

Probabilistic Forecasts for Hierarchical Time Series

Doctor of Philosophy

Pre-Submission Seminar

Candidate:

Puwasala Gamakumara

Supervisors:

Professor Rob J Hyndman
Professor George Athanasopoulos
Assoc Professor Anastasios Panagiotelis

Department of Econometrics and Business Statistics
Monash University
May, 2019

1 Introduction

Large collections of multiple time series often follow some aggregation structure. For example, the tourism flow of a country can be disaggregated according to a geographic hierarchy of states, zones and regions. Gross Domestic Product (GDP) of a country can be taken as another example since it is by definition the aggregation of other economic variables. To ensure aligned decision making, it is important that forecasts at the most disaggregated level add up to forecasts at more aggregated levels. This property is called “coherence”. On the other hand “reconciliation” is a process whereby incoherent forecasts are made coherent. Both of these concepts have been developed extensively for point forecasting. Generalising both of these concepts, particularly the latter, to probabilistic forecasting is a gap that we seek to address in this work. We do that by first providing a novel geometric interpretation to coherence and reconciliation in the point forecasting case. This can easily be generalised to probabilistic forecasting allowing us to derive further results for elliptical distributions as well as provide insight into forecast evaluation via multivariate scoring rules.

Traditional approaches to ensure coherent point forecasts produce first-stage forecasts at a single level of the hierarchy. To describe these we use the hierarchy in Figure 1 where the variable labeled *Tot* is the sum of the series *A* and series *B*, the series *A* is the sum of series *AA* and series *AB* and the series *B* is the sum of the series *BA* and *BB*. In the bottom-up approach (Dunn, Williams, and Dechaine, 1976), forecasts are produced at the most disaggregated level (series *AA*, *AB*, *BA* and *BB*) and then summed to recover forecasts for all higher-level series. Alternatively, in the top-down approach (Gross and Sohl, 1990), a top-level forecast is first produced (series *Tot*) and bottom-level forecasts are recovered by disaggregating the forecast using either historical or forecasted proportions. A middle-out approach is a hybrid between these two, that for the hierarchy in Figure 1 would produce first stage forecasts for series *A* and *B*.

In recent years, reconciliation methods introduced by Hyndman et al. (2011) have become increasingly popular. For these methods, first stage forecasts are independently produced for all series rather than series at a single level. Since these so-called ‘base’ forecasts are rarely coherent in practice, they are subsequently adjusted or ‘reconciled’ to ensure coherence. Note that we use coherence and reconciliation as distinct terms, in contrast to their at times ambiguous usage in the past. To date, reconciliation has typically been formulated as a regression problem with alternative reconciliation methods resembling different least squares estimators. These include Ordinary Least Squares OLS (Hyndman et al., 2011), Weighted Least Squares WLS (Athanasopoulos et al., 2017), and a Generalised Least Squares (GLS) estimator (Wickramasuriya, Athanasopoulos, and Hyndman, 2018) named MinT since it minimises the trace of the squared error matrix. These methods have been shown to outperform traditional alternatives across a range of simulated and real-world datasets (Athanasopoulos, Ahmed, and Hyndman, 2009; Van Erven and Cugliari, 2014; Wickramasuriya, Athanasopoulos, and Hyndman, 2018) since they use the information at all levels of the hierarchy and, in some sense, hedge against the risk of model misspecification at a single level.

A shortcoming of the existing literature is a focus on point forecasting despite an increased understanding over the past decade of the importance of providing a full predictive distribution for forecast uncertainty (see Gneiting and Katzfuss, 2014, and references therein). Indeed to the best of our knowledge, the work of Ben Taieb et al. (2017) and Jeon, Panagiotelis, and Petropoulos (2018) are the only paper to deal with coherent probabilistic forecasts Ben Taieb et al. (2017) proposed an algorithm to produce coherent probabilistic forecasts and applied it to UK electricity smart meter data. In their approach, a sample from the bottom level predictive distribution is first generated and then aggregated to obtain coherent probabilistic forecasts of the upper levels of the hierarchy. Hence this method is a bottom-up approach. They propose to first use the MinT algorithm to reconcile the means of the bottom level forecast distributions, and then a copula-based approach is employed to model the dependency structure of the hierarchy. The resulting multi-dimensional distribution is used to generating empirical forecast distributions for all bottom-level series. Thus, while Ben Taieb et al. (2017) provide coherent probabilistic forecasts, they do no forecast reconciliation of the distribu-

tions. In that sense, their approach is analogous to bottom-up point forecasting rather than forecast reconciliation.

Jeon, Panagiotelis, and Petropoulos (2018), on the other hand, is the only existing study that does reconciliation in probabilistic hierarchical forecasts. This method is based on cross-validation and it also does not assume any parametric distributions for predictive densities. They have applied this method to reconcile the probabilistic forecasts for temporal hierarchies of wind power in Crete. Although they argued that this approach works for cross-sectional hierarchies as well, they have not tested this in their work.

In contrast, the main objective of this thesis is to generalise both coherence and reconciliation from point to probabilistic forecasting. To facilitate this extension, we first provide a new interpretation of the point forecast reconciliation in terms of geometry. Next, we define the coherency and reconciliation based on the probabilistic forecast. We further introduce a parametric and a non-parametric approach for probabilistic forecast reconciliation. Finally, we apply these methods to two real word applications on Australian tourism flow forecasts and Australian GDP forecasts.

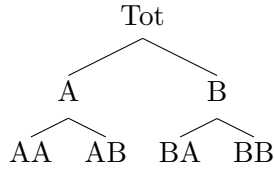


Figure 1: Two level hierarchical diagram

2 Overview of the Thesis

Forecasting hierarchical time series has been of great interest in many applications. In hierarchical time series, it is important to have “coherent” forecasts across the hierarchy. That is aggregates of the forecasts at lower levels should be equal to the forecasts at the upper levels of aggregation. While there is a rich literature on coherent hierarchical point forecasting, this research focuses on a probabilistic framework for coherent hierarchical forecasts through reconciliation. Structure of the research can be outlined as follows.

Chapter 1: Introduction

- Background and motivation to the study
- Objectives
- Outline of the succeeding chapters

Chapter 2: Literature review

This chapter contains a thorough literature review of existing hierarchical point forecasting methods. Further, we discuss the importance of probabilistic framework in time series. Also, this chapter will review existing scoring rules for evaluating multivariate probabilistic forecasts in time series.

Chapter 3: Hierarchical Forecast reconciliation

To facilitate the extension of point forecast reconciliation to probabilistic forecasting framework, we first provide a geometric interpretation of existing point reconciliation methods, re-framing them in terms of projections. In addition to being highly intuitive, this allows us to establish a number of

theoretical results. We prove two new theorems about point forecast reconciliation, the first showing that reconciliation via projections preserves the unbiasedness of base forecasts, while the second shows that reconciled forecasts dominate unreconciled forecasts via the distance reducing property of projections.

Further, this chapter discusses bias adjustments in incoherent forecasts. The projections preserve unbiasedness in reconciled forecasts only if the incoherent forecasts are unbiased. However, in this chapter, we show that projections still be used to reconcile biased incoherent forecasts after some bias adjustments. This will also produce unbiased coherent forecasts. We carry out an application on Australian prison data to show these conceptual things in this chapter.

We are planning to submit the work on this chapter to the International Journal of Forecasting.

Chapter 4: Probabilistic forecasts for hierarchical time series

Section 4.1

This chapter mainly focuses on probabilistic forecast reconciliation. We provide definitions of coherence and forecast reconciliation in the probabilistic setting and describe how these definitions lead to a reconciliation procedure that merely involves a change of basis and marginalisation. We show that probabilistic reconciliation via linear transformations can recover the true predictive distribution as long as the latter is in the elliptical class. We provide conditions for which this linear transformation is a projection, and although this projection cannot be feasibly estimated in practice, we provide a heuristic argument in favour of MinT reconciliation.

This chapter also covers the topic of forecast evaluation of probabilistic forecasts via scoring rules. In particular, we prove that for a coherent data generating process, the log score is not proper with respect to incoherent forecasts. Therefore we recommend the use of the energy score or variogram score for comparing reconciled to unreconciled forecasts. Two or more reconciled forecasts can be compared using log score, energy score or variogram score, although we show that comparisons should be made on the full hierarchy for the latter two scores.

Section 4.2

The second half of this chapter introduces a novel non-parametric framework to reconcile probabilistic forecasts. This involves producing future paths of the hierarchy by implicitly modeling the dependency structure of the hierarchy via bootstrapped training errors. Then these future paths will be reconciled to obtain coherent probabilistic forecasts. This method will be further evaluated via an extensive simulation study. A comprehensive explanation of this section is attached as the supplementary material to the paper I'm presenting in the pre-submission seminar.

Section 4.3

In the third section of this chapter, we apply both parametric and non-parametric probabilistic forecast reconciliation methods discussed in the previous section on producing coherent forecasts of Australian domestic tourism floor.

We are planning to submit the work on this chapter to the Journal of Computational Statistics and Data Analysis.

Chapter 5: Hierarchical forecasting on macroeconomic variables - An application to Australian GDP forecasts

GDP of any country calculated as the aggregation of other economic variables. Thus by definition, this forms hierarchical time series. Hence in this chapter, we apply the point and probabilistic forecast

reconciliation methods to forecast the Australian GDP. The results show that reconciliation improves the forecast performance in both point as well as probabilistic frameworks.

This chapter was submitted as a book chapter series in "Advanced Studies in Theoretical and Applied Econometrics" and it is now under review. The work on this chapter will be mainly presented in the pre-submission seminar.

Chapter 6: Conclusions

The final chapter of my thesis will consist of the concluding remarks of the whole thesis and possible future extensions.

