Metadata of the chapter that will be visualized online

Chapter Title	Hierarchical Forecasting	
Copyright Year	2019	
Copyright Holder	Springer Nature Switzerland AG	
Corresponding Author	Family Name	**Athanasopoulos**
	Particle	
	Given Name	**George**
	Suffix	
	Organization	Monash University
	Address	Caulfield, VIC, Australia
	Email	George.Athanasopoulos@monash.edu
Author	Family Name	**Gamakumara**
	Particle	
	Given Name	**Puwasala**
	Suffix	
	Organization	Monash University
	Address	Clayton, VIC, Australia
	Email	Puwasala.Gamakumara@monash.edu
Author	Family Name	**Panagiotelis**
	Particle	
	Given Name	**Anastasios**
	Suffix	
	Organization	Monash University
	Address	Caulfield, VIC, Australia
	Email	Anastasios.Panagiotelis@monash.edu
Author	Family Name	**Hyndman**
	Particle	
	Given Name	**Rob J.**
	Suffix	
	Organization	Monash University
	Address	Clayton, VIC, Australia
	Email	Rob.Hyndman@monash.edu
Author	Family Name	**Affan**
	Particle	
	Given Name	**Mohamed**
	Suffix	
	Organization	Monash University

	Address	Clayton, VIC, Australia
	Email	affan.momu@gmail.com

Abstract	Accurate forecasts of macroeconomic variables are crucial inputs into the decisions of economic agents and policy makers. Exploiting inherent aggregation structures of such variables, we apply forecast reconciliation methods to generate forecasts that are coherent with the aggregation constraints. We generate both point and probabilistic forecasts for the first time in the macroeconomic setting. Using Australian GDP we show that forecast reconciliation not only returns coherent forecasts but also improves the overall forecast accuracy in both point and probabilistic frameworks.

Chapter 21
Hierarchical Forecasting

George Athanasopoulos, Puwasala Gamakumara, Anastasios Panagiotelis, Rob J. Hyndman, and Mohamed Affan

21.1 Introduction

Accurate forecasting of key macroeconomic variables such as Gross Domestic Product (GDP), inflation, and industrial production, has been at the forefront of economic research over many decades. Early approaches involved univariate models or at best low dimensional multivariate systems. The era of big data has led to the use of regularisation and shrinkage methods such as dynamic factor models, Lasso, LARS, and Bayesian VARs, in an effort to exploit the plethora of potentially useful predictors now available. These predictors commonly also include the components of the variables of interest. For instance, GDP is formed as an aggregate of consumption, government expenditure, investment, and net exports, with each of these components also formed as aggregates of other economic variables. While the macroeconomic forecasting literature regularly uses such sub-indices as predictors, it does so in ways that fail to exploit accounting identities that describe known deterministic relationships between macroeconomic variables.

In this chapter we take a different approach. Over the past decade there has been a growing literature on forecasting collections of time series that follow aggregation constraints, known as hierarchical time series. Initially the aim of this literature was to ensure that forecasts adhered to aggregation constraints thus ensuring aligned decision-making. However, in many empirical settings the forecast reconciliation methods designed to deal with this problem have also been shown to improve

---

G. Athanasopoulos (✉) · A. Panagiotelis
Monash University, Caulfield, VIC, Australia
e-mail: George.Athanasopoulos@monash.edu; Anastasios.Panagiotelis@monash.edu

P. Gamakumara · R. J. Hyndman · M. Affan
Monash University, Clayton, VIC, Australia
e-mail: Puwasala.Gamakumara@monash.edu; Rob.Hyndman@monash.edu

AQ1

P. Fuleky (ed.), *Macroeconomic Forecasting in the Era of Big Data*,
Advanced Studies in Theoretical and Applied Econometrics 52,
https://doi.org/10.1007/978-3-030-31150-6_21

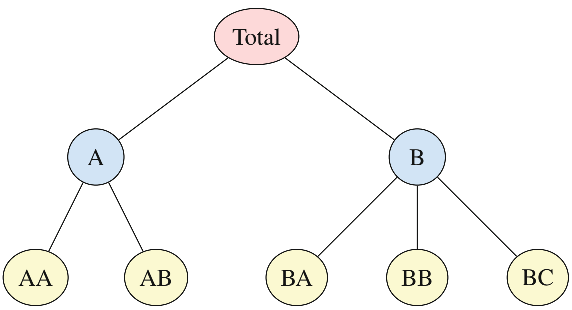
forecast accuracy. Examples include forecasting accidents and emergency admissions (Athanasopoulos, Hyndman, Kourentzes, & Petropoulos, 2017), mortality rates (Shang & Hyndman, 2017), prison populations (Athanasopoulos, Steel, & Weatherburn, 2019), retail sales (Villegas & Pedregal, 2018), solar energy (Yagli, Yang, & Srinivasan, 2019; Yang, Quan, Disfani, & Liu, 2017), tourism demand (Athanasopoulos, Ahmed, & Hyndman, 2009; Hyndman, Ahmed, Athanasopoulos, & Shang, 2011; Wickramasuriya, Athanasopoulos, & Hyndman, 2018), and wind power generation (Zhang & Dong, 2019). Since both aligned decision-making and forecast accuracy are key concerns for economic agents and policy makers we propose the application of state-of-the-art forecast reconciliation methods to macroeconomic forecasting. To the best of our knowledge the only application of forecast reconciliation methods to macroeconomics focuses on point forecasting for inflation (Capistrán, Constandse, & Ramos-Francia, 2010; Weiss, 2018).

The remainder of this chapter is set out as follows: Section 21.2 introduces the concept of hierarchical time series, i.e., collections of time series with known linear constraints, with a particular emphasis on macroeconomic examples. Section 21.3 describes state-of-the-art forecast reconciliation techniques for point forecasts, while Sect. 21.4 describes the more recent extension of these techniques to probabilistic forecasting. Section 21.5 describes the data used in our empirical case study, namely Australian GDP data, that is represented using two alternative hierarchical structures. Section 21.6 provides details on the setup of our empirical study including metrics used for the evaluation of both point and probabilistic forecasts. Section 21.7 presents results and Sect. 21.8 concludes providing future avenues for research that are of particular relevance to the empirical macroeconomist.

21.2 Hierarchical Time Series

To simplify the introduction of some notation we use the simple two-level hierarchical structure shown in Fig. 21.1. Denote as $y_{\mathrm{Tot},t}$ the value observed at time t for the most aggregate (Total) series corresponding to level 0 of the hierarchy. Below level 0, denote as $y_{i,t}$ the value of the series corresponding to node i, observed at time t.

Fig. 21.1 A simple two-level hierarchical structure



For example, $y_{\mathrm{A},t}$ denotes the tth observation of the series corresponding to node A at level 1, $y_{\mathrm{AB},t}$ denotes the tth observation of the series corresponding to node AB at level 2, and so on.

Let $\boldsymbol{y}_t = (y_{\mathrm{Tot},t}, y_{\mathrm{A},t}, y_{\mathrm{B},t}, y_{\mathrm{AA},t}, y_{\mathrm{AB},t}, y_{\mathrm{BA},t}, y_{\mathrm{BB},t}, y_{\mathrm{BC},t})'$ denote a vector containing observations across all series of the hierarchy at time t. Similarly denote as $\boldsymbol{b}_t = (y_{\mathrm{AA},t}, y_{\mathrm{AB},t}, y_{\mathrm{BA},t}, y_{\mathrm{BB},t}, y_{\mathrm{BC},t})'$ a vector containing observations only for the bottom-level series. In general, $\boldsymbol{y}_t \in \mathbb{R}^n$ and $\boldsymbol{b}_t \in \mathbb{R}^m$ where n denotes the number of total series in the structure, m the number of series at the bottom level, and $n > m$ always. In the simple example of Fig. 21.1, $n = 8$ and $m = 5$.

Aggregation constraints dictate that $y_{\mathrm{Tot}} = y_{\mathrm{A},t} + y_{\mathrm{B},t} = y_{\mathrm{AA},t} + y_{\mathrm{AB},t} + y_{\mathrm{BA},t} + y_{\mathrm{BB},t} + y_{\mathrm{BC},t}$, $y_{\mathrm{A},t} = y_{\mathrm{AA},t} + y_{\mathrm{AB},t}$ and $y_{\mathrm{B}} = y_{\mathrm{BA},t} + y_{\mathrm{BB},t} + y_{\mathrm{BC},t}$. Hence we can write

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t, \tag{21.1}$$

where

$$\boldsymbol{S} = \begin{pmatrix} 1\;1\;1\;1\;1 \\ 1\;1\;0\;0\;0 \\ 0\;0\;1\;1\;1 \\ \boldsymbol{I}_5 \end{pmatrix}$$

is an $n \times m$ matrix referred to as the *summing matrix* and $\boldsymbol{I}_m$ is an m-dimensional identity matrix. $\boldsymbol{S}$ reflects the linear aggregation constraints and in particular how the bottom-level series aggregate to levels above. Thus, columns of $\boldsymbol{S}$ span the linear subspace of $\mathbb{R}^n$ for which the aggregation constraints hold. We refer to this as the *coherent subspace* and denote it by $\mathfrak{s}$. Notice that pre-multiplying a vector in $\mathbb{R}^m$ by $\boldsymbol{S}$ will result in an n-dimensional vector that lies in $\mathfrak{s}$.

Property 21.1 A hierarchical time series has observations that are *coherent*, i.e., $\boldsymbol{y}_t \in \mathfrak{s}$ for all t. We use the term coherent to describe not just $\boldsymbol{y}_t$ but any vector in $\mathfrak{s}$.

Structures similar to the one shown in Fig. 21.1 can be found in macroeconomics. For instance, in Sect. 21.5 we consider two alternative hierarchical structures for the case of GDP and its components. However, while this motivating example involves aggregation constraints, the mathematical framework we use can be applied for any general linear constraints, examples of which are ubiquitous in macroeconomics. For instance, the trade balance is computed as exports minus imports, while the consumer price index is computed as a weighted average of sub-indices, which are in turn weighted averages of sub-sub-indices, and so on. These structures can also be captured by an appropriately designed $\boldsymbol{S}$ matrix.

An important alternative aggregation structure, also commonly found in macroeconomics, is one for which the most aggregate series is disaggregated by attributes of interest that are crossed, as distinct to nested which is the case for hierarchical

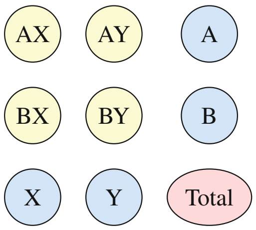


Fig. 21.2 A simple two-level grouped structure

time series. For example, industrial production may be disaggregated along the lines of geography or sector or both. We refer to this as a *grouped* structure. Figure 21.2 shows a simple example of such a structure. The Total series disaggregates into $y_{A,t}$ and $y_{B,t}$, but also into $y_{X,t}$ and $y_{Y,t}$, at level 1, and then into the bottom-level series, $\boldsymbol{b}_t = (y_{AX}, y_{AY}, y_{BX}, y_{BY})'$. Hence, in contrast to hierarchical structures, grouped time series do not naturally disaggregate in a unique manner.

An important implementation of aggregation structures are *temporal hierarchies* introduced by Athanasopoulos et al. (2017). In this case the aggregation structure spans the time dimension and dictates how higher frequency data (e.g., monthly) are aggregated to lower frequencies (e.g., quarterly, annual). There is a vast literature that studies the effects of temporal aggregation, going back to the seminal work of Amemiya and Wu (1972), Brewer (1973), Tiao (1972), Zellner and Montmarquette (1971) and others, including Hotta and Cardoso Neto (1993), Hotta (1993), Marcellino (1999), Silvestrini, Salto, Moulin, and Veredas (2008). The main aim of this work is to find the single best level of aggregation for modelling and forecasting time series. In this literature, the analyses, results (whether theoretical or empirical), and inferences, are extremely heterogeneous, making it very challenging to reach a consensus or to draw firm conclusions. For example, Rossana and Seater (1995) who study the effect of aggregation on several key macroeconomic variables state:

> Quarterly data do not seem to suffer badly from temporal aggregation distortion, nor are they subject to the construction problems affecting monthly data. They therefore may be the optimal data for econometric analysis.

