Hierarchical Forecasting

Name of First Author and Name of Second Author

1 Introduction

- Importance of coherency
- Point forecasting
- Probabilistic forecasting

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

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

Macroeconomic forecasts are crucial for economic and business activities of any economy. Therefore this area of study has a long history in literature. Econometricians have developed various approaches for getting reliable economic forecasts using macroeconomic data. However, the information of aggregation structure in

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

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

real data is limitedly used in literature. Moreover, having coherent forecasts will help the economists and policy makers for align decision making that impact for the whole economy. Therefore, our focus in this chapter is to introduce hierarchical forecasting methods for macroeconomic forecasting particularly for cross-sectional hierarchical data structures.

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

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

- Importance of coherency
- Point forecasting
- Probabilistic forecasting

2 Hierarchical time series

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

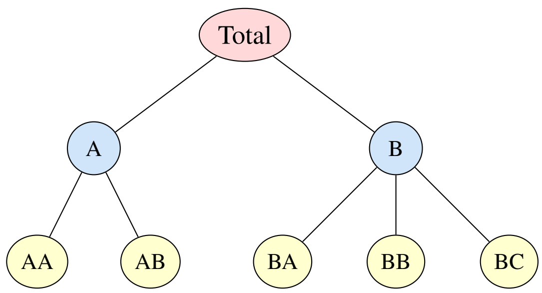

Fig. 1 A simple two-level hierarchical structure.

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

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

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

where

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

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

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

Structures similar to the one portrayed in Figure 1 can be found in macroeconomics. For instance in Section **??** we consider the case of GDP and its components. However, while this motivating example involves aggregation constraints, the mathematical framework we use can be applied for any general linear constraints, examples of which are ubiquitous in macroeconomics. For instance, the trade balance is computed as exports minus imports, while the consumer price index is computed as a weighted average of sub-indices, which are in turn weighted averages of sub-sub-indices and so on. These structures can also be captured by an appropriately designed $\boldsymbol{S}$ matrix.

An important alternative aggregation structure also commonly found in macroeconomics, is one for which the most aggregate series is disaggregated by attributes of interest that are crossed, as distinct to nested which is the case for hierarchical 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, $\boldsymbol{b}_t = (y_{AX}, y_{AY}, y_{BX}, y_{BY})'$. Hence, in contrast to hierarchical, grouped time series do not naturally disaggregate in a unique manner.

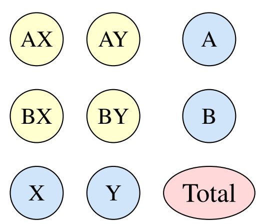

Fig. 2 A simple two-level grouped structure.

An important implementation of aggregation structures are *temporal hierarchies* introduced by **?**. In this case the aggregation structure spans the time dimension and dictates how higher frequency data (e.g., monthly) are aggregated to lower frequencies. There is a vast literature that studies the effects of temporal aggregation, going back to the seminal work of **????** and others such as, **????**. The main aim of this work is to find the single most optimum 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 some concrete conclusions. For example, **?** 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

?. **?** 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 overwhelming majority of the literature concentrates on a single level of temporal aggregation (there are some notable exceptions such as, **??**). **?** 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. This is an example of the benefits of forecast reconciliation to which we now turn out attention to.

3 Point forecasting

A requirement when forecasting hierarchial 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 $\tilde{\boldsymbol{y}}_{T+h|T}$, stacked in the same order as $\boldsymbol{y}_t$ and generated using information up to and including time T, are said to be *coherent* if $\tilde{\boldsymbol{y}}_{T+h|T} \in \mathfrak{s}$.

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

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

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 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 $\hat{\boldsymbol{b}}_{T+h|T} \in \mathbb{R}^m$, a set of

h-step ahead forecasts for the bottom-level series. For the simple hierarchical structure of Figure 1, $\hat{\boldsymbol{b}}_{T+h|T} = (\hat{y}_{AA,T+h|T}, \hat{y}_{AB,T+h|T}, \hat{y}_{BA,T+h|T}, \hat{y}_{BB,T+h|T}, \hat{y}_{BC,T+h|T})$, 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,

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

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

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,

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

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

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. Using such top-down approaches, is limited as it does not allow to capture and model individual series characteristics. To overcome this limitation, **?** introduced a new top-down approach which disaggregates the top-level forecasts based on proportions of forecasts rather than the historical data and show evidence that this method outperforms the conventional top-down approaches. However, a limitation of all top-down and by implication middle-out approaches that follow next, is that they introduce bias to the forecasts. We discuss this in detail in Section 3.2 that follows.

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 upwards. For series below the middle level, coherent forecasts are generated using a top-down approach by disaggregating the middle-level forecasts downwards. As mentioned above As the since the middle-out approach involves generating top-down forecasts, it also introduces bias to the forecasts. We discuss this in detail in Section 3.2 that follows.

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, ignoring any aggregation constraints, forecasts for all the series across all levels of the hierarchy are generated. We refer to these as *base* forecasts and denote them by $\hat{\boldsymbol{y}}_{T+h|T}$. In general, base forecasts will not be coherent. An example of an exception is when a simple method such as a random walk is used to generate naïve base forecasts for all the series in the hierarchy. In this case coherent the nature of the data is extended to the forecasts.

In a second step, base forecasts are reconciled, in an ex-post adjustment, so that they become coherent. This is achieved by projecting the base forecasts $\hat{\boldsymbol{y}}_{T+h|T}$ onto the coherent subspace $\mathfrak{s}$, via a projection matrix $\boldsymbol{SG}$, resulting in a set of coherent forecasts $\tilde{\boldsymbol{y}}_{T+h|T}$. More specifically,

$$\tilde{\boldsymbol{y}}_{T+h|T} = \boldsymbol{SG}\hat{\boldsymbol{y}}_{T+h|T}, \tag{2}$$

where $\boldsymbol{G}$ is an $m \times n$ matrix that maps $\hat{\boldsymbol{y}}_{T+h|T}$ to the $\mathbb{R}^m$ space, producing a set of coherent forecasts for the bottom-level, which are in turn mapped to the coherent subspace by the summing matrix $\boldsymbol{S}$ as defined in (1). We restrict our attention to projections on $\mathfrak{s}$ in which case $\boldsymbol{SGS} = \boldsymbol{S}$. This ensures that unbiasedness is preserved, i.e., for a set of unbiased base forecasts reconciled forecasts will also be unbiased.

Note that all single-level approaches discussed so far can also be represented by (2) using appropriately designed $\boldsymbol{G}$ matrices, however not all of these will be projections. For example for the bottom-up approach, $\boldsymbol{G} = \big(\boldsymbol{0}_{(m\times n-m)}\ \boldsymbol{I}_m\big)$ in which case $\boldsymbol{SGS} = \boldsymbol{S}$. For any top-down approach $\boldsymbol{G} = \big(\boldsymbol{p}\ \boldsymbol{0}_{(m\times n-1)}\big)$, for which case $\boldsymbol{SGS} \neq \boldsymbol{S}$.

3.2.1 Optimal MinT reconciliation

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

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

$$\hat{\boldsymbol{e}}_{T+h|T} = \boldsymbol{y}_{T+h|T} - \hat{\boldsymbol{y}}_{T+h|T} \tag{3}$$

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

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

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

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

is the unique solution that minimises the $\text{tr}[\boldsymbol{SGW}_h\boldsymbol{S}'\boldsymbol{G}']$ subject to $\boldsymbol{SGS} = \boldsymbol{S}$. MinT is optimal in the sense that given a set of unbiased base forecasts, it returns a set of best linear unbiased reconciled forecasts, using as $\boldsymbol{G}$ the unique solution that minimises the trace (hence MinT) of the variance of the forecast error of the reconciled forecasts. A significant advantage of the MinT reconciliation solution is that it is the first to incorporate the full correlation structure of the hierarchy via $\boldsymbol{W}_h$. However, estimating $\boldsymbol{W}_h$ is challenging, especially for $h > 1$. **?** present possible alternative estimators for $\boldsymbol{W}_h$ and show that these lead to different $\boldsymbol{G}$ matrices. We summarise these below.

- Set $\boldsymbol{W}_h = k_h\boldsymbol{I}_n$, for all h, where $k_h > 0$ is a proportionality constant. This most simplifying assumption returns $\boldsymbol{G} = (\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$ so that the base forecasts are orthogonally projected onto the coherent subspace $\mathfrak{s}$. Hence, we refer to this as the OLS projection which minimises the Euclidean distance between $\hat{\boldsymbol{y}}_{T+h|T}$ and $\tilde{\boldsymbol{y}}_{T+h|T}$.
  ? also come to the OLS solution, however from the perspective of the following regression model. They propose to reconcile unbiased base forecasts through

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

  where $\beta_{T+h|T} = E[\boldsymbol{b}_{T+h}|\boldsymbol{b}_1, \ldots.., \boldsymbol{b}_T]$ is the unknown conditional mean of the bottom-level series and $\boldsymbol{\varepsilon}_{T+h|T}$ is the coherence or reconciliation error with mean zero and variance $\boldsymbol{V}$. Hence, the OLS solution leads to the same projection matrix $\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$. We should note that using the usual GLS estimator in this context is not possible as $\boldsymbol{V}$ is not identifiable as shown by **?**.
  The OLS solution would be optimal under the conditions that the base forecast errors are uncorrelated and equivariant across all levels of the aggregation structure. However due to differences in scale across levels due to aggregation, the later is impossible to be satisfied.
  I think I just leave the one from above A disadvantage of the OLS solution, further to not accounting for the correlations across series, is that the homoscedastic diagonal entries do not account for the scale differences between the levels of the hierarchy due to aggregation. However OLS does well in practice because as

discussed it minimises the Euclidean distance and blah blah. Not sure how much we want to say here.