Note:

According to the initial plan of my thesis, the first project consisted of the new definitions for point and probabilistic forecast reconciliation followed by the parametric approach. Following the feedback we received from the editors of the journal we submitted this paper, we decided to split this paper and have the first paper completely on point forecast reconciliation including some additional work on bias adjustment of point forecasts and an application to Australian prison data. All this work contains in the current working paper "Geometric View on Hierarchical Forecast Reconciliation" and this will be included as the third chapter of my thesis.

All definitions about probabilistic forecasts followed by the parametric approach are now included in the first half of the second project (Chapter 4 in my thesis) and the working paper on this is "Probabilistic Forecasts for Hierarchical Time Series"

3 Time plan for the thesis completion

June 2019 - July 2019

- Finishing the analysis of Australian prison data
- Completing the write-up of chapter 3
- Expecting to submit to IJF

July 2019 - August 2019

- Completing the write-up of chapter 4
- Expecting to submit to JBES

August 2019 - October 2019

- Completing the Introduction, Background review and Conclusion chapters and submit the thesis

Table 1: Time-line for the completion

	Thesis Chapter	Task description	Time duration	Progress
1 and 2.	Introduction and Background Review	Writing the chapter.	September/2019 - October/2019	40% Completed
3.	Hierarchical forecast reconciliation	Completing the paper.	May/2019 - July/2019	75% Completed
4.	Probabilistic forecast reconciliation for hierarchical time series	Completing the paper.	June/2019 - August/2019	90% Completed
5.	Application	Forecasting Australian GDP		100% Completed

Hierarchical forecasting on macroeconomic variables - An application to Australian GDP forecasts

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

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.

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

George Athanasopoulos
Monash University, Caulfield VIC 3162, Australia e-mail: George.Athanasopoulos@monash.edu

Puwasala Gamakumara
Monash University, Clayton VIC 3800, Australia, e-mail: Puwasala.Gamakumara@monash.edu

Anastasios Panagiotelis
Monash University, Caulfield VIC 3162, Australia e-mail: Anastasios.Panagiotelis@monash.edu

Rob J Hyndman
Monash University, Clayton VIC 3800, Australia e-mail: Rob.Hyndman@monash.edu

Mohamed Affan
e-mail: affan.momu@gmail.com

regularization 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 forecast accuracy. Examples include forecasting accidents and emergency admissions (Athanasopoulos et al. 2017), mortality rates (Shang & Hyndman 2017), prison populations (Athanasopoulos et al. 2019), retail sales (Villegas & Pedregal 2018), solar energy (Yang et al. 2017, Yagli et al. 2019), tourism demand (Athanasopoulos et al. 2009, Hyndman et al. 2011, Wickramasuriya et al. 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 et al. 2010, Weiss 2018).

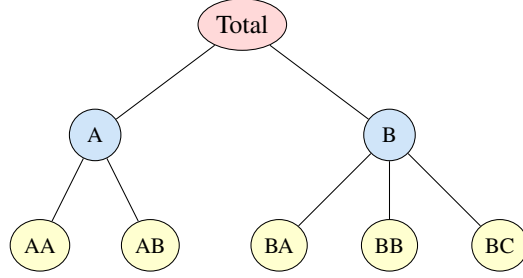
The remainder of the chapter is set out as follows. Section 2 introduces the concept of hierarchical time series, i.e. collections of time series with known linear constraints, with a particular emphasis on macroeconomic examples. Section 3 describes state-of-the-art forecast reconciliation techniques for point forecasts, while Section 4 describes the more recent extension of these techniques to probabilistic forecasting. Section 5 describes the data used in our empirical case study, namely Australian GDP data, that is represented using two alternative hierarchical structures. Section 6 provides details on the setup of our empirical study including metrics used for the evaluation of both point and probabilistic forecasts. Section 7 presents results and Section 8 concludes providing future avenues for research that are of particular relevance to the empirical macroeconomist.

2 Hierarchical time series

To simplify the introduction of some notation we use the simple two-level hierarchical structure shown in Figure 1. Denote as $y_{\text{Tot},t}$ 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 . For example, $y_{A,t}$ denotes the t th observation of the series corresponding to node A at level 1, $y_{AB,t}$ denotes the t th observation of the series corresponding to node AB at level 2, and so on.

Fig. 1 A simple two-level hierarchical structure.



Let $\mathbf{y}_t = (y_{\text{Tot},t}, y_{A,t}, y_{B,t}, y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}, y_{BC,t})'$ denote a vector containing observations across all series of the hierarchy at time t . Similarly denote as $\mathbf{b}_t = (y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}, y_{BC,t})'$ a vector containing observations only for the bottom-level series. In general, $\mathbf{y}_t \in \mathbb{R}^n$ and $\mathbf{b}_t \in \mathbb{R}^m$ 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 Figure 1, $n = 8$ and $m = 5$.

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

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t, \quad (1)$$

where

$$\mathbf{S} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ \mathbf{I}_5 \end{pmatrix}$$

is an $n \times m$ matrix referred to as the *summing matrix* and \mathbf{I}_m is an m -dimensional identity matrix. \mathbf{S} reflects the linear aggregation constraints and in particular how the bottom-level series aggregate to levels above. Thus, columns of \mathbf{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 \mathbf{S} will result in an n -dimensional vector that lies in \mathfrak{s} .

Property 1. A hierarchical time series has observations that are *coherent*, i.e., $\mathbf{y}_t \in \mathfrak{s}$ for all t . We use the term coherent to describe not just \mathbf{y}_t but any vector in \mathfrak{s} .

Structures similar to the one shown in Figure 1 can be found in macroeconomics. For instance, in Section 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 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 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 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, $\mathbf{b}_t = (y_{AX}, y_{AY}, y_{BX}, y_{BY})'$. Hence, in contrast to hierarchical structures, grouped time series do not naturally disaggregate in a unique manner.

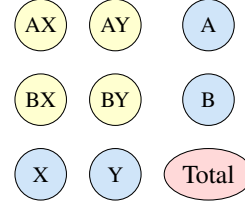


Fig. 2 A simple two-level grouped structure.

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 Zellner & Montmarquette (1971), Amemiya & Wu (1972), Tiao (1972), Brewer (1973) and others, including Hotta (1993), Hotta & Cardoso Neto (1993), Marcellino (1999), Silvestrini et al. (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 & 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 & 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 et al. 2011, Kourentzes et al. 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.

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

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 Figure 3, where $y_{\text{Tot}} = y_A + y_B$. While base forecasts could lie anywhere in \mathbb{R}^3 , the realisations and coherent forecasts lie in a two dimensional subspace $\mathfrak{s} \subset \mathbb{R}^3$.

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 the rest of the structure.

3.1.1 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 et al. 1976). In general, this consists of first generating $\hat{\mathbf{b}}_{T+h|T} \in \mathbb{R}^m$, a set of h -step-ahead forecasts for the bottom-level series. For the simple hierarchical structure of Figure 1, $\hat{\mathbf{b}}_{T+h|T} = (\hat{y}_{AA,T+h|T}, \hat{y}_{AB,T+h|T}, \hat{y}_{BA,T+h|T}, \hat{y}_{BB,T+h|T}, \hat{y}_{BC,T+h|T})$, 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

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

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.

3.1.2 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

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

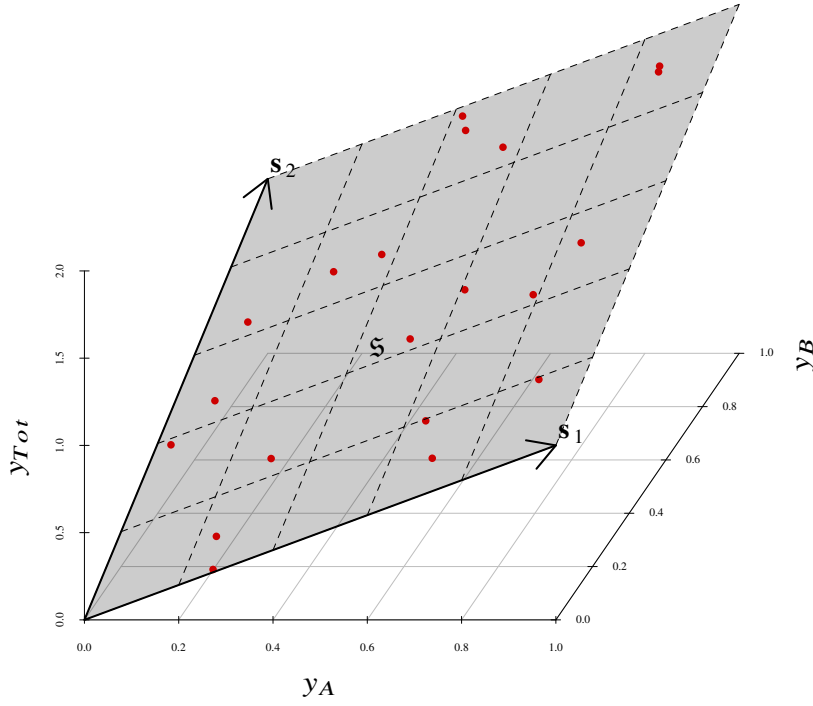


Fig. 3 Representation of a coherent subspace in a three dimensional hierarchy where $y_{\text{Tot}} = y_A + y_B$. The coherent subspace is depicted as a gray two dimensional plane labelled \mathfrak{s} . Note that the columns of $\tilde{\mathbf{s}}_1 = (1, 1, 0)'$ and $\tilde{\mathbf{s}}_2 = (1, 0, 1)'$ form a basis for \mathfrak{s} . The red points lying on \mathfrak{s} can be either realisations or coherent forecasts.

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

Traditionally, proportions have been calculated based on the observed historical data. Gross & 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 Section 3.2.

3.1.3 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. Since the middle-out approach involves generating top-down forecasts, it also introduces bias to the forecasts.