A similar conclusion is reached by Nijman and Palm (1990). Silvestrini et al. (2008) consider forecasting French cash state deficit and provide empirical evidence of forecast accuracy gains from forecasting with the aggregate model rather than aggregating forecasts from the disaggregate model.

The vast majority of this literature concentrates on a single level of temporal aggregation (although there are some notable exceptions such as Andrawis, Atiya, and El-Shishiny (2011), Kourentzes, Petropoulos, and Trapero (2014)). Athanasopoulos et al. (2017) show that considering multiple levels of aggregation via temporal hierarchies and implementing forecast reconciliation approaches rather than single-level approaches results in substantial gains in forecast accuracy across all levels of temporal aggregation.

21.3 Point Forecasting

A requirement when forecasting hierarchical time series is that the forecasts adhere to the same aggregation constraints as the observed data; i.e., they are coherent.

Definition 21.1 A set of h-step-ahead forecasts $\tilde{\boldsymbol{y}}_{T+h|T}$, stacked in the same order as $\boldsymbol{y}_t$ and generated using information up to and including time T, are said to be *coherent* if $\tilde{\boldsymbol{y}}_{T+h|T} \in \mathfrak{s}$.

Hence, coherent forecasts of lower level series aggregate to their corresponding upper level series and vice versa.

Let us consider the smallest possible hierarchy with two bottom-level series, depicted in Fig. 21.3, where $y_{\text{Tot}} = y_{\text{A}} + y_{\text{B}}$. While base forecasts could lie anywhere in $\mathbb{R}^3$, the realisations and coherent forecasts lie in a two dimensional subspace $\mathfrak{s} \subset \mathbb{R}^3$.

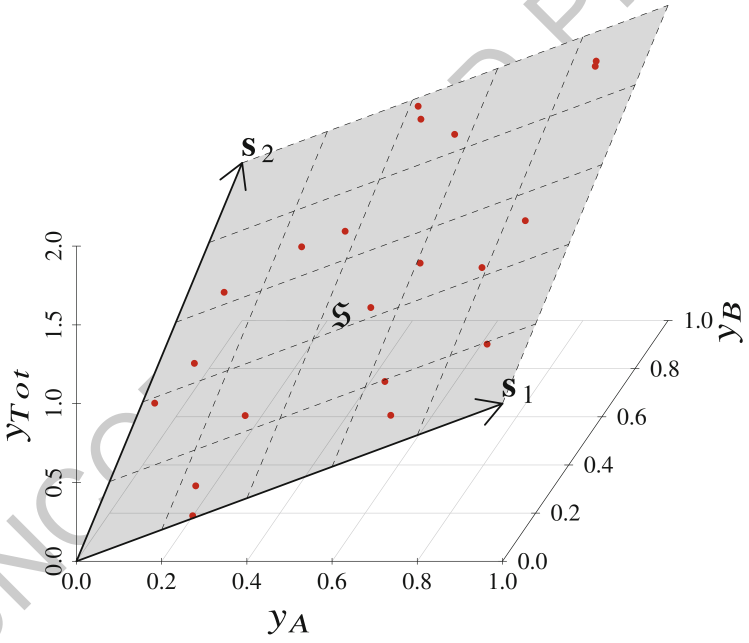


Fig. 21.3 Representation of a coherent subspace in a three dimensional hierarchy where $y_{\text{Tot}} = y_{\text{A}} + y_{\text{B}}$. The coherent subspace is depicted as a grey two dimensional plane labelled $\mathfrak{s}$. Note that the columns of $\mathbf{s}_1 = (1, 1, 0)'$ and $\mathbf{s}_2 = (1, 0, 1)'$ form a basis for $\mathfrak{s}$. The red points lying on $\mathfrak{s}$ can be either realisations or coherent forecasts

21.3.1 Single-Level Approaches

A common theme across all traditional approaches for forecasting hierarchical time series is that a single level of aggregation is first selected and forecasts for that level are generated. These are then linearly combined to generate a set of coherent forecasts for the rest of the structure.

Bottom-Up

In the *bottom-up* approach, forecasts for the most disaggregate level are first generated. These are then aggregated to obtain forecasts for all other series of the hierarchy (Dunn, Williams, & Dechaine, 1976). In general, this consists of first generating $\hat{\boldsymbol{b}}_{T+h|T} \in \mathbb{R}^m$, a set of h-step-ahead forecasts for the bottom-level series. For the simple hierarchical structure of Fig. 21.1, $\hat{\boldsymbol{b}}_{T+h|T} = (\hat{y}_{\text{AA},T+h|T}, \hat{y}_{\text{AB},T+h|T}, \hat{y}_{\text{BA},T+h|T}, \hat{y}_{\text{BB},T+h|T}, \hat{y}_{\text{BC},T+h|T})$, where $\hat{y}_{i,T+h|T}$ is the h-step-ahead forecast of the series corresponding to node i. A set of coherent forecasts for the whole hierarchy is then given by

$$\tilde{\boldsymbol{y}}^{\text{BU}}_{T+h|T} = \boldsymbol{S}\hat{\boldsymbol{b}}_{T+h|T}.$$

Generating bottom-up forecasts has the advantage of no information being lost due to aggregation. However, bottom-level data can potentially be highly volatile or very noisy and therefore challenging to forecast.

Top-Down

In contrast, *top-down* approaches involve first generating forecasts for the most aggregate level and then disaggregating these down the hierarchy. In general, coherent forecasts generated from top-down approaches are given by

$$\tilde{\boldsymbol{y}}^{\text{TD}}_{T+h|T} = \boldsymbol{S}\boldsymbol{p}\hat{y}_{\text{Tot},T+h|T},$$

where $\boldsymbol{p} = (p_1, \ldots, p_m)'$ is an m-dimensional vector consisting of a set of proportions which disaggregate the top-level forecast $\hat{y}_{\text{Tot},T+h|T}$ to forecasts for the bottom-level series; hence $\boldsymbol{p}\hat{y}_{\text{Tot},T+h|T} = \hat{\boldsymbol{b}}_{T+h|T}$. These are then aggregated by the summing matrix $\boldsymbol{S}$.

Traditionally, proportions have been calculated based on the observed historical data. Gross and Sohl (1990) present and evaluate twenty-one alternative approaches. The most convenient attribute of these approaches is their simplicity. Generating a set of coherent forecasts involves only modelling and generating forecasts for the most aggregate top-level series. In general, such top-down approaches seem to produce quite reliable forecasts for the aggregate levels and they are useful with low count data. However, a significant disadvantage is the loss of information due to aggregation. A limitation of such top-down approaches is that characteristics of lower level series cannot be captured. To overcome this, Athanasopoulos et al. (2009) introduced a new top-down approach which disaggregates the top-level based on proportions of forecasts rather than the historical data and showed that this method outperforms the conventional top-down approaches. However, a limitation of all top-down approaches is that they introduce bias to the forecasts even when the top-level forecast itself is unbiased. We discuss this in detail in Sect. 21.3.2.

Middle-Out

A compromise between bottom-up and top-down approaches is the middle-out approach. It entails first forecasting the series of a selected middle level. For series above the middle level, coherent forecasts are generated using the bottom-up approach by aggregating the middle-level forecasts. For series below the middle level, coherent forecasts are generated using a top-down approach by disaggregating the middle-level forecasts. Similarly to the top-down approach it is useful for when bottom-level data is low count. Since the middle-out approach involves generating top-down forecasts, it also introduces bias to the forecasts.

21.3.2 Point Forecast Reconciliation

All approaches discussed so far are limited to only using information from a single level of aggregation. Furthermore, these ignore any correlations across levels of a hierarchy. An alternative framework that overcomes these limitations is one that involves forecast *reconciliation*. In a first step. forecasts for all the series across all levels of the hierarchy are computed, ignoring any aggregation constraints. We refer to these as *base* forecasts and denote them by $\hat{\boldsymbol{y}}_{T+h|T}$. In general, base forecasts will not be coherent, unless a very simple method has been used to compute them such as for naïve forecasts. In this case, forecasts are simply equal to a previous realisation of the data and they inherit the property of coherence.

The second step is an adjustment that reconciles base forecasts so that they become coherent. In general, this is achieved by mapping the base forecasts $\hat{\boldsymbol{y}}_{T+h|T}$ onto the coherent subspace $\mathfrak{s}$ via a matrix $\boldsymbol{SG}$, resulting in a set of coherent forecasts $\tilde{\boldsymbol{y}}_{T+h|T}$. Specifically,

$$\tilde{\boldsymbol{y}}_{T+h|T} = \boldsymbol{SG}\hat{\boldsymbol{y}}_{T+h|T}, \tag{21.2}$$

where $\boldsymbol{G}$ is an $m \times n$ matrix that maps $\hat{\boldsymbol{y}}_{T+h|T}$ to $\mathbb{R}^m$, producing new forecasts for the bottom level, which are in turn mapped to the coherent subspace by the summing matrix $\boldsymbol{S}$. We restrict our attention to projections on $\mathfrak{s}$ in which case $\boldsymbol{SGS} = \boldsymbol{S}$.

This ensures that unbiasedness is preserved, i.e., for a set of unbiased base forecasts reconciled forecasts will also be unbiased.

Note that all single-level approaches discussed so far can also be represented by (21.2) using appropriately designed $\boldsymbol{G}$ matrices, however, not all of these will be projections. For example, for the bottom-up approach, $\boldsymbol{G} = \begin{pmatrix}\boldsymbol{0}_{(m\times n-m)} & \boldsymbol{I}_m\end{pmatrix}$ in which case $\boldsymbol{SGS} = \boldsymbol{S}$. For any top-down approach $\boldsymbol{G} = \begin{pmatrix}\boldsymbol{p} & \boldsymbol{0}_{(m\times n-1)}\end{pmatrix}$, for which $\boldsymbol{SGS} \neq \boldsymbol{S}$.

Optimal MinT Reconciliation

Wickramasuriya et al. (2018) build a unifying framework for much of the previous literature on forecast reconciliation. We present here a detailed outline of this approach and in turn relate it to previous significant contributions in forecast reconciliation.

Assume that $\hat{\boldsymbol{y}}_{T+h|T}$ is a set of unbiased base forecasts, i.e., $\mathrm{E}_{1:T}(\hat{\boldsymbol{y}}_{T+h|T}) = \mathrm{E}_{1:T}[\boldsymbol{y}_{T+h} \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_T]$, the true mean with the expectation taken over the observed sample up to time T. Let

$$\hat{\boldsymbol{e}}_{T+h|T} = \boldsymbol{y}_{T+h|T} - \hat{\boldsymbol{y}}_{T+h|T} \tag{21.3}$$

denote a set of base forecast errors with $\mathrm{Var}(\hat{\boldsymbol{e}}_{T+h|T}) = \boldsymbol{W}_h$, and

$$\tilde{\boldsymbol{e}}_{T+h|T} = \boldsymbol{y}_{T+h|T} - \tilde{\boldsymbol{y}}_{T+h|T}$$

denote a set of coherent forecast errors. Lemma 1 in Wickramasuriya et al. (2018) shows that for any matrix $\boldsymbol{G}$ such that $\boldsymbol{SGS} = \boldsymbol{S}$, $\mathrm{Var}(\tilde{\boldsymbol{e}}_{T+h|T}) = \boldsymbol{SGW}_h\boldsymbol{S}'\boldsymbol{G}'$. Furthermore Theorem 1 shows that

$$\boldsymbol{G} = (\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1} \tag{21.4}$$

is the unique solution that minimises the trace of $\boldsymbol{SGW}_h\boldsymbol{S}'\boldsymbol{G}'$ subject to $\boldsymbol{SGS} = \boldsymbol{S}$. MinT is optimal in the sense that given a set of unbiased base forecasts, it returns a set of best linear unbiased reconciled forecasts, using as $\boldsymbol{G}$ the unique solution that minimises the trace (hence MinT) of the variance of the forecast error of the reconciled forecasts.

