_i}|\breve{Y}_{T+h,i} - y_{T+h,i}| - \frac{1}{2}E_{\breve{F}_i}|\breve{Y}_{T+h,i} - \breve{Y}^*_{T+h,i}|, \qquad (24)$$

Table 1 Scoring rules to evaluate multivariate forecast densities. $\breve{\mathbf{y}}_{T+h}$ and $\breve{\mathbf{y}}^*_{T+h}$ be two independent random vectors from the coherent forecast distribution $\breve{\mathbf{F}}$ with the density function $\breve{\mathbf{f}}(\cdot)$ at time $T+h$ and $\mathbf{y}_{T+h}$ is the vector of realizations. Further $\breve{Y}_{T+h,i}$ and $\breve{Y}_{T+h,j}$ are ith and jth components of the vector $\breve{\mathbf{Y}}_{T+h}$. Moreover, the variogram score is given for order p where, w_{ij} are non-negative weights.

Scoring rule	Expression
Log score	$\text{LS}(\breve{\mathbf{F}}, \mathbf{y}_{T+h}) = -\log \breve{\mathbf{f}}(\mathbf{y}_{T+h})$
Energy score	$\text{eS}(\breve{\mathbf{Y}}_{T+h}, \mathbf{y}_{T+h}) = E_{\breve{\mathbf{F}}}\|\breve{\mathbf{Y}}_{T+h} - \mathbf{y}_{T+h}\|^\alpha - \frac{1}{2}E_{\breve{\mathbf{F}}}\|\breve{\mathbf{Y}}_{T+h} - \breve{\mathbf{Y}}^*_{T+h}\|^\alpha, \quad \alpha \in (0,2]$
Variogram score	$\text{VS}(\breve{\mathbf{F}}, \mathbf{y}_{T+h}) = \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} w_{ij}\Big(|y_{T+h,i} - y_{T+h,j}|^p - E_{\breve{\mathbf{F}}}|\breve{Y}_{T+h,i} - \breve{Y}_{T+h,j}|^p\Big)^2$

where $\breve{Y}_{T+h,i}$ and $\breve{Y}^*_{T+h,i}$ are two independent copies from the ith reconciled marginal forecast distribution $\tilde{F}_i$ of the hierarchy and $y_{T+h,i}$ is the ith realization from the true marginal distribution G_i. As in multivariate scoring rules, the expectations can be approximated by the sample average.

6.3 Comparison between different forecasting methods

We are mainly interested to evaluate the hierarchical forecasts in two aspects. One is to examine whether the having coherent forecasts is improving the forecast accuracy. This can be evaluated by comparing the incoherent forecasts with any coherent forecasts in both point as well as probabilistic framework. Secondly we are interested in finding the best reconciliation method by comparing reconciled forecast from different reconciliation methods.

For any comparison we use Skill score as defined in (Gneiting & Raftery 2007). For a given forecasting method, evaluated by a particular scoring rule $S(\cdot)$, the skill score can be calculated as follows,

$$Ss[S_B(\cdot)] = \frac{S_B(\boldsymbol{Y},\boldsymbol{y})^{\text{ref}} - S_B(\boldsymbol{\breve{Y}},\boldsymbol{y})}{S_B(\boldsymbol{Y},\boldsymbol{y})^{\text{ref}}} \times 100\%, \tag{25}$$

where $S_B(\cdot)$ is average score over B replicates and $S_B(\boldsymbol{Y},\boldsymbol{y})^{\text{ref}}$ is the average score of the reference forecasting methods. Thus $Ss[S_B(\cdot)]$ gives the percentage improvement of the preferred forecasting method relative to the reference method. Any positive value indicates that method is superior to the reference method, whereas any negative value of $Ss[S_B(\cdot)]$ indicate that the method we compared is poor than the reference method.

7 Empirical study

- GDP of Australia

- How GDP is measured

  - Production, Income and Expenditure approach
  - Explain the hierarchy

- Issues with data

  - Why use current price rather than constant price (This is why we ignores production approach)
  - What is statistical discrepancy
  - Frequency of data
  - Does the data satisfies coherency

- Forecasting methods

  - Add a figure of all time series in income approach. Explain why traditional methods might not work using these time series as forecasting one layer of the hierarchy will ignore the structural information of individual time series to be used in the forecasts.
  - Explain ETS and ARIMA forecasting methods briefly
  - Hierarchical forecasting