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

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

is the unbiased sample estimator of the in-sample one-step-ahead base forecast errors as defined in (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).
I have added this here because I think it is important and useful - we need to be careful to clarify that we used variance scaling when we get to the application.
An alternative diagonal scaling estimator is to set $\boldsymbol{W}_h = k_h \boldsymbol{\Lambda}$, for all h, where $k_h > 0$ and $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{S}\boldsymbol{1})$ with $\boldsymbol{1}$ being a unit column vector of dimension n. This was proposed by **?** for temporal hierarchies and assumes that each of the bottom-level base forecast errors has a variance k_h and are uncorrelated between nodes. Each element of the diagonal $\boldsymbol{\Lambda}$ matrix contains the number of forecast error variances contributing to that aggregation level. This estimator depends only on the aggregation structure and is therefore referred to as a WLS estimator applying structural scaling. Its advantage over OLS is that it assumes equivariant forecast errors only at the bottom-level of the structure and not across all levels. It is particularly useful in cases where forecast errors are not available; for example, in cases where the base forecasts are generated by judgemental forecasting.
- Set $\boldsymbol{W}_h = k_h \hat{\boldsymbol{W}}_1$, for all h, where $k_h > 0$, the unrestricted sample covariance estimator for $h = 1$. Although this is relatively simple to obtain and provides a good solution for small hierarchies, it does not provide reliable results as m grows compared to T. This is referred to this as the MinT(Sample) estimator.
- Set $\boldsymbol{W}_h = k_h \hat{\boldsymbol{W}}_1^D$, for all h, where $k_h > 0$, $\hat{\boldsymbol{W}}_1^D = \lambda_D \text{diag}(\hat{\boldsymbol{W}}_1) + (1-\lambda_D)\hat{\boldsymbol{W}}_1$ is a shrinkage estimator with diagonal target, and shrinkage intensity parameter

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

where $\hat{r}_{ij}$ is the ijth element of $\hat{\boldsymbol{R}}_1$, the 1-step-ahead sample correlation matrix as proposed by Schäfer & Strimmer (2005). Hence, off-diagonal elements of $\hat{\boldsymbol{W}}_1$ are shrunk towards zero while diagonal elements (variances) remain unchanged. This is referred to as the MinT(Shrink) estimator.

4 Hierarchical probabilistic forecasting

Point forecasts are limited since they provide no indication of uncertainty around the forecast. A richer description of forecast uncertainty can be obtained by providing a "probabilistic forecasts", that is a full density for the target of interest. For a review of probabilistic forecasts, and methods for evaluating such forecasts known as *scoring rules* see (Gneiting & Katzfuss 2014). In recent years, the use of probabilistic forecasts and their evaluation via scoring rules has become pervasive in macroeconomic forecasting, for example need to find some references that use scoring rules for macro forecasting. Check Bayesian macro guys like Koop Korobilis, Josh Chan also Mike Smith's work with Shaun Vahey.

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

In practice this definition suggests two approaches. For some parametric distributions, for instance the multivariate normal, it may be possible to derive a reconciled probabilistic forecast analytically. However, in macroeconomic forecasting, non-standard distributions such as bimodal distribution are often required to take different policy regimes into account worth checking if any (marginal) predictives are bimodal before we include this statement. In such cases a non-parametric approach based on bootstrapping in-sample errors proposed Gamakumara et al. (2018) can be used as long as a sample from the predictive distribution is available. Each of these scenarios is now covered in detail.

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

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

$$\tilde{\boldsymbol{y}}_{T+h|T} = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}_{T+h|T}, \tag{5}$$

and

$$\tilde{\Sigma}_{T+h|T} = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{\Sigma}}_{T+h|T}\boldsymbol{G}'\boldsymbol{S}'. \tag{6}$$

There are several options for obtaining the base probabilistic forecast and in particular the variance covariance matrix $\hat{\boldsymbol{\Sigma}}$. One option is to fit multivariate models level by level or for the hierarchy as a whole leading respectively to a $\hat{\boldsymbol{\Sigma}}$ that is block diagonal or dense. Another alternative is to fit univariate models for each individual series in which case $\hat{\boldsymbol{\Sigma}}$ is a diagonal matrix. Due to the large number of series under investigation here we consider the latter option. However we emphasise that correlation will enter the probabilistic forecast after reconciliation. The reconciled probabilistic forecast will ultimately depending on the choice of $\boldsymbol{G}$; the same choices of $\boldsymbol{G}$ matrices used in section 3 are be used here.

Need to check with Puwasala that base forecasts have diagonal sigma hat

4.2 Probabilistic forecast reconciliation in the non-parametric framework

In many applications, including macroeconomic forecasting, it may not reasonable to assume Gaussian predictive distributions. Therefore, non-parametric approaches has been widely used for probabilistic forecasts in different disciplines. For example, ensemble forecasting in weather applications (Gneiting & Raftery (2005), Gneiting & Katzfuss (2014), Gneiting et al. (2008)), bootstrap based approaches (Manzan & Zerom (2008), Vilar & Vilar (2013)). Check/replace these references with references that show heavy tails/skewness in macro applications.

Due to these concerns, we employ a reconciliation method proposed by Gamakumara et al. (2018) that does not make parametric assumptions about the predictive distribution. An important result that this method exploits is that applying methods for point forecast reconciliation to the draws from incoherent base predictive distribution results in a sample from the reconciled predictive distribution. This process, is summarised

1. Fit univariate models to each series in the hierarchy over a training set from $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_T$.
2. For each series compute h-step ahead point forecasts, for all h up to H. Collect these into a $n\times H$ matrix $\hat{\boldsymbol{Y}} := (\hat{\boldsymbol{y}}_{T+1|T},\ldots,\hat{\boldsymbol{y}}_{T+H|T})$, where $\hat{\boldsymbol{y}}_{T+h|T}$ is a $n\times 1$ vector of h-step point forecasts for all series in the hierarchy.
3. Compute one-step ahead in-sample forecasting errors. Collect these into an $n\times T$ matrix $\hat{\boldsymbol{E}} = (\hat{\boldsymbol{e}}_1,\hat{\boldsymbol{e}}_2,\ldots\ldots,\hat{\boldsymbol{e}}_T)$, where the $n\times 1$ vector $\hat{\boldsymbol{e}}_t = \boldsymbol{y}_t - \hat{\boldsymbol{y}}_{t|t-1}$. Here, $\hat{\boldsymbol{y}}_{t|t-1}$ is a vector of forecasts made for time t using information up to and including $t-1$. Information from $t=1,\ldots,T$ will be used to train the model used to form these forecasts.

4. Block bootstrap from $\hat{\boldsymbol{E}}$, that is choose H consecutive columns of $\hat{\boldsymbol{E}}$ at random, repeating this process B times. Denote the $n \times H$ matrix obtained at iteration b as $\hat{\boldsymbol{E}}^b$ for b=1,...,B.
5. For all b, compute $\hat{\boldsymbol{\Upsilon}}^b := \hat{\boldsymbol{Y}} + \hat{E}^b$. Each row of $\hat{\boldsymbol{\Upsilon}}^b$ is a sample path of h forecasts for a single series. Each column of $\hat{\boldsymbol{\Upsilon}}^b$ is a realisation from the joint predictive distribution at a particular horizon.
6. For each $b = 1, \dots, B$ select the h^{th} column of $\hat{\boldsymbol{\Upsilon}}^b$ and stack these to form a $n \times B$ matrix $\hat{\boldsymbol{\Upsilon}}_{T+h|T}$
7. For a given $\boldsymbol{G}$ matrix and for each $h = 1, \dots, H$ compute $\tilde{\boldsymbol{\Upsilon}}_{T+h|T} = \boldsymbol{SG}\hat{\boldsymbol{\Upsilon}}_{T+h|T}$. Each column of $\tilde{\boldsymbol{\Upsilon}}_{T+h|T}$ is a realisation from the joint h-step ahead reconciled predictive distribution.

Check with Puwasala that this is exactly what she has done. Notation may need work to bring in line with previous sections.

5 Empirical Study: Australian GDP

In our empirical data 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 can be disaggregated into additional series, which themselves could be targets of interests to forecasters. Each of these measures are aggregates of economic variables which are also targets of interests to the macroeconomic forecaster. This suggests a hierarchical approach to forecasting could be used to improve forecasts of all series in the hierarchy including headline GDP.

We concentrate on the Income and Expenditure approaches as nominal data are available only for these two. We focus only on 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 details on the data we refer the reader to (Australian Bureau of Statistics 2018).

5.1 Income approach

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

$$\begin{aligned} GDP(I) = {} & \textit{Compensation of employees} + \textit{Gross operating surplus} \\ & + \textit{Gross mixed income} + \textit{Taxes on production and imports} \\ & - \textit{Subsidies on production and imports} + \textit{Statistical decrepency (I)} \end{aligned}$$

Hierarchy shown in figure 12 in appendix 1 reflects how these are further disaggregated.

5.2 Expenditure approach

In the expenditure approach, the 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). Underline equation is,

$$\begin{aligned} GDP(E) = {} & \textit{Final consumption expenditure} + \textit{Gross fixed capital formation} \\ & + \textit{Changes in inventories} + \textit{Exports of goods and services} \\ & - \textit{Imports of goods and services} + \textit{Statistical decrepency (E)} \end{aligned}$$

Figures 10, 12 and 13 depict these hierarchies. In each hierarchy, the most aggregate level is denoted in grey whereas, the most disaggregate level is denoted in red. Intermediate levels are denoted in orange and blue. Levels denoted in orange continues to disaggregate further and these are separately depicted in different tree diagrams. Further, a description of each series in these hierarchies along with the series ID assigned by the ABS is given in the tables 1, 2, 3 and 4 in appendix 1.

Following subsections will give a brief description of income and expenditure hierarchies.

Associated hierarchical structure is given in figure 13, 14 and 15 in appendix 1.

Income and expenditure hierarchies consist 16 and 81 series respectively. All quarterly data for these series were obtained from the ABS and used to estimate coherent forecasts for Australian GDP along with its disaggregate components. In the following section we describe the hierarchical forecasting methods that we are using to get these forecasts.

- GDP of Australia

- How GDP is measured
  - Production, Income and Expenditure approach
  - Explain the hierarchy
- Issues with data
  - Why use current price rather than constant price (This is why we ignores production approach)
  - What is statistical discrepancy
  - Frequency of data
  - Does the data satisfies coherency
- Forecasting methods
  - Add a figure of all time series in income approach. Explain why traditional methods might not work using these time series as forecasting one layer of the hierarchy will ignore the structural information of individual time series to be used in the forecasts.
  - Explain ETS and ARIMA forecasting methods briefly
  - Hierarchical forecasting

In this empirical study we apply above discussed hierarchical methods to obtain coherent point and probabilistic forecasts for Australian GDP from income and expenditure approach along with the forecasts for its disaggregate components.