A significant advantage of the MinT reconciliation solution is that it is the first to incorporate the full correlation structure of the hierarchy via $\boldsymbol{W}_h$. However, estimating $\boldsymbol{W}_h$ is challenging, especially for $h > 1$. Wickramasuriya et al. (2018) present possible alternative estimators for $\boldsymbol{W}_h$ and show that these lead to different $\boldsymbol{G}$ matrices. We summarise these below.

- Set $\boldsymbol{W}_h = k_h\boldsymbol{I}_n$ for all h, where $k_h > 0$ is a proportionality constant. This simple assumption returns $\boldsymbol{G} = (\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$ so that the base forecasts are orthogonally projected onto the coherent subspace $\mathfrak{s}$ minimising the Euclidean distance between $\hat{\boldsymbol{y}}_{T+h|T}$ and $\tilde{\boldsymbol{y}}_{T+h|T}$. Hyndman et al. (2011) come to the same solution, however, from the perspective of the following regression model

$$\hat{\boldsymbol{y}}_{T+h|T} = \boldsymbol{S}\boldsymbol{\beta}_{T+h|T} + \boldsymbol{\varepsilon}_{T+h|T},$$

where $\boldsymbol{\beta}_{T+h|T} = \mathrm{E}[\boldsymbol{b}_{T+h} \mid \boldsymbol{b}_1, \ldots, \boldsymbol{b}_T]$ is the unknown conditional mean of the bottom-level series and $\boldsymbol{\varepsilon}_{T+h|T}$ is the coherence or reconciliation error with mean zero and variance $\boldsymbol{V}$. The OLS solution leads to the same projection matrix $\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$, and due to this interpretation we continue to refer to this reconciliation method as OLS. A disadvantage of the OLS solution is that the homoscedastic diagonal entries do not account for the scale differences between the levels of the hierarchy due to aggregation. Furthermore, OLS does not account for the correlations across series.

- Set $\boldsymbol{W}_h = k_h\mathrm{diag}(\hat{\boldsymbol{W}}_1)$ for all h ($k_h > 0$), where

$$\hat{\boldsymbol{W}}_1 = \frac{1}{T}\sum_{T=1}^{T} \hat{\boldsymbol{e}}_t\hat{\boldsymbol{e}}_t'$$

is the unbiased sample estimator of the in-sample one-step-ahead base forecast errors as defined in (21.3). Hence this estimator scales the base forecasts using the variance of the in-sample residuals and is therefore described and referred to as a weighted least squares (WLS) estimator applying variance scaling. A similar estimator was proposed by Hyndman et al. (2019).

An alternative WLS estimator is proposed by Athanasopoulos et al. (2017) in the context of temporal hierarchies. Here $\boldsymbol{W}_h$ is proportional to $\mathrm{diag}(\boldsymbol{S}\mathbf{1})$ where $\mathbf{1}$ is a unit column vector of dimension n. Hence the weights are proportional to the number of bottom-level variables required to form an aggregate. For example, in the hierarchy of Fig. 21.1, the weights corresponding to the Total, series A and series B are proportional to 5, 2 and 3 respectively. This weighting scheme depends only on the aggregation structure and is referred to as structural scaling. Its advantage over OLS is that it assumes equivariant forecast errors only at the bottom level of the structure and not across all levels. It is particularly useful in cases where forecast errors are not available; for example, in cases where the base forecasts are generated by judgemental forecasting.

- Set $\boldsymbol{W}_h = k_h\hat{\boldsymbol{W}}_1$ for all h ($k_h > 0$) to be proportional to the unrestricted sample covariance estimator for $h = 1$. Although this is relatively simple to obtain and provides a good solution for small hierarchies, it does not provide reliable results as m grows compared to T. This is referred to this as the MinT(Sample) estimator.

- Set $\boldsymbol{W}_h = k_h \hat{\boldsymbol{W}}_1^D$ for all h ($k_h > 0$), where $\hat{\boldsymbol{W}}_1^D = \lambda_D \mathrm{diag}(\hat{\boldsymbol{W}}_1) + (1 - \lambda_D)\hat{\boldsymbol{W}}_1$ is a shrinkage estimator with diagonal target and shrinkage intensity parameter

$$\hat{\lambda}_D = \frac{\sum_{i \neq j} \hat{\mathrm{Var}}(\hat{r}_{ij})}{\sum_{i \neq j} \hat{r}_{ij}^2},$$

  where $\hat{r}_{ij}$ is the (i, j)th element of $\hat{\boldsymbol{R}}_1$, the one-step-ahead sample correlation matrix as proposed by Schäfer and Strimmer (2005). Hence, off-diagonal elements of $\hat{\boldsymbol{W}}_1$ are shrunk towards zero while diagonal elements (variances) remain unchanged. This is referred to as the MinT(Shrink) estimator.

21.4 Hierarchical Probabilistic Forecasting

A limitation of point forecasts is that they provide no indication of uncertainty around the forecast. A richer description of forecast uncertainty can be obtained by providing a probabilistic forecast, also commonly referred to as a density forecast. For a review of probabilistic forecasts, and *scoring rules* for evaluating such forecasts, see Gneiting and Katzfuss (2014). This chapter and Chapter 24 respectively provide comprehensive summaries of methods for constructing density forecasts and predictive accuracy tests for both point and density forecasts. In recent years, the use of probabilistic forecasts and their evaluation via scoring rules has become pervasive in macroeconomic forecasting, some notable (but non-exhaustive) examples are Geweke and Amisano (2010), Billio, Casarin, Ravazzolo, and Van Dijk (2013), Carriero, Clark, and Marcellino (2015) and Clark and Ravazzolo (2015).

The literature on hierarchical probabilistic forecasting is still an emerging area of interest. To the best of our knowledge the first attempt to even define coherence in the setting of probabilistic forecasting is provided by Taieb, Taylor, and Hyndman (2017) who define a coherent forecast in terms of a convolution. An equivalent definition due to Gamakumara, Panagiotelis, Athanasopoulos, and Hyndman (2018) defines a coherent probabilistic forecast as a probability measure on the coherent subspace $\mathfrak{s}$. Gamakumara et al. (2018) also generalise the concept of forecast reconciliation to the probabilistic setting.

Definition 21.2 Let $\mathcal{A}$ be a subset[1] of $\mathfrak{s}$ and let $\mathcal{B}$ be all points in $\mathbb{R}^n$ that are mapped onto $\mathcal{A}$ after premultiplication by $\boldsymbol{SG}$. Letting $\hat{\nu}$ be a base probabilistic forecast for the full hierarchy, the coherent measure $\tilde{\nu}$ reconciles $\hat{\nu}$ if $\tilde{\nu}(\mathcal{A}) = \hat{\nu}(\mathcal{B})$ for all $\mathcal{A}$.

---

[1] Strictly speaking $\mathcal{A}$ is a Borel set.

In practice this definition leads to two approaches. For some parametric distributions, for instance the multivariate normal, a reconciled probabilistic forecast can be derived analytically. However, in macroeconomic forecasting, non-standard distributions such as bimodal distributions are often required to take different policy regimes into account. In such cases a non-parametric approach based on bootstrapping in-sample errors proposed Gamakumara et al. (2018) can be used. These scenarios are now covered in detail.

21.4.1 *Probabilistic Forecast Reconciliation in the Gaussian Framework*

In the case where the base forecasts are probabilistic forecasts characterised by elliptical distributions, Gamakumara et al. (2018) show that reconciled probabilistic forecasts will also be elliptical. This is particularly straightforward for the Gaussian distribution which is completely characterised by two moments. Letting the base probabilistic forecasts be $\mathcal{N}(\hat{\boldsymbol{y}}_{T+h|T}, \hat{\boldsymbol{\Sigma}}_{T+h|T})$, then the reconciled probabilistic forecasts will be $\mathcal{N}(\tilde{\boldsymbol{y}}_{T+h|T}, \tilde{\boldsymbol{\Sigma}}_{T+h|T})$, where

$$\tilde{\boldsymbol{y}}_{T+h|T} = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}_{T+h|T} \tag{21.5}$$

$$\text{and} \qquad \tilde{\boldsymbol{\Sigma}}_{T+h|T} = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{\Sigma}}_{T+h|T}\boldsymbol{G}'\boldsymbol{S}'. \tag{21.6}$$

There are several options for obtaining the base probabilistic forecasts and in particular the variance covariance matrix $\hat{\boldsymbol{\Sigma}}$. One option is to fit multivariate models either level by level or for the hierarchy as a whole leading respectively to a $\hat{\boldsymbol{\Sigma}}$ that is block diagonal or dense. Another option is to fit univariate models for each individual series in which case $\hat{\boldsymbol{\Sigma}}$ is a diagonal matrix. A third option that we employ here is to obtain $\hat{\boldsymbol{\Sigma}}$ using in-sample forecast errors, in a similar vein to how $\hat{\boldsymbol{W}}_1$ is estimated in the MinT method. Here the same shrinkage estimator described in Sect. 21.3.2 is used. The reconciled probabilistic forecast will ultimately depend on the choice of $\boldsymbol{G}$; the same choices of $\boldsymbol{G}$ matrices used in Sect. 21.3 can be used.

21.4.2 *Probabilistic Forecast Reconciliation in the Non-parametric Framework*

In many applications, including macroeconomic forecasting, it may not be reasonable to assume Gaussian predictive distributions. Therefore, non-parametric approaches have been widely used for probabilistic forecasts in different disciplines. For example, ensemble forecasting in weather applications (Gneiting, 2005; Gneiting & Katzfuss, 2014; Gneiting, Stanberry, Grimit, Held, & Johnson, 2008), and bootstrap-based approaches (Manzan & Zerom, 2008; Vilar & Vilar, 2013). In macroeconomics, Cogley, Morozov, and Sargent (2005) discuss the importance of allowing for skewness in density forecasts and more recently Smith and Vahey (2016) discuss this issue in detail.