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 $\hat{\mathbf{y}}_{T+h|T}$. 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 $\hat{\mathbf{y}}_{T+h|T}$ onto the

coherent subspace \mathfrak{s} via a matrix \mathbf{SG} , resulting in a set of coherent forecasts $\tilde{\mathbf{y}}_{T+h|T}$. Specifically,

$$\tilde{\mathbf{y}}_{T+h|T} = \mathbf{SG}\hat{\mathbf{y}}_{T+h|T}, \quad (2)$$

where \mathbf{G} is an $m \times n$ matrix that maps $\hat{\mathbf{y}}_{T+h|T}$ to \mathbb{R}^m , producing new forecasts for the bottom level, which are in turn mapped to the coherent subspace by the summing matrix \mathbf{S} . We restrict our attention to projections on \mathfrak{s} in which case $\mathbf{SGS} = \mathbf{S}$. 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 (2) using appropriately designed \mathbf{G} matrices, however not all of these will be projections. For example for the bottom-up approach, $\mathbf{G} = (\mathbf{0}_{(m \times n-m)} \mathbf{I}_m)$ in which case $\mathbf{SGS} = \mathbf{S}$. For any top-down approach $\mathbf{G} = (\mathbf{p} \mathbf{0}_{(m \times n-1)})$, for which $\mathbf{SGS} \neq \mathbf{S}$.

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

$$\hat{\mathbf{e}}_{T+h|T} = \mathbf{y}_{T+h|T} - \hat{\mathbf{y}}_{T+h|T} \quad (3)$$

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

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

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

$$\mathbf{G} = (\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h^{-1} \quad (4)$$

is the unique solution that minimises the trace of $\mathbf{SGW}_h\mathbf{S}'\mathbf{G}'$ subject to $\mathbf{SGS} = \mathbf{S}$. 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 \mathbf{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 \mathbf{W}_h . However, estimating \mathbf{W}_h is challenging, especially for $h > 1$. Wickramasuriya et al. (2018) present possible alternative estimators for \mathbf{W}_h and show that these lead to different \mathbf{G} matrices. We summarise these below.

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

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

where $\boldsymbol{\beta}_{T+h|T} = E[\mathbf{b}_{T+h} | \mathbf{b}_1, \dots, \mathbf{b}_T]$ 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 \mathbf{V} . The OLS solution leads to the same projection matrix $\mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'$, 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 $\mathbf{W}_h = k_h \text{diag}(\hat{\mathbf{W}}_1)$ for all h ($k_h > 0$), where

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

is the unbiased sample estimator of the in-sample one-step-ahead base forecast errors as defined in (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. (2016).

An alternative WLS estimator is proposed by Athanasopoulos et al. (2017) in the context of temporal hierarchies. Here \mathbf{W}_h is proportional to $\text{diag}(\mathbf{S}\mathbf{1})$ 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 Figure 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 $\mathbf{W}_h = k_h \hat{\mathbf{W}}_1$ 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 $\mathbf{W}_h = k_h \hat{\mathbf{W}}_1^D$ for all h ($k_h > 0$), where $\hat{\mathbf{W}}_1^D = \lambda_D \text{diag}(\hat{\mathbf{W}}_1) + (1 - \lambda_D) \hat{\mathbf{W}}_1$ is a shrinkage estimator with diagonal target and shrinkage intensity parameter

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

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

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 & Katzfuss 2014). 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 & Amisano (2010), Billio et al. (2013), Carriero et al. (2015) and Clark & 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 et al. (2017) who define a coherent forecast in terms of a convolution. An equivalent definition due to Gamakumara et al. (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 2. Let \mathcal{A} be a subset¹ of \mathfrak{s} and let \mathcal{B} be all points in \mathbb{R}^n that are mapped onto \mathcal{A} after premultiplication by \mathbf{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} .

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.

¹ Strictly speaking \mathcal{A} is a Borel set

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 $\mathcal{N}(\hat{\mathbf{y}}_{T+h|T}, \hat{\mathbf{\Sigma}}_{T+h|T})$, then the reconciled probabilistic forecasts will be $\mathcal{N}(\tilde{\mathbf{y}}_{T+h|T}, \tilde{\mathbf{\Sigma}}_{T+h|T})$, where

$$\tilde{\mathbf{y}}_{T+h|T} = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_{T+h|T} \quad (5)$$

$$\text{and} \quad \tilde{\mathbf{\Sigma}}_{T+h|T} = \mathbf{S}\mathbf{G}\hat{\mathbf{\Sigma}}_{T+h|T}\mathbf{G}'\mathbf{S}'. \quad (6)$$

There are several options for obtaining the base probabilistic forecasts and in particular the variance covariance matrix $\hat{\mathbf{\Sigma}}$. One option is to fit multivariate models either level by level or for the hierarchy as a whole leading respectively to a $\hat{\mathbf{\Sigma}}$ that is block diagonal or dense. Another option is to fit univariate models for each individual series in which case $\hat{\mathbf{\Sigma}}$ is a diagonal matrix. A third option that we employ here is to obtain $\hat{\mathbf{\Sigma}}$ using in-sample forecast errors, in a similar vein to how $\hat{\mathbf{W}}_1$ is estimated in the MinT method. Here the same shrinkage estimator described in Section 3.2 is used. The reconciled probabilistic forecast will ultimately depend on the choice of \mathbf{G} ; the same choices of \mathbf{G} matrices used in Section 3 can be used.

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 et al. 2008), and bootstrap based approaches (Manzan & Zerom 2008, Vilar & Vilar 2013). In macroeconomics, Cogley et al. (2005) discuss the importance of allowing for skewness in density forecasts and more recently Smith & 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 (Please refer to the attached supplementary materials for a detailed explanation of this approach). 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 $t = 1, \dots, T$. Let these models denote M_1, \dots, M_n .

2. Compute one-step-ahead in-sample forecast errors. Collect these into an $n \times T$ matrix $\hat{\mathbf{E}} = (\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \dots, \hat{\mathbf{e}}_T)$, where the n -vector $\hat{\mathbf{e}}_t = \mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}$. Here, $\hat{\mathbf{y}}_{t|t-1}$ 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 $\hat{\mathbf{E}}$; that is choose H consecutive columns of $\hat{\mathbf{E}}$ at random, repeating this process B times. Denote the $n \times H$ matrix obtained at iteration b as $\hat{\mathbf{E}}^b$ for $b = 1, \dots, B$.
4. For all b , compute $\hat{\mathbf{Y}}^b = \{\hat{\mathbf{y}}_1^b, \dots, \hat{\mathbf{y}}_n^b\}' \in \mathbb{R}^{n \times H} : \hat{y}_{i,h}^b = f(M_i, \hat{e}_{i,h}^b)$ where, $f(\cdot)$ is a function of fitted univariate model in step 1 and associated error. i.e., $\hat{y}_{i,h}$ is a sample path simulated from fitted model M_i for i^{th} series and error approximated by the corresponding block bootstrapped sample error $\hat{e}_{i,h}^b$ which is the $(i, h)^{\text{th}}$ element of $\hat{\mathbf{E}}^b$. Each row of $\hat{\mathbf{Y}}^b$ is a sample path of h forecasts for a single series. Each column of $\hat{\mathbf{Y}}^b$ is a realisation from the joint predictive distribution at a particular horizon.
5. For each $b = 1, \dots, B$, select the h^{th} column of $\hat{\mathbf{Y}}^b$ and stack these to form an $n \times B$ matrix $\hat{\mathbf{Y}}_{T+h|T}$.
6. For a given \mathbf{G} matrix and for each $h = 1, \dots, H$, compute $\tilde{\mathbf{Y}}_{T+h|T} = \mathbf{S}\mathbf{G}\hat{\mathbf{Y}}_{T+h|T}$. Each column of $\tilde{\mathbf{Y}}_{T+h|T}$ is a realisation from the joint h -step-ahead reconciled predictive distribution.

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 are 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).

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):

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

Figure 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.

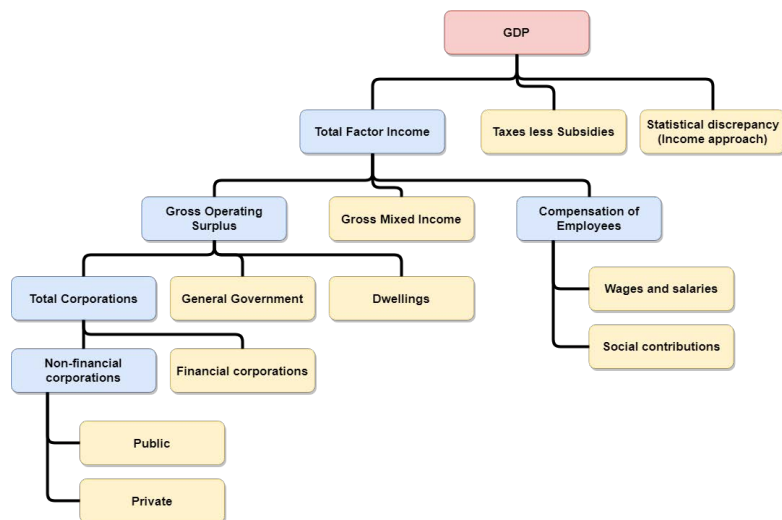


Fig. 4 Hierarchical structure of the income approach for GDP. The pink cell contains GDP the most aggregate series. The blue cells contain intermediate-level series and the yellow cells correspond to the most disaggregate bottom-level series.

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

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

Figures 5, 6 and 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 1, 2, 3 and 4 in the Appendix.

Figure 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.