Let us first observe the time series plots for each approach. Figure 3 and 4 depicts these plots for income and expenditure hierarchies respectively. The upper panel of each figure shows the time series of all aggregate level series whereas lower panel shows the bottom level series. We can see that different series reflect different characteristics. For example, in the expenditure hierarchy, some series reflects an upward trend while some others reflect a downward trend or no trend at all. Further some series are having seasonal pattern whereas some does not have any seasonality. Moreover the bottom level series reflects some noise level compared to aggregate series in both hierarchies. Therefore, coherent forecasts through traditional methods such as top-down or bottom-up methods would not be accurate as they will ignore part of the information in generating forecasts. Thus forecast reconciliation is important in this empirical study.

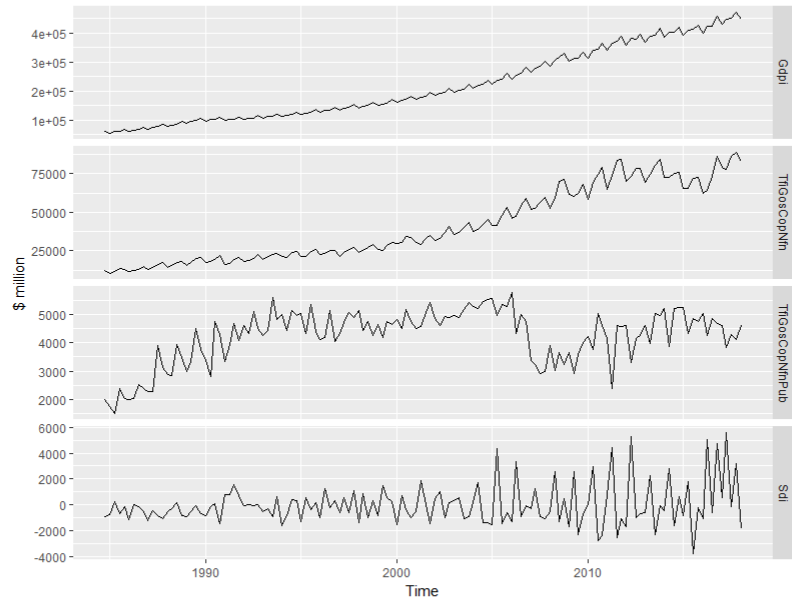

Fig. 3 Time plots for series from different levels of income hierarchy.

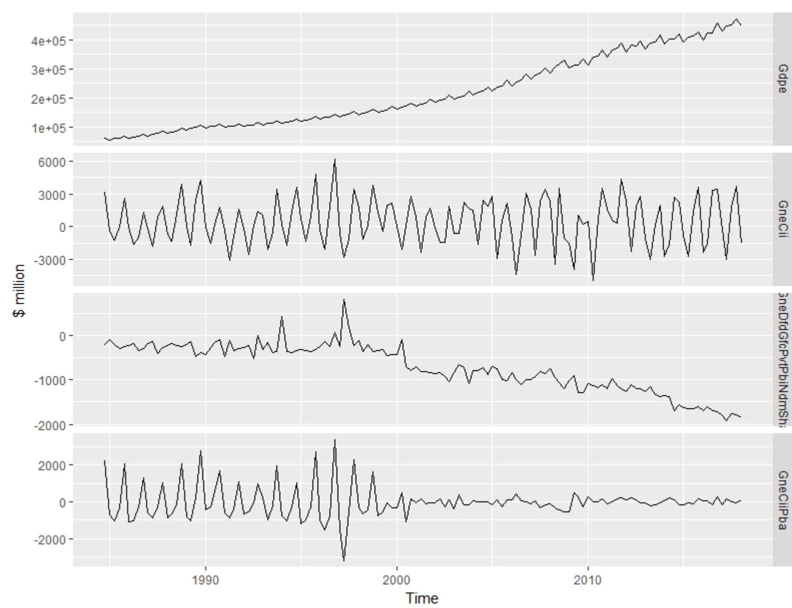

Fig. 4 Time plots for series from different levels of expenditure hierarchy.

6. Experimental Design and - setup: Time frame of Rolling window, - alternative methods: ARIMA/base vs naïve , BU, OLS, WLS, MinT(Shrink). - evaluation

- Point: MASE, MASE - Probabilistic: CRPS, Energy Score Skill score.

7. Results

6 Methodology

We now demonstrate the potential for reconciliation methods to improve forecast accuracy for Australian GDP data. We consider forecasts from $h=1$ quarter ahead forecasts up to $h=4$ quarter ahead using an *expanding* window. First, the training sample is set to QX of XXXX to QX of XXXX and forecasts are produced for QX of XXXX to QX of XXXX. Then the training window is expanded one period ahead, i.e. from QX of XXXX to QX of XXXX with forecasts produces for QX of XXXX to QX of XXXX. All up this leads to XX forecasts.

Need to get these dates off Puwasala. Also last section may need to be changed

6.1 Models

The first task in forecasting time series is to obtain base forecasts for all series in the hierarchy. In the case of the income approach this necessitates forecasting XX time series while in the case of the expenditure approach forecasts for XX time series must be obtained. As such our focus was on a methodology that was fast but flexible. We consider simple univariate ARIMA models, where model order is selected via a combination of unit root testing and AIC using an algorithm developed by XXX and implement in the auto.arima function in XXX. Cite this to Rob's satisfaction. A similar approach was also undertaken using the ETS framework to produce base forecasts. This had minimal impact on our conclusions with respect to forecast reconciliation methods, and in most cases ARIMA forecats outperformed ETS forecasts. Consequently, results for ETS models are excluded but are available from the authors upon request again do we put this in an appendix?. 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 regresssion. However, **?** 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 forecast reconciliation approaches that we consider are bottom up, OLS, WLS with What scaling did we use 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 reconciled are compared to a naïve benchmark. Since the data are not deseasonalised, the naïve benchmark is a seasonal random walk, i.e. the forecast for GDP (or one of its components) is the realised GDP in the same quarter of the previous year. The naïve forecast is by construction coherent and therefore does not need to be reconciled.

6.2 Evaluation

For evaluating point forecasts we consider two metrics, the Root Mean Squared Error (RMSE) and the Mean Absolute Scaled Error (MASE). The absolute scaled error is defined as

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

where $\breve{e}_{t+h}$[2] is the difference between any forecast and the realisation 44444 check m is and change since m is bottom level dimension. 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. As such 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 **?**.

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

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

where $\boldsymbol{y}_{T+h}$ is the realisation at time $T+h$, $\alpha \in (0,2]$. What did we use for alpha?. The expectations can be evaluated numerically as long as a sample from $\breve{F}$ is available which is the case for all methods we employ. An advantage of using energy score is that in the univariate case it simplifies to the commonly used cumulative rank probability score (CRPS) given by

$$\mathrm{CRPS}(\breve{F}_i, y_{i,T+h}) = E_{\breve{F}_i}|\breve{y}_{i,T+h} - y_{i,T+h}| - \frac{1}{2}E_{\breve{F}_i}|\breve{y}_{i,T+h} - \breve{y}^{*}_{i,T+h}|, \tag{7}$$

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

As an alternative to the energy score, log score and variogram scores were also considered. The log score was disregarded since Gamakumara et al. (2018) prove that the log score is improper with respect to the class of incoherent probabilistic

[2] breve is used instead of a hat or tilde to denote that this can be the error either a base or reconciled forecast

forecasts when the true DGP is coherent. The variogram score gave similar results to the energy score; variogram score results are omitted for brevity but are available from the authors upon request. or we put them in an appendix

Discuss results

7 Results

7.1 Base forecasts

Due to the different features in each time series a variety of ARIMA models were selected to be used as base models. Some generalisations about models. Figure 5 gives some indication of the performance of these base forecasting models relative to the naïve forecast over different horizons. All results in this Figure and in subsequent figures are presented as the percentage changes in a forecasting metric relative to the base forecast, a measure known in the forecasting literature as *skill scores*. Skill scores are computed such that positive values represent an improvement in forecasting performance over the base forecast and negative values represent a deterioration in forecast quality. In Figure 5, panels on the left refer to results for the Income hierarchy with panels on the right referring to the expenditure hierarchy. The top panels summarise results over all series in the hierarchy How exactly do we do this. Do we compute MSE for each series then add? Do we compute skill score for each series and average these?. The clear result is that base forecasts are more accurate than naïve forecasts, however as the forecasting horizon increases, the difference becomes smaller. This is to be expected since the naïve model here is a seasonal random walk, for horizons $h < 4$ the forecast from an ARIMA model is based on more recent information.

Similar results are obtained when MASE is used as the metric for evaluating forecast quality. However one disadvantage of the base forecasts relative to the naïve forecast is that base forecasts are no longer coherent. As such we now turn out attention to investigating whether reconciliation methods can lead to further improvements in forecast accuracy relative to base forecasts.