Due to these concerns, we employ the bootstrap method proposed by Gamakumara et al. (2018) that does not make parametric assumptions about the predictive distribution. An important result exploited by this method is that applying point forecast reconciliation to the draws from an incoherent base predictive distribution, results in a sample from the reconciled predictive distribution. We summarise this process below:

1. Fit univariate models to each series in the hierarchy over a training set from $t = 1, \ldots, T$. Let these models denote $M_1, \ldots, M_n$.
2. Compute one-step-ahead in-sample forecast errors. Collect these into an $n \times T$ matrix $\hat{\boldsymbol{E}} = (\hat{\boldsymbol{e}}_1, \hat{\boldsymbol{e}}_2, \ldots, \hat{\boldsymbol{e}}_T)$, where the n-vector $\hat{\boldsymbol{e}}_t = \boldsymbol{y}_t - \hat{\boldsymbol{y}}_{t|t-1}$. Here, $\hat{\boldsymbol{y}}_{t|t-1}$ is a vector of forecasts made for time t using information up to and including time $t-1$. These are called in-sample forecasts since while they depend only on past values, information from the entire training sample is used to estimate the parameters for the models on which the forecasts are based.
3. Block bootstrap from $\hat{\boldsymbol{E}}$; that is, choose H consecutive columns of $\hat{\boldsymbol{E}}$ at random, repeating this process B times. Denote the $n \times H$ matrix obtained at iteration b as $\hat{\boldsymbol{E}}^b$ for $b = 1, \ldots, B$.
4. For all b, compute $\hat{\boldsymbol{\Upsilon}}^b = \{\hat{\boldsymbol{\gamma}}_1^b, \ldots, \hat{\boldsymbol{\gamma}}_n^b\}' \in \mathbb{R}^{n \times H} : \hat{\gamma}_{i,h}^b = f(M_i, \hat{e}_{i,h}^b)$ where, $f(.)$ is a function of fitted univariate model in step 1 and associated error. That is, $\hat{\gamma}_{i,h}$ is a sample path simulated from fitted model M_i for ith series and error approximated by the corresponding block bootstrapped sample error $\hat{e}_{i,h}^b$ which is the (i, h)th element of $\hat{\boldsymbol{E}}^b$. Each row of $\hat{\boldsymbol{\Upsilon}}^b$ is a sample path of h forecasts for a single series. Each column of $\hat{\boldsymbol{\Upsilon}}^b$ is a realisation from the joint predictive distribution at a particular horizon.
5. For each $b = 1, \ldots, B$, select the hth column of $\hat{\boldsymbol{\Upsilon}}^b$ and stack these to form an $n \times B$ matrix $\hat{\boldsymbol{\Upsilon}}_{T+h|T}$.
6. For a given $\boldsymbol{G}$ matrix and for each $h = 1, \ldots, H$, compute $\tilde{\boldsymbol{\Upsilon}}_{T+h|T} = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{\Upsilon}}_{T+h|T}$. Each column of $\tilde{\boldsymbol{\Upsilon}}_{T+h|T}$ is a realisation from the joint h-step-ahead reconciled predictive distribution.

21.5 Australian GDP

In our empirical application we consider Gross Domestic Product (GDP) of Australia with quarterly data spanning the period 1984:Q4–2018:Q3. The Australian Bureau of Statistics (ABS) measures GDP using three main approaches namely Production, Income, and Expenditure. The final GDP figure is obtained as an average of these three figures. Each of these measures is aggregates of economic variables which are also targets of interests for the macroeconomic forecaster. This suggests a hierarchical approach to forecasting could be used to improve forecasts of all series in the hierarchy including the headline GDP.

We concentrate on the Income and Expenditure approaches as nominal data are available only for these two. We restrict our attention to nominal data due to the fact that real data are constructed via a chain price index approach with different price deflators used for each series. As a result, real GDP data are not coherent—the aggregate series is not a linear combination of the disaggregate series. For similar reasons we do not use seasonally adjusted data; the process of seasonal adjustment results in data that are not coherent. Finally, although there is a small statistical discrepancy between each series and the headline GDP figure, we simply treat this statistical discrepancy, which is also published by the ABS, as a time series in its own right. For further of the details on the data please refer to Australian Bureau of Statistics (2018).

AQ3

21.5.1 Income Approach

Using the income approach, GDP is calculated by aggregating all income flows. In particular, GDP at purchaser's price is the sum of all factor incomes and taxes, minus subsidies on production and imports (Australian Bureau of Statistics, 2015):

$$\begin{aligned} GDP = {} & Gross\ operating\ surplus + Gross\ mixed\ income \\ & + Compensation\ of\ employees \\ & + Taxes\ less\ subsidies\ on\ production\ and\ imports \\ & + Statistical\ discrepancy\ (I). \end{aligned}$$

Figure 21.4 shows the full hierarchical structure capturing all components aggregated to form GDP using the income approach. The hierarchy has two levels of aggregation below the top-level, with a total of $n = 16$ series across the whole structure and $m = 10$ series at the bottom level.

21.5.2 Expenditure Approach

In the expenditure approach, GDP is calculated as the aggregation of final consumption expenditure, gross fixed capital formation (GFCF), changes in inventories of finished goods, work-in-progress, and raw materials and the value of exports

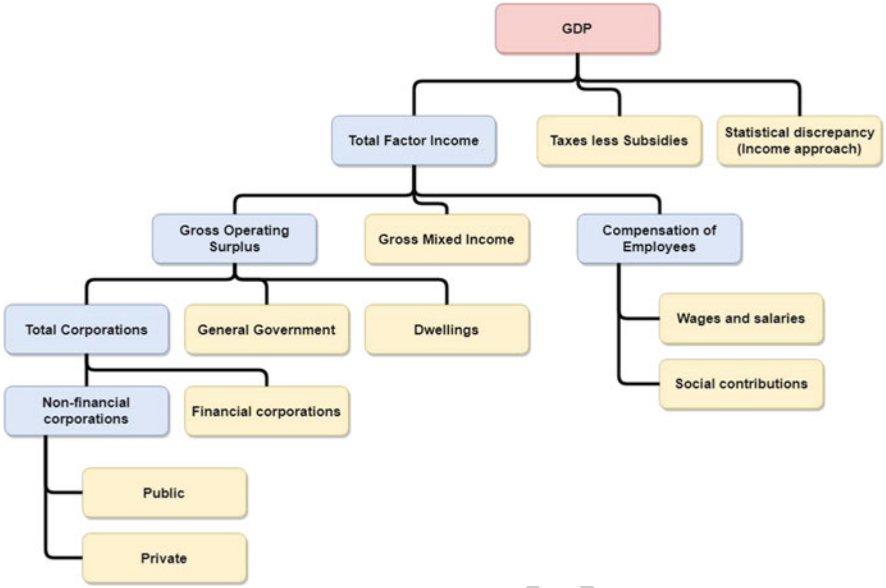


Fig. 21.4 Hierarchical structure of the income approach for GDP. The pink cell contains GDP the most aggregate series. The blue cells contain intermediate-level series and the yellow cells correspond to the most disaggregate bottom-level series

less imports of the goods and services (Australian Bureau of Statistics, 2015). The underlying equation is:

$$\begin{aligned} GDP &= Final\ consumption\ expenditure + Gross\ fixed\ capital\ formation \\ &\quad + Changes\ in\ inventories + Trade\ balance + Statistical\ discrepancy\ (E). \end{aligned}$$

Figures 21.5, 21.6, and 21.7 show the full hierarchical structure capturing all components aggregated to form GDP using the expenditure approach. The hierarchy has three levels of aggregation below the top-level, with a total of $n = 80$ series across the whole structure and $m = 53$ series at the bottom level. Descriptions of each series in these hierarchies along with the series ID assigned by the ABS are given in the Tables 21.1, 21.2, 21.3, and 21.4 in the Appendix.

Figure 21.8 displays time series from the income and expenditure approaches. The top panel shows the most aggregate GDP series. The panels below show series from lower levels for the income hierarchy (left panel) and the expenditure hierarchy (right panel). The plots show the diverse features of the time series with some displaying positive and others negative trending behaviour, some showing no trends but possibly a cycle, and some having a strong seasonal component. These highlight the need to account for and model all information and diverse signals from each series in the hierarchy, which can only be achieved through a forecast reconciliation approach.

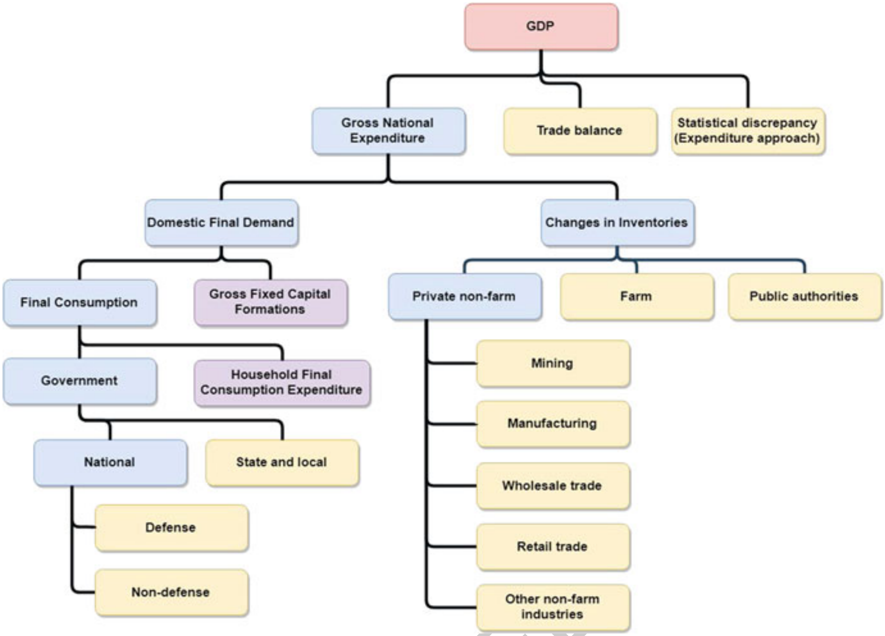


Fig. 21.5 Hierarchical structure of the expenditure approach for GDP. The pink cell contains GDP, the most aggregate series. The blue and purple cells contain intermediate-level series with the series in the purple cells further disaggregated in Figs. 21.6 and 21.7. The yellow cells contain the most disaggregate bottom-level series

21.6 Empirical Application Methodology

We now demonstrate the potential for reconciliation methods to improve forecast accuracy for Australian GDP. We consider forecasts from $h = 1$ quarter ahead up to $h = 4$ quarters ahead using an *expanding* window. First, the training sample is set from 1984:Q4 to 1994:Q3 and forecasts are produced for 1994:Q4 to 1995:Q3. Then the training window is expanded by one quarter at a time, i.e., from 1984:Q4 to 2017:Q4 with the final forecasts produced for the last available observation in 2018:Q1. This leads to 94 1-step-ahead, 93 2-steps-ahead, 92 3-steps-ahead, and 91 4-steps-ahead forecasts available for evaluation.

21.6.1 Models

The first task in forecast reconciliation is to obtain base forecasts for all series in the hierarchy. In the case of the income approach, this necessitates forecasting $n = 16$ separate time series while in the case of the expenditure approach, forecasts

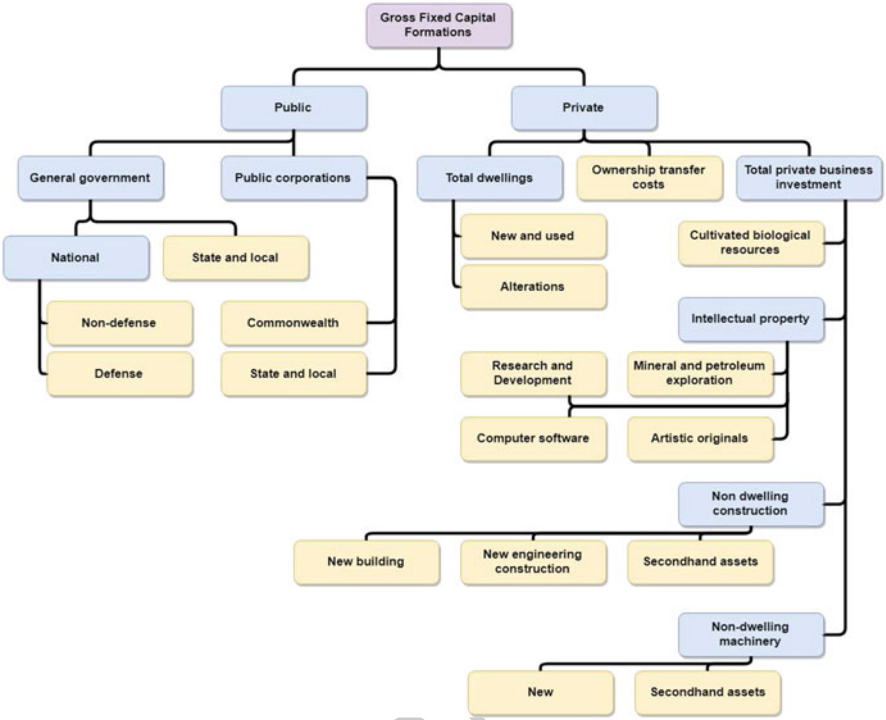


Fig. 21.6 Hierarchical structure for Gross Fixed Capital Formations under the expenditure approach for GDP, continued from Fig. 21.5. Blue cells contain intermediate-level series and the yellow cells correspond to the most disaggregate bottom-level series

for $n = 80$ separate time series must be obtained. Given the diversity in these time series discussed in Sect. 21.5, we focus on an approach that is fast but also flexible. We consider simple univariate ARIMA models, where model order is selected via a combination of unit root testing and the AIC using an algorithm developed by Hyndman, Koehler, Ord, and Snyder (2008) and implemented in the `auto.arima()` function in Hyndman, Lee, and Wang (2019). A similar approach was also undertaken using the ETS framework to produce base forecasts (Hyndman & Khandakar, 2008). Using ETS models to generate base forecasts had minimal impact on our conclusions with respect to forecast reconciliation methods and in most cases ARIMA forecasts were found to be more accurate than ETS forecasts. Consequently for brevity, we have excluded presenting the results for ETS models. However, these are available from github[2] and are discussed in detail in Gamakumara (2019). We note that a number of more complicated approaches could have been used to obtain base forecasts including multivariate models such

---

[2] The relevant github repository is https://github.com/PuwasalaG/Hierarchical-Book-Chapter.