7.1 Point forecasting

7.1.1 Income approach

Table 2 Summary of point forecast in income approach

| | ETS | | | | | | | | ARIMA | | | | | | | |
| | MSE | | | | MASE | | | | MSE | | | | MASE | | | |
Method	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
Most aggregate level																
Benchmark	−418.74	−135.45	−36.30	−3.10	−127.28	−69.39	−28.70	−21.03	−420.06	−128.25	−39.11	−15.49	−140.89	−67.24	−38.84	−28.69
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bottom-up	−17.93	−3.68	7.80	9.35	−5.62	−3.49	−4.36	−15.49	−9.15	7.01	10.74	13.18	−13.13	−1.12	−6.25	−9.78
OLS	−0.18	3.45	6.05	6.70	0.05	0.96	3.47	1.50	2.12	2.46	2.54	1.92	−0.83	1.84	0.75	0.62
WLS	−4.05	5.59	11.57	13.09	−0.67	2.96	3.32	−2.27	2.41	6.80	9.35	9.71	−4.14	1.41	1.05	−1.78
MinT Shrink	−0.72	5.86	9.33	9.70	3.25	5.14	4.63	−1.15	7.06	9.05	8.63	9.73	−0.14	0.93	1.23	0.39
Average across all aggregate levels																
Benchmark	−320.17	−101.63	−29.68	−2.32	−96	−40.28	−11.83	−1.92	−298.21	−86.44	−19.78	0.71	−78.18	−36.49	−11.83	−2.91
Base	0.00	0.00	0.00	0.00	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bottom-up	−7.64	−2.49	0.73	1.55	−2	0.00	−2.15	−1.92	1.61	8.32	5.18	9.05	1.82	1.35	−1.08	0.00
OLS	3.16	3.29	2.91	2.92	2	2.78	2.15	1.92	4.79	4.33	3.05	3.37	1.82	1.35	1.08	1.94
WLS	3.61	4.84	4.39	4.45	4	4.17	2.15	1.92	7.89	8.49	6.94	8.15	3.64	2.70	1.08	1.94
MinT Shrink	6.46	5.18	2.39	1.54	6	4.17	3.23	2.88	12.69	11.96	6.76	8.69	7.27	5.41	2.15	4.85
Average across bottom level																
Benchmark	−200.06	−71.72	−21.93	−0.40	−61.02	−30.67	−11.11	−2.02	−191.76	−64.02	−7.71	6.65	−63.79	−32.43	−13.64	−4.12
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bottom-up	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OLS	0.66	−2.57	−3.50	−3.73	−23.73	−26.67	−18.89	−12.12	−2.88	−5.33	−2.55	−4.91	−20.69	−24.32	−22.73	−20.62
WLS	6.18	3.62	1.38	0.28	1.69	0.00	−1.11	0.00	2.43	−0.25	2.93	0.29	0.00	−1.35	−1.14	0.00
MinT Shrink	8.78	4.34	0.58	−1.43	3.39	1.33	1.11	1.01	4.81	1.57	2.15	0.47	0.00	−1.35	−1.14	0.00
Average across all levels																
Benchmark	−295.40	−96.29	−28.41	−2.01	−71.43	−33.78	−10.99	−1.98	−276.70	−82.54	−17.73	1.68	−68.42	−33.78	−12.22	−4.04
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bottom-up	−6.06	−2.05	0.61	1.31	0.00	0.00	−1.10	0.00	1.28	6.88	4.30	7.57	1.75	0.00	0.00	0.00
OLS	2.64	2.24	1.86	1.87	−14.29	−14.86	−10.99	−6.93	3.24	2.65	2.09	2.01	−12.28	−14.86	−13.33	−12.12
WLS	4.14	4.62	3.89	3.79	3.57	1.35	0.00	0.99	6.79	6.97	6.25	6.86	1.75	0.00	0.00	1.01
MinT Shrink	6.94	5.03	2.09	1.07	5.36	2.70	2.20	1.98	11.10	10.15	5.98	7.34	3.51	1.35	0.00	1.01