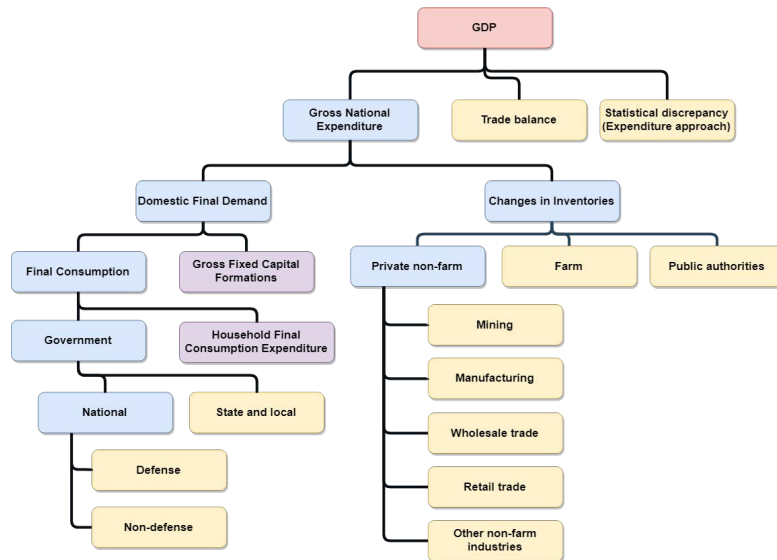


Fig. 5 Hierarchical structure of the expenditure approach for GDP. The pink cell contains GDP, the most aggregate series. The blue and purple cells contain intermediate-level series with the series in the purple cells further disaggregated in Figures 6 and 7. The yellow cells contain the most disaggregate bottom-level series.

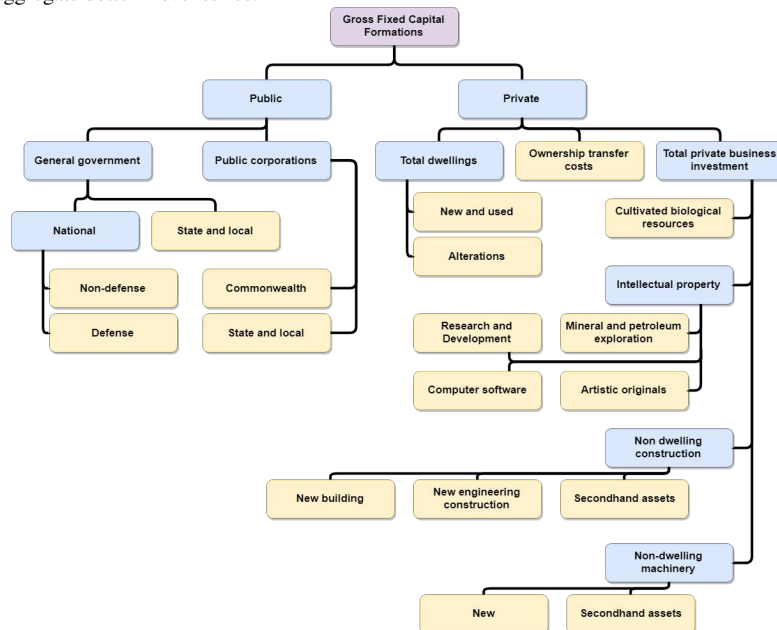


Fig. 6 Hierarchical structure for Gross Fixed Capital Formations under the expenditure approach for GDP, continued from Figure 5. Blue cells contain intermediate-level series and the yellow cells correspond to the most disaggregate bottom-level series.

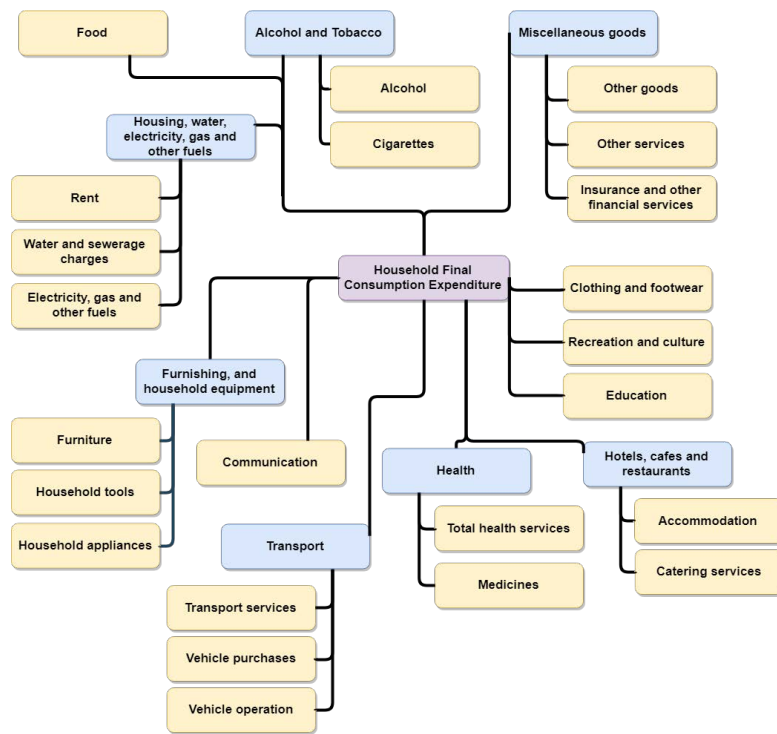


Fig. 7 Hierarchical structure for Household Final Consumption Expenditure under the expenditure approach for GDP, continued from Figure 5. Blue cells contain intermediate-level series and the yellow cells correspond to the most disaggregate bottom-level series.

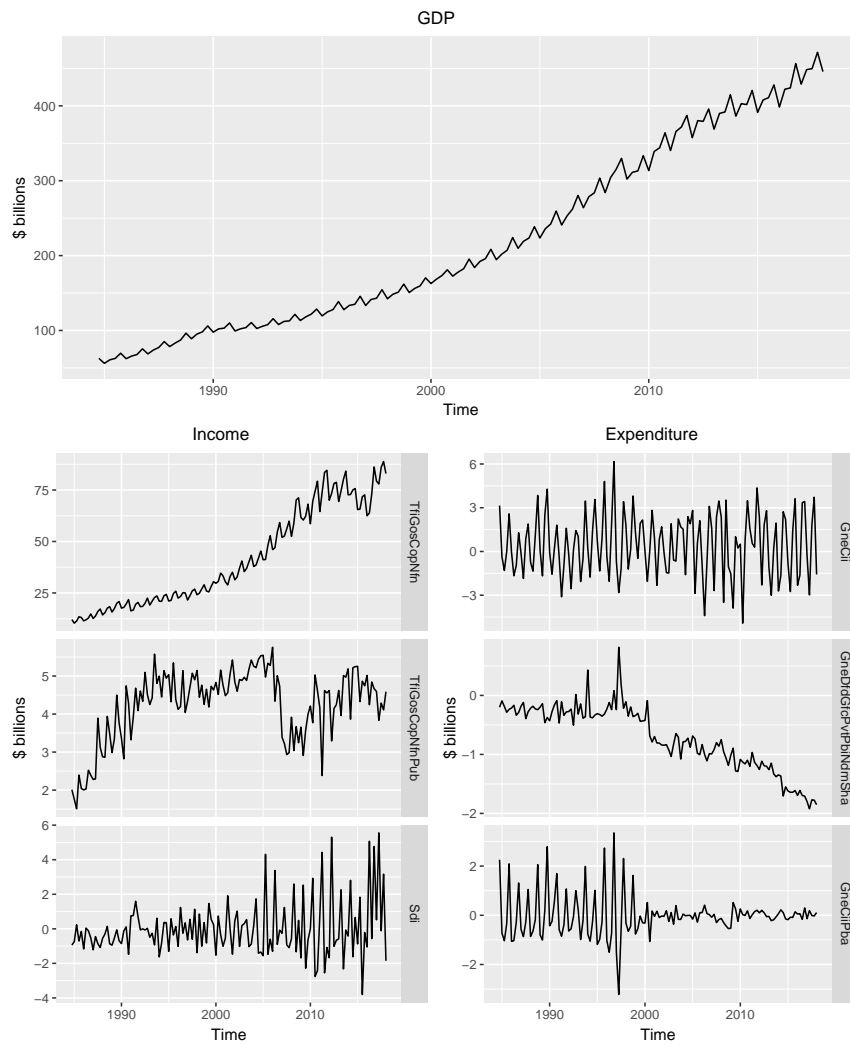


Fig. 8 Time plots for series from different levels of income and expenditure hierarchies.

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.

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 for $n = 80$ separate time series must be obtained. Given the diversity in these time series discussed in Section 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 & Khandakar (2008) and implemented in the `auto.arima()` function in Hyndman et al. (2019). A similar approach was also undertaken using the ETS framework to produce base forecasts (Hyndman et al. 2008). This 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 as vector autoregressions, and models and methods that handle a large number of predictors such as factor models or least angle regression. However, Panagiotelis et al. (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

² The relevant github repository is [PuwasalaG/Hierarchical-Book-Chapter](#)

previous year. The naïve forecasts are by construction coherent and therefore do not need to be reconciled.

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

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

where \check{e}_{t+h} is the difference between any forecast and the realisation³, 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 & Athanasopoulos (2018).

Forecast accuracy of probabilistic forecasts can be evaluated using scoring rules (Gneiting & Katzfuss 2014). Let \check{F} be a probabilistic forecast and let $\check{y} \sim \check{F}$ 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

$$eS(\check{F}_{T+h|T}, \mathbf{y}_{T+h}) = E_{\check{F}} \|\check{y}_{T+h} - \mathbf{y}_{T+h}\|^\alpha - \frac{1}{2} E_{\check{F}} \|\check{y}_{T+h} - \check{y}_{T+h}^*\|^\alpha,$$

where \mathbf{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 \check{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

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

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

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.

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)

7 Results

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 income hierarchy, some series require seasonal differencing while other did not. Furthermore the AR orders vary from 0–3, the MA orders from 0–2, and their seasonal counterparts SAR from 0–2 and SMA from 0–1. Figure 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.

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.