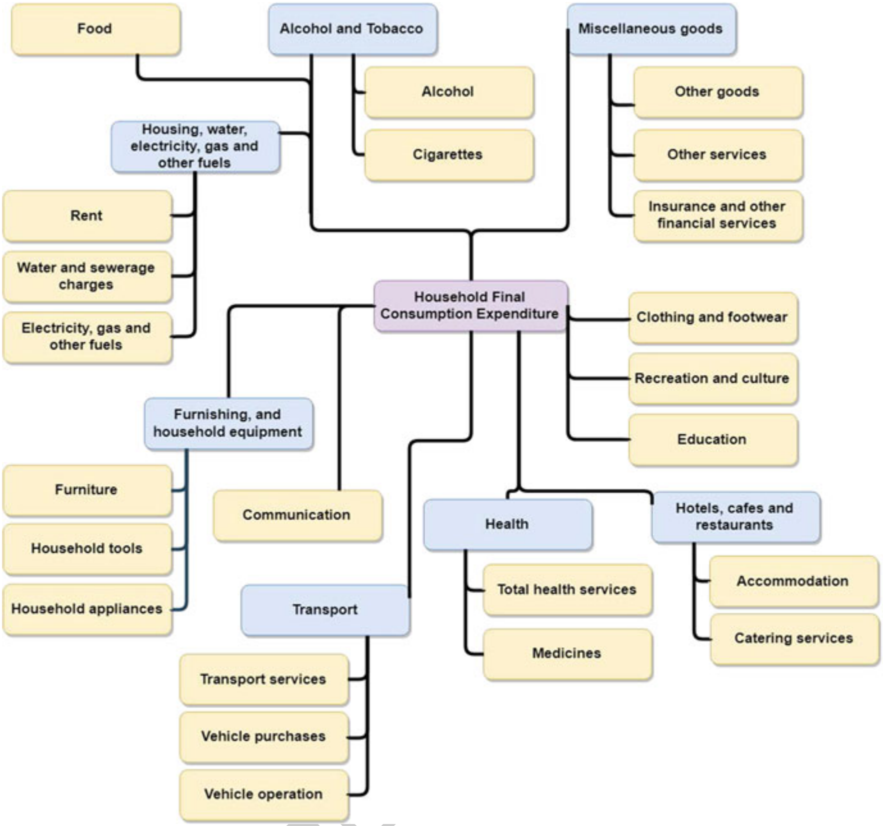


Fig. 21.7 Hierarchical structure for Household Final Consumption Expenditure under the expenditure approach for GDP, continued from Fig. 21.5. Blue cells contain intermediate-level series and the yellow cells correspond to the most disaggregate bottom-level series

as vector autoregressions, and models and methods that handle a large number of predictors such as factor models or least angle regression. However, Panagiotelis, Athanasopoulos, Hyndman, Jiang, and Vahid (2019) show that univariate ARIMA models are highly competitive for forecasting Australian GDP even compared to these methods, and in any case our primary motivation is to demonstrate the potential of forecast reconciliation.

The hierarchical forecasting approaches we consider are bottom-up, OLS, WLS with variance scaling and the MinT(Shrink) approach. The MinT(Sample) approach was also used but due to the size of the hierarchy, forecasts reconciled via this approach were less stable. Finally, all forecasts (both base and coherent) are compared to a seasonal naïve benchmark (Hyndman & Athanasopoulos, 2018); i.e., the forecast for GDP (or one of its components) is the realised GDP in the same quarter of the previous year. The naïve forecasts are by construction coherent and therefore do not need to be reconciled.

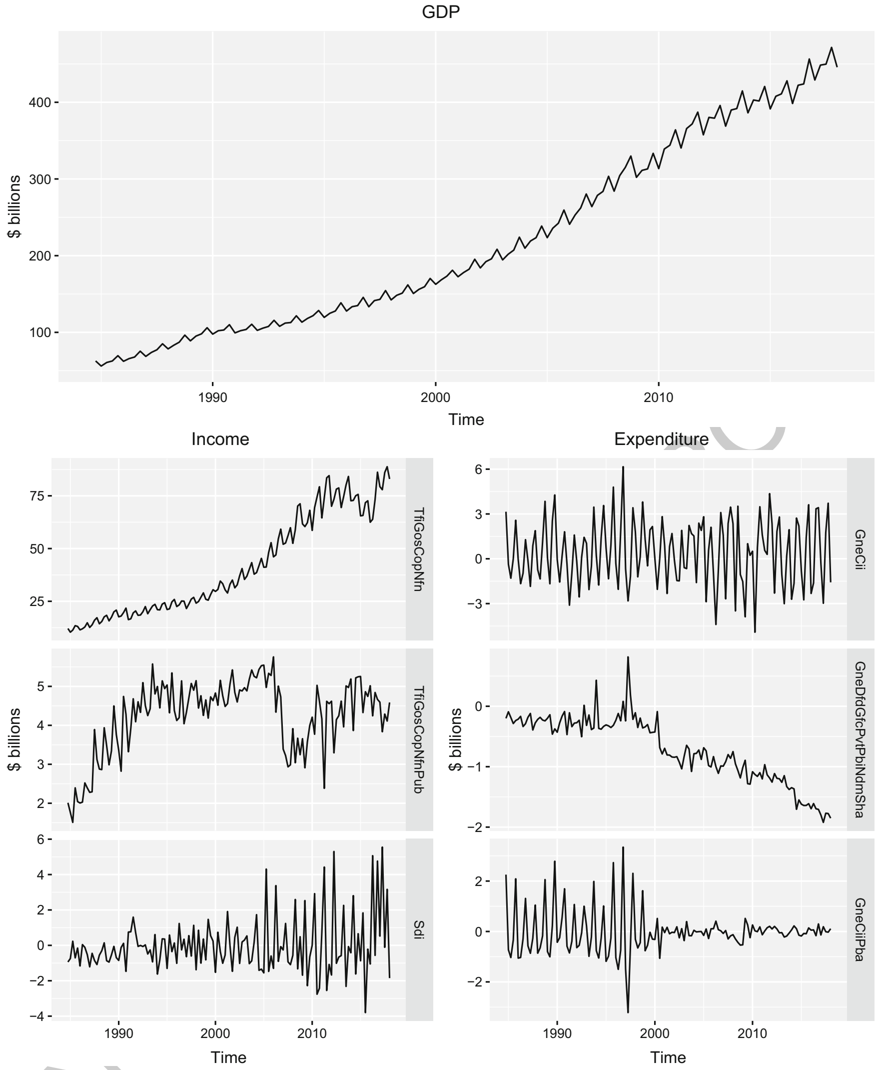


Fig. 21.8 Time plots for series from different levels of income and expenditure hierarchies

21.6.2 Evaluation

For evaluating point forecasts we consider two metrics, the Mean Squared Error (MSE) and the Mean Absolute Scaled Error (MASE) calculated over the expanding window. The absolute scaled error is defined as

$$q_{T+h} = \frac{|\breve{e}_{T+h|T}|}{(T-4)^{-1}\sum_{t=5}^{T}|y_t - y_{t-4}|},$$

where $\breve{e}_{t+h}$ is the difference between any forecast and the realisation,[3] and 4 is used due to the quarterly nature of the data. An advantage of using MASE is that it is a scale independent measure. This is particularly relevant for hierarchical time series, since aggregate series by their very nature are on a larger scale than disaggregate series. Consequently, scale dependent metrics may unfairly favour methods that perform well for the aggregate series but poorly for disaggregate series. For more details on different point forecast accuracy measures, refer to Chapter 3 of Hyndman and Athanasopoulos (2018).

Forecast accuracy of probabilistic forecasts can be evaluated using scoring rules (Gneiting & Katzfuss, 2014). Let $\breve{F}$ be a probabilistic forecast and let $\breve{\boldsymbol{y}} \sim \breve{F}$ where a breve is again used to denote that either base forecasts or coherent forecasts can be evaluated. The accuracy of multivariate probabilistic forecasts will be measured by the energy score given by

$$eS(\breve{F}_{T+h|T}, \boldsymbol{y}_{T+h}) = \mathrm{E}_{\breve{F}} \|\breve{\boldsymbol{y}}_{T+h} - \boldsymbol{y}_{T+h}\|^{\alpha} - \frac{1}{2}\mathrm{E}_{\breve{F}} \|\breve{\boldsymbol{y}}_{T+h} - \breve{\boldsymbol{y}}^{*}_{T+h}\|^{\alpha} ,$$

where $\boldsymbol{y}_{T+h}$ is the realisation at time $T+h$, and $\alpha \in (0, 2]$. We set $\alpha = 1$, noting that other values of α give similar results. The expectations can be evaluated numerically as long as a sample from $\breve{F}$ is available, which is the case for all methods we employ. An advantage of using energy scores is that in the univariate case it simplifies to the commonly used cumulative rank probability score (CRPS) given by

$$\mathrm{CRPS}(\breve{F}_i, y_{i,T+h}) = \mathrm{E}_{\breve{F}_i} |\breve{y}_{i,T+h} - y_{i,T+h}| - \frac{1}{2}\mathrm{E}_{\breve{F}_i} |\breve{y}_{i,T+h} - \breve{y}^{*}_{i,T+h}|,$$

where the subscript i is used to denote that CRPS measures forecast accuracy for a single variable in the hierarchy.

Alternatives to the energy score were also considered, namely log scores and variogram scores. The log score was disregarded since Gamakumara et al. (2018) prove that the log score is improper with respect to the class of incoherent probabilistic forecasts when the true DGP is coherent. The variogram score gave similar results to the energy score; these results are omitted for brevity but are available from github and are discussed in detail in Gamakumara (2019).

21.7 Results

21.7.1 *Base Forecasts*

Due to the different features in each time series, a variety of ARIMA and seasonal ARIMA models were selected for generating base forecasts. For example, in the

---

[3] Breve is used instead of a hat or tilde to denote that this can be the error for either a base or reconciled forecast.