7.1.2 Expenditure approach

Table 3 Summary of point forecast in expenditure approach

	ETS								ARIMA							
	MSE				MASE				MSE				MASE			
Method	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
Most aggregate level																
Benchmark	−418.74	−135.45	−36.30	−3.10	−127.28	−69.39	−28.70	−21.03	−420.06	−128.25	−39.11	−15.49	−140.89	−67.24	−38.84	−28.69
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bottom-up	−24.62	−6.94	13.89	19.17	−6.68	−2.54	3.99	−2.73	−44.70	−16.08	0.96	7.78	−18.84	−1.99	−6.89	−8.10
OLS	8.32	8.19	11.30	11.18	6.06	5.67	6.74	4.30	5.37	5.72	7.59	7.70	0.43	3.56	3.38	4.25
WLS	8.75	10.69	17.96	18.90	8.48	7.91	9.37	4.83	0.97	5.08	10.24	11.50	0.15	2.91	2.11	3.12
MinT Shrink	9.80	10.20	16.71	18.44	9.02	6.77	9.30	5.88	2.45	3.64	8.33	8.62	0.42	0.89	0.94	1.78
Average across all aggregate levels																
Benchmark	−342.16	−149.39	−58.77	−21.04	−56.72	−32.10	−14.58	−2.78	−311.20	−147.72	−57.31	−28.38	−50.00	−30.49	−14.58	−5.71
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bottom-up	−9.67	1.15	6.67	10.92	2.99	4.94	4.17	5.56	−28.47	−15.65	−8.36	−2.91	−1.43	−1.22	1.04	0.00
OLS	9.39	9.12	9.29	8.96	1.49	1.23	1.04	0.93	7.67	5.77	6.24	6.01	2.86	1.22	2.08	1.90
WLS	12.33	14.06	13.73	14.24	5.97	6.17	5.21	5.56	5.56	5.02	7.25	8.18	4.29	3.66	4.17	3.81
MinT Shrink	13.99	16.04	14.26	14.47	7.46	7.41	6.25	5.56	7.71	4.56	7.32	7.20	5.71	3.66	5.21	3.81
Average across bottom level																
Benchmark	−167.10	−69.04	−24.00	−3.48	−53.01	−27.45	−11.76	−2.27	−136.49	−43.61	−10.04	6.21	−49.41	−26.21	−12.71	−2.27
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bottom-up	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OLS	−1.06	−5.53	−11.11	−15.14	−14.46	−13.73	−16.81	−17.42	3.40	2.61	−3.02	−4.06	−15.29	−11.65	−11.86	−9.85
WLS	1.91	−0.99	−4.84	−7.77	0.00	0.00	0.00	−1.52	6.53	7.03	1.43	0.90	1.18	0.00	0.00	0.00
MinT Shrink	2.36	−2.43	−6.15	−8.77	1.20	0.00	0.00	−1.52	5.37	7.14	0.48	−0.96	1.18	0.00	0.00	0.00
Average across all levels																
Benchmark	−301.11	−132.24	−52.19	−17.98	−55.84	−28.42	−12.61	−2.42	−268.67	−122.65	−47.54	−21.54	−50.0	−27.08	−12.61	−3.25
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0	0.00	0.00	0.00
Bottom-up	−7.41	0.91	5.41	9.02	0.00	1.05	0.90	1.61	−21.54	−11.88	−6.64	−2.34	0.0	0.00	0.00	0.00
OLS	6.94	5.99	5.43	4.77	−10.39	−9.47	−11.71	−12.10	6.63	5.01	4.33	4.02	−10.0	−8.33	−7.21	−6.50
WLS	9.89	10.84	10.22	10.41	1.30	2.11	0.90	0.81	5.79	5.50	6.05	6.74	2.5	1.04	1.80	0.81
MinT Shrink	11.26	12.10	10.40	10.42	2.60	2.11	1.80	0.81	7.14	5.18	5.91	5.58	2.5	2.08	1.80	0.81

7.2 Probabilistic forecasting: Gaussian approach

7.2.1 Income approach

Table 4 Skill scores with respect to energy score and variogram score for multivariate Gaussian forecast distribution

	ETS								ARIMA							
	ES				VS				ES				VS			
Method	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
Benchmark	−96.14	−36.28	−6.59	5.26	−156.30	−50.32	−9.15	6.00	−88.32	−32.30	−3.07	6.61	−145.36	−40.81	0.21	13.59
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bottom-up	−0.09	0.84	−1.11	−2.53	1.08	2.99	2.51	2.44	2.24	3.06	0.81	1.09	4.11	6.63	3.45	4.93
OLS	1.87	1.93	1.33	1.41	1.01	−0.62	−0.60	0.00	1.62	1.63	1.18	1.06	−0.68	−0.31	0.02	−0.31
WLS	3.85	4.00	2.05	1.16	7.41	5.61	3.62	2.64	3.51	3.77	2.62	2.68	6.31	6.69	5.86	5.54
MinT Sample	4.58	−0.14	−2.41	−2.78	8.48	−1.36	−8.09	−13.39	−0.44	−4.27	−5.78	−2.11	−0.65	−0.55	−6.82	−4.65
MinT Shrink	5.43	4.54	2.40	1.61	9.84	6.16	3.39	1.13	5.59	3.93	2.67	4.20	8.83	9.20	6.28	6.62