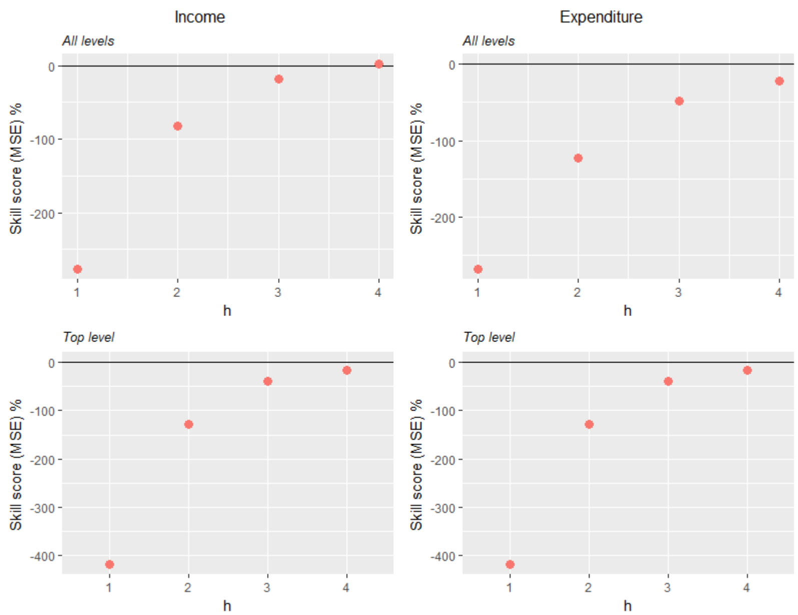

Fig. 5 Skill scores for naive forecasts (relative to base forecasts) using mean square error as a forecast metric. Top panels refer to results summarised over all series, bottom panels only refer to the top-level GDP series. Left panels refer to the income hierarchy, right panels to the expenditure hierarchy

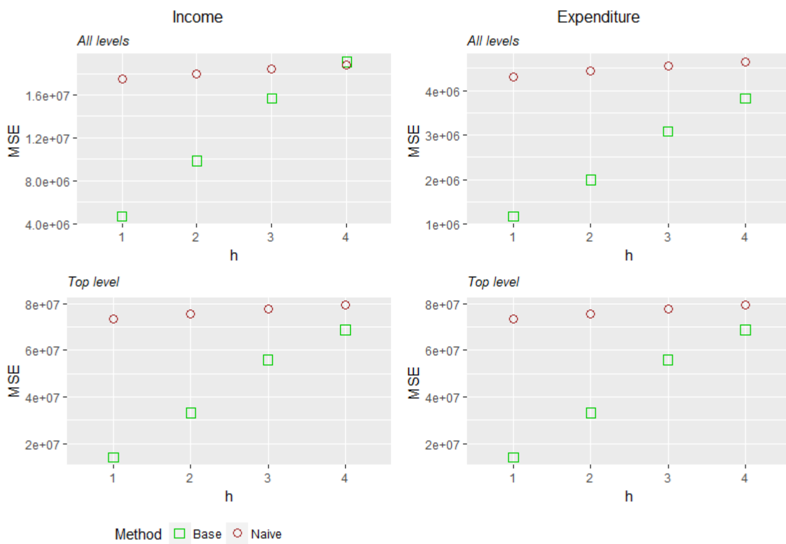

Fig. 6 Mean squared errors for naive and base forecasts. Top panels refer to results summarised over all series, bottom panels only refer to the top-level GDP series. Left panels refer to the income hierarchy, right panels to the expenditure hierarchy

7.2 Point Forecast Reconciliation

We now turn our attention to an evaluation of point forecasts obtained using different reconciliation methods. The top row of Figure 7 shows skill scores based on the MSE and MASE. These are aggregated over all series for both the income hierarchy and expenditure hierarchy and over different forecast horizons Puwasala to confirm whether we are looking at a skill score based on average MSE or an average of skill scores. Results are displayed as skill scores relative to base forecasts. We conclude that reconciliation methods generally improve forecast accuracy relative to base forecasts regardless of the hierarchy used, the forecasting horizon, the forecasting metric used to evaluate forecasts or the specific 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 differences between reconciliation methods we break down these results by different levels of each hierarchy. The second row of Figure XXX shows the forecasting performance a single series, namely the GDP which represents the top level of both hierarchies. The third row shows results for all series excluding those on 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, the second is that bottom up relatively poorly on aggregate series. These two features are particularly exacerbated for the larger Ependiture hierarchy. These results are consistent with other findings in the forecast reconciliation literature see for instance XXX.

Fig. 7 Skill score point forecasts from different reconciliation methods (with reference to base forecasts). Left two panels refer to skill score using MSE for income and expenditure hierarchies. Similarly right two panels refer to skill score using MASE. First row refer to results summarised over all series, second row refers to top-level GDP series, third row to aggregate levels and last row to the bottom level.

7.3 Probabilistic Forecast Reconciliation

We now turn our attention towards results for probabilistic forecasts. Figure 8 shows skill scores based on the energy score which as a multivariate score summarises forecast accuracy over the entire hierarchy. The top panels refer to assuming Gaussian probabilistic forecasts as described in Section 4.1 while the bottom panels refer to the method described in Section 4.2. The left panels correspond to the income hierarchy, the right panels 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 in MinT, 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 other 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 9 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.

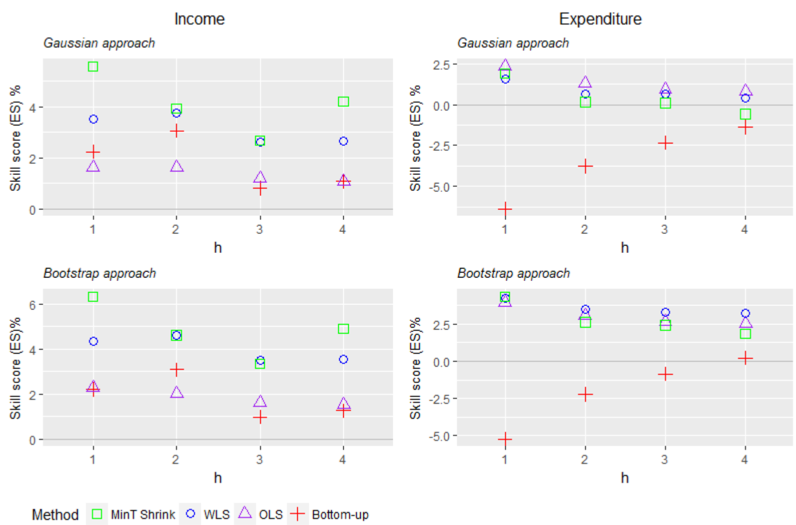

Fig. 8 Skill score for multivariate probabilistic forecasts from different reconciliation methods (with reference to base forecasts) using energy score. Top panel refers to the results for Gaussian approach and bottom panel refers to the non-parametric bootstrap approach. Left panel refers to the income hierarchy and right panel to the expenditure hierarchy.

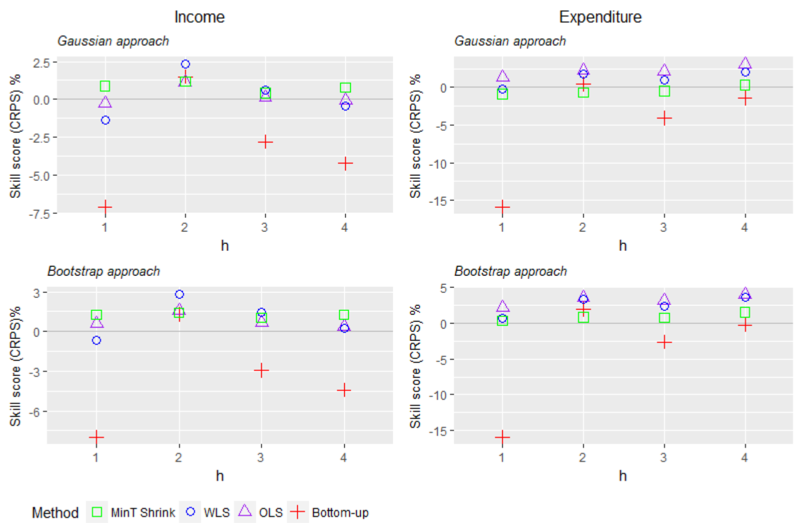

Fig. 9 Skill score for probabilistic forecasts of top-level GDP from different reconciliation methods (with reference to base forecasts) using CRPS. Top panel refers to the results for Gaussian approach and bottom panel refers to the non-parametric bootstrap approach. Left panel refers to the income hierarchy and right panel to the expenditure hierarchy.

Possible discuss skill scores in more detail especially w.r.t how they are combined

8 Conclusions

- Reconciliation can improve forecasts in macro context. This is true when we want to forecast all series but also when we only care about a single series.
- Bottom up which is the most simple and obvious way is not the best. Particularly if we only care about top level
- MinT works really well. Optimality is based on minimising trace of covariance matrix of reconciled errors but it good performance is robust to different evaluation metrics and even probabilistic forecasts.
- Above point may not be true for larger hierarchies.
- Future Avenues:
  - Scope to use methods for more complex structures (grouped/temporal)
  - Focus on constructing base forecasts using base models. More sophisticated approaches could be used. Will gains still be there or still be modest.
  - Non-linear constraints.

Appendix

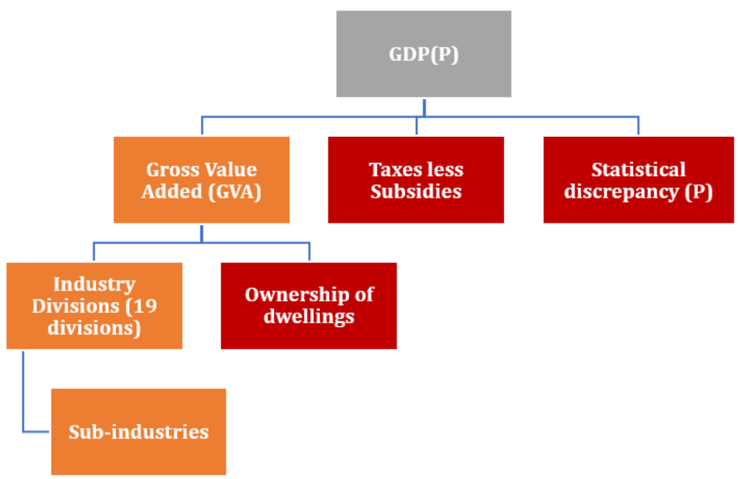

Fig. 10 Hierarchy of production approach.