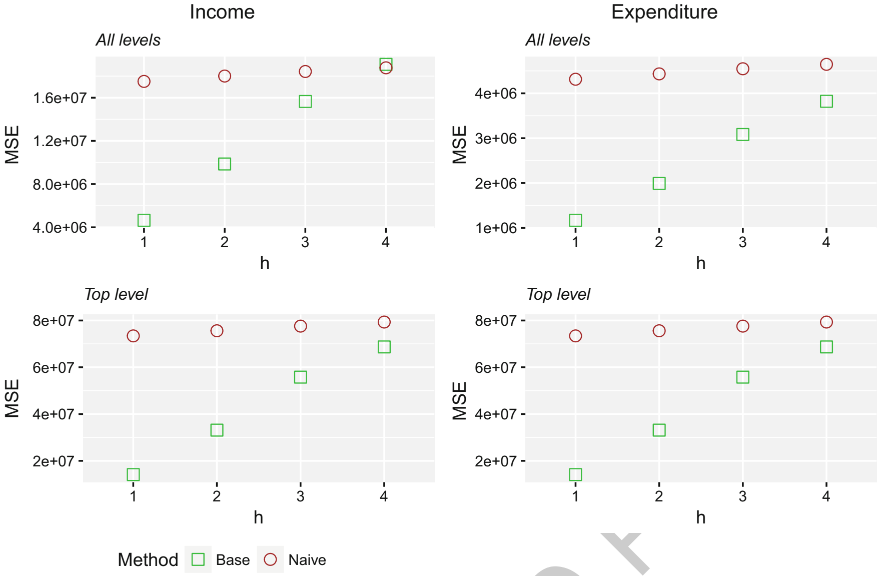


Fig. 21.9 Mean squared errors for naïve and ARIMA base forecasts. Top panels refer to results summarised over all series while bottom panels refer to results for the top-level GDP series. Left panels refer to the income hierarchy and right panels to the expenditure hierarchy

income hierarchy, some series require seasonal differencing while other did not. Furthermore the AR orders vary from 0 to 3, the MA orders from 0 to 2, and their seasonal counterparts SAR from 0 to 2 and SMA from 0 to 1. Figure 21.9 compares the accuracy of the ARIMA base forecasts to the seasonal naïve forecasts over different forecast horizons. The panels on the left show results for the Income hierarchy while the panels on the right show the results for the Expenditure hierarchy. The top panels summarise results over all series in the hierarchy, i.e., we calculate the MSE for each series and then average over all series. The bottom panels show the results for the aggregate level GDP.

The clear result is that base forecasts are more accurate than the naïve forecasts, however, as the forecasting horizon increases, the differences become smaller. This is to be expected since the naïve model here is a seasonal random walk, and for horizons $h < 4$, forecasts from an ARIMA model are based on more recent information. Similar results are obtained when MASE is used as the metric for evaluating forecast accuracy.

One disadvantage of the base forecasts relative to the naïve forecasts is that base forecasts are not coherent. As such we now turn our attention to investigating whether reconciliation approaches can lead to further improvements in forecast accuracy relative to the base forecasts.

21.7.2 *Point Forecast Reconciliation*

We now turn our attention to evaluating the accuracy of point forecasts obtained using the different reconciliation approaches as well as the single-level bottom-up approach. All results in subsequent figures are presented as the percentage changes in a forecasting metric relative to base forecasts, a measure known in the forecasting literature as *skill scores*. Skill scores are computed such that positive values represent an improvement in forecasting accuracy over the base forecasts while negative values represent a deterioration.

Figures 21.10 and 21.11 show skill scores using MSE and MASE respectively. The top row of each figure shows skill scores based on averages over all series. We conclude that reconciliation methods generally improve forecast accuracy relative to base forecasts regardless of the hierarchy used, the forecasting horizon, the forecast error measure or the reconciliation method employed. We do, however, note that while all reconciliation methods improve forecast performance, MinT(Shrink) is the best forecasting method in most cases.

To further investigate the results we break down the skill scores by different levels of each hierarchy. The second row of Figs. 21.10 and 21.11 shows the skill scores for a single series, namely GDP which represents the top-level of both hierarchies. The third row shows results for all series excluding those of the bottom level, while the final row shows results for the bottom-level series only. Here, we see two general features. The first is that OLS reconciliation performs poorly on the bottom-level series, and the second is that bottom-up performs relatively poorly on aggregate series. The two features are particularly exacerbated for the larger expenditure hierarchy. These results are consistent with other findings in the forecast reconciliation literature (see for instance Athanasopoulos et al., 2017; Wickramasuriya et al., 2018).

21.7.3 *Probabilistic Forecast Reconciliation*

We now turn our attention towards results for probabilistic forecasts. Figure 21.12 shows results for the energy score which as a multivariate score summarises forecast accuracy over the entire hierarchy. Once again all results are presented

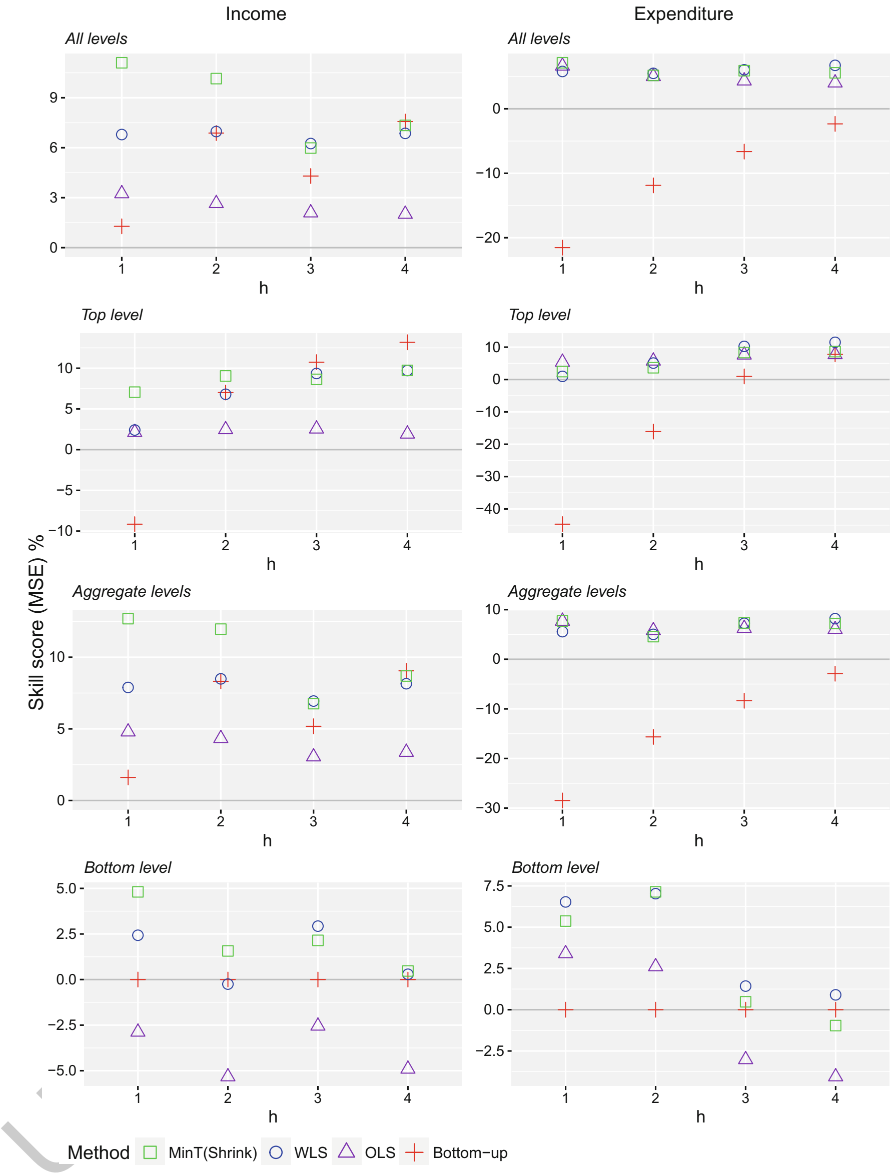


Fig. 21.10 Skill scores for point forecasts from alternative methods (with reference to base forecasts) using MSE. The left panels refer to the income hierarchy while the right panels refer to the expenditure hierarchy. The first row refers to results summarised over all series, the second row to top-level GDP series, the third row to aggregate levels, and the last row to the bottom level

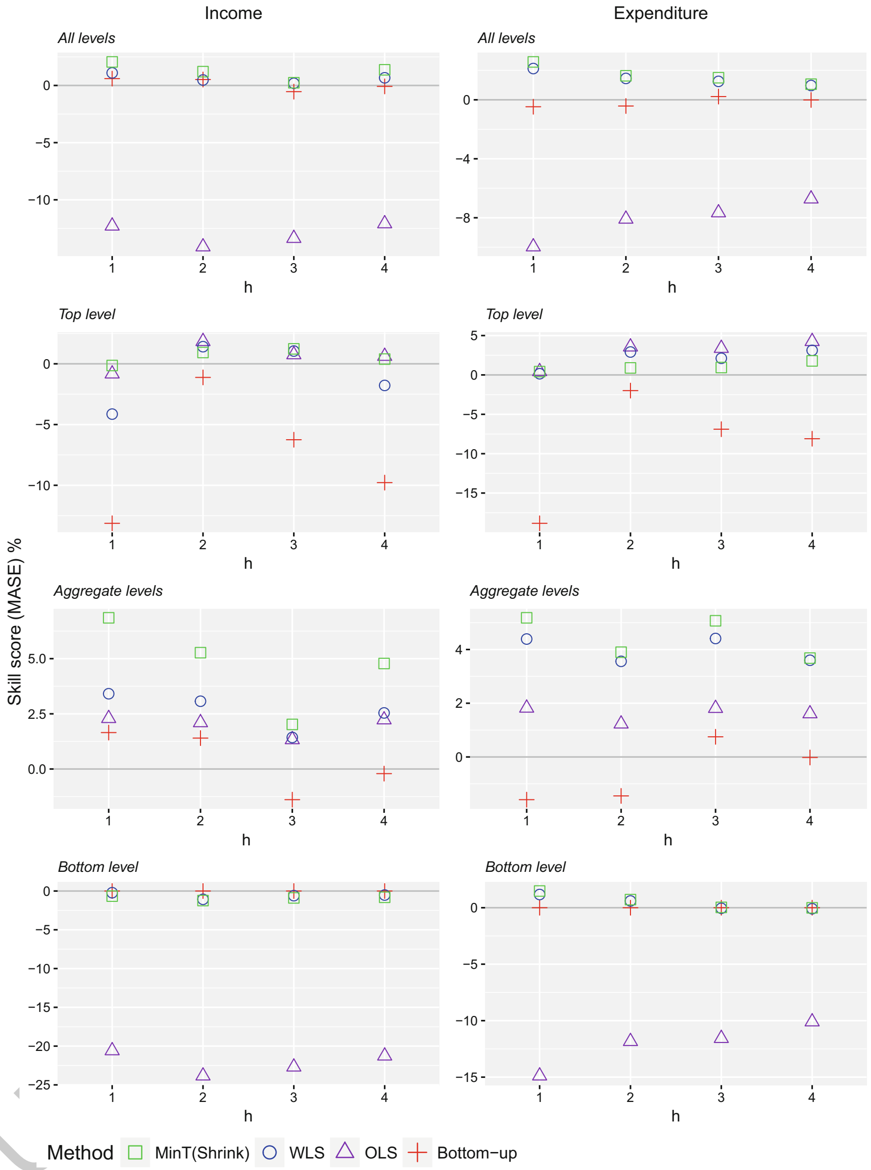


Fig. 21.11 Skill scores for point forecasts from different reconciliation methods (with reference to base forecasts) using MASE. The left two panels refer to the income hierarchy and the right two panels to the expenditure hierarchy. The first row refers to results summarised over all series, the second row to top-level GDP series, the third row to aggregate levels, and the last row to the bottom level