Table 5 Skill scores with respect to log score for multivariate Gaussian forecast distribution

	ETS				ARIMA			
Method	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
Bottom-up	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OLS	−2.62	5.65	20.17	34.21	−1.71	7.24	21.16	34.33
WLS	−0.26	−1.22	−1.60	−1.81	−0.40	−1.52	−1.70	−2.04
MinT Sample	−3.43	−10.18	−15.72	−20.50	−4.98	−11.25	−15.98	−18.74
MinT Shrink	−0.16	−1.60	−2.40	−2.91	−0.82	−2.35	−2.91	−3.31

Table 6 Skill scores for univariate Gaussian forecast distributions of individual series of the hierarchy

	ETS								ARIMA							
	CRPS				LS				CRPS				LS			
Method	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
Most aggregate level																
Benchmark	−129.45	−55.66	−14.67	−4.17	−5.53	11.47	28.94	36.32	−133.77	−50.95	−19.67	−7.94	−4.15	15.67	31.67	37.55
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OLS	−0.82	1.61	2.64	1.09	−1.27	−1.57	−1.10	−1.18	−0.25	1.11	0.15	−0.06	−0.73	−1.07	−1.37	−1.82
WLS	−1.80	2.88	2.92	−2.21	−2.55	−2.78	−1.76	−2.72	−1.39	2.32	0.65	−0.41	−1.18	−0.27	1.27	1.40
MinT Sample	2.88	0.92	−0.49	−4.32	−2.22	−5.82	−12.33	−17.64	−4.91	−9.54	−9.82	−6.08	−3.92	−10.06	−13.91	−13.34
MinT Shrink	1.01	4.31	3.22	−1.83	−1.95	−2.81	−3.96	−6.27	0.89	1.17	0.43	0.78	−1.11	−1.45	−1.69	−1.87
Average across all aggregate levels																
Benchmark	−108.81	−41.06	−7.27	4.76	−6.09	10.39	23.15	30.46	−97.82	−37.19	−5.84	4.74	−5.03	11.88	24.56	30.37
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OLS	1.43	1.86	1.48	1.33	−1.12	−2.70	−4.01	−4.50	1.29	1.27	0.53	1.15	−1.21	−2.91	−4.50	−5.15
WLS	2.47	2.69	0.97	−0.11	−1.73	−3.80	−6.31	−7.52	2.45	2.07	0.50	1.49	−1.61	−3.65	−5.77	−6.11
MinT Sample	3.19	−2.24	−3.30	−3.63	−1.73	−7.01	−14.84	−18.70	−1.34	−6.29	−8.33	−3.23	−4.72	−10.96	−18.30	−18.60
MinT Shrink	4.02	3.13	1.40	0.59	−1.52	−4.22	−8.03	−9.58	5.22	2.49	0.51	3.26	−1.61	−4.40	−8.44	−8.88
Average across all bottom level																
Benchmark	−74.79	−29.85	−4.80	5.13	−8.12	9.06	28.41	43.29	−72.44	−28.79	−4.24	4.85	−7.85	10.01	29.56	44.08
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OLS	−6.37	−7.66	−5.73	−3.82	−2.79	6.22	21.18	34.84	−7.87	−9.25	−7.89	−8.12	−2.28	6.57	19.63	32.19
WLS	1.64	0.61	0.28	0.20	−0.13	−0.84	−1.07	−0.90	−0.31	−0.35	0.24	0.26	−0.94	−1.14	−1.15	−1.15
MinT Sample	0.72	−3.14	−2.34	−4.16	−1.65	−5.48	−8.13	−10.65	−6.11	−8.48	−9.33	−6.84	−3.54	−7.92	−10.66	−10.88
MinT Shrink	3.45	1.61	1.58	1.00	0.13	−1.05	−1.72	−2.13	0.20	−0.86	−0.39	0.78	−0.76	−2.09	−2.75	−2.80
Average across all levels																
Benchmark	−97.58	−37.60	−6.54	4.87	−7.31	9.64	26.30	38.47	−89.67	−34.63	−5.37	4.77	−6.69	10.85	27.54	38.98
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OLS	−1.15	−1.08	−0.65	−0.17	−2.09	2.41	10.97	20.04	−1.65	−1.94	−1.95	−1.54	−1.73	2.47	9.79	18.27
WLS	2.20	2.04	0.77	−0.02	−0.81	−2.12	−3.11	−3.41	1.56	1.33	0.42	1.13	−2.09	−2.91	−3.00	−3.00
MinT Sample	2.38	−2.52	−3.02	−3.78	−1.74	−6.17	−10.82	−13.66	−2.87	−6.96	−8.63	−4.28	−4.04	−9.23	−13.70	−13.74
MinT Shrink	3.83	2.66	1.46	0.71	−0.58	−2.41	−4.20	−4.96	3.60	1.47	0.24	2.54	−1.04	−3.04	−5.05	−5.05