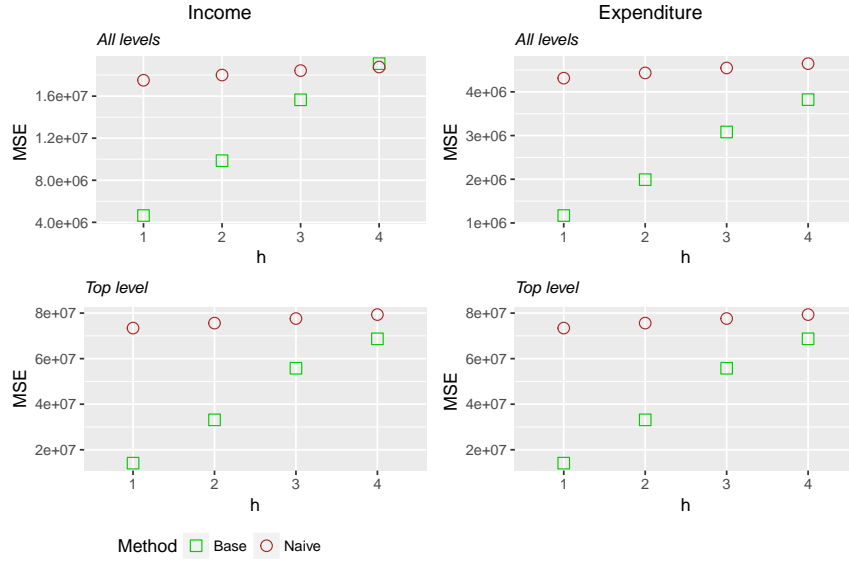


Fig. 9 Mean squared errors for naïve and ARIMA base forecasts. Top panels refer to results summarised over all series while bottom panels refer to results for the top-level GDP series. Left panels refer to the income hierarchy and right panels to the expenditure hierarchy.

Figures 10 and 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 Figures 10 and 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)

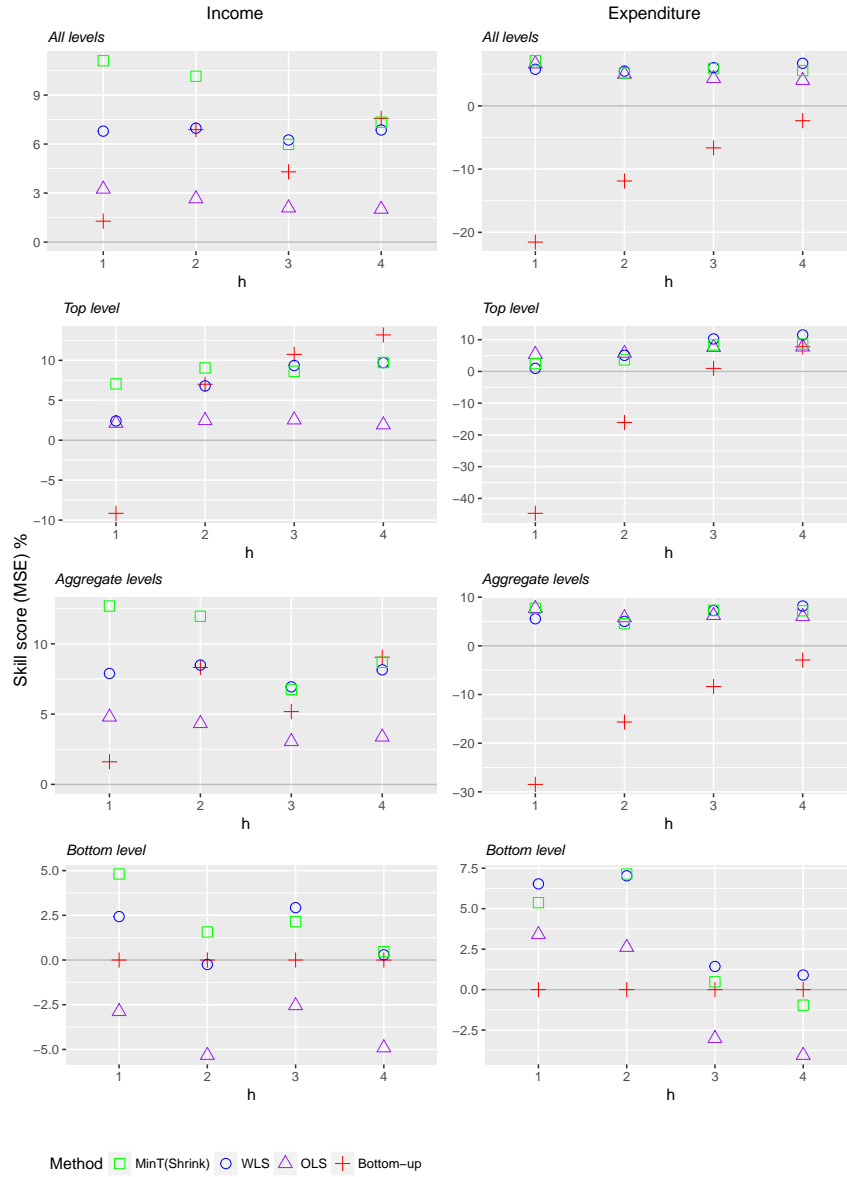


Fig. 10 Skill scores for point forecasts from alternative methods (with reference to base forecasts) using MSE. The left panels refer to the income hierarchy while the right panels refer to the expenditure hierarchy. The first row refers to results summarised over all series, the second row to top-level GDP series, the third row to aggregate levels, and the last row to the bottom level.

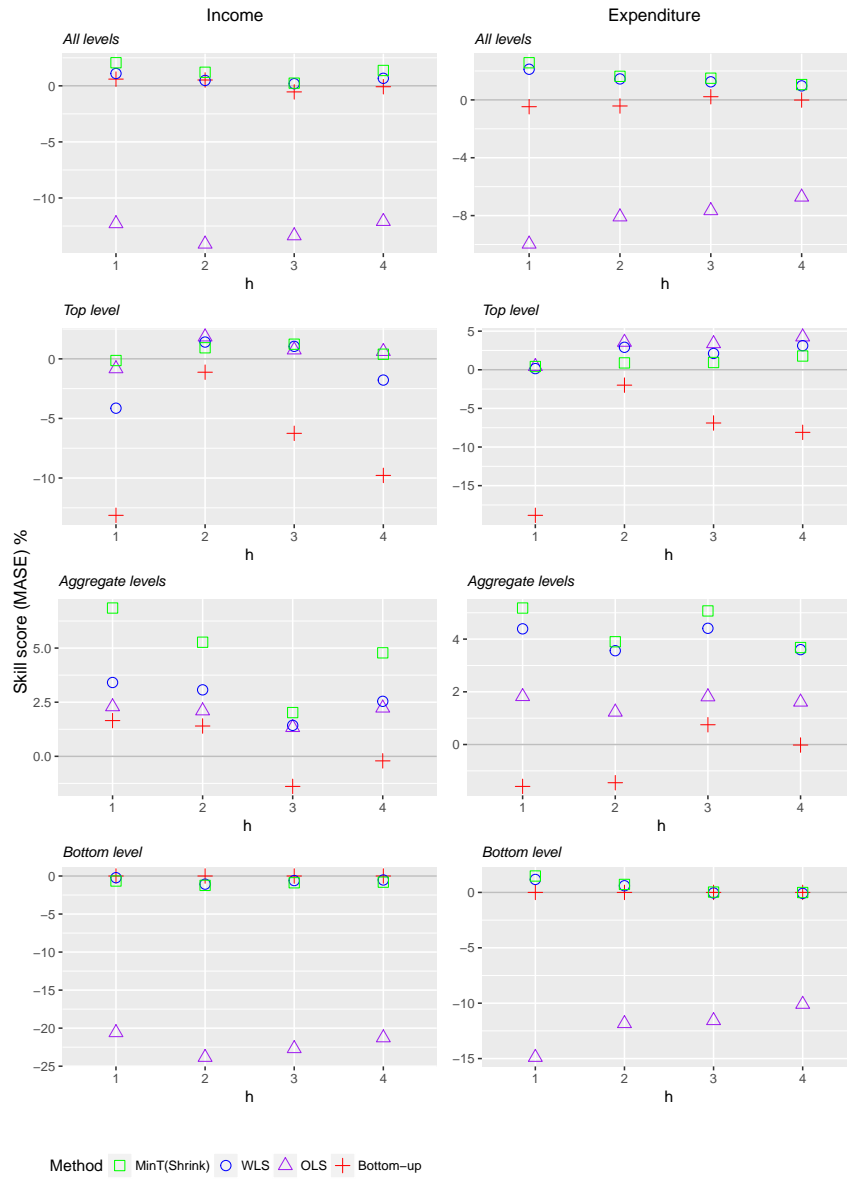


Fig. 11 Skill scores for point forecasts from different reconciliation methods (with reference to base forecasts) using MASE. The left two panels refer to the income hierarchy and the right two panels refer to the expenditure hierarchy. The first row refers to results summarised over all series, the second row to top-level GDP series, the third row to aggregate levels, and the last row to the bottom level.

7.3 Probabilistic Forecast Reconciliation

We now turn our attention towards results for probabilistic forecasts. Figure 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 as skill scores relative to base forecasts. The top panels refer to results assuming Gaussian probabilistic forecasts as described in Section 4.1 while the bottom panels refer to the non-parametric bootstrap method described in Section 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 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, Figure 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.

