Understanding Forecast Reconciliation

Ross Hollyman[a,*], Fotios Petropoulos[a], Michael E. Tipping[b]

[a] *School of Management, University of Bath, United Kingdom*
[b] *Institute for Mathematical Innovation, University of Bath, United Kingdom*

Abstract

A series of recent papers introduce the concept of Forecast Reconciliation, a process by which independently generated forecasts of a collection of linearly related time series are reconciled via the introduction of accounting aggregations that naturally apply to the data. Aside from its clear presentational and operational virtues, the reconciliation approach generally improves the accuracy of the combined forecasts. In this paper, we examine the mechanisms by which this improvement is generated by re-formulating the reconciliation problem as a combination of direct forecasts of each time series with additional *indirect* forecasts derived from the linear constraints. Our work establishes a direct link between the nascent Forecast Reconciliation literature and the extensive work on Forecast Combination. In the original hierarchical setting, our approach clarifies for the first time how unbiased forecasts for the entire collection can be generated from base forecasts made at any level of the hierarchy, and we illustrate more generally how simple robust combined forecasts can be generated in any multivariate setting subject to linear constraints. In an empirical example, we show that simple combinations of such forecasts generate significant improvements in forecast accuracy where it matters most: Where noise levels are highest and the forecasting task is at its most challenging.

Keywords: forecasting, forecast combinations, unbiasedness, top-down, hierarchies

1. Introduction

Multivariate time series where observations are grouped into a hierarchy occur in many contexts. For example, individual companies within the economy can be grouped by industry, then

*Corresponding author

Email addresses: rah98@bath.ac.uk (Ross Hollyman), f.petropoulos@bath.ac.uk (Fotios Petropoulos), mt821@bath.ac.uk (Michael E. Tipping)

by broader sector and finally into an economy-wide whole. In an international setting, economic or financial statistics may be reported by industry/sector and in parallel by country or region, giving rise to multiple hierarchies. Within a company, sales forecasts might be grouped by individual product, product category, and region. In a tourism context, one might be interested in aggregations of data by geographic region and perhaps by purpose or mode of travel. Forecasts often need to be made at different levels of a hierarchy, and may be of interest to, and indeed generated by, different constituencies within an organisation. Such forecasts may be made using different methods, driven both by the data at hand and by the needs and perhaps the statistical sophistication of each particular interest group. Whilst historical data will naturally aggregate over the hierarchy, forecasts will usually do so only by chance. Management decisions taken on the basis of inconsistent and contradictory forecasts are likely to inherit the same shortcomings.

Historically this problem has been addressed either by making ad-hoc or heuristic adjustments to existing sets of forecasts, or perhaps more systematically by building forecasts at a single level in the hierarchy and making adjustments to fill out the other levels. A 'Top Down' approach focuses on forecasting the top level and applying proportions to fill out lower part of the hierarchy. These proportions can be modelled in several ways, but essentially produce conditional forecasts given the forecast of the top level; for three such approaches, see Athanasopoulos et al. (2009). A 'Bottom Up' process builds forecasts at the bottom level and aggregates, and a 'Middle Out' approach splits proportionally downwards and aggregates upwards. All these three approaches used in isolation fail to take account of the information embedded in data at other levels within the hierarchy. Bottom level forecasts are often statistically noisy and volatile but may contain rich information, perhaps on trend or seasonality, which is lost in aggregation. Aggregate data are less noisy due to a portfolio effect: They represent combinations of noisy but less than perfectly correlated series, and may well reveal trends which are difficult or impossible to discern from the bottom level series alone.

In a series of recent papers, Hyndman et al. (2011), Hyndman et al. (2016), Wickramasuriya et al. (2019) develop a principled, systematic, and scalable approach to forecast reconciliation (FR) in such scenarios. The approach combines forecasts from different levels of the hierarchy to form a new, reconciled set of forecasts which conform to the natural aggregations constraints of the hierarchy. Empirical results show that not only do the reconciled forecasts 'add up' (a property which Panagiotelis et al. (2020) refers to as coherency), but are usually more accurate than the original base forecasts.

The original insight that a combination of two or more forecasts reduces error variance dates back several decades, and in particular, to Bates and Granger (1969). Whilst FR is clearly a form of forecast combination, the matrix algebra which allows FR to scale efficiently can tend to obscure the detail of precisely which forecasts are being combined. As Panagiotelis et al. (2020) note "*This setup can be counter-intuitive since a vector comprised of forecasts from different time series models is also assumed to be the dependent variable in a regression model*". Panagiotelis et al. (2020) go on to emphasise the importance of the linear aggregation constraints to the FR problem, and present its solution in a multidimensional geometric setting. We share with these authors the view that the key elements of FR are linear constraints within a multivariate setting, but adopt a different expositional approach: We combine forecasts *implied* by the constraints with direct base forecasts for each bottom level series. By doing so, we form a direct link between the rapidly growing literature on FR and the extensive body of work on Forecast Combination (FC).

A portfolio perspective is the clearest way to understand the principles underlying forecast combination. The mathematics underlying both settings are identical, hinging on the equation for the sums of random variables. A portfolio formed of two (less than perfectly correlated) securities benefits from a diversification effect: It can have lower volatility and will offer a more attractive risk/return trade-off than that offered by either security in isolation. The portfolio analogy readily extends to a hierarchical setting. Consider the example of a regional stock market index, perhaps 'North America' comprising the US and Canada. The regional index is simply a portfolio of the two countries indices in known proportions. The equivalent economic exposure to Canada can be created *either* via a long position in Canada *or* a long position in the regional index and a short position in the US country index. From a forecasting perspective, and in a situation where there exists a 'Total' series made up of two constituents, the analogous process creates two independent forecasts for a given disaggregate series: the first is the direct forecast, and the second an indirect forecast, made up of the forecast for the aggregate series *less* the forecast of other constituent. The mathematics of this setting exactly mirror the index portfolio situation and underpin the perspective adopted in the remainder of this paper.

A further important result emerges from viewing the FR problem from a FC perspective.

Whilst the literature shows that top-down forecasts can often be more accurate than those made on a bottom-up basis, Hyndman et al. (2011) show that the approach typically used to generate them induces a bias in the reconciled forecasts when the input is a vector of unbiased unconditional base forecasts. We provide insight as to why the approach in the literature is problematic, and our approach makes it possible to reconcile top-down forecasts in a manner that eliminates this bias.

A key lesson from the extensive literature on FC is that simple combining techniques (i.e., those which do not depend on the estimation of forecast error covariances) often perform as well or better than more statistically sophisticated approaches. Forecast combination achieves its largest gains in the most challenging forecasting tasks: Where forecasts are inherently poor; which is precisely where the statistical assumptions underlying more sophisticated techniques break down. In financial applications, portfolio optimisation exercises taking large scale covariance matrices as inputs are notoriously difficult to calibrate, for this reason the optimisation process is often disparagingly referred to by experienced practitioners as 'error maximisation'. The work of Black (1992) spawned a literature detailing a shrinkage based approach to this problem. For these reasons, simple equally weighted combinations are good default priors for any new combination based approach. Elliott and Timmermann (2013) set out the conditions under which equal-weight approaches are optimal. The 'MinT' FR methodology of Wickramasuriya et al. (2019) is elegant, easy to apply and scales efficiently to very large hierarchies, but is dependent on the estimation of a high dimensional variance/covariance matrix. We develop an alternative, but equally scalable approach based on equally weighted combinations of forecasts derived from different hierarchical levels, which performs at its best on the noisiest bottom level forecasts.

The following section briefly reviews the relevant literature on FR and FC. We then move on to examine the FR problem from a FC perspective, and show how to generate unbiased top-down and middle-out forecasts for a hierarchy. We then examine simple combinations of such forecasts, which we apply to a large sample of tourism data. Finally, we offer some implications for decision making, and our conclusions.

2. Background and Literature Review

2.1. Forecast Reconciliation

The foundational ideas underlying the papers reviewed here are set out in Hyndman et al. (2011), Hyndman et al. (2016), Wickramasuriya et al. (2019). Assume we have a group of series grouped together into some hierarchy so that there are in total m series, of which n series comprise the bottom level of the hierarchy. If $\mathbf{y}$ is a vector of all the m series at time t, and $\mathbf{b}$ is an n vector of **only** the bottom-level observations for the same time period, then $\mathbf{y} = \mathbf{S}\mathbf{b}$ where $\mathbf{S}$ represents a $m \times n$ 'summing matrix' which describes the hierarchical structure of the data and is key to the FR process. The m rows of matrix $\mathbf{S}$ represent the exposure of each series in the collection (at all levels) to the n bottom level or disaggregate series contained in $\mathbf{b}$.

Now assume at time t we have available an m vector of (unbiased) h-steps-ahead forecasts $\hat{\mathbf{y}}$. These are referred to as **base** forecasts. The idea is to estimate a new set of *reconciled* forecasts for the n bottom level series β so that the complete set of m reconciled forecasts is given by $\tilde{\mathbf{y}} = \mathbf{S}\beta$. Hyndman et al. (2011) and Hyndman et al. (2016) write the regression model as

$$\hat{\mathbf{y}} = \mathbf{S}\beta + \varepsilon \tag{1}$$

and choose β to minimise the aggregate reconciliation error ε, representing the difference between the original forecasts $\hat{\mathbf{y}}$ and the reconciled forecasts $\tilde{\mathbf{y}}$, so the estimated residuals from (1) are given by $\tilde{\varepsilon} = \hat{\mathbf{y}} - \tilde{\mathbf{y}}$. For forecasting purposes, however, the main objective is the production of accurate (but reconciled) forecasts, i.e., to choose β so as to minimise some function of the reconciled forecast errors ϵ, i.e., to write

$$\tilde{\mathbf{y}} = \mathbf{S}\beta + \epsilon. \tag{2}$$

So the estimated residuals are $\tilde{\epsilon} = \mathbf{y} - \tilde{\mathbf{y}}$. This is the approach taken in Wickramasuriya et al. (2019), who show that all linear reconciliation approaches can be written as

$$\tilde{\mathbf{y}} = \mathbf{S}\mathbf{G}\hat{\mathbf{y}} \tag{3}$$

for some $n \times m$ matrix $\mathbf{G}$. The role of $\mathbf{G}$ is to map the base forecasts $\hat{\mathbf{y}}$ to reconciled bottom-level forecasts (β) which are then aggregated by $\mathbf{S}$ to produce reconciled mid- and top-level forecasts. The matrix $\mathbf{G}$ can be chosen to reconcile forecasts in the various ways described above, or can be estimated based on the error covariance matrix of $\hat{\mathbf{y}}$ to form an 'optimal' forecast reconciliation, weighting forecasts at different hierarchical levels. Wickramasuriya et al. (2019) present 'MinT' – an efficient, scalable FR algorithm which adopts this approach. For the special case of Bottom Up, $\beta = \hat{\mathbf{b}}$ in which $\hat{\mathbf{b}}$ is the vector of the bottom-level base forecasts. The Top Down case can be obtained when $\tilde{\mathbf{y}} = \mathbf{S}\mathbf{q}\hat{y}_{\mathrm{T}}$ where $\mathbf{q}$ is a vector of proportions by which the top-level base forecast, $\hat{y}_{\mathrm{T}}$, is disaggregated to the bottom-level series (Hyndman and Athanasopoulos, 2018). This equation suggests $\tilde{y}_{\mathrm{T}} = \hat{y}_{\mathrm{T}}$, where $\tilde{y}_{\mathrm{T}}$ represents the top-level reconciled forecast.