7.2.2 Expenditure approach

Table 7 Skill scores with respect to energy score and variogram score for multivariate Gaussian forecast distribution

	ETS								ARIMA							
	ES				VS				ES				VS			
Method	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
Benchmark	−93.6361	−44.9789	−13.7444	1.0137	−91.7685	−44.2491	−12.5531	4.1985	−84.6833	−43.3660	−14.8151	−2.6141	−79.4035	−38.5125	−11.4300	2.2691
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bottom-up	−1.05	1.81	3.01	2.92	1.81	4.10	5.14	7.08	−6.37	−3.76	−2.32	−1.36	−2.25	−1.35	0.57	1.21
OLS	2.63	2.00	1.59	1.16	−0.41	−2.42	−3.94	−5.01	2.34	1.32	0.91	0.83	0.77	−0.27	−2.16	−2.43
WLS	3.93	3.66	2.84	1.91	3.89	3.38	3.06	2.37	1.56	0.67	0.64	0.39	4.55	3.89	2.87	2.76
MinT Shrink	4.37	4.02	2.88	1.85	5.00	2.88	2.23	0.73	1.91	0.17	0.11	−0.57	4.23	3.81	2.82	2.17

Table 8 Skill scores with respect to log score for multivariate Gaussian forecast distribution

	ETS				ARIMA			
Method	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
Bottom-up	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OLS	0.12	2.92	6.09	9.12	0.53	4.69	9.20	13.10
WLS	−3.26	−5.20	−6.58	−6.95	−3.05	−4.56	−5.51	−5.90
MinT Shrink	−3.47	−5.80	−7.59	−8.22	−3.48	−5.33	−6.47	−7.06

Table 9 Skill scores for univariate Gaussian forecast distributions of individual series of the hierarchy

| | ETS | | | | | | | | ARIMA | | | | | | | |
| | CRPS | | | | LS | | | | CRPS | | | | LS | | | |
Method	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
Most aggregate level																
Benchmark	−130.00	−58.82	−12.94	−4.28	−5.53	11.24	29.09	36.32	−133.49	−49.65	−20.19	−8.14	−4.15	15.78	31.60	37.55
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OLS	4.24	3.26	4.98	3.17	−1.44	−4.49	−5.58	−7.27	1.39	2.29	2.09	3.08	−2.14	−4.19	−5.14	−5.59
WLS	5.87	3.75	7.82	2.85	−2.85	−7.56	−7.39	−9.31	−0.22	1.83	1.00	2.05	−3.73	−5.77	−5.50	−5.18
MinT Shrink	5.72	2.98	7.43	4.21	−3.36	−9.06	−10.51	−12.34	−0.87	−0.69	−0.46	0.31	−4.19	−8.94	−10.48	−11.27
Average across all aggregate levels																
Benchmark	−70.80	−55.79	−22.73	−5.75	11.95	1.74	12.44	22.13	−66.95	−55.09	−23.68	−9.08	11.26	1.09	11.34	19.55
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OLS	2.25	1.65	1.12	0.74	−1.35	−3.26	−4.78	−4.81	2.07	0.52	0.47	0.11	−0.68	−3.28	−4.07	−4.36
WLS	4.13	3.80	2.27	0.97	−11.85	−8.79	−13.68	−17.23	0.89	−0.62	0.29	−0.41	−12.82	−8.85	−12.79	−16.23
MinT Shrink	4.78	4.90	2.65	1.15	−12.91	−9.45	−15.31	−19.68	1.54	−0.96	0.19	−0.98	−14.56	−10.49	−15.12	−19.28
Average across all bottom level																
Benchmark	−21.37	−6.39	−22.58	−19.45	21.03	36.61	17.12	12.00	−18.89	−3.17	−18.96	−16.45	27.52	45.46	18.07	14.67
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OLS	−4.64	−6.43	−6.11	−7.67	12.98	21.02	3.57	6.95	−3.16	−3.09	−4.04	−4.66	21.14	33.84	4.57	10.33
WLS	−0.06	−1.30	−1.58	−3.08	−3.02	−3.76	−2.94	−2.91	0.44	0.13	−0.36	−0.65	−2.22	−2.34	−2.39	−2.07
MinT Shrink	0.64	−1.56	−2.05	−3.54	−4.12	−5.36	−4.20	−3.92	0.54	0.19	−0.79	−1.52	−3.60	−4.06	−3.84	−3.37
Average across all levels																
Benchmark	−50.77	−35.31	−22.68	−9.74	17.94	27.04	15.46	16.15	−47.50	−33.24	−22.13	−11.32	22.25	34.67	15.72	16.57
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OLS	−0.54	−1.70	−1.20	−1.71	8.03	14.31	0.51	2.22	−0.05	−1.00	−1.01	−1.33	13.99	24.84	1.52	4.62
WLS	2.43	1.69	1.03	−0.21	−6.14	−5.18	−6.82	−8.58	0.71	−0.30	0.08	−0.48	−5.63	−3.93	−6.09	−7.63
MinT Shrink	3.10	2.22	1.14	−0.22	−7.14	−6.50	−8.24	−10.19	1.14	−0.48	−0.13	−1.14	−7.14	−5.58	−7.81	−9.54