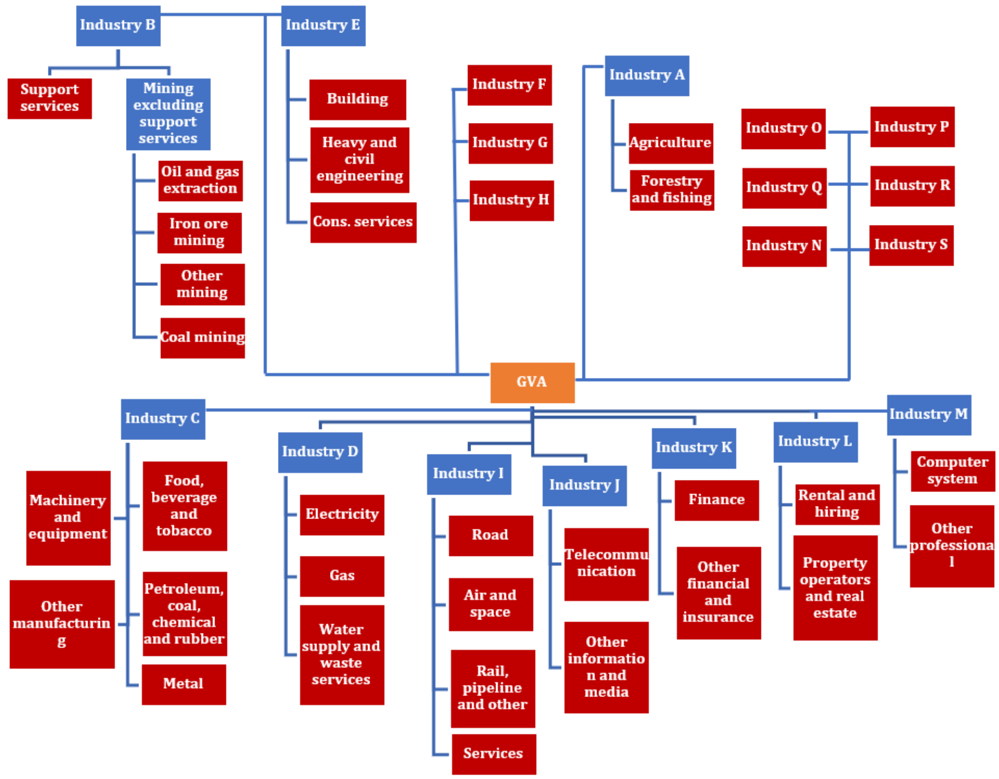

Fig. 11 Hierarchy of GVA under production approach.

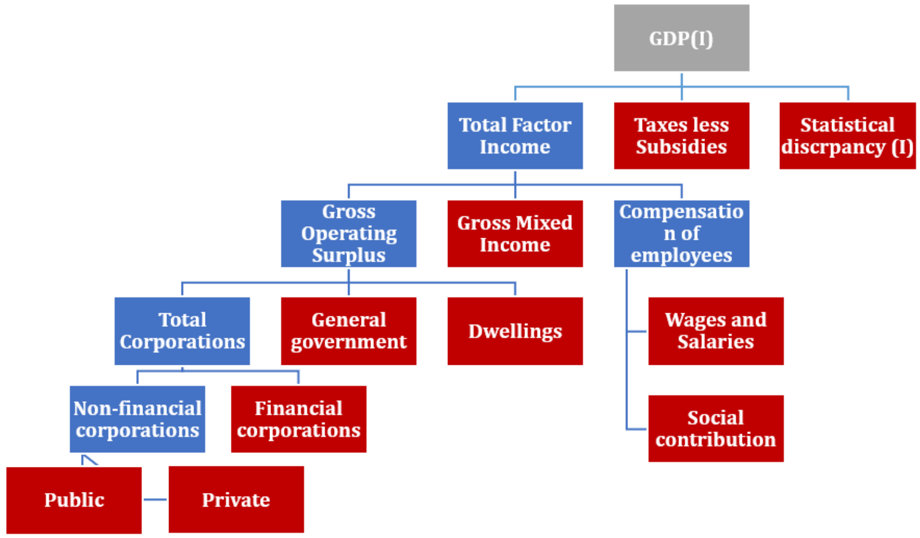

Fig. 12 Hierarchy of income approach.

$$GDP(E) = Final\ consumption\ expenditure + Gross\ fixed\ capital\ formation + Changes\ in\ inventories + Exports\ of\ goods\ and\ services - Imports\ of\ goods\ and\ services + Statistical\ decrepency\ (E)$$

Associated hierarchical structure is given in figure 13, 14 and 15.

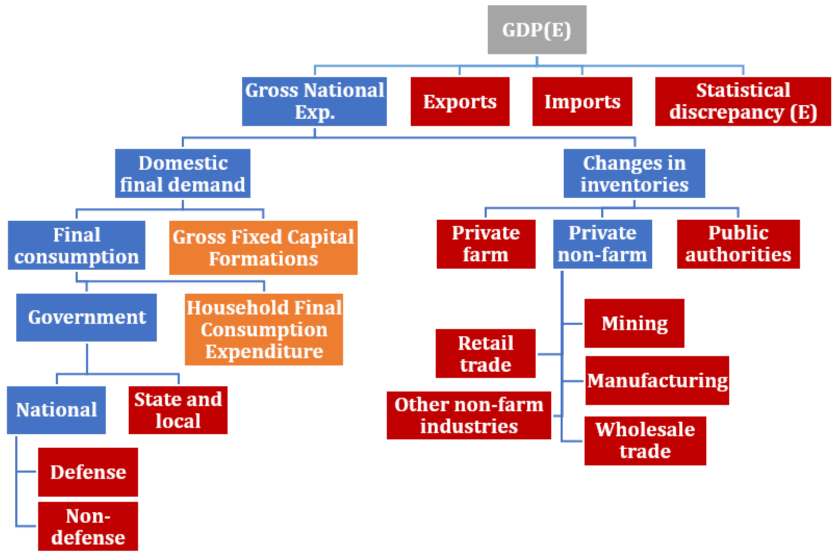

Fig. 13 Hierarchy of expenditure approach.

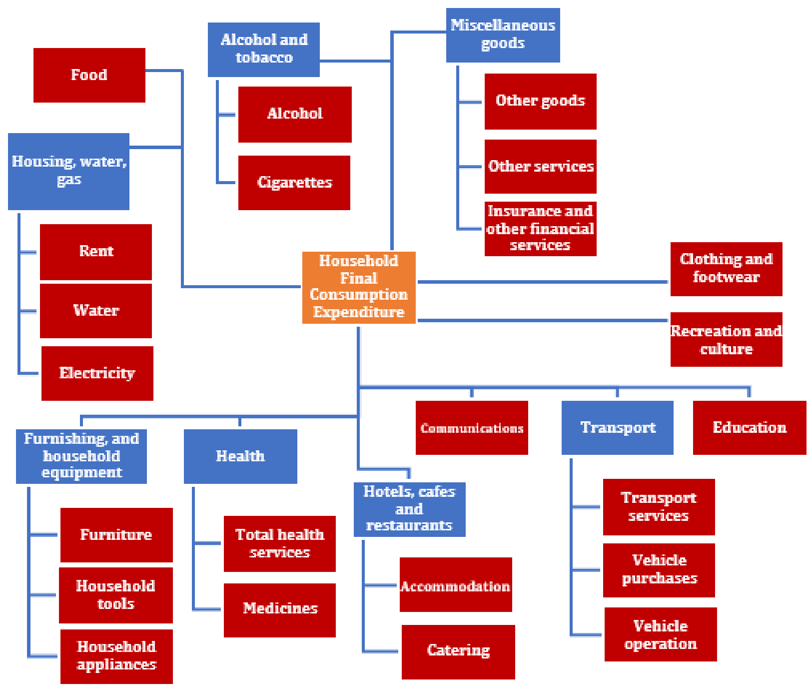

Fig. 14 Household final consumption expenditure under expenditure approach.

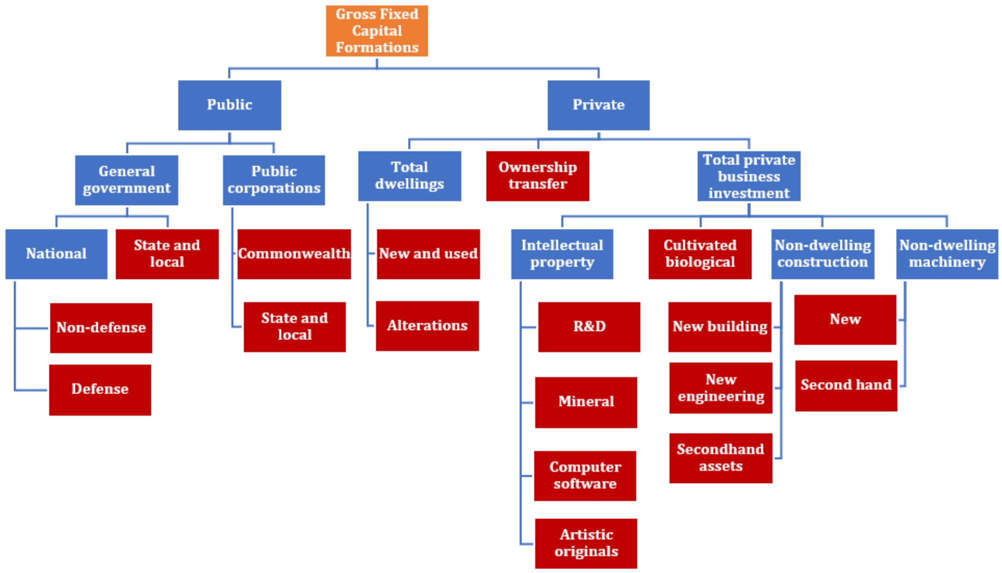

Fig. 15 Gross fixed capital formation (GFCF) under expenditure approach.