The fundamental ideas set out above have been extended in several directions from the original cross-sectional setting. Perhaps most significantly, Athanasopoulos et al. (2017) adopt the $\mathbf{S}$ matrix approach temporally to time-series where forecasts are required to aggregate over a number of future periods. Separate forecasts are made using underlying data aggregated at different frequencies (eg. quarterly, half-yearly, and annually) which are then combined using similar techniques to those discussed here. Clearly, such series will occur in many contexts, and the modelling trade-offs are similar in nature to those which occur in the cross-section; lower-level data is plentiful but noisy, whereas higher-level aggregates are easier to model and trends may be easier to discern. Temporal FR allows all the available data to be used. Spiliotis et al. (2019) refine and develop the temporal approach, pointing out that particular care must be taken when seasonal data are aggregated. The same paper also discusses the situation where more than one forecast is available for a particular layer of the hierarchy.

Most authors adopt what might be termed an 'optimal structural scaling approach' to temporal reconciliation, accounting only for the differing variances of data observed at different frequencies within the same 'optimal' framework of Wickramasuriya et al. (2019). Nystrup et al. (2020) move away from the diagonal of the variance/covariance matrix, proposing several methods to deal with autocorrelation in the underlying series, although, as pointed out in the paper, significant error autocorrelation would suggest that base forecasts might be improved before the application of reconciliation. These authors find that the most significant benefits of their approach accrue to most inaccurate forecasts, a point to which we return below. Nystrup et al. (2020) also explicitly allude to the connection between FR and FC, a linkage we make explicit in this paper.

Spiliotis et al. (2020) and Kourentzes and Athanasopoulos (2019) examine the impact of combining temporal and cross-sectional hierarchical approaches. Spiliotis et al. (2020) explore the effects of sequential reconciliation, applying cross-sectional and temporal reconciliation in series. Neither of these approaches, as Kourentzes and Athanasopoulos (2019) point out, result in entirely coherent forecasts across both hierarchies. The latter authors reconcile temporally, and then reconcile in the cross-section at each time horizon within the temporal hierarchy. A consensus reconciliation matrix is then derived by averaging across the separate temporal levels. The authors mention simultaneous cross sectional and temporal reconciliation, but foresee difficulty in correctly designing a large **S** matrix and estimating **W**, the m dimensional variance/covariance matrix of forecast errors, in such a context.

In a series of papers Shang (2017), Shang and Haberman (2017) and Shang and Hyndman (2017) apply hierarchical forecasting to functional time series, where a function needs to vary coherently across hierarchies. Their application is to forecasts of mortality rates, which are modelled as a smooth function of age. Mortality statistics are often broken down by region, sex, and socio-economic grouping, and it is desirable that future forecasts of mortality rates (also at different time horizons) be presented in a coherent way across hierarchies.

Emphasising the scalability of the MinT approach, Ben Taieb et al. (2020) apply the approach to a very large hierarchy of short time horizon electricity smart meter data, and produce probabilistic forecasts for the hierarchy. A number of other authors consider utilising FR to generate probabilistic rather than point forecasts, although most authors fail to account for uncertainty in base model parameters, and all except Park and Nassar (2014) attempt to generate probabilistic forecasts whilst working in a frequentist framework. A complication in generating probabilistic time-series forecasts is accounting for dependence amongst series. Athanasopoulos et al. (2019) outline a procedure similar in principle to 'Filtered Historical Simulation' (FHS) (see, for example, Gurrola-Perez and Murphy (2015)) which has been used to simulate multivariate financial returns in a risk management context. FHS samples from historic errors residual to fitted models for volatility and correlation, and samples 'strips' of h period errors to account for temporal as well as cross-sectional dependence. Ben Taieb et al. (2017) propose a procedure involving copulas to measure series dependency, reordering of quantile forecasts, and LASSO estimation to generate probabilistic forecasts. Jeon et al. (2019) use a simpler cross-validation based approach, which also addresses the issue of estimation error in **W**. Pennings and van Dalen (2017) develop the FR approach in several directions, integrating it more closely in to the forecasting process, and derive forecasting benefits from doing so.

2.2. Forecast Combination

The initial idea of Forecast Combination dates to the late 1960's, perhaps most notably to Bates and Granger (1969). An extensive literature has built upon the topic, which is comprehensively reviewed and summarised in Elliott and Timmermann (2013). The reader is referred to that article for a comprehensive survey of the topic. An initial motivation for FC was a situation where base forecasts are generated with the aid of private information not available to the analyst, making it impossible to build an meta model. Whilst combining information sets and building such a model may be theoretically optimal, in practice there exist several examples where combinations of simpler sub-models are significantly more effective; see Elliott and Timmermann (2013) for a comprehensive discussion. Rapach et al. (2010) compare the performance of combined forecasts of Stock Market Returns, demonstrating that a 'kitchen sink' model built on the basis of all potential predictors at once performs particularly poorly, whereas combinations of several univariate models are considerably more effective. Elliott and Timmermann (2013) outline several circumstances where FC might be expected to add value, in particular, where:

- The nature of the underlying data generating process is unknown or uncertain;
- The parameters of the model are unknown or uncertain;
- The parameters (or indeed the model) are subject to breaks or changes over time;
- The underlying forecasts are based on different loss functions (thinking perhaps of incentives within different parts of an organisation)

Clearly such situations are commonplace and describe most if not all forecasting problems in management, economic, and social settings. For any variable of interest there are (at least) three different ways of forecasting using combinations: (a) combining forecasts from different univariate models, (b) decomposition of the series into components which are modelled and recombined, and (c) modelling with different sets of predictors and combining the forecasts.

The idea of combining Exponential Smoothing forecasts with those of ARIMA models lay behind the original work of Bates and Granger (1969), and Stock and Watson (1998) examine the performance of several univariate models for the prediction of US Economic time series, finding that combinations of simple univariate models perform better than any other single approach. Other examples of combining forecasts from different univariate models include the combination of forecasts from a particular family of models (see, for example, Kolassa, 2011; Kourentzes et al., 2019) and simple or weighted combinations across multiple different families (Petropoulos and Svetunkov, 2020; Montero-Manso et al., 2020).

Decomposition-based methods include the temporal hierarchies as described above, and the theta method (Assimakopoulos and Nikolopoulos, 2000; Fiorucci et al., 2016) where the seasonally-adjusted signal is separated into long- and short-term variations. In the same category we could place bagging-related approaches (Bergmeir et al., 2016; Petropoulos et al., 2018), where the original series is used to produce several bootstrap series; these bootstraps are forecasted independently (using the same or different models), and, finally, the forecasts from all bootstraps are aggregated (combined).

Rapach et al. (2010) is an example of using an array of different predictors, and Elliott (2015) take approach (c) to a logical (frequentist) conclusion and compute and average over forecasts using all possible subsets of models from a group of predictors.

Economics and finance dominate in terms of applications of FC, but some interesting examples in different sectors include Bordignon et al. (2013) and Graefe et al. (2014). In addition to combination strategies, a large variety of combining methodologies and weighting schemes exist, including the inverse MSE approach of Bates and Granger (1969), regression models (Granger and Ramanathan, 1984; Diebold and Pauly, 1990) and principal components models (Bansal et al., 2015; Stock, 2002). These can be static or time-varying and based on in-sample or out-of-sample forecast errors. Stock and Watson (2004) and Rapach et al. (2010) compare several approaches on economic and financial time series respectively. Rapach et al. (2019) select and then combine sets of predictors; the same principles underlie Bayesian approaches which average across a number of different models, using the 'spike and slab' model of George and McCulloch (1993) and Mitchell and Beauchamp (1988). Whilst effective on small to medium sized problems, these approaches based on discrete random variables struggle to deal with large numbers of potential predictors where combinatorial complexity affects performance. More recently approaches, such as Carvalho et al. (2010) and Thomson et al. (2019), perform well on large problems. From the limited set of examples outlined above, it is clear that the FC literature is substantial, and researchers will doubtless find several ideas within it that can be usefully applied to reconciliation problems.

A criticism often levelled at FC based approaches is that if adding a new model improves forecast accuracy, then the original model was clearly miss-specified. The direct implication of such an argument is modelling oversight, rather than inherent uncertainty in the problem at hand. Were the same argument to be made in a financial context, along the lines of diversification being a crutch for inadequate security analysts, most would rightly suspect behavioural overconfidence on the part of the critic. The empirical success of FC suggests that a little humility goes a long way in improving forecast accuracy. We now propose a FC-based approach to the reconciliation of a simple hierarchy.