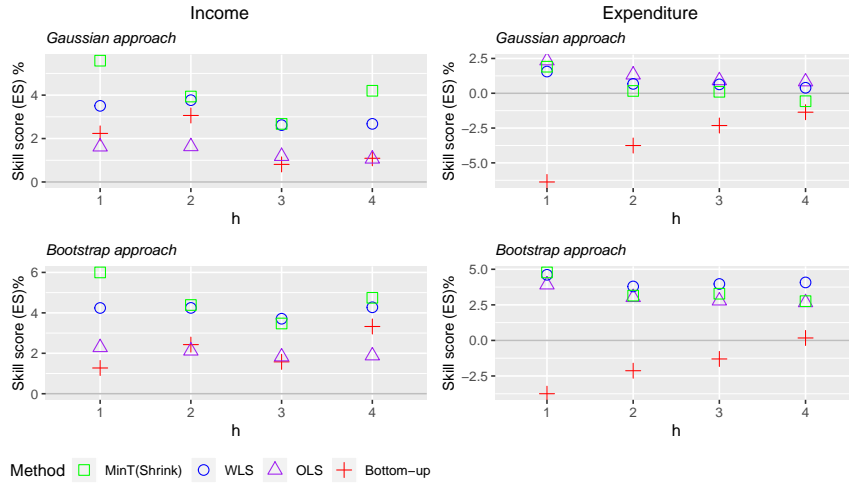


Fig. 12 Skill scores for multivariate probabilistic forecasts from different reconciliation methods (with reference to base forecasts) using energy scores. The top panels refer to the results for the Gaussian approach and the bottom panels to the non-parametric bootstrap approach. Left panels refer to the income hierarchy and right panels to the expenditure hierarchy.

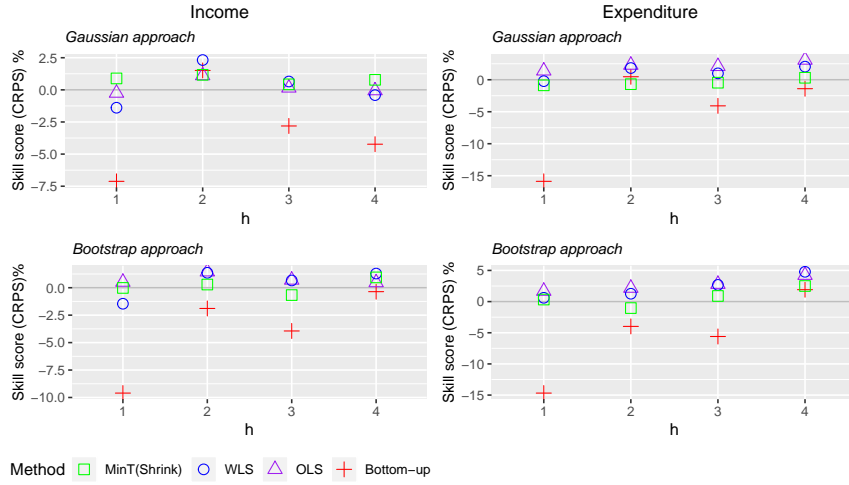


Fig. 13 Skill scores for probabilistic forecasts of top-level GDP from different reconciliation methods (with reference to base forecasts) using CRPS. Top panels refer to the results for Gaussian approach and bottom panels refer to the non-parametric bootstrap approach. The left panel refers to the income hierarchy and the right panel to the expenditure hierarchy.

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.

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.
- 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.
- Athanasopoulos, G., Ahmed, R. A. & Hyndman, R. J. (2009), 'Hierarchical forecasts for Australian domestic tourism', *International Journal of Forecasting* **25**(1), 146–166.
- 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.
- Australian Bureau of Statistics (2015), 'Australian system of national accounts: Concepts, sources and methods'. Cat 5216.0.
- 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.
- 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.
- Gamakumara, P. (2019), Probabilistic Forecasts in Hierarchical Time Series, PhD thesis, Monash University.
- Gamakumara, P., Panagiotelis, A., Athanasopoulos, G. & Hyndman, R. J. (2018), Probabilistic forecasts in hierarchical time series, Working paper 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.
- Gneiting, T. & Katzfuss, M. (2014), 'Probabilistic forecasting', *Annual Review of Statistics and Its Application* **1**, 125–151.
- Gneiting, T., Stanberry, L. I., Grimit, E. P., Held, L. & Johnson, N. A. (2008), 'Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds', *Test* **17**(2), 211–235.
- Gross, C. W. & Sohl, J. E. (1990), 'Disaggregation methods to expedite product line forecasting', *Journal of Forecasting* **9**(3), 233–254.
- 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.
- 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.
- Hyndman, R. J. & Athanasopoulos, G. (2018), *Forecasting: Principles and Practice*, OTexts.
URL: <https://OTexts.com/fpp2>
- Hyndman, R. J., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., Yasmeeen, F., R Core Team, Ihaka, R., Reid, D., Shaub, D., Tang, Y. & Zhou, Z. (2019), *forecast: Forecasting Functions for Time Series and Linear Models*. Version 8.5.
URL: <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*, Springer-Verlag, Berlin.
- 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.
- 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.
- Manzan, S. & Zerom, D. (2008), 'A bootstrap-based non-parametric forecast density', *International Journal of Forecasting* **24**(3), 535–550.
- 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* (forthcoming).
- 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.
- 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.
- 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', pp. 1–30.
- Tiao, G. C. (1972), 'Asymptotic behaviour of temporal aggregates of time series', *Biometrika* **59**(3), 525–531.
- Vilar, J. A. & Vilar, J. A. (2013), 'Time series clustering based on nonparametric multidimensional forecast densities', *Electronic Journal of Statistics* **7**(1), 1019–1046.
- 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.
- Weiss, C. (2018), *Essays in Hierarchical Time Series Forecasting and Forecast Combination*, PhD thesis, 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* **145**(9), 1–45.
- Yagli, G. M., Yang, D. & Srinivasan, D. (2019), 'Reconciling solar forecasts: Sequential reconciliation', *Solar Energy* **179**, 391–397.
- Yang, D., Quan, H., Disfani, V. R. & Liu, L. (2017), 'Reconciling solar forecasts: Geographical hierarchy', *Solar Energy* **146**, 276–286.
- 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).

Appendix

Table 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 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.
GneDfdFceGvtSnI	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
GneDfdGfcPubPcpSnI	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
GneDfdGfcPubGvtSnI	A2302554X	Gen. gov. - State and local; GFCF
GneDfdGfcPubGvt	A2302555A	Gen. gov.; GFCF
GneDfdGfcPub	A2302556C	Plc.; GFCF
GneDfdGfc	A2302557F	All sectors; GFCF

Table 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 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
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

Supplementary materials

4 A non-parametric bootstrap approach for probabilistic forecast reconciliation

The second half of the fourth chapter of my thesis introduces a novel non-parametric approach to reconcile hierarchical probabilistic forecasts. In this section, I provide a detailed explanation of this methodology followed by a comprehensive simulation study.

The proposed method initially involves obtaining probabilistic forecasts without considering the aggregation constraints. That is we first get the incoherent probabilistic forecasts. Then we reconcile these to obtain the coherent probabilistic forecasts of the hierarchy.

4.1 Incoherent probabilistic forecasts

We now explain the process of obtaining incoherent sample paths. First we fit appropriate univariate models for each series in the hierarchy based on the training data $\mathbf{y}_{1:T}$. Then we compute 1-step-ahead training errors as $e_{i,t} = y_{i,t} - \hat{y}_{i,t}$ for $i = 1, \dots, n$ and $t = 1, \dots, T$ where $\hat{y}_{i,t} = E(y_{i,t} | y_{i,1:t-1})$. These training errors will be then stored in a matrix $\mathbf{I}_{(T \times n)} = (\mathbf{e}_1, \dots, \mathbf{e}_T)'$ where $\mathbf{e}_t = (e_{1,t}, \dots, e_{n,t})$ stored in the same order as $\hat{\mathbf{y}}_t$ for $t = 1, \dots, T$. Next we block bootstrap a sample of size H from \mathbf{I} . That is, we randomly select consecutive H rows from \mathbf{I} and store in a matrix $\mathbf{I}_{(H \times n)}^b = (\mathbf{e}_1^b, \dots, \mathbf{e}_H^b)'$ and repeat this for $b = 1, \dots, B$.

Finally, we generate the h -step-ahead future paths using the fitted univariate models conditioning on the past observations. We incorporate bootstrapped training errors as the error series for generating future paths. This will implicitly model the contemporaneous correlation structure of the hierarchy. Further, the consecutive (block) training errors will model the serial correlation of each series. To explain this process more explicitly consider the following example.

Example 1 Suppose we have fitted an $ARMA(p, q)$ model for the i^{th} series of the hierarchy by using data $y_{i,1} : y_{i,T}$. i.e.

$$\begin{aligned} y_{i,t} &= \alpha_1 y_{i,t-1} + \alpha_2 y_{i,t-2} + \dots + \alpha_p y_{i,t-p} + \beta_1 \epsilon_{i,t-1} + \beta_2 \epsilon_{i,t-2} + \dots + \beta_q \epsilon_{i,t-q} + \epsilon_{i,t}, \\ y_{i,t} &= (\alpha_1 + \alpha_2 L + \dots + \alpha_p L^{p-1}) y_{i,t-1} + (\beta_1 + \beta_2 L + \dots + \beta_q L^{q-1}) \epsilon_{i,t-1} + \epsilon_{i,t}, \end{aligned}$$

where L is the well known lag or backshift operator. Then the h -step-ahead b^{th} future path for i^{th} series can be produced as,

$$\hat{y}_{i,T+h}^b = (\hat{\alpha}_1 + \hat{\alpha}_2 L + \dots + \hat{\alpha}_p L^{p-1}) y_{i,T+h-1} + (\hat{\beta}_1 + \hat{\beta}_2 L + \dots + \hat{\beta}_q L^{q-1}) \epsilon_{i,T+h-1} + e_{i,h}^b.$$

where, $e_{i,h}^b$ is the $(h \times i)^{\text{th}}$ element from \mathbf{I}^b ,

$$y_{i,T+h-1} = \begin{cases} y_{i,1} : y_{i,T} & \text{for } T+h-1 \leq T \\ \hat{y}_{i,T+1}^b : \hat{y}_{i,T+h-1}^b & \text{for } T+h-1 > T \end{cases}$$

and

$$\epsilon_{i,T+h-1} = \begin{cases} \epsilon_{i,1} : \epsilon_{i,T} & \text{for } T+h-1 \leq T \\ e_{i,1}^b : e_{i,h-1}^b & \text{for } T+h-1 > T \end{cases}.$$

Once we get the h -step-ahead sample path for all n series in the hierarchy, we stack them in the same order as \mathbf{y}_t . Repeating the same process for $b = 1, \dots, B$ we will get a set of bootstrapped

h -step-ahead future paths of size B . We denote this as $\hat{\mathbf{Y}}_{T+h} = (\hat{\mathbf{y}}_{T+h}^1, \dots, \hat{\mathbf{y}}_{T+h}^B)'$ where each row of $\hat{\mathbf{Y}}_{T+h}$ represent the h -step-ahead sample paths for a single series and columns represent that for corresponding series of the hierarchy.