Table 1 Variables, Series IDs and their descriptions for 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.
GneDfdFceGvtSnl	A2302526R	Gen. gov. - State and local; Final consumption exp,
GneDfdFceGvt	A2302527T	Gen. gov.; Final consumption exp.
GneDfdFce	A2302529W	All sectors; Final consumption exp.
GneDfdGfcPvtTdwNnu	A2302543T	Pvt.; Gross fixed capital formation (GFCF)
GneDfdGfcPvtTdwAna	A2302544V	Pvt.; GFCF - Dwellings - Alterations and additions
GneDfdGfcPvtTdw	A2302545W	Pvt.; GFCF - Dwellings - Total
GneDfdGfcPvtOtc	A2302546X	Pvt.; GFCF - Ownership transfer costs
GneDfdGfcPvtPbiNdcNbd	A2302533L	Pvt. GFCF - Non-dwelling construction - New building
GneDfdGfcPvtPbiNdcNec	A2302534R	Pvt.; GFCF - Non-dwelling construction - New engineering construction
GneDfdGfcPvtPbiNdcSha	A2302535T	Pvt.; GFCF - Non-dwelling construction - Net purchase of second hand assets
GneDfdGfcPvtPbiNdc	A2302536V	Pvt.; GFCF - Non-dwelling construction - Total
GneDfdGfcPvtPbiNdmNew	A2302530F	Pvt.; GFCF - Machinery and equipment - New
GneDfdGfcPvtPbiNdmSha	A2302531J	Pvt.; GFCF - Machinery and equipment - Net purchase of second hand assets
GneDfdGfcPvtPbiNdm	A2302532K	Pvt.; GFCF - Machinery and equipment - Total
GneDfdGfcPvtPbiCbr	A2716219R	Pvt.; GFCF - Cultivated biological resources
GneDfdGfcPvtPbiIprRnd	A2716221A	Pvt.; GFCF - Intellectual property products - Research and development
GneDfdGfcPvtPbiIprMnp	A2302539A	Pvt.; GFCF - Intellectual property products - Mineral and petroleum exploration
GneDfdGfcPvtPbiIprCom	A2302538X	Pvt.; GFCF - Intellectual property products - Computer software
GneDfdGfcPvtPbiIprArt	A2302540K	Pvt.; GFCF - Intellectual property products - Artistic originals
GneDfdGfcPvtPbiIpr	A2716220X	Pvt.; GFCF - Intellectual property products Total
GneDfdGfcPvtPbi	A2302542R	Pvt.; GFCF - Total private business investment
GneDfdGfcPvt	A2302547A	Pvt.; GFCF
GneDfdGfcPubPcpCmw	A2302548C	Plc. corporations - Commonwealth; GFCF
GneDfdGfcPubPcpSnl	A2302549F	Plc. corporations - State and local; GFCF
GneDfdGfcPubPcp	A2302550R	Plc. corporations; GFCF Total
GneDfdGfcPubGvtNatDef	A2302551T	Gen. gov. - National; GFCF - Defence
GneDfdGfcPubGvtNatNdf	A2302552V	Gen. gov. - National ; GFCF - Non-defence
GneDfdGfcPubGvtNat	A2302553W	Gen. gov. - National ; GFCF Total
GneDfdGfcPubGvtSnl	A2302554X	Gen. gov. - State and local; GFCF
GneDfdGfcPubGvt	A2302555A	Gen. gov.; GFCF
GneDfdGfcPub	A2302556C	Plc.; GFCF
GneDfdGfc	A2302557F	All sectors; GFCF

Table 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

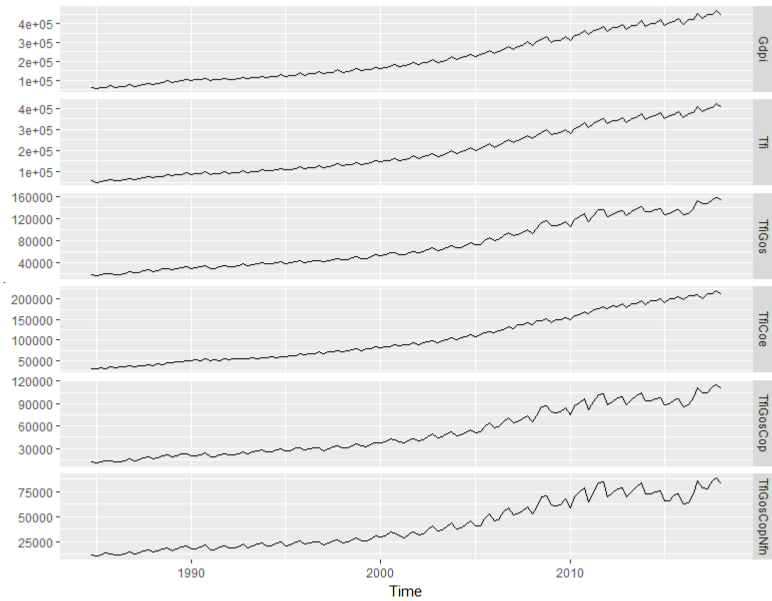

Fig. 16 All aggregate level series of income hierarchy.

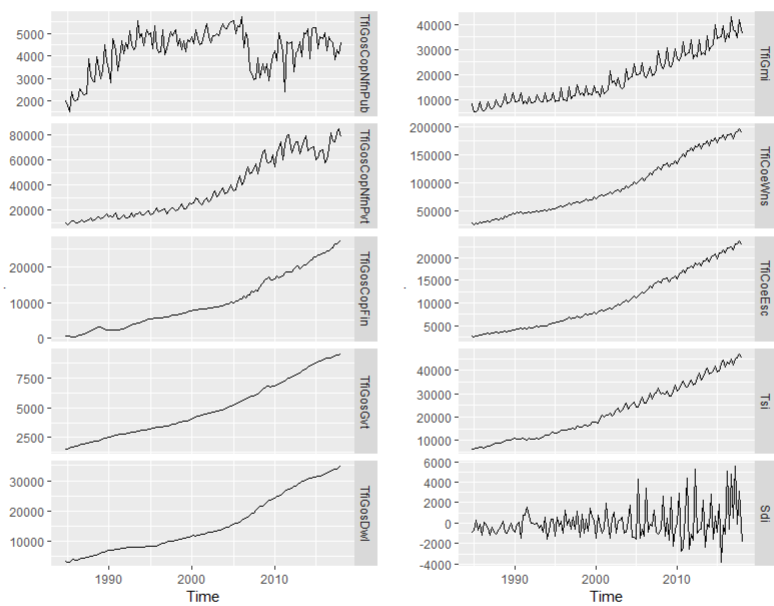

Fig. 17 All bottom level series of income hierarchy.

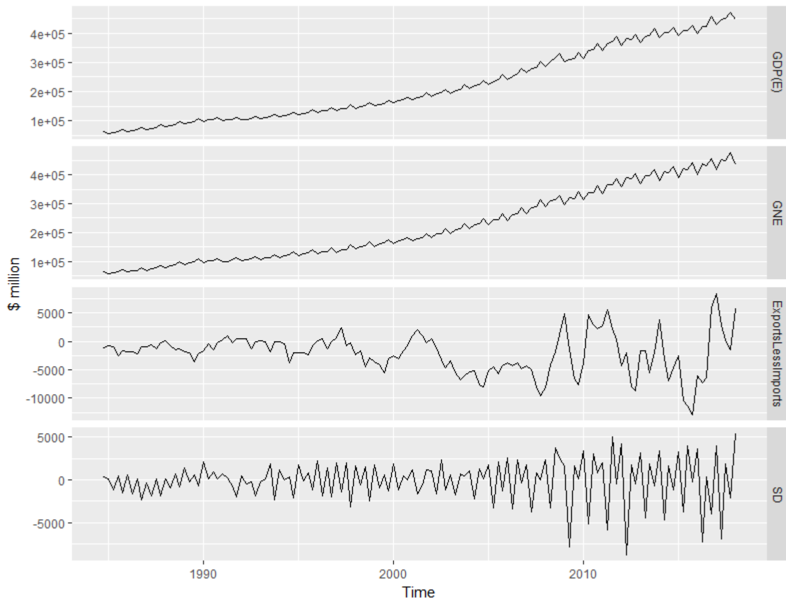

Fig. 18 GDP(E), GNE, Experts less Imports and Statistical discrepancy in expenditure hierarchy.

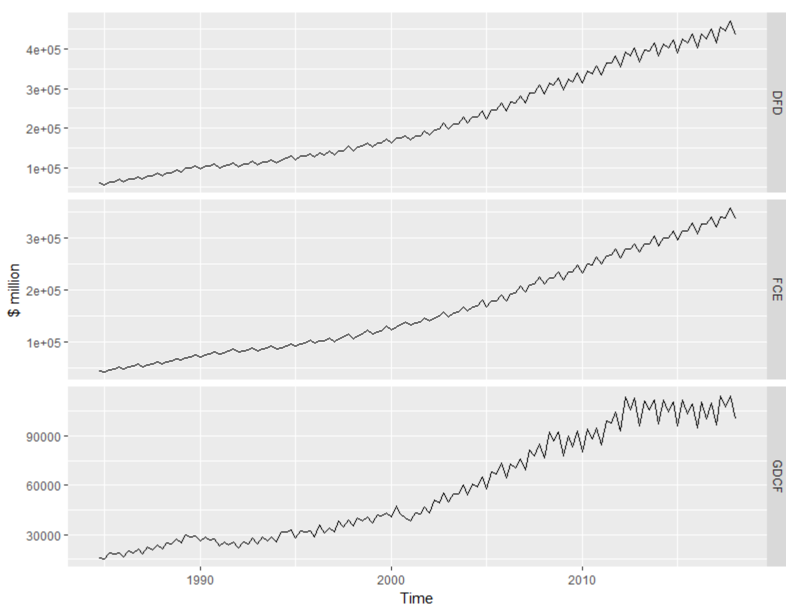

Fig. 19 Domestic Final Demand, Final Consumption Expenditure and Gross Fixed Capital Formations in expenditure hierarchy.

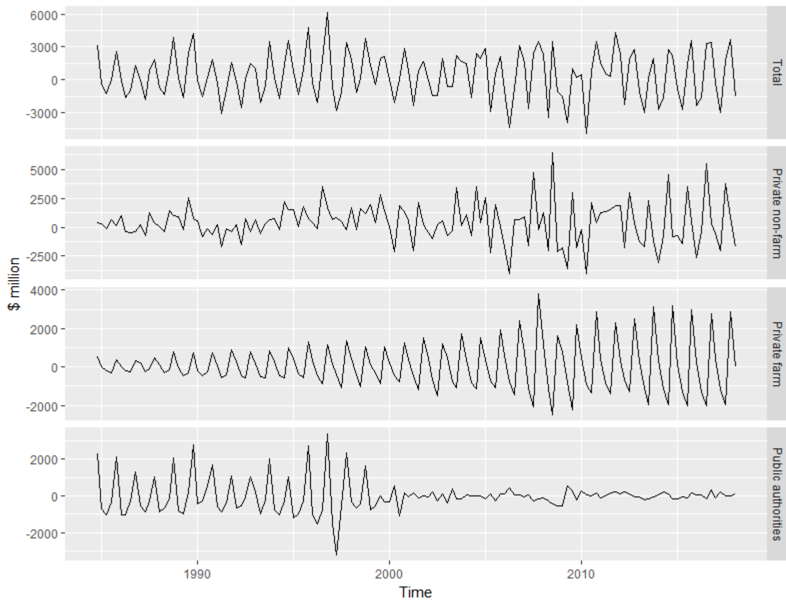

Fig. 20 Total changes in inventory, Private non-farm, Private farm and Public authorities in expenditure hierarchy.