3. Hierarchical Reconciliation as Forecast Combination

Consider initially a simple hierarchy composed of three series, two bottom-level ($n = 2$) or disaggregate time series A and B, and a total, T, such that $\mathrm{T} = \mathrm{A} + \mathrm{B}$. The total number of series in this simple hierarchy is $m = 3$. We assume that forecasts are available for T, A, and B, along with an estimate of the covariance matrix of forecast errors, $\mathbf{W}$. The base or unreconciled forecasts for the individual series are $\hat{\mathbf{y}} = [\hat{y}_{\mathrm{T}}, \hat{y}_{\mathrm{A}}, \hat{y}_{\mathrm{B}}]'$, representing the inputs to the FR process. The process itself is characterised by two matrices $\mathbf{G}$ and $\mathbf{S}$, and the output is a set of reconciled forecasts $\tilde{\mathbf{y}}$ such that

$$\tilde{\mathbf{y}} = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}. \tag{4}$$

In earlier literature $\mathbf{G}$ (or sometimes $\mathbf{P}$) is referred to in different ways, we refer to it interchangeably as a reconciliation or weighting matrix; as we shall see that it plays the role of a weighting matrix in *all* reconciliation approaches based on (4), be they top-down, bottom-up, middle-out or ‘optimal’. The summing matrix for this simple hierarchy is

$$\mathbf{S} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{5}$$

while the weighting matrix $\mathbf{G}$, is of order 2×3 containing the reconciliation weights:

$$\mathbf{G} = \begin{bmatrix} g_{\mathrm{AT}} & g_{\mathrm{AA}} & g_{\mathrm{AB}} \\ g_{\mathrm{BT}} & g_{\mathrm{BA}} & g_{\mathrm{BB}} \end{bmatrix}, \tag{6}$$

with g_{XY} denoting the contribution weight of the *base* forecast of node Y to the *reconciled* forecast of node X. For the reconciled forecasts to be unbiased, $\mathbf{GS} = \mathbf{I}_n$ (Wickramasuriya et al., 2019), where $\mathbf{I}_n$ is an identity matrix of order n. This implies the following set of constraints:

$$g_{\text{AT}} + g_{\text{AA}} = 1, \qquad g_{\text{AT}} + g_{\text{AB}} = 0,$$
$$g_{\text{BT}} + g_{\text{BA}} = 0, \qquad g_{\text{BT}} + g_{\text{BB}} = 1.$$

The constraints can be incorporated directly into the matrix $\mathbf{G}$, so equation (6) becomes:

$$\mathbf{G} = \begin{bmatrix} (1-g_{\text{AA}}) & g_{\text{AA}} & (g_{\text{AA}}-1) \\ (1-g_{\text{BB}}) & (g_{\text{BB}}-1) & g_{\text{BB}} \end{bmatrix}. \tag{7}$$

Equation (4) can now be written as

$$\begin{bmatrix} \tilde{y}_{\text{T}} \\ \tilde{y}_{\text{A}} \\ \tilde{y}_{\text{B}} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} (1-g_{\text{AA}}) & g_{\text{AA}} & (g_{\text{AA}}-1) \\ (1-g_{\text{BB}}) & (g_{\text{BB}}-1) & g_{\text{BB}} \end{bmatrix} \begin{bmatrix} \hat{y}_{\text{T}} \\ \hat{y}_{\text{A}} \\ \hat{y}_{\text{B}} \end{bmatrix}, \tag{8}$$

which can be re-arranged as

$$\begin{aligned} \tilde{y}_{\text{T}} &= (2 - g_{\text{AA}} - g_{\text{BB}})\hat{y}_{\text{T}} + (g_{\text{AA}} + g_{\text{BB}} - 1)(\hat{y}_{\text{A}} + \hat{y}_{\text{B}}), \\ \tilde{y}_{\text{A}} &= g_{\text{AA}}\hat{y}_{\text{A}} + (1 - g_{\text{AA}})(\hat{y}_{\text{T}} - \hat{y}_{\text{B}}), \\ \tilde{y}_{\text{B}} &= g_{\text{BB}}\hat{y}_{\text{B}} + (1 - g_{\text{BB}})(\hat{y}_{\text{T}} - \hat{y}_{\text{A}}). \end{aligned} \tag{9}$$

Therefore, for the hierarchy described by $\mathbf{S}$ in (5), two alternative forecasts are available for each of the time series A, B, and T as a consequence of the aggregation constraints implied by the hierarchy. Consider series A, for which the direct base forecast is $\hat{y}_{\text{A}}$. Because A = T − B, and we have available forecasts for T and B, an additional implicit forecast for A is available as $(\hat{y}_{\text{T}} - \hat{y}_{\text{B}})$. As per Bates and Granger (1969), if the direct and implicit forecasts are unbiased and not perfectly correlated, a weighted combination of the two benefits from a diversification effect, such that the combined forecast $\tilde{y}_{\text{A}} = g_{\text{AA}}\hat{y}_{\text{A}} + (1 - g_{\text{AA}})(\hat{y}_{\text{T}} - \hat{y}_{\text{B}})$ will also be unbiased and will have lower error variance than either $\hat{y}_{\text{A}}$ or $(\hat{y}_{\text{T}} - \hat{y}_{\text{B}})$ alone. Panagiotelis et al. (2020) approach this problem from a geometric perspective, but for most audiences we believe that our explanation of FR, based on combining forecasts (and utilising a portfolio analogy) will be easier to understand.

Most empirical work in FC adopts a Mean Squared Error (MSE) loss function which is implicit in our discussion above. Elliott and Timmermann (2013) and the references therein discuss the choice of loss function in FC in some detail. It is worth reiterating the point made by the same authors; the variance of the combined forecast is *always* less than or equal that of the individual forecasts, so under MSE loss FC never detracts value.

Equation (8) shows that $\mathbf{G}$ can be regarded as a weighting matrix, of a form constrained by the hierarchy, which combines the base forecasts from the hierarchy to give reconciled forecasts at the most disaggregate level ($\mathbf{G}\hat{\mathbf{y}}$). These new forecasts are in turn aggregated by $\mathbf{S}$ to produce the complete set of reconciled forecasts $\tilde{\mathbf{y}}$. $\mathbf{G}$ in equation (7) represents the general form of the weighting matrix for a hierarchy defined by the $\mathbf{S}$ in (5), with g_{AA} and g_{BB} representing the weights on the direct base forecasts of A and B in the forecast combination equations (9). We set out our approach in this way in order to understand the mechanics underlying FR; for anything other than trivial hierarchies, using the matrices $\mathbf{S}$ and $\mathbf{G}$ directly will be much more tractable. Appendix B sets out the forecast combination based approach for a more complex three level hierarchy.

The structure of (7) and (9) allows us to explore some useful special cases:

- If $g_{\mathrm{AA}} = g_{\mathrm{BB}} = 1$, then $\mathbf{G}$ becomes the bottom-up reconciliation matrix, similar to Hyndman et al. (2011):
$$\mathbf{G}_{\mathrm{BU}} = \begin{bmatrix} \mathbf{0}_{n\times(m-n)} & \mathbf{I}_n \end{bmatrix}. \tag{10}$$
- If the weights on the base forecast of T in the combining equations for the disaggregate series sum to 1, i.e. in this case $(1 - g_{\mathrm{AA}}) + (1 - g_{\mathrm{BB}}) = 1$, we can see that $\tilde{y}_{\mathrm{T}} = \tilde{y}_{\mathrm{A}} + \tilde{y}_{\mathrm{B}} = \hat{y}_{\mathrm{T}}$, that is the base forecast and reconciled forecast at the top level of the hierarchy are equal. Therefore the constraint, $\sum_j g_{j\mathrm{T}} = 1$ for $j = 1...n$ can be used to generate top-down forecasts. To our knowledge, all top-down weighting matrices previously suggested in the literature resulted in biased reconciled forecasts. For example, Hyndman et al. (2011), Panagiotelis et al. (2020) and references therein employ a top-down weighting matrix of the form:
$$\mathbf{G} = \begin{bmatrix} \mathbf{q} & \mathbf{0}_{n\times(m-1)} \end{bmatrix}, \tag{11}$$
where $\mathbf{q}$ is a vector of size n which proportions the top level forecast across the disaggregate series. The same authors point out that weighting matrices of this form can never satisfy the constraint $\mathbf{GS} = \mathbf{I}_n$ required for the resulting forecasts to be unbiased. $\mathbf{G}$ in the form set out in (7) with $(1-g_{\mathrm{AA}})+(1-g_{\mathrm{BB}}) = 1$ allows for unbiased top-down (or middle-out) forecast reconciliation.

- Weighted combinations of valid $\mathbf{G}$ matrices, for example of the form

$$\mathbf{G} = w_1\mathbf{G}_1 + w_2\mathbf{G}_2 + ... + w_k\mathbf{G}_k \tag{12}$$

with $\sum w_i = 1$ and $0 \leq w \leq 1$ are also valid $\mathbf{G}$ matrices, satisfying $\mathbf{GS} = \mathbf{I}_n$ and can be applied to a vector of base forecasts $\hat{\mathbf{y}}$ to generate simple and robust FR strategies. $\mathbf{G}_i$ might for instance represent top-down, bottom-up, or middle-out weighting matrices.

If we can reasonably estimate the forecast error covariance matrix of $\hat{\mathbf{y}}$, $\mathbf{W}$, then we can write the top row vector of $\mathbf{G}$ in (7), corresponding to the combining equation for series A, as $\mathbf{g}_{\mathrm{A}} = \left[(1-g_{\mathrm{AA}}) \quad g_{\mathrm{AA}} \quad (g_{\mathrm{AA}}-1)\right]$ and minimise the variance of $\tilde{y}_{\mathrm{A}}$, $\mathrm{Var}(\tilde{y}_{\mathrm{A}}) = \mathbf{g}_{\mathrm{A}}\mathbf{W}\mathbf{g}_{\mathrm{A}}^{\mathrm{T}}$, to analytically calculate the optimal weight g_{AA}. This is mathematically the same exercise as calculating the minimum variance portfolio of two securities given their estimated variances and covariances. The optimal weight g_{BB} can be derived by direct analogy. Alternatively we can simultaneously estimate g_{AA} and g_{BB}, by minimising $\mathrm{tr}\left(\mathbf{GWG}^{\mathrm{T}}\right)$, where $\mathrm{tr}(\cdot)$ is the trace operator that sums the variances of $\tilde{y}_{\mathrm{A}}$ and $\tilde{y}_{\mathrm{B}}$ which appear on the diagonal of $\mathbf{GWG}^{\mathrm{T}}$. The MinT FR approach of Wickramasuriya et al. (2019) is likewise obtained by choosing the weighting matrix to minimise $\mathrm{tr}\left(\mathbf{GWG}^{\mathrm{T}}\right)$ using the constraints written out separately as $\mathbf{GS} = \mathbf{I}_n$.

4. Combinations of Unbiased Forecasts

In section 3, we showed that reconciliation matrices can be constructed to generate unbiased, coherent forecasts corresponding to any level of a hierarchy. In this section we examine how such reconciliation matrices can be constructed in practice without requiring the estimation of large covariance matrices, and form averages of these separate components which can be used to reconcile vectors of forecasts in the usual framework. We then suggest a simple modification to this process which we believe will generate improved forecasts in many settings.

4.1. Simple Estimates of G

'Bottom-up reconciled' forecasts correspond to the case where reconciled forecasts ($\tilde{y}_j$) are equal to base forecasts ($\hat{y}_j$) for all of the bottom level series. 'Top-down reconciled' forecasts are those where $\tilde{y}_{\mathrm{T}}$ is equal to the base forecast $\hat{y}_{\mathrm{T}}$ for the top-level aggregate forecast, and other hierarchical levels are filled out by proportion. 'Middle-out reconciled' forecasts can be constructed in a similar way. Generally, we call *level-reconciled* forecasts those were the reconciled forecasts coincide with the base forecasts of a particular level.