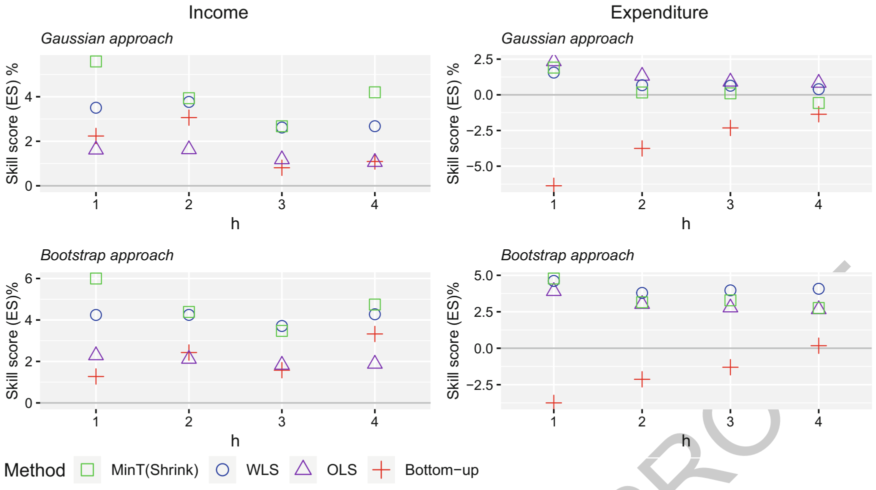


Fig. 21.12 Skill scores for multivariate probabilistic forecasts from different reconciliation methods (with reference to base forecasts) using energy scores. The top panels refer to the results for the Gaussian approach and the bottom panels to the non-parametric bootstrap approach. Left panels refer to the income hierarchy and right panels to the expenditure hierarchy

as skill scores relative to base forecasts. The top panels refer to results assuming Gaussian probabilistic forecasts as described in Sect. 21.4.1 while the bottom panels refer to the non-parametric bootstrap method described in Sect. 21.4.2. The left panels correspond to the income hierarchy while the right panels correspond to the expenditure hierarchy. For the income hierarchy, all methods improve upon base forecasts at all horizons. In nearly all cases the best performing reconciliation method is MinT(Shrink), a notable result since the optimal properties for MinT have thus far only been established theoretically in the point forecasting case. For the larger expenditure hierarchy results are a little more mixed. While bottom-up tends to perform poorly, all reconciliation methods improve upon base forecasts (with the single exception of MinT(Shrink) in the Gaussian framework four quarters ahead). Interestingly, OLS performs best under the assumption of Gaussianity—this may indicate that OLS is a more robust method under model misspecification but further investigation is required.

Finally, Fig. 21.13 displays the skill scores based on the cumulative ranked probability score for a single series, namely top-level GDP. The cause of the poor performance of bottom-up reconciliation as a failure to accurately forecast aggregate series is apparent here.

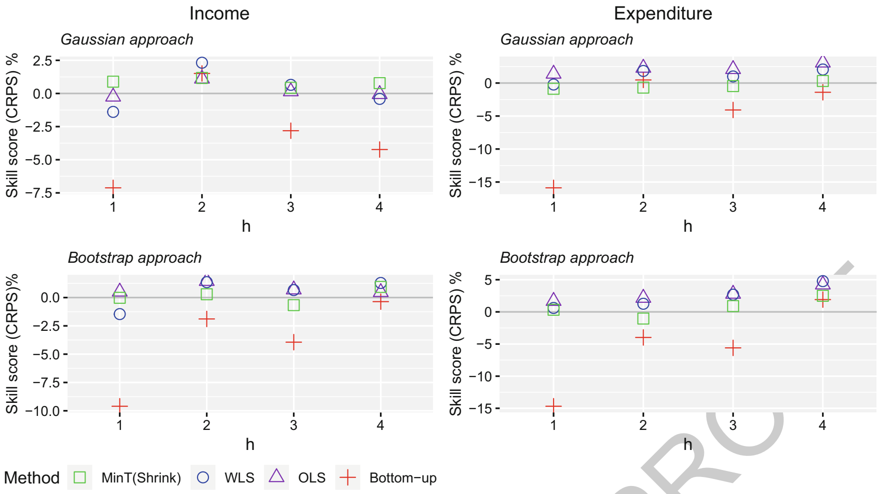


Fig. 21.13 Skill scores for probabilistic forecasts of top-level GDP from different reconciliation methods (with reference to base forecasts) using CRPS. Top panels refer to the results for Gaussian approach and bottom panels refer to the non-parametric bootstrap approach. The left panel refers to the income hierarchy and the right panel to the expenditure hierarchy

21.8 Conclusions

In the macroeconomic setting, we have demonstrated the potential for forecast reconciliation methods to not only provide coherent forecasts, but to also improve overall forecast accuracy. This result holds for both point forecasts and probabilistic forecasts, for the two different hierarchies we consider and over different forecasting horizons. Even where the objective is to only forecast a single series, for instance top-level GDP, the application of forecast reconciliation methods improves forecast accuracy.

By comparing results from different forecast reconciliation techniques we draw a number of conclusions. Despite its simplicity, the single-level bottom-up approach can perform poorly at more aggregated levels of the hierarchy. Meanwhile, when forecast accuracy at the bottom level is evaluated, OLS tends to break down in some instances. Overall, the WLS and MinT(Shrink) methods (and particularly the latter) tend to yield the highest improvements in forecast accuracy. Similar results can be found in both simulations and the empirical studies of Athanasopoulos et al. (2017) and Wickramasuriya et al. (2018).

There are a number of open avenues for research in the literature on forecast reconciliation, some of which are particularly relevant to macroeconomic applications. First there is scope to consider more complex aggregation structures, for instance in addition to the hierarchies we have already considered, data on GDP and GDP components disaggregated along geographical lines are also available. This leads to a grouped aggregation structure. Also, given the substantial literature on the optimal frequency at which to analyse macroeconomic data, a study on forecasting GDP or other variables as a temporal hierarchy may be of interest. In this chapter we have only shown that reconciliation methods can be used to improve forecast accuracy when univariate ARIMA models are used to produce base forecasts. It will be interesting to evaluate whether such results hold when a multivariate approach, e.g., a Bayesian VAR or dynamic factor model, is used to generate base forecasts, or whether the gains from forecast reconciliation would be more modest. Finally, a current limitation of the forecast reconciliation literature is that it only applies to collections of time series that adhere to linear constraints. In macroeconomics there are many examples of data that adhere to non-linear constraints, for instance real GDP is a complicated but deterministic function of GDP components and price deflators. The extension of forecast reconciliation methods to non-linear constraints potentially holds great promise for continued improvement in macroeconomic forecasting.

Appendix

See Tables 21.1, 21.2, 21.3, 21.4.

Table 21.1 Variables, series IDs and their descriptions for the income approach

Variable	Series ID	Description
Gdpi	A2302467A	GDP(I)
Sdi	A2302413V	Statistical discrepancy (I)
Tsi	A2302412T	Taxes less subsidies (I)
TfiCoeWns	A2302399K	Compensation of employees; Wages and salaries
TfiCoeEsc	A2302400J	Compensation of employees; Employers' social contributions
TfiCoe	A2302401K	Compensation of employees
TfiGosCopNfnPvt	A2323369L	Private non-financial corporations; Gross operating surplus
TfiGosCopNfnPub	A2302403R	Public non-financial corporations; Gross operating surplus
TfiGosCopNfn	A2302404T	Non-financial corporations; Gross operating surplus
TfiGosCopFin	A2302405V	Financial corporations; Gross operating surplus
TfiGosCop	A2302406W	Total corporations; Gross operating surplus
TfiGosGvt	A2298711F	General government; Gross operating surplus
TfiGosDwl	A2302408A	Dwellings owned by persons; Gross operating surplus
TfiGos	A2302409C	All sectors; Gross operating surplus
TfiGmi	A2302410L	Gross mixed income
Tfi	A2302411R	Total factor income

Table 21.2 Variables, series IDs and their descriptions for expenditure approach

Variable	Series ID	Description
Gdpe	A2302467A	GDP(E)
Sde	A2302566J	Statistical discrepancy(E)
Exp	A2302564C	Exports of goods and services
Imp	A2302565F	Imports of goods and services
Gne	A2302563A	Gross national exp.
GneDfdFceGvtNatDef	A2302523J	Gen. gov.—National; Final consumption exp.—Defence
GneDfdFceGvtNatNdf	A2302524K	Gen. gov.—National; Final consumption exp.—Non-defence
GneDfdFceGvtNat	A2302525L	Gen. gov.—National; Final consumption exp.
GneDfdFceGvtSnl	A2302526R	Gen. gov.—State and local; Final consumption exp,
GneDfdFceGvt	A2302527T	Gen. gov.; Final consumption exp.
GneDfdFce	A2302529W	All sectors; Final consumption exp.
GneDfdGfcPvtTdwNnu	A2302543T	Pvt.; Gross fixed capital formation (GFCF)
GneDfdGfcPvtTdwAna	A2302544V	Pvt.; GFCF—Dwellings—Alterations and additions
GneDfdGfcPvtTdw	A2302545W	Pvt.; GFCF—Dwellings—Total
GneDfdGfcPvtOtc	A2302546X	Pvt.; GFCF—Ownership transfer costs
GneDfdGfcPvtPbiNdcNbd	A2302533L	Pvt. GFCF—Non-dwelling construction—New building
GneDfdGfcPvtPbiNdcNec	A2302534R	Pvt.; GFCF—Non-dwelling construction—New engineering construction
GneDfdGfcPvtPbiNdcSha	A2302535T	Pvt.; GFCF—Non-dwelling construction—Net purchase of second hand assets
GneDfdGfcPvtPbiNdc	A2302536V	Pvt.; GFCF—Non-dwelling construction—Total
GneDfdGfcPvtPbiNdmNew	A2302530F	Pvt.; GFCF—Machinery and equipment—New
GneDfdGfcPvtPbiNdmSha	A2302531J	Pvt.; GFCF—Machinery and equipment—Net purchase of second hand assets
GneDfdGfcPvtPbiNdm	A2302532K	Pvt.; GFCF—Machinery and equipment—Total
GneDfdGfcPvtPbiCbr	A2716219R	Pvt.; GFCF—Cultivated biological resources
GneDfdGfcPvtPbiIprRnd	A2716221A	Pvt.; GFCF—Intellectual property products—Research and development
GneDfdGfcPvtPbiIprMnp	A2302539A	Pvt.; GFCF—Intellectual property products—Mineral and petroleum exploration
GneDfdGfcPvtPbiIprCom	A2302538X	Pvt.; GFCF—Intellectual property products—Computer software
GneDfdGfcPvtPbiIprArt	A2302540K	Pvt.; GFCF—Intellectual property products—Artistic originals
GneDfdGfcPvtPbiIpr	A2716220X	Pvt.; GFCF—Intellectual property products Total
GneDfdGfcPvtPbi	A2302542R	Pvt.; GFCF—Total private business investment
GneDfdGfcPvt	A2302547A	Pvt.; GFCF
GneDfdGfcPubPcpCmw	A2302548C	Plc. corporations—Commonwealth; GFCF

(continued)

Table 21.2 (continued)

Variable	Series ID	Description
GneDfdGfcPubPcpSnl	A2302549F	Plc. corporations—State and local; GFCF
GneDfdGfcPubPcp	A2302550R	Plc. corporations; GFCF Total
GneDfdGfcPubGvtNatDef	A2302551T	Gen. gov.—National; GFCF—Defence
GneDfdGfcPubGvtNatNdf	A2302552V	Gen. gov.—National; GFCF—Non-defence
GneDfdGfcPubGvtNat	A2302553W	Gen. gov.—National; GFCF Total
GneDfdGfcPubGvtSnl	A2302554X	Gen. gov.—State and local; GFCF
GneDfdGfcPubGvt	A2302555A	Gen. gov.; GFCF
GneDfdGfcPub	A2302556C	Plc.; GFCF
GneDfdGfc	A2302557F	All sectors; GFCF