We note that $\hat{\mathbf{Y}}_{T+h}$ is an empirical sample from the incoherent probability distribution of the hierarchy. However, it is very unlikely that $\hat{\mathbf{Y}}_{T+h}$ lies in the coherent subspace. Thus it requires the reconciliation which will be discussed in the next subsection.

4.2 Reconciliation of incoherent future paths

To reconcile the incoherent sample paths, we follow the definition of reconciliation. We project each sample path in $\hat{\mathbf{Y}}_{T+h}$ to the coherent subspace via the projection $\mathbf{S}\mathbf{G}$. i.e. for any \mathbf{G} we can write,

$$\tilde{\mathbf{y}}_{T+h}^b = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_{T+h}^b, \quad (1)$$

consequently we have,

$$\tilde{\mathbf{Y}}_{T+h} = \mathbf{S}\mathbf{G}\hat{\mathbf{Y}}_{T+h}', \quad (2)$$

where, each row in $\tilde{\mathbf{Y}}_{T+h}'$ represent a single reconciled sample path. Further $\tilde{\mathbf{Y}}_{T+h}'$ form an empirical sample from the reconciled forecast distribution of the hierarchy. Any \mathbf{G} matrix introduced in point forecast reconciliation can also be used for this sample path reconciliation. However, in the following subsection we discuss a method to find \mathbf{G} that is optimal for probabilistic forecasts.

4.3 Optimal reconciliation of incoherent future paths

Let us now propose to find an optimal \mathbf{G} for reconciling future paths by minimising a proper multivariate scoring rule. The respective objective function can be written as,

$$\underset{\mathbf{G}_h}{\operatorname{argmin}} \quad \mathbb{E}_Q[S(\mathbf{S}\mathbf{G}_h\hat{\mathbf{Y}}_{T+h}', \mathbf{y}_{T+h})], \quad (3)$$

where S is a proper scoring rule that follows

$$\mathbb{E}_Q[S(Q, \omega)] \leq \mathbb{E}_Q[S(P, \omega)], \quad (4)$$

where $\mathbb{E}_Q[S(P, \omega)]$ is the expected score under the true distribution Q (Gneiting and Katzfuss, 2014; Gneiting et al., 2008).

We use the subscript h on \mathbf{G} in (3) to emphasis distinct \mathbf{G} matrices for different forecast horizons. The energy score given in Table (2) is a proper scoring rule. Approximating the expectations in this equation through sample means we can write,

$$\mathbb{E}S(\mathbf{S}\mathbf{G}_h\hat{\mathbf{Y}}_{T+h}', \mathbf{y}_{T+h}) \approx \frac{1}{B} \sum_{b=1}^B \|\mathbf{S}\mathbf{G}_h\hat{\mathbf{y}}_{T+h,j}^b - \mathbf{y}_{T+h}\| - \frac{1}{2(B-1)} \sum_{b=1}^{B-1} \|\mathbf{S}\mathbf{G}_h(\hat{\mathbf{y}}_{T+h,j}^b - \hat{\mathbf{y}}_{T+h,j}^{b+1})\|. \quad (5)$$

where B is the empirical sample size from the coherent forecast distribution. Now we can rewrite the objective function in (3) as,

$$\underset{\mathbf{G}}{\operatorname{argmin}} \frac{1}{N} \sum_{j=1}^N \left\{ \frac{1}{B} \sum_{b=1}^B \|\mathbf{S}\mathbf{G}_h\hat{\mathbf{y}}_{T+h,j}^b - \mathbf{y}_{T+h,j}\| - \frac{1}{2(B-1)} \sum_{b=1}^{B-1} \|\mathbf{S}\mathbf{G}_h(\hat{\mathbf{y}}_{T+h,j}^b - \hat{\mathbf{y}}_{T+h,j}^{b+1})\| \right\} \quad (6)$$

where, the expectation \mathbb{E}_Q is approximated through the sample mean over $\{\mathbb{E}S(\mathbf{S}\mathbf{G}_h\hat{\mathbf{Y}}_{T+h,1}', \mathbf{y}_{T+h,1}), \dots, \mathbb{E}S(\mathbf{S}\mathbf{G}_h\hat{\mathbf{Y}}_{T+h,N}', \mathbf{y}_{T+h,N})\}$. Since this is hard to solve analytically, we use a numerical optimization methods to estimate the matrix \mathbf{G}_h that minimizes above objective function and thus obtain the optimally reconciled future paths.

4.3.1 Reparameterisation of \mathbf{G}

We consider different parameterisations when estimating the optimal \mathbf{G} via above optimisation process. Let,

$$\mathbf{G}_h = (\mathbf{S}'\mathbf{W}_h\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h. \quad (7)$$

This structure for \mathbf{G}_h will ensure $\mathbf{S}\mathbf{G}_h$ is a projection matrix and it projects each sample path onto \mathfrak{s} .

Method 1 Minimising the objective function in (6) over \mathbf{W}_h . This solves an unconstrained optimisation problem

Method 2 Consider the cholesky decomposition of \mathbf{W}_h . i.e. let $\mathbf{W}_h = \mathbf{R}_h'\mathbf{R}_h$ where \mathbf{R}_h is an upper triangular matrix. Thus minimising (6) over \mathbf{R}_h

Method 3 Similar to Method 2, minimising (6) over cholesky decomposition of \mathbf{W}_h but restricted for scaling. i.e., $\mathbf{W}_h = \mathbf{R}_h'\mathbf{R}$ s.t $\mathbf{i}'\mathbf{W}_h\mathbf{i} = 1$ where $\mathbf{i} = (1, 0, \dots, 0)'$

Method 4 Minimising (6) over \mathbf{G}_h such that $\mathbf{G}_h\mathbf{S} = \mathbf{I}$. This constraint will ensure that $\mathbf{S}\mathbf{G}_h$ is a projection onto \mathfrak{s}

4.4 Simulation study

We now turn our attention to comparing different reconciliation methods with optimal reconciliation in a simulation setting.

4.4.1 Data generating process (DGP)

For the data generating process, we consider the hierarchy is given in Figure 1, comprising two aggregation levels with four bottom-level series. Each bottom-level series will be generated first, and then summed to obtain the data for the upper-level series.

Initially, the error series were generated for four bottom-level series. To impose dependency among these series, errors were generated from the Gumbel copula model with beta margins. A two dimensional Gumbel copula is given by,

$$C_\theta(u_1, u_2) = \exp\{ -[(-\ln(u_1))^\theta + (-\ln(u_2))^\theta]^{1/\theta} \}.$$

We generate random variates $\{u_{AA}, u_{AB}\}$ from $C_{\theta=10}(\cdot)$ and $\{u_{BA}, u_{BB}\}$ from $C_{\theta=8}(\cdot)$ for series $\{AA, AB\}$ and $\{BA, BB\}$ respectively. Next we generate the quantiles from beta distributions with

Table 2: Scoring rules to evaluate multivariate forecast densities. $\check{\mathbf{Y}}_{T+h}$ and $\check{\mathbf{Y}}_{T+h}^*$ be two independent random vectors from the coherent forecast distribution $\check{\mathbf{F}}$ with the density function $\check{f}(\cdot)$ at time $T+h$ and \mathbf{y}_{T+h} is the vector of realizations. Further $\check{Y}_{T+h,i}$ and $\check{Y}_{T+h,j}$ are i th and j th components of the vector $\check{\mathbf{Y}}_{T+h}$. Further the variogram score is given for order p where, w_{ij} are non-negative weights.

Scoring rule	Expression	Reference
Log score	$\text{LS}(\check{\mathbf{F}}, \mathbf{y}_{T+h}) = -\log \check{f}(\mathbf{y}_{T+h})$	Gneiting and Raftery (2007)
Energy score	$\text{eS}(\check{\mathbf{Y}}_{T+h}, \mathbf{y}_{T+h}) = \mathbb{E}_{\check{\mathbf{F}}} \ \check{\mathbf{Y}}_{T+h} - \mathbf{y}_{T+h}\ ^\alpha - \frac{1}{2} \mathbb{E}_{\check{\mathbf{F}}} \ \check{\mathbf{Y}}_{T+h} - \check{\mathbf{Y}}_{T+h}^*\ ^\alpha, \quad \alpha \in (0, 2]$	Gneiting et al. (2008)
Variogram score	$\text{VS}(\check{\mathbf{F}}, \mathbf{y}_{T+h}) = \sum_{i=1}^n \sum_{j=1}^n w_{ij} \left(y_{T+h,i} - y_{T+h,j} ^p - \mathbb{E}_{\check{\mathbf{F}}} \check{Y}_{T+h,i} - \check{Y}_{T+h,j} ^p \right)^2$	Scheuerer and Hamill (2015)

shape parameters $\alpha = 1$ and $\beta = 3$ correspond to $\{u_{AA}, u_{AB}, u_{BA}, u_{BB}\}$. Denote these quantiles as $\{\varepsilon_{AA}, \varepsilon_{AB}, \varepsilon_{BA}, \varepsilon_{BB}\}$.