Hyndman et al. (2011) write the bottom-up reconciliation matrix directly as equation (10). They also suggest adopting a top-down reconciliation matrix of the form in (11), but point out that such a weighting matrix can never satisfy the constraint $\mathbf{GS} = \mathbf{I}_n$ required for the resulting forecasts to be unbiased. We can also see that this is the case from the examination of the combining equations (9). The point is also made from a geometric perspective by Panagiotelis et al. (2020). As we note above, equation (4), with $(1 - g_{AA}) + (1 - g_{BB}) = 1$ gives unbiased top-down reconciled forecasts if the base forecasts are unbiased, i.e. $\tilde{y}_T = \tilde{y}_A + \tilde{y}_B = \hat{y}_T$. The issue of bias induced by equation (11) is well flagged in the literature; indeed, the casual reader might easily gain the mistaken impression that *all* top-down forecasts are inherently biased.

Taking for example series A in the simple hierarchy of section 3, from (9) we see that $\tilde{y}_A = g_{AA}\hat{y}_A + (1 - g_{AA})(\hat{y}_T - \hat{y}_B)$. In a forecast combination context, we know (Bates and Granger, 1969; Elliott and Timmermann, 2013) that the optimal value for g_{AA} is given by

$$g_{AA} = \frac{\sigma^2_{\text{T-B}} - \sigma_{\text{A,T-B}}}{\sigma^2_{\text{T-B}} + \sigma^2_A - 2\sigma_{\text{A,T-B}}} \quad \text{and} \quad (1 - g_{AA}) = \frac{\sigma^2_A - \sigma_{\text{A,T-B}}}{\sigma^2_{\text{T-B}} + \sigma^2_A - 2\sigma_{\text{A,T-B}}}, \tag{13}$$

where σ_A is the standard deviation of the forecast errors for series A and $\sigma_{\text{A,T-B}}$ is the covariance between the forecast errors of the direct forecast (A) and the indirect forecast (T − B).

It is common in FC exercises to ignore covariance terms and to write

$$(1 - g_{AA}) = \frac{\sigma^2_A}{\sigma^2_{\text{T-B}} + \sigma^2_A}. \tag{14}$$

In the case of top-down forecasts, $\sum_{j=1}^{n}(1 - g_{jj}) = 1$ and $(1 - g_{jj})$ is given by equation (14). Alternatively, taking $\tilde{y}_A = g_{AA}\hat{y}_A + (1 - g_{AA})(\hat{y}_T - \hat{y}_B)$ and rearranging, we get

$$\tilde{y}_A = \hat{y}_A + (1 - g_{AA})(\hat{y}_T - \hat{y}_A - \hat{y}_B), \tag{15}$$

which demonstrates that reconciled forecast is given by adding a proportion $(1 - g_{AA})$ of the total cross hierarchy reconciliation error $(\hat{y}_T - \hat{y}_A - \hat{y}_B)$ to the base forecast of A. In practice, we simply set for each bottom level forecast j

$$(1 - g_j) = \frac{\sigma^2_j}{\sum_{j=1}^{n} \sigma^2_j}, \tag{16}$$

which avoids the need to calculate $\hat{y}_{\text{T-B}}$. The two approaches give virtually identical results in practice.

Although equation (11) produced reasonable forecasts in the work of Athanasopoulos et al. (2009), it is a flawed approach as it fails to account for the forecast error variances of the bottom level series in updating their expected reconciled values conditional on the forecast for T. We expect our TD approach to outperform the top down approach suggested in the literature which disregards the prior variance of A and B. The resulting top-down forecasts can still be expressed as $\mathbf{q}y_{\text{T}}$ where $\mathbf{q}$ is a $m-n$ vector of proportions (Hyndman and Athanasopoulos, 2018), but the proportions in $\mathbf{q}$ *do not* simply represent the historical proportions of the bottom series to the top level.

Whilst it is a relatively straightforward matter to write down the weighting matrix $\mathbf{G}$ for a simple hierarchy, this is not the case when the number of series becomes large and the hierarchy complex. Appendix A sets out a simple procedure to calculate the weighting matrix $\mathbf{G}$ for any hierarchy at any given level of aggregation. Once a set of $\mathbf{G}$ matrices have been generated we can take their simple weighted average and reconcile the vector of base forecasts as per equation 12:

$$\tilde{\mathbf{y}} = \mathbf{S}(w_1\mathbf{G}_1 + w_2\mathbf{G}_2 + \ldots + w_k\mathbf{G}_k)\hat{\mathbf{y}} = \mathbf{S}(w_1\mathbf{G}_1\hat{\mathbf{y}} + w_2\mathbf{G}_2\hat{\mathbf{y}} + \ldots + w_k\mathbf{G}_k\hat{\mathbf{y}}). \tag{17}$$

4.2. Combined Conditional Coherent Forecasts

Such top-down/middle-out reconciliation matrices fit in to the existing framework of equation (4), where a vector of m existing base forecasts is pre-multiplied by $\mathbf{SG}$ to produce reconciled forecasts which are now coherent but ‘close’ (Panagiotelis et al., 2020) to the base forecasts. This similarity between base and reconciled forecasts is an advantage of forecast *reconciliation*, but may actually be a disadvantage from a forecast *combination* perspective. The full benefits of combining forecasts are obtained when forecasts are less correlated and more diverse. Can we capture these benefits whilst holding on to the attractive and now familiar forecast reconciliation framework?

A (valid) criticism of the ‘top-down’ approach we advocate above is that we are not producing truly ‘top-down’ forecasts – but actually a linear combination of forecasts of all hierarchical levels, constructed so that the reconciliation process does not change the top level base forecast. In this regard our top-down approach outlined above is similar to the ‘Top Down Forecast Proportions’ (TDFP) approach of Athanasopoulos et al. (2009). We can address this issue by building a new set of top-down forecasts which do not incorporate the base forecasts at the lower levels. These forecasts take a top-level base forecast (y_{T}), and require a model, or at least some assumptions, to estimate the bottom levels *conditional* on the forecast of T, i.e. in the simple example above we are interested in the conditional expectations $\mathbb{E}(\mathrm{A} \mid \mathrm{T})$ and $\mathbb{E}(\mathrm{B} \mid \mathrm{T})$.

$\mathbb{E}(\mathrm{A} \mid \mathrm{T})$ is of course a regression of A on T, which suggests a number of ways of building such conditional forecasts. Models for $\mathbb{E}(\mathrm{A} \mid \mathrm{T})$ and $\mathbb{E}(\mathrm{B} \mid \mathrm{T})$ may or may not produce a set of forecasts for the hierarchy which are coherent, but if this is not the case they can easily be reconciled with an appropriate **G** matrix, produced as described above, which does not change the value of the base forecast at the top level. If this exercise is repeated at each level of the hierarchy (producing conditional forecasts based on the base forecasts of a particular hierarchical level), the result will be several sets of reconciled forecasts. These forecasts are *conditional* on the base forecasts at each level and *coherent* in the usual sense that they aggregate across the hierarchy. We propose to then *combine* them to form *Combined Conditional Coherent* (CCC) forecasts.

Effectively, we can produce multiple $\hat{\mathbf{y}}$ vectors, apply the corresponding **G** matrices, and combine the reconciled forecasts:

$$\tilde{\mathbf{y}} = \mathbf{S}(w_1\mathbf{G}_1\hat{\mathbf{y}}_1 + w_2\mathbf{G}_2\hat{\mathbf{y}}_2 + \ldots + w_k\mathbf{G}_k\hat{\mathbf{y}}_k). \tag{18}$$

The set of k forecasts (where k is the number of levels in the hierarchy) form a relatively diverse pool or ensemble of coherent forecasts which when combined deliver the full diversification benefit of forecast combination.

In our empirical work below we use a very simple approach to produce the conditional forecasts; essentially a 'seasonal average' model. We calculate the calendar month mean and variance of each bottom level series. Our 'base' forecasts for the top level of the hierarchy then comprised of the forecast value of the top level, and the historic mean values (of the respective month) for the bottom-level series. These base forecasts are then reconciled via a top-down **G** matrix as described above, with the forecast error variances required to estimate **G** being the sample variances of the base series. Clearly several richer and potentially more accurate conditional models are possible.

The information set at time t for our CCC forecasts generated at level i in the hierarchy is comprised of the base forecasts for the time series in level i and the sample data for the series of the hierarchy. As such, the forecasts are directly comparable with the two top-down forecasting approaches that are based on historical proportions ('HP1' and 'HP2') of Athanasopoulos et al. (2009). The third top-down approach (TDFP) considered by these authors augments the historical data with forecasts of the bottom-level series. The TDFP approach therefore has more in common with equation (17) whereas the approach described in this section is similar in nature to HP1 and HP2.

As discussed above, a key finding of the FC literature is that simple combination methods are difficult to beat in practice and often outperform more complex approaches. In the words of Elliott and Timmermann (2013): *"Equally weighted combinations occupy a special place in the forecast combination literature"*. These authors cite several empirical examples where simple schemes have proven superior to approaches based on variance/covariance matrix estimation, and discuss the theoretical conditions where such an approach may be optimal. It is not hard to see why this might be the case, especially in a multivariate context: equal weighting requires no further parameter estimation, whilst variance/covariance schemes requires the estimation of a large matrix that is itself conditional on the parameters of the underlying forecasting models. On this basis, where it is hard to argue *a priori* for the superiority of any particular forecasting approach, and/or where the models under consideration have similar error variance, equal weighting is the sensible default option. In a hierarchical forecasting context, Abouarghoub et al. (2018) again show that an equally weighted forecast combination is effective in forecasting freight rates in the context of international shipping data, and provides more accurate forecasts than an 'optimal' approach. In an organisational context, it is sometimes the case that more than one forecast is produced for some hierarchical levels. These situations are easily handled within our framework, which potentially allows for several forecasts for any hierarchical level. In section 5, we apply just such an equally weighted forecast combination approach to a substantial data set from the Australian tourism sector.

4.3. General Multivariate Forecasting with Linear Constraints

We echo the views of Panagiotelis et al. (2020) that approaches developed to deal with hierarchical time series apply more broadly to any multivariate collection subject to linear constraints. Equations (9) and (14) comprise a very simple and robust approach to forecasting in these settings. For the example above, no $\mathbf{S}$ matrix is required and we can simply write $\mathrm{T} = \mathrm{A} + \mathrm{B}$ (or $\mathrm{A} = \mathrm{T} - \mathrm{B}$) where there is no clear hierarchy of variables. We then produce base forecasts $\hat{y}_j$ of each component $j \in [\mathrm{T}, \mathrm{A}, \mathrm{B}]$ and conditional forecasts of the other variables given each j and reconcile these if necessary. The final set of forecasts for the entire collection is simply a weighted combination $\sum w_j \tilde{\mathbf{y}}_j$ with $\sum w_j = 1$.