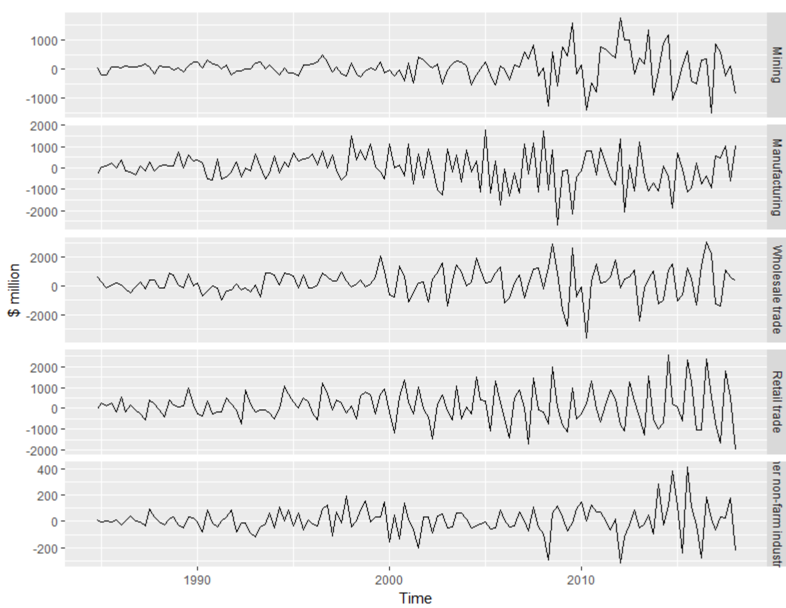

Fig. 21 Disaggregation of private non-farm in expenditure hierarchy.

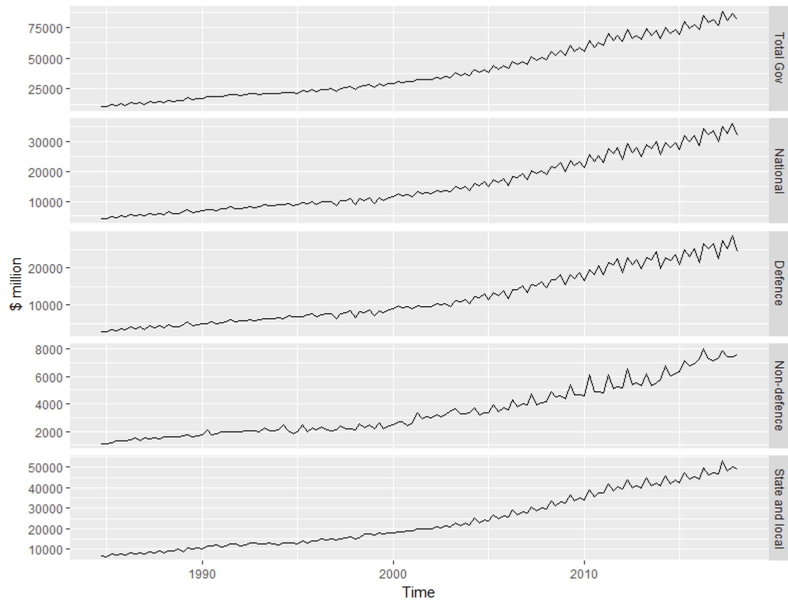

Fig. 22 Disaggregation of government final consumption expenditure.

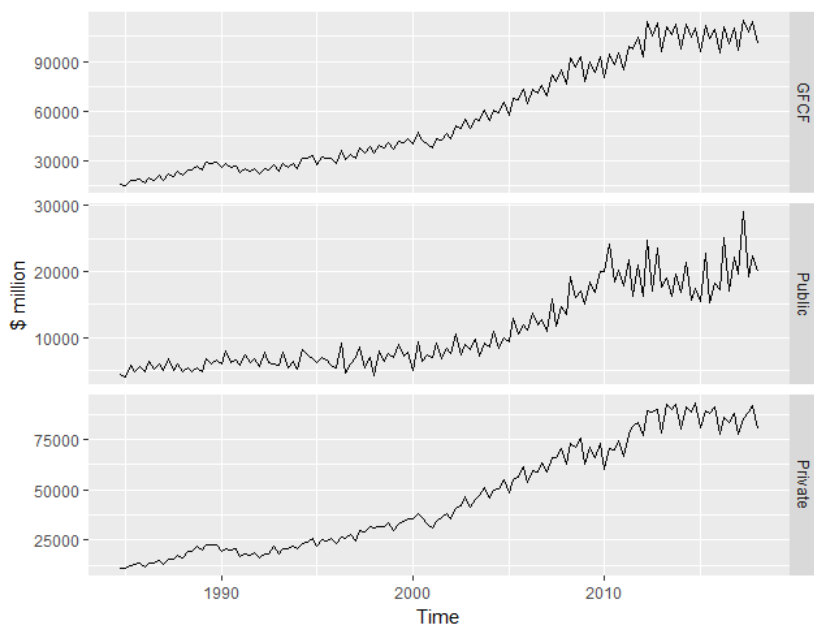

Fig. 23 Public, private and total fixed capital formations.

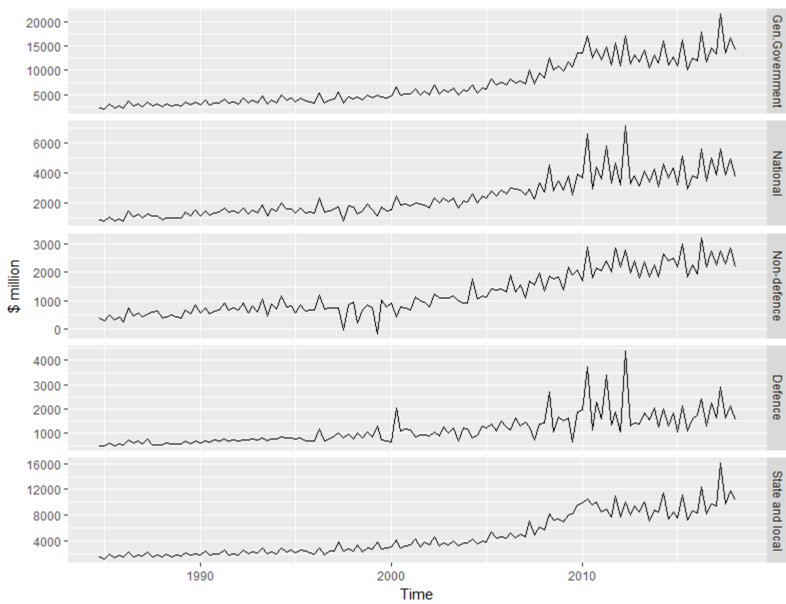

Fig. 24 Disaggregation of general government of Gross fixed capital formations.

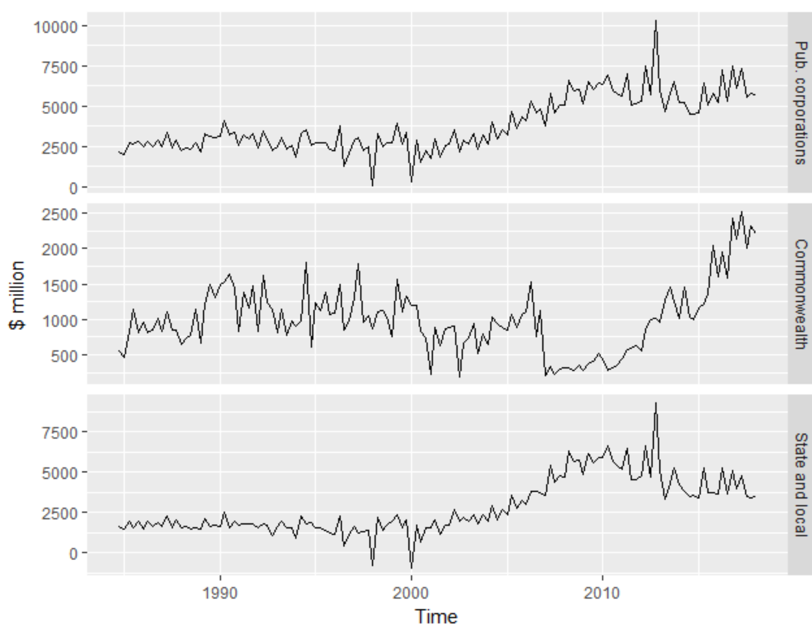

Fig. 25 Disaggregation of public corporations of Gross fixed capital formations.

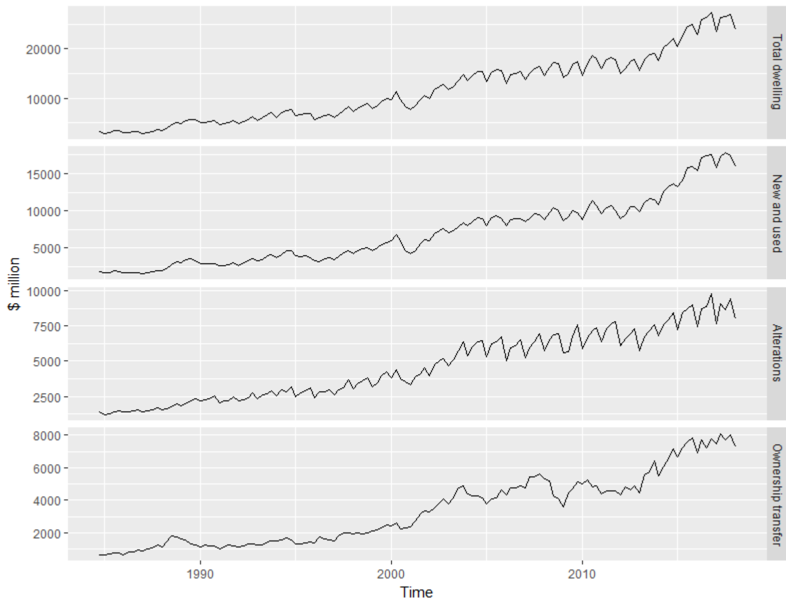

Fig. 26 Disaggregation of total dwelling and ownership transfer.