7.3 Probabilistic forecasting: Non-parametric bootstrap approach

7.3.1 Income approach

Table 10 Skill scores with respect to energy score and variogram score for multivariate forecast distribution

| | ETS | | | | | | | | ARIMA | | | | | | | |
| | ES | | | | VS | | | | ES | | | | VS | | | |
Method	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
Benchmark	−103.31	−39.64	−8.52	3.84	−164.35	−52.47	−9.68	6.05	−95.27	−36.15	−5.41	4.93	−150.24	−41.37	0.33	14.06
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bottom-up	0.44	1.11	−0.99	−2.29	1.38	3.24	2.69	2.65	2.21	3.09	0.98	1.26	4.13	6.66	3.67	5.20
OLS	2.35	2.25	1.77	1.85	0.72	−1.08	−0.74	−0.11	2.29	2.02	1.61	1.55	−0.91	−0.43	−0.14	−0.37
WLS	4.63	4.69	2.90	2.03	7.26	5.43	3.72	2.66	4.35	4.59	3.51	3.55	6.52	6.76	6.01	5.61
MinT Sample	5.52	0.75	−1.62	−2.32	8.63	−1.24	−8.06	−13.44	0.20	−3.64	−5.26	−1.72	−0.49	−0.43	−6.76	−4.58
MinT Shrink	6.34	5.16	3.18	2.27	9.84	6.06	3.34	1.05	6.32	4.61	3.34	4.88	8.79	9.20	6.34	6.79

Table 11 Skill scores for univariate forecast distributions of individual series of the hierarchy

Method	ETS (CRPS)				ARIMA (CRPS)			
	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
Most aggregate level								
Benchmark	−139.03	−60.56	−17.03	−5.22	−144.75	−56.38	−22.61	−9.99
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bottom-up	−6.01	−2.62	−2.92	−12.05	−7.99	1.29	−2.93	−4.41
OLS	−0.13	1.60	3.20	1.62	0.59	1.55	0.67	0.36
WLS	−0.95	3.20	3.80	−0.75	−0.63	2.79	1.46	0.27
MinT Sample	3.29	1.26	−0.09	−4.15	−4.68	−9.35	−9.40	−5.79
MinT Shrink	1.61	4.31	3.91	−0.74	1.28	1.43	1.04	1.28
Average across all aggregate levels								
Benchmark	−116.05	−44.38	−8.96	3.46	−104.93	−41.23	−8.22	3.04
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OLS	2.12	2.22	1.99	1.84	2.18	1.75	0.99	1.69
WLS	3.28	3.20	1.68	0.71	3.22	2.75	1.30	2.20
MinT Sample	4.02	−1.54	−2.53	−3.34	−0.69	−5.80	−7.97	−3.03
MinT Shrink	4.84	3.53	2.07	1.09	5.93	3.13	1.02	3.81
Average across all bottom level								
Benchmark	−81.67	−32.82	−6.70	3.68	−78.46	−31.91	−6.22	3.38
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OLS	−6.79	−8.46	−6.40	−4.31	−8.31	−10.13	−8.85	−8.94
WLS	1.91	1.17	0.80	0.74	0.05	0.17	0.68	0.65
MinT Sample	1.32	−2.31	−1.66	−3.82	−5.67	−7.72	−8.88	−6.51
MinT Shrink	3.91	2.19	1.99	1.30	0.36	−0.37	−0.04	1.06
Average across all levels								
Benchmark	−104.77	−40.82	−8.29	3.52	−96.44	−38.38	−7.63	3.14
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OLS	−0.81	−1.07	−0.48	0.05	−1.18	−1.88	−1.91	−1.40
WLS	2.83	2.58	1.42	0.72	2.20	1.96	1.12	1.75
MinT Sample	3.13	−1.78	−2.27	−3.48	−2.29	−6.38	−8.24	−4.04
MinT Shrink	4.53	3.11	2.05	1.15	4.14	2.06	0.71	3.01