5. Reconciled Forecasts for Australian Domestic Tourism

We explore an empirical application of various forecast reconciliation approaches to a substantial set of time series from the Tourism sector. The data describe the number of nights spent away from home, classified both by destination and purpose of travel for domestic tourism with Australia between 1998 and 2016 [1]. The data comprises $n = 304$ disaggregate series, which are aggregated both geographically and by purpose of travel, resulting in 8 hierarchical levels and a total of $m = 555$ monthly time series. We follow Wickramasuriya et al. (2019) and build base forecasts for each series over a 96-month rolling window, using Exponential Smoothing models (on default settings) from Rob Hyndman's automated R package *forecast*[2]. Wickramasuriya et al. (2019) have shown that the 'MinT' approach performs well on this data set, generating meaningful improvements in forecast accuracy. We focus on Exponential Smoothing Models (ETS), as the original authors show that this methodology produces substantially more accurate base forecasts than ARIMA based models in this setting, therefore presenting a more challenging environment for forecast combination techniques we consider.

Base forecasts, in-sample fitted values, and model errors from each ETS model are calculated for each forecast month, and out of sample forecasts out to a 12-month horizon are produced using a variety of reconciliation approaches:

- Base Forecasts – unreconciled forecasts for each level of the hierarchy
- Bottom Up (BU) – Forecasts from the most disaggregate series, aggregated via the summing matrix $\mathbf{S}$.
- Top Down (TD) – Forecasts from the top level of the hierarchy, reconciled using the process described in Section 4 and Appendix A.

[1] The data was downloaded from https://robjhyndman.com/publications/mint/

[2] https://pkg.robjhyndman.com/forecast/

- Top Down Historic Proportions (TDHP) – Forecasts from the top level of the hierarchy, reconciled using the 'HP2' approach of Athanasopoulos et al. (2009).

- OLS – The original FR model of Hyndman et al. (2011). A simple approach which assumes that the base forecasts are independent and identically distributed; an assumption clearly violated in the data. OLS is calculated by setting $\mathbf{W} = \mathbf{I}$ in the MinT reconciliation approach of Wickramasuriya et al. (2019).

- WLSv – Again based on equation MinT but with the error variances of each series taken into account, i.e., the forecast errors are assumed to be unbiased and independent but with differing variances. $\mathbf{W}$ is set equal to a diagonal matrix of the variances of the base forecast errors.

- MinTShrink – The optimal approach of Wickramasuriya et al. (2019) based on a shrinkage estimator of the covariance matrix of forecast errors.

- CCC – an equally weighted average of 8 forecasts for each series derived from 8 levels of the hierarchy. We combine 6 middle-out approaches, plus bottom-up and top-down. For the Top Down and Middle Out forecasts which, along with bottom-up forecasts comprise the CCC approach, the reconciled forecasts for each level are produced by allocating reconciliation error in proportion to the variance of each disaggregate series measured over the preceding 96 months. The variances are calculated on a seasonally adjusted basis, i.e., by using the mean for each calendar month as the location parameter in the variance calculation. This results in 8 sets of forecasts, each corresponding to the forecast information at a particular level of the hierarchy. The $\mathbf{G}$ matrix for each i is applied to generate bottom level forecasts, which are then aggregated for the entire hierarchy via the usual $\mathbf{S}$ matrix. Final CCC forecasts are formed as an equally weighted average of these eight forecasts.

Table 1 sets out the results of the out of sample forecasting exercise. The table shows the average RMSE errors at each level of the hierarchy expressed as a proportion of the RMSE of the base forecasts, less 1. Therefore positive numbers in the table represent the percentage deterioration in forecast relative to base forecasts, whereas a negative numbers represent the percentage improvement relative to base.

		Forecast horizon													
		All purposes							By purpose of travel						
	Approach	1	2	3	6	12	1-6	1-12	1	2	3	6	12	1-6	1-12
Australia	BU	10.8	5.3	3.8	4.9	6.8	5.7	6.5	7.8	3.6	3.3	4.5	4.5	4.5	4.5
	TD	-0.0	-0.0	-0.0	-0.0	**-0.0**	-0.0	-0.0	10.8	11.3	10.9	8.6	3.0	9.9	5.0
	TDHP	-0.0	0.0	0.0	0.0	0.0	0.0	0.0	16.8	17.4	16.6	13.1	4.5	15.4	7.7
	OLS	**-1.3**	-1.5	-1.6	-0.7	0.6	-1.3	**-0.2**	**-0.5**	-0.2	-0.2	-0.4	**-1.0**	-0.4	**-0.5**
	WLSv	2.0	-0.6	-1.1	0.7	3.2	0.0	1.9	1.1	-0.3	-0.4	0.3	0.6	-0.1	0.4
	MinTShrink	-0.0	**-1.7**	**-2.4**	**-0.9**	1.3	**-1.5**	-0.1	-0.4	**-1.1**	**-1.2**	**-0.6**	-0.4	**-1.0**	-0.5
	CCC	1.1	-1.0	-1.4	0.4	2.8	-0.4	1.5	1.9	1.3	1.1	0.9	-0.0	1.0	0.2
States	BU	3.1	1.9	2.7	4.5	3.9	2.9	3.8	2.3	1.9	0.7	1.5	0.6	1.6	1.6
	TD	2.5	4.4	5.6	5.6	2.4	4.6	3.8	5.1	5.8	5.0	4.4	-0.8	4.9	1.9
	TDHP	3.0	5.0	6.1	6.1	2.5	5.1	4.0	7.7	8.5	7.4	6.5	0.2	7.3	3.3
	OLS	-2.2	-1.4	-0.7	**-0.4**	**-0.7**	-1.1	-0.9	-1.1	-0.7	-1.0	-1.2	-2.5	-0.9	-1.5
	WLSv	-2.2	-1.5	-0.6	0.5	0.4	-1.0	-0.2	-1.2	-0.7	-1.5	-1.4	-2.7	-1.2	-1.6
	MinTShrink	-2.8	-1.8	**-1.0**	-0.2	-0.5	**-1.4**	**-1.2**	**-1.7**	**-1.0**	**-1.8**	**-1.7**	-3.1	**-1.5**	-2.0
	CCC	**-2.9**	**-1.9**	-1.0	0.2	0.0	-1.4	-0.4	-1.3	-0.6	-1.4	-1.4	**-3.6**	-1.3	**-2.2**
Zones	BU	1.7	1.2	1.2	2.1	2.4	1.8	2.9	0.9	0.3	0.1	0.6	1.7	0.6	1.5
	TD	0.0	0.8	0.6	-0.1	-2.2	0.5	-0.6	0.8	0.9	0.3	0.1	-2.0	0.4	-1.1
	TDHP	0.4	1.2	0.9	0.3	-2.2	0.8	-0.5	2.1	2.3	1.6	1.2	-1.5	1.6	-0.3
	OLS	-1.2	-0.9	-1.1	-1.6	-1.4	-1.1	-1.1	-1.1	-1.1	-1.1	-1.1	-1.1	-1.1	-1.2
	WLSv	-1.5	-1.3	-1.3	-1.4	-1.0	-1.2	-0.7	-1.5	-1.5	-1.7	-1.5	-1.4	-1.6	-1.4
	MinTShrink	-2.1	-1.8	-1.7	-1.9	-1.7	-1.7	-1.5	-1.9	-1.9	-2.0	-1.9	-1.8	-2.0	-1.9
	CCC	**-2.8**	**-2.6**	**-2.7**	**-2.7**	**-2.6**	**-2.5**	**-2.1**	**-2.4**	**-2.4**	**-2.8**	**-2.7**	**-3.1**	**-2.6**	**-2.9**
Regions	BU	0.9	-0.5	0.4	1.3	2.3	0.7	1.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	TD	-1.6	-1.9	-1.4	-1.8	-2.3	-1.7	-2.2	-0.7	-0.4	-0.8	-1.5	-3.6	-1.0	-2.9
	TDHP	-0.5	-0.8	-0.5	-0.9	-1.7	-0.7	-1.5	0.5	0.8	0.3	-0.4	-3.0	0.1	-2.1
	OLS	-1.1	-1.7	-1.5	-1.4	-1.1	-1.4	-1.4	-1.0	-0.7	-0.8	-1.0	-1.6	-1.0	-1.5
	WLSv	-1.6	-2.4	-1.9	-1.5	-0.8	-1.8	-1.4	-1.3	-1.1	-1.3	-1.4	-2.0	-1.4	-2.0
	MinTShrink	-1.9	-2.7	-2.2	-1.9	-1.2	-2.1	-1.9	-1.7	-1.5	-1.6	-1.8	-2.4	-1.8	-2.3
	CCC	**-3.3**	**-3.9**	**-3.4**	**-3.4**	**-2.8**	**-3.5**	**-3.3**	**-2.9**	**-2.7**	**-3.0**	**-3.4**	**-4.5**	**-3.2**	**-4.2**

Table 1: The left-hand columns represent results grouped geographically. Right-hand side columns set out results for the same geographies split by purpose of travel. The data becomes more disaggregate reading downwards and to the right, with the bottom right hand side representing the most disaggregate forecasts, and overall totals in the top left corner. The numbers represent the percentage increase or decrease in RMSE of reconciled forecasts relative to the base forecasts. Bold entries identify the best performing approaches. All base forecasts were generated using the ETS routine of the R package 'forecast'.