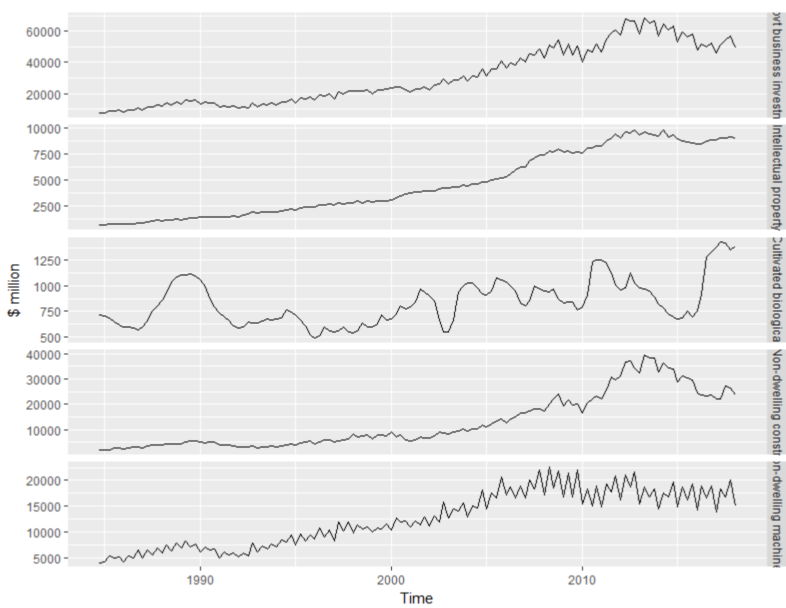

Fig. 27 Main disaggregation of total private business investments.

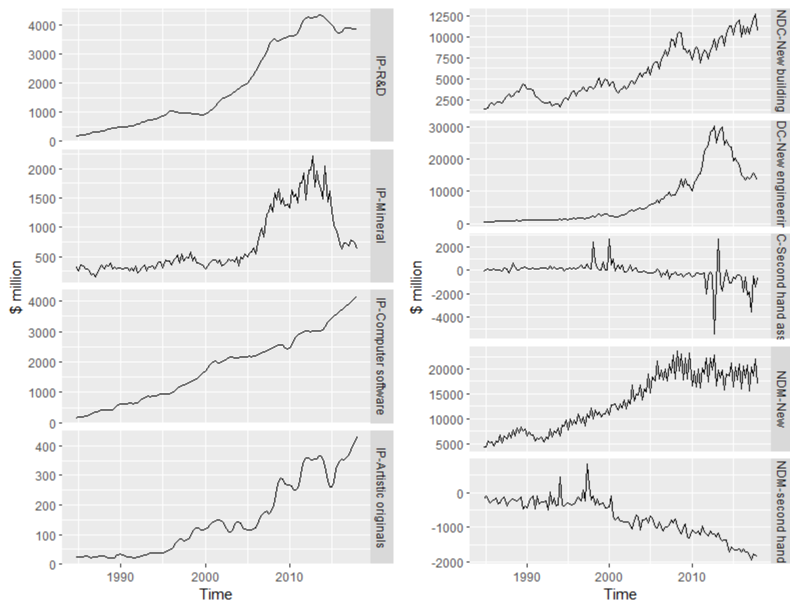

Fig. 28 Remaining disaggregation of total private business investments.

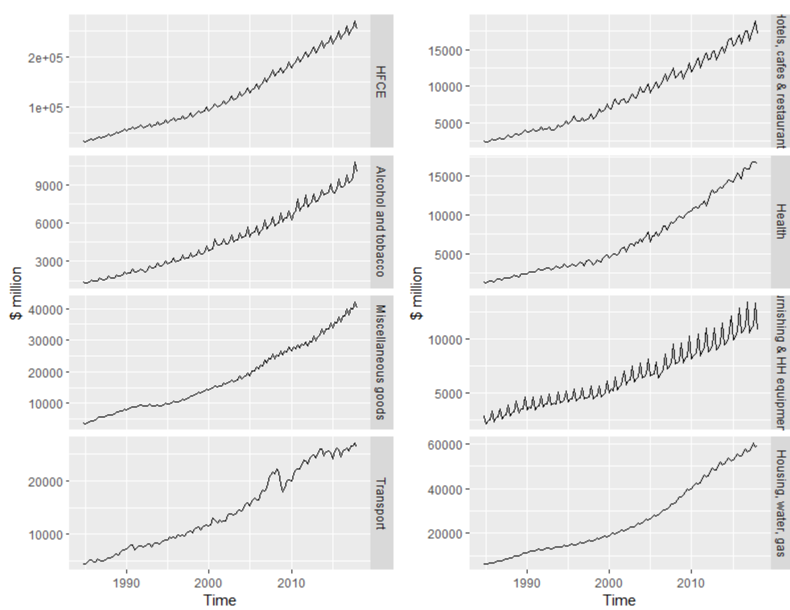

Fig. 29 Main disaggregation of household final consumption expenditure.

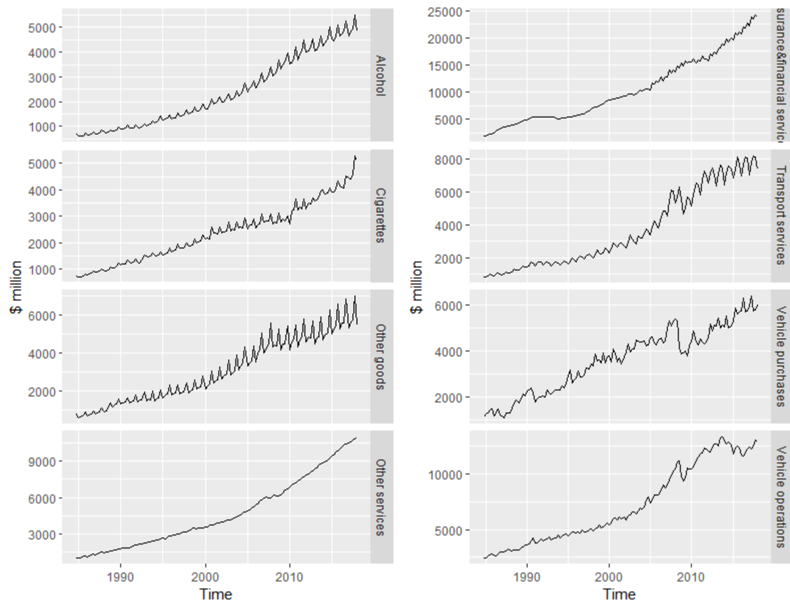

Fig. 30 Disaggregation of household final consumption expenditure.

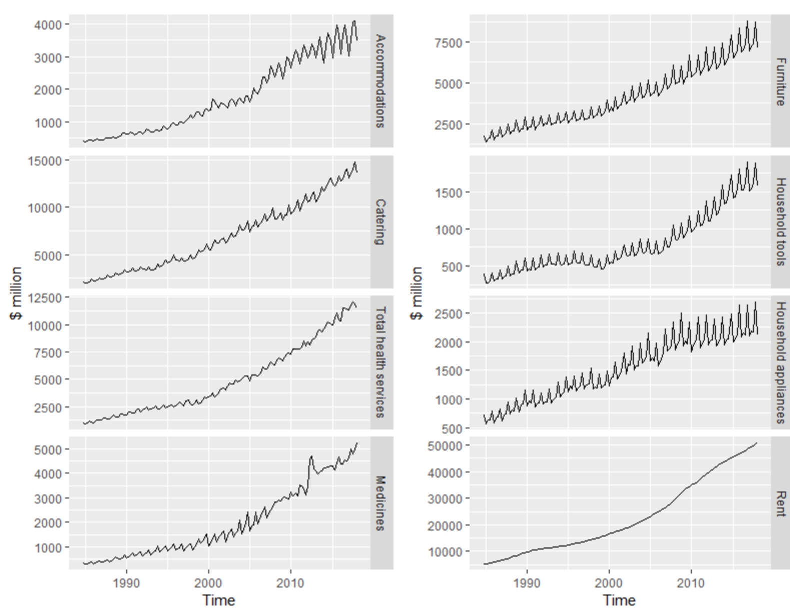

Fig. 31 Disaggregation of household final consumption expenditure.

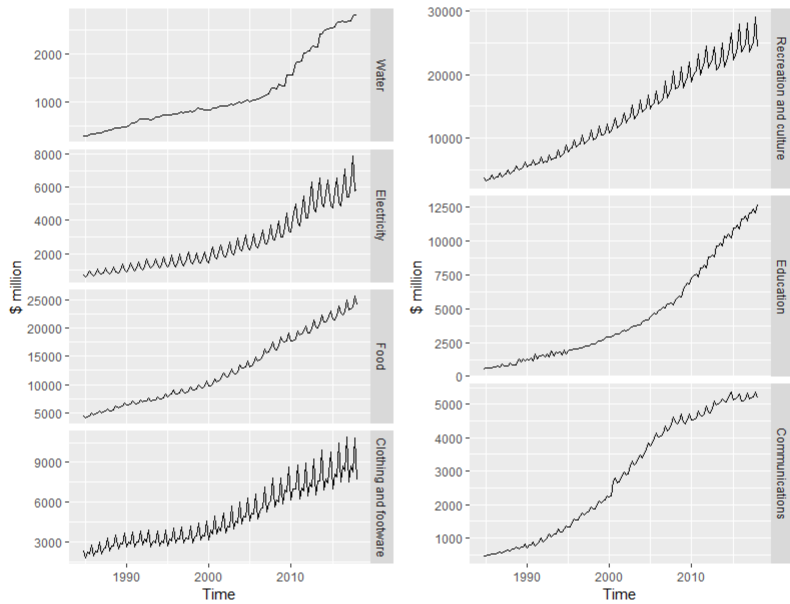

Fig. 32 Disaggregation of household final consumption expenditure.

References

Australian Bureau of Statistics (2015), Australian System of National Accounts: Concepts, Sources and Methods, Technical report.
URL: *http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/5216.02015?OpenDocument*

Australian Bureau of Statistics (2018), Australian National Accounts : National Income, Expenditure and Product, Technical report.
URL: *http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/5206.0Sep 2018?OpenDocument*

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

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

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

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

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

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

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

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

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

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