Then $\{w_{AA,t}, w_{AB,t}, w_{BA,t}, w_{BB,t}\}$ were generated from ARIMA(p, d, q) processes, where (p, q) and d take integers from $\{1, 2\}$ and $\{0, 1\}$ respectively with equal probabilities. $\{\varepsilon_{AA}, \varepsilon_{AB}, \varepsilon_{BA}, \varepsilon_{BB}\}$ were incorporated as the errors driving these ARIMA processes. The parameters for the AR and MA components are randomly and uniformly generated from $[0.3, 0.5]$ and $[0.3, 0.7]$ respectively.

In practice, hierarchical time series likely to have much noisier series at lower levels of aggregation. Following the method proposed by Wickramasuriya, Athanasopoulos, and Hyndman (2018), we replicate this feature in our simulations by generating the bottom-level series $\{y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}\}$ as follows:

$$\begin{aligned} y_{AA,t} &= w_{AA,t} + u_t - 0.5v_t, \\ y_{AB,t} &= w_{AB,t} - u_t - 0.5v_t, \\ y_{BA,t} &= w_{BA,t} + u_t + 0.5v_t, \\ y_{BB,t} &= w_{BB,t} - u_t + 0.5v_t, \end{aligned}$$

where $u_t \sim \mathcal{N}(0, \sigma_u^2)$ and $v_t \sim \mathcal{N}(0, \sigma_v^2)$. The aggregate series in the middle-level are given by:

$$\begin{aligned} y_{A,t} &= w_{AA,t} + w_{AB,t} - v_t, \\ y_{B,t} &= w_{BA,t} + w_{BB,t} + v_t, \end{aligned}$$

and the total series is given by

$$y_{Tot,t} = w_{AA,t} + w_{AB,t} + w_{BA,t} + w_{BB,t}.$$

To ensure the disaggregate series are noisier than the aggregate series, we choose σ_u^2 and σ_v^2 such that

$$\text{Var}(\varepsilon_{AA,t} + \varepsilon_{AB,t} + \varepsilon_{BA,t} + \varepsilon_{BB,t}) \leq \text{Var}(\varepsilon_{AA,t} + \varepsilon_{AB,t} - v_t) \leq \text{Var}(\varepsilon_{AA,t} + u_t - 0.5v_t).$$

Similar inequalities hold when $\varepsilon_{AA,t}$ is replaced by $\varepsilon_{AB,t}$, $\varepsilon_{BA,t}$ and $\varepsilon_{BB,t}$ in the third term. The variance covariance matrix of $\{\varepsilon_{AA}, \varepsilon_{AB}, \varepsilon_{BA}, \varepsilon_{BB}\}$ is,

$$\Sigma = \begin{pmatrix} 0.0388 & 0.0385 & 0.0010 & 0.0010 \\ 0.0385 & 0.0390 & 0.0008 & 0.0008 \\ 0.0010 & 0.0008 & 0.0387 & 0.0377 \\ 0.0010 & 0.0008 & 0.0377 & 0.0381 \end{pmatrix}.$$

Thus we choose $\sigma_u^2 = 10$ and $\sigma_v^2 = 7$ to satisfy the above constraints.

4.4.2 Simulation set up for optimal reconciliation

1. Generate time series with 2500 data points in each, corresponds to the hierarchy in Figure 1 based on the DGP explained.
2. Consider a rolling window of 600 observations. We refer to it the outer rolling window.
 - i. Inside this outer rolling window consider an inner rolling window of $T = 500$ observations.
 - ii. Using inner rolling window, fit univariate ARIMA models to each series in the hierarchy.
 - iii. Based on these fitted models, we generate $B = 1000$ of $h = 3$ step-ahead incoherent future paths incorporating bootstrap errors as described in Sub section 4.1. Thus we get $\{\hat{\mathbf{Y}}_{T+1,j=1}, \hat{\mathbf{Y}}_{T+2,j=1}, \hat{\mathbf{Y}}_{T+3,j=1}\}$.

Table 3: Energy scores (ES) and variogram scores (VS) for probabilistic forecasts from different reconciliation methods. Bottom row represent the scores for base forecasts which are not coherent. The smaller the scores, the better the forecasts are.

Reconciliation	h=1		h=2		h=3	
method	ES	VS	ES	VS	ES	VS
Optimal(Method-1)	5.36	1.21	5.51	1.27	5.83	1.38
Optimal(Method-2)	5.37	1.21	5.53	1.27	5.83	1.37
Optimal(Method-3)	5.37	1.21	5.53	1.27	5.83	1.37
Optimal(Method-4)	5.38	1.21	5.54	1.27	5.83	1.38
MinT(Shrink)	5.33	1.19	5.50	1.26	5.77	1.34
WLS	5.43	1.23	5.60	1.30	5.89	1.40
OLS	5.51	1.23	5.70	1.30	5.98	1.40
<i>Base</i>	<i>5.71</i>	<i>1.28</i>	<i>5.94</i>	<i>1.37</i>	<i>6.27</i>	<i>1.49</i>

- iv. Repeat step (iii) for $j = 1, \dots, N$ where $N = 100$, by rolling the inner window by one step ahead at a time.
 - v. Collect $\{\hat{\mathbf{Y}}_{T+h,j=1}, \dots, \hat{\mathbf{Y}}_{T+h,j=100}\}$ for $h = 1, \dots, 3$ into a separate arrays of matrices.
 - vi. For each forecast horizon h , estimate the optimal \mathbf{G}_h that will reconcile $\{\hat{\mathbf{Y}}_{T+h,j=1}, \dots, \hat{\mathbf{Y}}_{T+h,j=100}\}$ by minimising the average energy score as explained in Subsection 4.3. Also use different reparameterisation of \mathbf{G}_h as explained in Subsection 4.3.1. Denote this as \mathbf{G}_h^{Opt} .
 - vii. Roll the inner rolling window another one step ahead and repeat step (ii) and (iii). Let these future paths denoted by $\hat{\mathbf{Y}}_{T+h}$ for $h = 1, 2, 3$.
 - viii. Compute $\tilde{\mathbf{Y}}_{T+h} = \mathbf{S}\mathbf{G}_h\hat{\mathbf{Y}}'_{T+h}$ for $h = 1, 2, 3$ using \mathbf{G}_h^{Opt} as well as using other \mathbf{G} matrices used in point forecast reconciliation.
3. Repeat Step 2 for 1000 times by rolling the outer rolling window one step ahead at each time. Collect 1000 of these reconciled future paths, $\tilde{\mathbf{Y}}_{T+h}$, from different reconciliation methods for $h = 1, 2, 3$ and evaluate the forecasting performances.

4.4.3 Results and discussion

Following the simulation process explained before, the probabilistic forecasts from different reconciliation methods were generated. Energy score and variogram scores were calculated to assess the predictive performance of these forecasts. Results are presented in Table 3.

Mann-Whitney test for location comparison claimed that the ES and VS for all reconciled forecasts are significantly lower than that of base/incoherent forecasts. Thus it is evident that all reconciliation methods produce coherent probabilistic forecasts with improved predictive ability than incoherent forecasts. In addition to that, MinT(Shrink) and Optimal methods are having similar prediction accuracy as there is no significant difference between the scores from these reconciliation methods. However, the optimal reconciliation required high computational cost and it will increase for larger hierarchies. Further, it requires sufficient data points for the learning process of \mathbf{G} . Thus we claim to use the MinT \mathbf{G} for reconciling bootstrapped future paths for two reasons. Firstly it is computationally effective than the optimal method and secondly, it produces accurate probabilistic forecasts at least as good as Optimal method.

References

- Athanasopoulos, G., R. A. Ahmed, and R. J. Hyndman (2009). Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting* **25**(1), 146–166.
- Athanasopoulos, G., R. J. Hyndman, N. Kourentzes, and F. Petropoulos (2017). Forecasting with temporal hierarchies. *European Journal of Operational Research* **262**(1), 60–74.
- Ben Taieb, S., R. Huser, R. J. Hyndman, and M. G. Genton (2017). Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression. *IEEE Transactions on Smart Grid* **7**(5), 2448–2455.
- Dunn, D. M., W. H. Williams, and T. L. Dechaine (1976). Aggregate Versus Subaggregate Models in Local Area Forecasting. *Journal of American Statistical Association* **71**(353), 68–71.
- Gneiting, T. and M. Katzfuss (2014). Probabilistic Forecasting. *Annual Review of Statistics and Its Application* **1**, 125–151.
- Gneiting, T. and A. E. Raftery (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association* **102**(477), 359–378.
- Gneiting, T., L. I. Stanberry, E. P. Grimit, L. Held, and N. A. Johnson (2008). Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *Test* **17**(2), 211–235.
- Gross, C. W. and J. E. Sohl (1990). Disaggregation methods to expedite product line forecasting. *Journal of Forecasting* **9**(3), 233–254.
- Hyndman, R. J., R. A. Ahmed, G. Athanasopoulos, and H. L. Shang (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics and Data Analysis* **55**(9), 2579–2589.
- Jeon, J., A. Panagiotelis, and F. Petropoulos (2018). “Reconciliation of probabilistic forecasts with an application to wind power”.
- Scheuerer, M. and T. M. Hamill (2015). Variogram-Based Proper Scoring Rules for Probabilistic Forecasts of Multivariate Quantities *. *Monthly Weather Review* **143**(4), 1321–1334.
- Van Erven, T. and J. Cugliari (2014). *Game-Theoretically Optimal reconciliation of contemporaneous hierarchical time series forecasts*. Ed. by A. Antoniadis, X. Brossat, and J. M. Poggi, pp. 297–317.
- Wickramasuriya, S. L., G. Athanasopoulos, and R. J. Hyndman (2018). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association* **145**9, 1–45.