		Forecast horizon													
		All purposes							By purpose of travel						
	Approach	1	2	3	6	12	1-6	1-12	1	2	3	6	12	1-6	1-12
Australia	BU	12.3	8.0	5.3	5.4	9.2	7.5	8.9	12.1	7.5	6.9	6.8	6.7	7.2	6.5
	TD	-0.0	-0.0	-0.0	-0.0	**-0.0**	-0.0	-0.0	12.1	11.2	10.7	9.6	6.0	10.1	7.1
	TDHP	-0.0	0.0	0.0	-0.0	0.0	-0.0	0.0	16.7	15.8	14.9	12.8	7.0	14.0	9.0
	OLS	**-1.2**	**-1.0**	-1.0	-1.0	0.5	**-1.0**	**-0.2**	**2.6**	0.9	1.2	0.8	**0.5**	1.2	0.7
	WLSv	1.9	1.3	-0.2	0.5	4.2	0.9	3.0	4.5	2.2	2.1	1.1	2.6	1.8	2.1
	MinTShrink	-0.1	0.2	**-1.5**	**-1.1**	1.9	-0.5	0.8	2.7	**0.9**	**0.5**	**-0.3**	0.9	**0.4**	**0.5**
	CCC	0.9	0.7	-0.7	-0.1	3.6	0.3	2.5	5.1	3.6	3.1	2.2	2.6	2.8	2.4
States	BU	4.3	1.5	2.6	3.7	4.3	2.9	4.1	1.0	0.0	-0.2	0.6	-0.7	0.4	-0.2
	TD	1.5	1.4	2.1	2.2	0.4	1.8	1.5	1.6	1.1	0.5	0.7	-2.4	0.9	-1.0
	TDHP	3.3	3.0	3.8	3.3	0.2	3.3	1.8	2.3	1.6	1.0	1.2	-2.2	1.4	-0.7
	OLS	-0.8	-0.4	0.3	0.0	**-1.5**	-0.3	-0.8	2.5	1.8	2.0	1.3	1.0	1.9	1.0
	WLSv	-0.8	-1.7	-1.0	-0.2	0.5	-1.0	0.1	-1.6	-2.2	-2.4	-2.1	-2.7	-2.0	-2.5
	MinTShrink	-1.2	-2.0	-1.6	-0.9	-0.5	-1.5	**-0.9**	-2.1	**-2.6**	-2.9	-2.4	-3.1	-2.4	-2.9
	CCC	**-1.8**	**-2.6**	**-2.0**	**-1.1**	-0.5	**-2.0**	-0.8	**-2.2**	-2.6	**-3.2**	**-2.7**	**-4.2**	**-2.7**	**-3.5**
Zones	BU	1.1	0.5	0.5	1.6	1.8	1.1	1.5	-0.6	-0.6	-0.6	-0.3	0.1	-0.5	-0.1
	TD	-1.1	-1.3	-1.7	-1.7	-2.7	-1.5	-2.1	-0.4	-0.6	-0.7	-0.7	-1.7	-0.7	-1.4
	TDHP	-1.1	-1.2	-1.6	-1.9	**-3.0**	-1.5	-2.4	-0.3	-0.5	-0.6	-0.7	-1.9	-0.6	-1.5
	OLS	-0.2	-0.6	-0.9	-0.8	-1.2	-0.6	-1.0	6.9	6.5	6.8	5.7	6.5	6.6	5.9
	WLSv	-1.8	-2.0	-1.9	-1.5	-1.0	-1.7	-1.2	-1.5	-1.5	-1.5	-1.4	-1.1	-1.5	-1.3
	MinTShrink	-2.2	-2.4	-2.4	-2.0	-1.5	-2.2	-1.8	-1.8	-1.7	-1.7	-1.6	-1.3	-1.7	-1.5
	CCC	**-3.1**	**-3.3**	**-3.5**	**-3.1**	-2.8	**-3.2**	**-2.9**	**-2.6**	**-2.5**	**-2.5**	**-2.5**	**-2.6**	**-2.6**	**-2.6**
Regions	BU	-1.6	-1.9	-1.5	-1.5	-1.0	-1.6	-1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	TD	-2.7	-2.9	-2.7	-3.3	-3.8	-3.0	-3.5	0.6	0.4	0.2	-0.1	-1.0	0.2	-0.5
	TDHP	-2.7	-2.9	-2.7	-3.4	-3.8	-3.1	-3.5	0.4	0.3	0.2	-0.2	-1.2	0.1	-0.8
	OLS	-0.3	-0.6	-0.7	-1.1	-1.3	-0.8	-1.1	17.6	17.6	17.9	16.3	15.9	17.4	16.0
	WLSv	-2.8	-2.9	-2.8	-2.8	-2.5	-2.9	-2.6	-0.2	-0.1	-0.2	-0.2	-0.4	-0.2	-0.3
	MinTShrink	-3.1	-3.2	-3.1	-3.3	-2.9	-3.2	-3.0	-0.4	-0.3	-0.4	-0.4	-0.6	-0.4	-0.5
	CCC	**-4.0**	**-4.2**	**-4.1**	**-4.3**	**-4.1**	**-4.2**	**-4.1**	**-0.8**	**-0.8**	**-0.9**	**-1.1**	**-1.6**	**-1.0**	**-1.4**

Table 2: The left-hand columns represent results grouped geographically. Right-hand side columns set out results for the same geographies split by purpose of travel. The numbers represent the percentage increase or decrease in MASE (using the seasonal naive forecast) of reconciled forecasts relative to the base forecasts. Bold entries identify the best performing approaches.

The clear conclusion from Table 1 is the overwhelming superiority of CCC at the bottom levels of the hierarchy. CCC is the best performing methodology for the bottom half (4 levels) of the hierarchy at all time horizons. Wickramasuriya et al. (2019) describe how forecasts at the bottom of the hierarchy are particularly difficult to generate, and how even the normally robust ETS procedures fail to identify important features of the disaggregated level data. Reinforcing the message from previous research, both in FR and FC, it is here, where accurate forecasting becomes more difficult, that the benefits from combining forecasts become greater. At the same time, however, the reliance on optimal approaches on variance/covariance estimation would seem to make them less effective than simpler methodologies.

Table 2 sets out the Mean Absolute Scaled Error of Hyndman and Koehler (2006) for the various forecasting methods at different hierarchical levels. Again, the superiority of CCC at lower hierarchical levels is very clear. By contrast, each of the optimal methods struggles to deal with the increased levels of noise towards the bottom of the hierarchy. MinT generally performs better than WLS suggesting that estimation of error variance is at least as problematic as covariance terms for this data set. Averaging across the 8 hierarchical levels in Table 2 leads to Table 3. The superiority of the CCC approach is clearly shown in this summary table. In a non-tabulated result, we calculated a simple combination of top-down and bottom-up forecasts only, which performed creditably although not as strongly as CCC.

The original FC paper of Bates and Granger (1969) suggested an 'optimal' forecast combination approach, but many researchers have found that simplified approaches work better in practice. Our results echo these findings in a FR setting.

6. Implications for Decision Making

FR techniques bring two key benefits to the forecasting process; firstly they generate forecasts which are coherent (the levels 'add up' in the same way as historical data), and secondly they generally improve forecast accuracy via the well documented benefits of combining forecasts.

A further practical benefit of the FR approach as outlined in the literature is its applicability to any set of base forecasts, although the optimal approaches developed to date are demanding in terms of the quality and quantity of data necessary to operate the process. Whilst base forecast error variances/covariances may be readily available when forecasting approaches are entirely quantitative at each level of the hierarchy, this will not be the case in many real world situations that might otherwise benefit from FR techniques. In this paper we show that FR can be just as effective (and indeed more so) without using any information on the forecast error variances or co-variances. The benefits of FR are therefore made available to a potentially much wider audience.

Combinations of forecasts will generally be more accurate than base forecasts as they help to diversify forecast error, and assuming that top down forecasts are generally less volatile than those lower down the hierarchy, apply a degree of shrinkage to bottom up forecasts. Such benefits are likely to apply both to judgemental and quantitative forecasts (see Summers and Pritchett, 2014, for an interesting economic example of bias in judgemental forecasts) and as much to forecasts made by experts as generalists (see Miller and Gelman, 2020, for a link between behavioural biases and a Bayesian approach to learning).

In an organisational context, embedding a FR process can democratise the forecasting process taking in information from different sources across an organisation. The implications of our work would suggest that:

- Forecasts should be made independently at different hierarchical levels, either by using appropriate quantitative or judgemental methods;
- The methods described above should be used to map forecasts made at top and middle levels of the hierarchy on to bottom level forecasts;
- These forecasts should be combined together (using equal weights unless there exists a compelling reason to do otherwise) and summed to provide coherent forecasts for the hierarchy.
- In general, combined forecasts are a clear mathematical illustration of the benefits of what might be termed a 'democratic' approach: Less correlated forecasts from diverse viewpoints across an organisation make for better forecasts once pooled together, but are potentially catastrophic if allowed to exist in parallel. 'Groupthink' results in a less diversified pool of base forecasts and poorer outcomes.
- Managers should be made aware of the general result that pooled forecasts generally outperform the best individual forecast *even when the identity of the best individual model is known beforehand...*

7. Conclusion

Hierarchical forecast reconciliation is a powerful management tool that generates coherent forecasts of large collections of time series in an efficient and principled way. The techniques are broadly applicable in many sectors of the economy and deserve to become an important component of the forecasting toolbox. Firstly the process produces a single set of forecasts that can systematically incorporate insight from different parts of an organisation, taking into account knowledge and expertise embedded at each level. Forecasts generated via reconciliation techniques inherit the benefits of combined forecasts, which have been validated in many empirical settings, and indeed have often been shown to outperform even the best *ex ante* forecasting model.

We showed how to write simple reconciliation problems as forecast combinations, and explain how the corresponding weighting matrix is constructed. We showed how to calculate statistically unbiased reconciled forecasts from any level of a hierarchy, given unbiased base forecasts, and confirm that simple equally weighted combinations of these forecasts significantly outperform optimised methodologies on a benchmark data set, especially in the noisiest sections of the hierarchy. In future work we plan to extend our approach to generate probabilistic forecasts within a Bayesian framework and to examine different combination methodologies.

The gains from combination are particularly valuable as they are often at their greatest when improvements are needed most: Where existing forecasts are of poor quality. Forecast combination can readily be explained to non-technical audiences via a portfolio diversification analogy, and can act as a subtle behavioural 'nudge' to forecasting processes, adding beneficial diversification, and an element of shrinkage as a counterweight to any tendency towards overconfidence. These benefits are well-documented in an extensive literature spanning several decades, and our paper forms a bridge from this to the more recent work on hierarchical reconciliation.

Acknowledgements

We would like to thank Rob J. Hyndman, George Athanasopoulos, and Anastasios Panagiotelis for their useful feedback on earlier versions of the manuscript. We would also like to thank the two anonymous reviewers for their comments which helped us improve the manuscript.

	Forecast horizon						
Approach	1	2	3	6	12	1-6	1-12
BU	3.5	1.9	1.6	2.1	2.7	2.1	2.5
TD	1.4	1.1	1.0	0.8	-0.6	0.9	0.0
TDHP	2.2	2.0	1.8	1.4	-0.5	1.7	0.3
OLS	3.4	3.1	3.3	2.7	2.5	3.1	2.6
WLSv	-0.3	-0.9	-1.0	-0.8	0.0	-0.8	-0.3
MinTShrink	-1.0	-1.4	-1.6	-1.5	-0.8	-1.4	-1.1
CCC	**-1.1**	**-1.5**	**-1.7**	**-1.6**	**-1.1**	**-1.6**	**-1.2**

Table 3: Mean Absolute Scaled Errors from Table 2 averaged across each of the 8 levels of the hierarchy. Bold entries identify the best performing approach.