7.3.2 Expenditure approach

Table 12 Skill scores with respect to energy score and variogram score for multivariate forecast distribution

Method	ETS								ARIMA							
	ES				VS				ES				VS			
	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
Benchmark	−103.71	−50.21	−16.69	−1.06	−95.71	−45.37	−12.23	4.58	−93.65	−48.12	−17.52	−4.58	−84.33	−40.65	−11.04	2.55
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bottom-up	−1.16	1.74	3.00	3.23	1.49	3.84	4.86	7.07	−5.24	−2.23	−0.85	0.23	−2.22	−1.27	1.00	1.54
OLS	4.42	3.79	3.35	2.90	0.44	−1.80	−3.56	−4.66	3.92	3.04	2.65	2.54	0.90	−0.28	−1.93	−2.19
WLS	6.58	6.36	5.37	4.40	5.09	4.29	3.66	2.91	4.25	3.50	3.31	3.24	5.18	4.26	3.53	3.31
MinT Shrink	6.73	6.31	4.89	4.05	6.12	3.63	2.66	1.28	4.34	2.62	2.42	1.84	4.82	4.11	3.28	2.54

Table 13 Skill scores for univariate forecast distributions of individual series of the hierarchy

	ETS (CRPS)				ARIMA (CRPS)			
Method	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
				Most aggregate level				
Benchmark	−145.58	−62.25	−15.91	−5.28	−146.33	−54.60	−22.52	−9.98
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bottom-up	−9.78	−4.06	4.99	−0.41	−15.96	1.99	−2.61	−0.23
OLS	5.54	4.93	5.94	4.29	2.18	3.60	3.17	4.03
WLS	6.86	5.91	8.71	4.73	0.75	3.38	2.41	3.65
MinT Shrink	6.86	4.51	8.15	5.67	0.38	0.87	0.75	1.59
				Average across all aggregate levels				
Benchmark	−78.02	−61.91	−26.48	−7.91	−73.77	−60.75	−26.93	−11.29
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OLS	3.96	3.77	3.28	2.87	3.66	2.39	2.54	1.96
WLS	6.72	6.84	5.29	3.96	3.61	2.45	3.50	2.75
MinT Shrink	7.07	7.35	5.21	3.84	3.92	1.72	2.90	1.66
				Average across all bottom level				
Benchmark	−21.66	−6.40	−23.56	−20.25	−19.81	−3.71	−20.12	−17.64
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OLS	−4.76	−6.87	−6.02	−7.84	−3.47	−3.37	−3.99	−4.75
WLS	1.17	−0.43	−0.73	−2.24	1.18	0.81	0.43	0.07
MinT Shrink	1.68	−0.89	−1.40	−2.87	1.05	0.61	−0.25	−1.03
				Average across all levels				
Benchmark	−55.07	−38.75	−25.54	−11.54	−51.91	−36.69	−24.68	−13.23
Base	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OLS	0.41	−0.67	0.27	−0.28	0.78	−0.04	0.39	−0.08
WLS	4.46	3.81	3.35	2.14	2.63	1.76	2.49	1.93
MinT Shrink	4.87	3.91	3.08	1.87	2.76	1.25	1.86	0.84

References

Athanasopoulos, G., Ahmed, R. A. & Hyndman, R. J. (2009), 'Hierarchical forecasts for Australian domestic tourism', *International Journal of Forecasting* **25**(1), 146–166.

Ben Taieb, S., Huser, R., Hyndman, R. J. & Genton, M. G. (2017), 'Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression', *IEEE Transactions on Smart Grid* **7**(5), 2448–2455.

Dunn, D. M., Williams, W. H. & Dechaine, T. L. (1976), 'Aggregate Versus Sub-aggregate Models in Local Area Forecasting', *Journal of American Statistical Association* **71**(353), 68–71.