Table 21.3 Variables, series IDs and their descriptions for changes in inventories—expenditure approach

Variable	Series ID	Description
GneCii	A2302562X	Changes in Inventories
GneCiiPfm	A2302560V	Farm
GneCiiPba	A2302561W	Public authorities
GneCiiPnf	A2302559K	Private; Non-farm Total
GneCiiPnfMin	A83722619L	Private; Mining (B)
GneCiiPnfMan	A3348511X	Private; Manufacturing (C)
GneCiiPnfWht	A3348512A	Private; Wholesale trade (F)
GneCiiPnfRet	A3348513C	Private; Retail trade (G)
GneCiiPnfOnf	A2302273C	Private; Non-farm; Other non-farm industries

Table 21.4 Variables, series IDs and their descriptions for household final consumption—expenditure approach

Variable	Series ID	Description
GneDfdHfc	A2302254W	Household Final Consumption Expenditure
GneDfdFceHfcFud	A2302237V	Food
GneDfdFceHfcAbt	A3605816F	Alcoholic beverages and tobacco
GneDfdFceHfcAbtCig	A2302238W	Cigarettes and tobacco
GneDfdFceHfcAbtAlc	A2302239X	Alcoholic beverages
GneDfdFceHfcCnf	A2302240J	Clothing and footwear
GneDfdFceHfcHwe	A3605680F	Housing, water, electricity, gas and other fuels
GneDfdFceHfcHweRnt	A3605681J	Actual and imputed rent for housing
GneDfdFceHfcHweWsc	A3605682K	Water and sewerage charges
GneDfdFceHfcHweEgf	A2302242L	Electricity, gas and other fuel
GneDfdFceHfcFhe	A2302243R	Furnishings and household equipment
GneDfdFceHfcFheFnt	A3605683L	Furniture, floor coverings and household goods
GneDfdFceHfcFheApp	A3605684R	Household appliances

(continued)

Table 21.4 (continued)

Variable	Series ID	Description
GneDfdFceHfcFheTls	A3605685T	Household tools
GneDfdFceHfcHlt	A2302244T	Health
GneDfdFceHfcHltMed	A3605686V	Medicines, medical aids and therapeutic appliances
GneDfdFceHfcHltHsv	A3605687W	Total health services
GneDfdFceHfcTpt	A3605688X	Transport
GneDfdFceHfcTptPvh	A2302245V	Purchase of vehicles
GneDfdFceHfcTptOvh	A2302246W	Operation of vehicles
GneDfdFceHfcTptTsv	A2302247X	Transport services
GneDfdFceHfcCom	A2302248A	Communications
GneDfdFceHfcRnc	A2302249C	Recreation and culture
GneDfdFceHfcEdc	A2302250L	Education services
GneDfdFceHfcHcr	A2302251R	Hotels, cafes and restaurants
GneDfdFceHfcHcrCsv	A3605694V	Catering services
GneDfdFceHfcHcrAsv	A3605695W	Accommodation services
GneDfdFceHfcMis	A3605696X	Miscellaneous goods and services
GneDfdFceHfcMisOgd	A3605697A	Other goods
GneDfdFceHfcMisIfs	A2302252T	Insurance and other financial services
GneDfdFceHfcMisOsv	A3606485T	Other services

References

Amemiya, T., & Wu, R. Y. (1972). The effect of aggregation on prediction in the autoregressive model. *Journal of the American Statistical Association, 67*(339), 628–632. arXiv: 0026.

Andrawis, R. R., Atiya, A. F., & El-Shishiny, H. (2011). Combination of long term and short term forecasts, with application to tourism demand forecasting. *International Journal of Forecasting, 27*(3), 870–886. https://doi.org/10.1016/j.ijforecast.2010.05.019

Athanasopoulos, G., Ahmed, R. A., & Hyndman, R. J. (2009). Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting, 25*(1), 146–166. https://doi.org/10.1016/j.ijforecast.2008.07.004

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

Athanasopoulos, G., Steel, T., & Weatherburn, D. (2019). *Forecasting prison numbers: A grouped time series approach*. Melbourne: Monash University.

Australian Bureau of Statistics. (2015). Australian system of national accounts: Concepts, sources and methods. Cat 5216.0.

AQ4

Australian Bureau of Statistics. (2018). Australian national accounts: National income, expenditure and product. Cat 5206.0.

Billio, M., Casarin, R., Ravazzolo, F., & Van Dijk, H. K. (2013). Time-varying combinations of predictive densities using nonlinear filtering. *Journal of Econometrics, 177*(2), 213–232.

Brewer, K. (1973). Some consequences of temporal aggregation and systematic sampling for ARMA and ARMAX models. *Journal of Econometrics, 1*(2), 133–154. https://doi.org/10.1016/03044076(73)900158

Capistrán, C., Constandse, C., & Ramos-Francia, M. (2010). Multi-horizon inflation forecasts using disaggregated data. *Economic Modelling, 27*(3), 666–677.

Carriero, A., Clark, T. E., & Marcellino, M. (2015). Realtime nowcasting with a Bayesian mixed frequency model with stochastic volatility. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 178*(4), 837–862.

Clark, T. E., & Ravazzolo, F. (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics, 30*(4), 551–575.

Cogley, T., Morozov, S., & Sargent, T. J. (2005). Bayesian fan charts for UK inflation: Forecasting and sources of uncertainty in an evolving monetary system. *Journal of Economic Dynamics and Control, 29*(11), 1893–1925.

Dunn, D. M., Williams, W. H., & Dechaine, T. L. (1976). Aggregate versus subaggregate models in local area forecasting. *Journal of American Statistical Association, 71*(353), 68–71. https://doi.org/10.2307/2285732

Gamakumara, P. (2019). *Probabilistic forecasts in hierarchical time series* (Doctoral dissertation, Monash University).

Gamakumara, P., Panagiotelis, A., Athanasopoulos, G., & Hyndman, R. J. (2018). *Probabilistic forecasts in hierarchical time series* (Working paper No. 11/18). Monash University Econometrics & Business Statistics.

Geweke, J., & Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting, 26*(2), 216–230.

Gneiting, T. (2005). Weather forecasting with ensemble methods. *Science, 310*(5746), 248–249. https://doi.org/10.1126/science.1115255

Gneiting, T., & Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application, 1*, 125–151. https://doi.org/10.1146/annurev-statistics-062713-085831

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. https://doi.org/10.1007/s11749-008-0114-x

Gross, C. W., & Sohl, J. E. (1990). Disaggregation methods to expedite product line forecasting. *Journal of Forecasting, 9*(3), 233–254. https://doi.org/10.1002/for.3980090304

Hotta, L. K. (1993). The effect of additive outliers on the estimates from aggregated and disaggregated ARIMA models. *International Journal of Forecasting, 9*(1), 85–93. https://doi.org/10.1016/0169-2070(93)90056-S

Hotta, L. K., & Cardoso Neto, J. (1993). The effect of aggregation on prediction in autoregressive integrated moving-average models. *Journal of Time Series Analysis, 14*(3), 261–269.

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. https://doi.org/10.1016/j.csda.2011.03.006

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. OTexts. Retrieved from https://OTexts.com/fpp2

Hyndman, R. J., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., ... Zhou, Z. (2019). *Forecast: Forecasting functions for time series and linear models*. Version 8.5. Retrieved from https://CRAN.R-project.org/package=forecast

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software, 26*(3), 1–22.

Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: The state space approach*. Berlin: Springer.

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. https://doi.org/10.1016/j.csda.2015.11.007

Kourentzes, N., Petropoulos, F., & Trapero, J. R. (2014). Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting, 30*(2), 291–302. https://doi.org/10.1016/j.ijforecast.2013.09.006

Manzan, S., & Zerom, D. (2008). A bootstrap-based non-parametric forecast density. *International Journal of Forecasting, 24*(3), 535–550. https://doi.org/10.1016/j.ijforecast.2007.12.004

Marcellino, M. (1999). Some consequences of temporal aggregation in empirical analysis. *Journal of Business & Economic Statistics, 17*(1), 129–136.

Nijman, T. E., & Palm, F. C. (1990). Disaggregate sampling in predictive models. *Journal of Business & Economic Statistics, 8*(4), 405–415.

Panagiotelis, A., Athanasopoulos, G., Hyndman, R. J., Jiang, B., & Vahid, F. (2019). Macroeconomic forecasting for Australia using a large number of predictors. *International Journal of Forecasting, 35*(2), 616–633.

Rossana, R., & Seater, J. (1995). Temporal aggregation and economic times series. *Journal of Business & Economic Statistics, 13*(4), 441–451.

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), 1–30.

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*(2), 330–343. https://doi.org/10.1080/106018600.2016.1237877. arXiv: 1609.04222.

Silvestrini, A., Salto, M., Moulin, L., & Veredas, D. (2008). Monitoring and forecasting annual public deficit every month: The case of France. *Empirical Economics, 34*(3), 493–524. https://doi.org/10.1007/s00181-007-0132-7. arXiv: 0016.

Smith, M. S., & Vahey, S. P. (2016). Asymmetric forecast densities for us macroeconomic variables from a Gaussian copula model of cross-sectional and serial dependence. *Journal of Business & Economic Statistics, 34*(3), 416–434.

Taieb, S. B., Taylor, J. W., & Hyndman, R. J. (2017). Hierarchical probabilistic forecasting of electricity demand with smart meter data, 1–30.

Tiao, G. C. (1972). Asymptotic behaviour of temporal aggregates of time series. *Biometrika, 59*(3), 525–531. arXiv: 0027.

Vilar, J. A., & Vilar, J. A. (2013). Time series clustering based on nonparametric multidimensional forecast densities. *Electronic Journal of Statistics, 7*(1), 1019–1046. https://doi.org/10.1214/13-EJS800

Villegas, M. A., & Pedregal, D. J. (2018). Supply chain decision support systems based on a novel hierarchical forecasting approach. *Decision Support Systems, 114*, 29–36. https://doi.org/10.1016/j.dss.2018.08.003

Weiss, C. (2018). *Essays in hierarchical time series forecasting and forecast combination* (Doctoral dissertation, University of Cambridge).

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. https://doi.org/10.1080/01621459.2018.1448825

Yagli, G. M., Yang, D., & Srinivasan, D. (2019). Reconciling solar forecasts: Sequential reconciliation. *Solar Energy, 179*, 391–397. https://doi.org/10.1016/j.solener.2018.12.075

Yang, D., Quan, H., Disfani, V. R., & Liu, L. (2017). Reconciling solar forecasts: Geographical hierarchy. *Solar Energy, 146*, 276–286. https://doi.org/10.1016/j.solener.2017.02.010

Zellner, A., & Montmarquette, C. (1971). A study of some aspects of temporal aggregation problems in econometric analyses. *The Review of Economics and Statistics, 53*(4), 335–342.

Zhang, Y., & Dong, J. (2019). Least squares-based optimal reconciliation method for hierarchical forecasts of wind power generation. *IEEE Transactions on Power Systems* (forthcoming). https://doi.org/10.1109/TPWRS.2018.2868175

AQ5

AUTHOR QUERIES

AQ1. Please check if the affiliations are presented correctly.

AQ2. Please note that “Chapter 24” is cited in the text, but not provided in this book. Kindly check and update.

AQ3. Please check the hierarchy of the section headings and correct if necessary.

AQ4. Please provide the complete bibliographic details for Refs. Australian Bureau of Statistics (2015), Australian Bureau of Statistics (2018), and Taieb et al. (2017).

AQ5. Please update Ref. Zhang and Dong (2019), forthcoming.