Appendix A: Computing the weighting matrix

In order to quickly compute a top down combining matrix we utilise the following procedure:

1. For a given aggregate level k (perhaps composed of a number of aggregate series), choose the proportioning vector(s) for k. In the spirit of Bates and Granger (1969) we write a vector $\mathbf{p}$ proportional to the variance of the bottom level constituents from the in sample period. i.e.:

$$\mathbf{p} = \begin{bmatrix} 0_{m_k} & p_0 & \dots & p_n \end{bmatrix}$$

with $p_i = \sigma_i^2 / \sum_{j=1}^{n} \sigma_j^2$ for $i = 1 \dots n$, the bottom-level series, and m_k is the number of series on level k. The summing matrix for aggregate k is $\mathbf{S}_k$. $\mathbf{p}$ can be a matrix if there are more than one aggregate series in level k, see Appendix B below for an example of this case.

2. Augment $\mathbf{S}_k$ with $\mathbf{p}'$ to form matrix $\mathbf{A}_k$:

$$\mathbf{A}_k = \begin{bmatrix} \mathbf{p}' & \mathbf{S}_k \end{bmatrix}$$

3. Introduce $\bar{\mathbf{G}}_k$ which is the usual $\mathbf{G}$ matrix, but with an additional row representing the combination weights on the aggregate series of interest. For example, $\bar{\mathbf{G}}_k$ for a simple hierarchy where T = A + B:

$$\bar{\mathbf{G}}_k = \begin{bmatrix} g_{\mathrm{TT}} & g_{\mathrm{TA}} & g_{\mathrm{TB}} \\ g_{\mathrm{AT}} & g_{\mathrm{AA}} & g_{\mathrm{AB}} \\ g_{\mathrm{BT}} & g_{\mathrm{BA}} & g_{\mathrm{BB}} \end{bmatrix}, \tag{19}$$

4. Write $\bar{\mathbf{G}}_k \mathbf{A} = \mathbf{I}_{n+1}$ and solve for $\bar{\mathbf{G}}_k$. This step imposes the required conditions that the combining weights on the direct forecasts of A and B in the resulting weight matrix sum to one, i.e. $(1 - g_{\mathrm{AA}}) + (1 - g_{\mathrm{BB}}) = 1$ in the notation of equation (9). The bottom n rows of $\bar{\mathbf{G}}_k$ then form an unbiased top down weight matrix.

To illustrate this procedure, we utilize the simple hierarchy discussed above, assuming the variances of series A and B are estimated as .7 and .3 respectively. The vector $\mathbf{p}$ is then:

$$\mathbf{p} = \begin{bmatrix} 0.0 & 0.7 & 0.3 \end{bmatrix}.$$

The matrix $\mathbf{A}_k$ is

$$\mathbf{A}_k = \begin{bmatrix} 0.0 & 1.0 & 1.0 \\ 0.7 & 1.0 & 0.0 \\ 0.3 & 0.0 & 1.0 \end{bmatrix}$$

and setting

$$\bar{\mathbf{G}}_k \mathbf{A}_k = \begin{bmatrix} g_{\text{TT}} & g_{\text{TA}} & g_{\text{TB}} \\ g_{\text{AT}} & g_{\text{AA}} & g_{\text{AB}} \\ g_{\text{BT}} & g_{\text{BA}} & g_{\text{BB}} \end{bmatrix} \begin{bmatrix} 0.0 & 1.0 & 1.0 \\ 0.7 & 1.0 & 0.0 \\ 0.3 & 0.0 & 1.0 \end{bmatrix} = \mathbf{I}_3 \tag{20}$$

leads to

$$\bar{\mathbf{G}}_k = \begin{bmatrix} -1.0 & 1.0 & 1.0 \\ 0.7 & 0.3 & -0.7 \\ 0.3 & -0.3 & 0.7 \end{bmatrix} \tag{21}$$

and

$$\mathbf{G} = \begin{bmatrix} 0.7 & 0.3 & -0.7 \\ 0.3 & -0.3 & 0.7 \end{bmatrix}. \tag{22}$$

Appendix B: 3-Level Hierarchy

Assume a three-level hierarchy where T disaggregates to X and Y, while X disaggregates to A and B and Y disaggregates to C, D, and E. The **S** matrix is:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{23}$$

The $\mathbf{G}$ matrix is:

$$\begin{bmatrix} g_{\text{AT}} & g_{\text{AX}} & g_{AY} & g_{\text{AA}} & g_{\text{AB}} & g_{\text{AC}} & g_{\text{AD}} & g_{\text{AE}} \\ g_{\text{BT}} & g_{\text{BX}} & g_{\text{BY}} & g_{\text{BA}} & g_{\text{BB}} & g_{\text{BC}} & g_{\text{BD}} & g_{\text{BE}} \\ g_{\text{CT}} & g_{\text{CX}} & g_{\text{CY}} & g_{\text{CA}} & g_{\text{CB}} & g_{\text{CC}} & g_{\text{CD}} & g_{\text{CE}} \\ g_{\text{DT}} & g_{\text{DX}} & g_{\text{DY}} & g_{\text{DA}} & g_{\text{DB}} & g_{\text{DC}} & g_{\text{DD}} & g_{\text{DE}} \\ g_{\text{ET}} & g_{\text{EX}} & g_{\text{EY}} & g_{\text{EA}} & g_{\text{EB}} & g_{\text{EC}} & g_{\text{ED}} & g_{\text{EE}} \end{bmatrix}. \tag{24}$$

$\mathbf{GS} = \mathbf{I}_5$ implies:

$$\begin{bmatrix} g_{\text{AA}}+g_{\text{AT}}+g_{\text{AX}} & g_{\text{AB}}+g_{\text{AT}}+g_{\text{AX}} & g_{\text{AC}}+g_{\text{AT}}+g_{\text{AY}} & g_{\text{AD}}+g_{\text{AT}}+g_{\text{AY}} & g_{\text{AE}}+g_{\text{AT}}+g_{\text{AY}} \\ g_{\text{BA}}+g_{\text{BT}}+g_{\text{BX}} & g_{\text{BB}}+g_{\text{BT}}+g_{\text{BX}} & g_{\text{BC}}+g_{\text{BT}}+g_{\text{BY}} & g_{\text{BD}}+g_{\text{BT}}+g_{\text{BY}} & g_{\text{BE}}+g_{\text{BT}}+g_{\text{BY}} \\ g_{\text{CA}}+g_{\text{CT}}+g_{\text{CX}} & g_{\text{CB}}+g_{\text{CT}}+g_{\text{CX}} & g_{\text{CC}}+g_{\text{CT}}+g_{\text{CY}} & g_{\text{CD}}+g_{\text{CT}}+g_{\text{CY}} & g_{\text{CE}}+g_{\text{CT}}+g_{\text{CY}} \\ g_{\text{DA}}+g_{\text{DT}}+g_{\text{DX}} & g_{\text{DB}}+g_{\text{DT}}+g_{\text{DX}} & g_{\text{DC}}+g_{\text{DT}}+g_{\text{DY}} & g_{\text{DD}}+g_{\text{DT}}+g_{\text{DY}} & g_{\text{DE}}+g_{\text{DT}}+g_{\text{DY}} \\ g_{\text{EA}}+g_{\text{ET}}+g_{\text{EX}} & g_{\text{EB}}+g_{\text{ET}}+g_{\text{EX}} & g_{\text{EC}}+g_{\text{ET}}+g_{\text{EY}} & g_{\text{ED}}+g_{\text{ET}}+g_{\text{EY}} & g_{\text{EE}}+g_{\text{ET}}+g_{\text{EY}} \end{bmatrix} = \mathbf{I}_5.$$

Given the aggregation constraints embedded in the usual $\mathbf{S}$ matrix used in the literature, the base forecast of Y is not included in the combination forecast for series A or B, i.e. $p_{\text{AY}} = p_{\text{BY}} = 0$, likewise X for series C, D and E. The vector representing the top row of $\mathbf{G}$ can then be written as:

$$\begin{bmatrix} g_{\text{AT}} & (1-g_{\text{AT}}-g_{\text{AA}}) & 0 & g_{\text{AA}} & (g_{\text{AA}}-1) & -g_{\text{AT}} & -g_{\text{AT}} & -g_{\text{AT}} \end{bmatrix}$$

A little algebra leads to a forecast combination equation involving two unknown weights for $\tilde{y}_{\text{A}}$:

$$\tilde{y}_{\text{A}} = g_{\text{AA}}\hat{y}_{\text{A}} + (1-g_{\text{AA}}-g_{\text{AT}})(\hat{y}_{\text{X}}-\hat{y}_{\text{B}}) + g_{\text{AT}}(\hat{y}_{\text{T}} - \sum_{j\in[\text{B,C,D,E}]} \hat{y}_j) \tag{25}$$

The standard summing matrix $\mathbf{S}$ implies that the reconciled forecast equation for each disaggregate series is written as a combination of:

- the direct forecast,
- the middle level aggregate less the appropriate other disaggregate constituents,
- the total less all other disaggregate level series.

We reiterate the point made in Panagiotelis et al. (2020) that the aggregation constraints in complex hierarchies can be written in several different ways, corresponding to alternative forms of the **S** matrix. An alternative forecast for A in this instance could be written as (T – Y – B). Adding an additional aggregate row in to the summing matrix adds a new combining weight in to the reconciliation equation (25).

We now show how an equally weighted forecast can be constructed from such a hierarchy, assuming that the variances of the disaggregate series [A, B, C, D, E] are [.7, .3, .5, .1, .2].