Gamakumara, P., Panagiotelis, A., Athanasopoulos, G. & Hyndman, R. J. (2018), Probabilisitic Forecasts in Hierarchical Time Series.

Gel, Y., Raftery, A. E. & Gneiting, T. (2004), 'Calibrated Probabilistic Mesoscale Weather Field Forecasting', *Journal of the American Statistical Association* **99**(July), 575–583.

Gneiting, T. & Katzfuss, M. (2014), 'Probabilistic Forecasting', *Annual Review of Statistics and Its Application* **1**, 125–151.

Gneiting, T. & Raftery, A. E. (2005), 'Weather forecasting with ensemble methods', *Science* **310.5746**, 248–249.

Gneiting, T. & Raftery, A. E. (2007), 'Strictly Proper Scoring Rules, Prediction, and Estimation', *Journal of the American Statistical Association* **102**(477), 359–378.

Gneiting, T., Raftery, A. E., Westveld, A. H. & Goldman, T. (2005), 'Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation', *Monthly Weather Review* **133**(5), 1098–1118.

Gneiting, T., Stanberry, L. I., Grimit, E. P., Held, L. & Johnson, N. A. (2008), 'Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds', *Test* **17**(2), 211–235.

Gross, C. W. & Sohl, J. E. (1990), 'Disaggregation methods to expedite product line forecasting', *Journal of Forecasting* **9**(3), 233–254.

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G. & Shang, H. L. (2011), 'Optimal combination forecasts for hierarchical time series', *Computational Statistics and Data Analysis* **55**(9), 2579–2589.

Hyndman, R. J. & Athanasopoulos, G. (2018), *Forecasting: principles and practice*, OTexts.

Hyndman, R. J., Lee, A. J. & Wang, E. (2016), 'Fast computation of reconciled forecasts for hierarchical and grouped time series', *Computational Statistics and Data Analysis* **97**, 16–32.
URL: *http://dx.doi.org/10.1016/j.csda.2015.11.007*

Jeon, J., Panagiotelis, A. & Petropoulos, F. (2018), Reconciliation of probabilistic forecasts with an application to wind power.

Manzan, S. & Zerom, D. (2008), 'A bootstrap-based non-parametric forecast density', *International Journal of Forecasting* **24**(3), 535–550.

McSharry, P. E., Bouwman, S. & Bloemhof, G. (2005), 'Probabilistic forecasts of the magnitude and timing of peak electricity demand', *IEEE Transactions on Power Systems* **20**(2), 1166–1172.

Mirčetić, D., Nikoličić, S., Stojanović, D. & Maslarić, M. (2017), 'Modified top down approach for hierarchical forecasting in a beverage supply chain', *Transportation Research Procedia* **22**, 193–202.

Pinson, P., Madsen, H., Papaefthymiou, G. & Klöckl, B. (2009), 'From Probabilistic Forecasts to Wind Power Production', *Wind Energy* **12**(1), 51–62.

Schäfer, J. & Strimmer, K. (2005), 'A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics', *Statistical Applications in Genetics and Molecular Biology* **4**(1).
URL: *https://www.degruyter.com/view/j/sagmb.2005.4.issue-1/sagmb.2005.4.1.1175/sagmb.2005.4.1.1175.xml*

Scheuerer, M. & Hamill, T. M. (2015), 'Variogram-Based Proper Scoring Rules for Probabilistic Forecasts of Multivariate Quantities *', *Monthly Weather Review* **143**(4), 1321–1334.

Shang, H. L. (2017), 'Reconciling Forecasts of Infant Mortality Rates at National and Sub-National Levels: Grouped Time-Series Methods', *Population Research and Policy Review* **36**(1), 55–84.

Taieb, S. B., Taylor, J. W. & Hyndman, R. J. (2017), 'Hierarchical Probabilistic Forecasting of Electricity Demand with Smart Meter Data', pp. 1–30.

Van Erven, T. & Cugliari, J. (2014), *Game-Theoretically Optimal reconciliation of contemporaneous hierarchical time series forecasts*.

Vilar, J. A. & Vilar, J. A. (2013), 'Time series clustering based on nonparametric multidimensional forecast densities', *Electronic Journal of Statistics* **7**(1), 1019–1046.

Wickramasuriya, S. L., Athanasopoulos, G. & Hyndman, R. J. (2018), 'Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization', *Journal of the American Statistical Association* **1459**, 1–45.
URL: *https://www.tandfonline.com/doi/full/10.1080/01621459.2018.1448825*