Bottom Up Forecasts

These are trivially generated with the (5×8) matrix:

$$\mathbf{G}_0 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{26}$$

Middle Out Forecasts

The matrix $\mathbf{A}_\mathrm{M}$ for the middle level is as follows, incorporating the rows corresponding to X and Y from the **S** matrix, and the proportionate variances of the constituents of X and Y in the first two columns:

$$\mathbf{A}_\mathrm{M} = \begin{bmatrix} 0.0 & 0.0 & 1.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 1.0 & 1.0 \\ 0.70 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.30 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.62 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.12 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.26 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}. \tag{27}$$

We now solve $\bar{\mathbf{G}}_\mathrm{M}\mathbf{A}_\mathrm{M} = \mathbf{I}_7$ which gives:

$$\bar{\mathbf{G}}_M = \begin{bmatrix} -1.0 & 0.0 & 1.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & -1.0 & 0.0 & 0.0 & 1.0 & 1.0 & 1.0 \\ 0.7 & 0.0 & 0.3 & -0.7 & 0.0 & 0.0 & 0.0 \\ 0.3 & 0.0 & -0.3 & 0.7 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.63 & 0.0 & 0.0 & 0.37 & -0.63 & -0.63 \\ 0.0 & 0.12 & 0.0 & 0.0 & -0.12 & 0.88 & -0.12 \\ 0.0 & 0.25 & 0.0 & 0.0 & -0.25 & -0.25 & 0.75 \end{bmatrix} \tag{28}$$

Removing the top two rows and augmenting with a leading column vector of zeros leads to a (5×8) **G** matrix:

$$\mathbf{G}_M = \begin{bmatrix} 0.0 & 0.7 & 0.0 & 0.3 & -0.7 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.3 & 0.0 & -0.3 & 0.7 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.63 & 0.0 & 0.0 & 0.37 & -0.63 & -0.63 \\ 0.0 & 0.0 & 0.12 & 0.0 & 0.0 & -0.12 & 0.88 & -0.12 \\ 0.0 & 0.0 & 0.25 & 0.0 & 0.0 & -0.25 & -0.25 & 0.75 \end{bmatrix} \tag{29}$$

Top Down Forecasts

The matrix $\mathbf{A}_T$ for the top level is as follows, incorporating the row corresponding to T from the **S** matrix, and the proportionate variances of the disaggregate constituents in the first column:

$$\mathbf{A}_T = \begin{bmatrix} 0.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 \\ 0.39 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.17 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.28 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.06 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.11 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}. \tag{30}$$

Solving $\bar{\mathbf{G}}_T \mathbf{A}_T = \mathbf{I}_6$ and re arranging to give a (5×8) weight matrix gives:

$$\mathbf{G}_T = \begin{bmatrix} 0.39 & 0.0 & 0.0 & 0.61 & -0.39 & -0.39 & -0.39 & -0.39 \\ 0.17 & 0.0 & 0.0 & -0.17 & 0.83 & -0.17 & -0.17 & -0.17 \\ 0.28 & 0.0 & 0.0 & -0.28 & -0.28 & 0.72 & -0.28 & -0.28 \\ 0.06 & 0.0 & 0.0 & -0.06 & -0.06 & -0.06 & 0.94 & -0.06 \\ 0.11 & 0.0 & 0.0 & -0.11 & -0.11 & -0.11 & -0.11 & 0.89 \end{bmatrix}. \tag{31}$$

References

Abouarghoub, W., Nomikos, N., Petropoulos, F., 2018. On reconciling macro and micro energy transport forecasts for strategic decision making in the tanker industry. Transportation Research Part E: Logistics and Transportation Review 113, 225–238.

Assimakopoulos, V., Nikolopoulos, K., 2000. The Theta model: a decomposition approach to forecasting. International Journal of Forecasting 16, 521–530.

Athanasopoulos, G., Ahmed, R.A., Hyndman, a.o.J., 2009. Hierarchical forecasts for australian domestic tourism. International Journal of Forecasting 25, 146–166.

Athanasopoulos, G., Gamakumara, P., Panagiotelis, A., Hyndman, R., Affan, M., 2019. Hierarchical forecasting. IDEAS Working Paper Series from RePEc .

Athanasopoulos, G., Hyndman, R.J., Kourentzes, N., Petropoulos, F., 2017. Forecasting with temporal hierarchies. European Journal of Operational Research 262, 60–74.

Bansal, N., Strauss, J., Nasseh, A., 2015. Can we consistently forecast a firm's earnings? using combination forecast methods to predict the eps of dow firms. Journal of Economics and Finance 39, 1–22. doi:10.1007/s12197-012-9234-y.

Bates, J.M., Granger, C.W.J., 1969. The combination of forecasts. Journal of the Operational Research Society 20, 451–468.

Ben Taieb, S., Taylor, J.W., Hyndman, R.J., 2017. Coherent probabilistic forecasts for hierarchical time series. Proceedings of the 34th International Conference on Machine Learning , 3348–3357.

Ben Taieb, S., Taylor, J.W., Hyndman, R.J., 2020. Hierarchical probabilistic forecasting of electricity demand with smart meter data. Journal of the American Statistical Association 0, 1–17.

Bergmeir, C., Hyndman, R.J., Benítez, J.M., 2016. Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation. International Journal of Forecasting 32, 303–312. doi:10.1016/j.ijforecast.2015.07.002.

Black, F., 1992. Global portfolio optimization. Financial Analysts Journal 48, 28–44.

Bordignon, S., Bunn, D.W., Lisi, F., Nan, F., 2013. Combining day-ahead forecasts for british electricity prices. Energy Economics 35, 88–103.

Carvalho, C.M., Polson, N.G., Scott, J.G., 2010. The horseshoe estimator for sparse signals. Biometrika 97, 465–480.

Diebold, F.X., Pauly, P., 1990. The use of prior information in forecast combination. International Journal of Forecasting 6, 503–508.

Elliott, G., 2015. Complete subset regressions with large-dimensional sets of predictors. Journal of Economic Dynamics & Control 54, 86–111.

Elliott, G., Timmermann, A., 2013. Handbook of economic forecasting. Amsterdam : Elsevier, Amsterdam.

Fiorucci, J.A., Pellegrini, T.R., Louzada, F., Petropoulos, F., Koehler, A.B., 2016. Models for optimising the theta method and their relationship to state space models. International Journal of Forecasting 32, 1151–1161. doi:10.1016/j.ijforecast.2016.02.005.

George, E.I., McCulloch, R.E., 1993. Variable selection via Gibbs sampling. Journal of the American Statistical Association 88, 881–890.

Graefe, A., Armstrong, J.S., Jones Jr, R.J., Cuzán, A.G., 2014. Combining forecasts: An application to elections. International Journal of Forecasting 30, 43–54.

Granger, C.W.J., Ramanathan, R., 1984. Improved methods of combining forecasts. Journal of Forecasting 3, 197–204.

Gurrola-Perez, P., Murphy, D., 2015. Filtered historical simulation value-at-risk models and their competitors. Bank of England. Quarterly Bulletin Working Paper 525, 1–33.

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, 2579–2589.

Hyndman, R.J., Athanasopoulos, G., 2018. Forecasting: principles and practice. 2nd edition ed., OTexts.

Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. International Journal of Forecasting 22, 679–688.

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.

Jeon, J., Panagiotelis, A., Petropoulos, F., 2019. Probabilistic forecast reconciliation with applications to wind power and electric load. European Journal of Operational Research 279, 364–379.

Kolassa, S., 2011. Combining exponential smoothing forecasts using akaike weights. International Journal of Forecasting 27, 238–251. doi:10.1016/j.ijforecast.2010.04.006.

Kourentzes, N., Athanasopoulos, G., 2019. Cross-temporal coherent forecasts for australian tourism. Annals of Tourism Research 75, 393–409.

Kourentzes, N., Barrow, D., Petropoulos, F., 2019. Another look at forecast selection and combination: Evidence from forecast pooling. International Journal of Production Economics 209, 226–235. doi:10.1016/j.ijpe.2018.05.019.

Miller, J.B., Gelman, A., 2020. Laplace's theories of cognitive illusions, heuristics and biases. Statistical science 35, 159–170.

Mitchell, T.J., Beauchamp, J.J., 1988. Bayesian variable selection in linear regression. Journal of the American Statistical Association 83, 1023–1032.

Montero-Manso, P., Athanasopoulos, G., Hyndman, R.J., Talagala, T.S., 2020. FFORMA: Feature-based forecast model averaging. International Journal of Forecasting 36, 86–92. doi:10.1016/j.ijforecast.2019.02.011.

Nystrup, P., Lindström, E., Pinson, P., Madsen, H., 2020. Temporal hierarchies with autocorrelation for load forecasting. European Journal of Operational Research 280, 876–888. doi:10.1016/j.ejor.2019.07.061.

Panagiotelis, A., Gamakumara, P., Athanasopoulos, G., Hyndman, R.J., 2020. Forecast reconciliation: A geometric view with new insights on bias correction. International Journal of Forecatsing, To Appear .

Park, M., Nassar, M., 2014. Variational bayesian inference for forecasting hierarchical time series.

Pennings, C.L.P., van Dalen, J., 2017. Integrated hierarchical forecasting. European Journal of Operational Research 263, 412–418. doi:10.1016/j.ejor.2017.04.047.

Petropoulos, F., Hyndman, R.J., Bergmeir, C., 2018. Exploring the sources of uncertainty: Why does bagging for time series forecasting work? European Journal of Operational Research doi:10.1016/j.ejor.2018.01.045.

Petropoulos, F., Svetunkov, I., 2020. A simple combination of univariate models. International Journal of Forecasting 36, 110–115. doi:10.1016/j.ijforecast.2019.01.006.

Rapach, D.E., Strauss, J.K., Tu, J., Zhou, G., 2019. Industry return predictability: A machine learning approach. The Journal of Financial Data Science 1, 9–28.

Rapach, D.E., Strauss, J.K., Zhou, G., 2010. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. Review of Financial Studies 23, 821–862.

Shang, H., 2017. Reconciling forecasts of infant mortality rates at national and sub-national levels: Grouped time-series methods. Population Research and Policy Review 36, 55–84.

Shang, H., Haberman, S., 2017. Grouped multivariate and functional time series forecasting:an application to annuity pricing. Insurance: Mathematics and Economics 75, 166–179.

Shang, H.L., Hyndman, R.J., 2017. Grouped functional time series forecasting: An application to age-specific mortality rates. Journal of Computational and Graphical Statistics 26, 330–343.

Spiliotis, E., Petropoulos, F., Assimakopoulos, V., 2019. Improving the forecasting performance of temporal hierarchies. PLoS ONE 14, e0223422.

Spiliotis, E., Petropoulos, F., Kourentzes, N., Assimakopoulos, V., 2020. Cross-temporal aggregation: Improving the forecast accuracy of hierarchical electricity consumption. Applied Energy 261, 114339.

Stock, J.H., 2002. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167–1180. doi:10.1198/016214502388618960.

Stock, J.H., Watson, M.W., 1998. A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series. NBER Working Papers 6607. National Bureau of Economic Research, Inc.

Stock, J.H., Watson, M.W., 2004. Combination forecasts of output growth in a seven-country data set. Journal of Forecasting 23, 405–430.

Summers, L.H., Pritchett, L., 2014. Asiaphoria meets regression to the mean. NBER Working Paper Series .

Thomson, W., Jabbari, S., Taylor, A., Arlt, W., Smith, D., 2019. Simultaneous parameter estimation and variable selection via the logit-normal continuous analogue of the spike-and-slab prior. Journal of the Royal Society Interface 16.

Wickramasuriya, S.L., Athanasopoulos, G., Hyndman, R.J., 2019. Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. Journal of the American Statistical Association 114, 804–819.