Cross–Temporal Coherent Forecasts for Gross Domestic Product

G I Rathanayke [1]

Abstract

Timely and accurate forecasts aligning different views of economic agents are of utmost importance in macroeconomic forecasting to facilitate effective policy decisions. Thus, this study investigates the ability of a reconciliation approach to align different viewpoints regarding forecasts and thereby increasing the forecast performance specifically related to GDP forecasting. The proposed methodology is based on forecasting hierarchical time series which is a collection of time series that follow an inherent aggregation structure. The aggregation constraints can be cross-sectional or temporal dimension. Thus, this method attempts to reconcile forecasts so that they follow aggregation constraints in both dimensions. This property is referred to as cross-temporal coherency. As the initial step forecasts are obtained for each of the series in the cross-temporal hierarchy. These are referred to as base forecasts and are often incoherent. These forecasts are then revised so that they become cross-temporally coherent. This is referred to as cross-temporal forecast reconciliation. Empirical applications based on disaggregated economic activities of the production approach for Sri Lankan GDP reveal that this approach brings improvement in forecast accuracy by blending different viewpoints in a data driven way. These cross-temporal coherent forecasts align decisions within an organisation transparently towards one number by aligning short term forecasts with long term forecasts and aligning views at different levels within the GDP hierarchy. As the proposed method is independent of forecasting models different short term forecasting models and long term forecasting models can be used to reflect different viewpoints.

Key Words: *Cross-sectional aggregation, Temporal aggregation, Forecast combinations, Hierarchical time series, Forecast reconciliation*

JEL Classification: *C53; N15; F43*

1 The author is currently serving as a Senior Assistant Director of the Macroprudential Surveillance Department. Corresponding email: gayani_ishara@cbsl.lk. The views presented in this paper are those of the author and do not necessarily indicate the views of the Central Bank of Sri Lanka

1. Introduction

1.1. Background

Forecasting macroeconomic variables (especially Gross Domestic Product (GDP) and inflation) is a leading research topic in the current macroeconometric literature as the challenges faced evolve over time. Macroeconomic forecasts are of utmost importance for policy makers to make informed decisions. Particularly, to take proactive decisions rather than reactive decisions. For instance, an early forecast of a recession would assist the government to move towards an expansionary fiscal policy to mitigate the impact of a severe economic downturn. Moreover, a forecast of inflation dropping under the target level of a central bank would give them an early indication to go for easing of monetary policy to stimulate the economy to bring the inflation rate back to the target at the right time. The timing of policy decisions is crucial as it is well known and universally accepted that the impact of monetary policy and fiscal policy decisions are transmitted with a lag. This highlights the importance of accurate forecasts as policy decisions must be timed in such a way that their impact is transmitted to the economy when it is required in order to obtain the intended results. In other words, policies are implemented today for forecasted future economic situations. Thus, improving the reliability and accuracy of macroeconomic forecasts is vital at this stage. Ample sophisticated forecasting models have been developed over time in macroeconomic forecasting literature, both in univariate and multivariate settings. The most prominent models include Dynamic Stochastic General Equilibrium models (DSGE), Dynamic factor models, VAR, and Bayesian VAR. In these models, GDP is commonly modeled in aggregate form. Focus on modeling and forecasting disaggregated subcomponents of GDP either based on the demand side or the production side is very limited. However, this area has recieved growing attention in recent years with studies such as Hahn and Skudelny (2008); Barhoumi et al. (2012); Esteves (2013); Higgins (2014); and Heinisch and Scheufele (2018) which mainly focus on exploring and comparing the accuracy gain of direct GDP forecasting and disaggregated GDP forecasting using a bottom-up approach. This approach involves forecasting the most disaggregated series and simply adding them to form forecasts of the aggregated series. The bottom-up approach has the strength in a way that it does not lose information due to aggregation. However, it only uses information from a single level of aggregation and ignores any correlations between the components and aggregates. In addition, this will perform poorly if the disaggregated series have low signal to noise ratio.

1.2. Hierarchical time series

A collection of time series that follows an aggregation constraint is referred to as a hierarchical time series (Hyndman and Athanasopoulos (2018)). For example, contemporaneous aggregation of GDP components in the production front which is a supply oriented decomposition of the value added by economic activities based on the national accounting methodology (European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations, and World Bank (2009)) is a cross-sectional hierarchy with aggregation constraints imposed via national accounting identities. There is growing literature which focuses on forecasting such a collection of hierarchical time series with the aim of ensuring that forecasts adhered to the aggregation constraints across the hierarchy. That is, the sum of the disaggregates should be equal to the corresponding aggregates. If we consider cross-sectional dimensions in the context of the GDP hierarchy, the sum of the forecasts of the economic activities should add up to the forecast of GDP. This property is referred to as coherency (Kourentzes and Athanasopoulos (2019)). Forecasts that are generated separately for each series in the hierarchy are base forecasts. These forecasts may not follow the aggregation constraints of the hierarchy except in the case where forecasts are generated by a simple naïve method. The process that adjusts these incoherent base forecasts to be in line with the aggregation constraints in the hierarchy is known as forecast reconciliation. Forecast reconciliation with cross-sectional hierarchies will align lower-level operational forecast with strategic forecast at higher levels.

1.3. Temporal hierarchies

As explained by Athanasopoulos et al. (2017) temporal hierarchy can be computed for any time series by using non overlapping temporal aggregations. For example, if GDP series is observed in quarterly frequency, we can compute semi-annual and annual levels to form the temporal hierarchy. Forecast reconciliation with temporal hierarchies will align short term forecasts with long term forecasts.

These forecast reconciliation methods have been proven to produce coherent forecasts that adhere to the aggregation structure and improve forecasting accuracy (Hyndman et al. (2011); Athanasopoulos, Ahmed, and Hyndman (2009); Hyndman, Lee, and Wang (2016); Wickramasuriya, Athanasopoulos, and Hyndman (2019); Athanasopoulos et al. (2017); Kourentzes and Athanasopoulos (2019); Athanasopoulos et al. (2020)). However, most of these studies focus on cross-sectional reconciliation or temporal reconciliation separately. To the best of my knowledge, the only studies that consider both these dimensions of reconciliation are those of Kourentzes and Athanasopoulos (2019) which introduce a two step method to generate cross-temporal coherent forecasts for Australian tourism, and Spiliotis et al. (2020) which attempts to sequentially combine multiple temporal aggregation with cross-sectional hierarchies related to electricity consumption.

1.4. Forecasting cross-temporal hierarchical time series

Forecasting cross-temporal hierarchical time series is challenging as forecasts need to adhere to both cross-sectional and temporal aggregation constraints. This is referred to as cross-temporal coherency. This property is important as it enables the forecasts to reflect real features of data. Further, coherent forecasts will enable aligned policy direction with one unique view.

In the context of GDP forecasting, it is vital to have forecasts that adhere to both cross sectional and temporal aggregation constraints for aligned decision making with one unique view on the future economic path. A recent study by Athanasopoulos et al. (2020) focuses on the application of cross-sectional forecast reconciliation using income and expenditure approach national accounting identities. However, to the best of my knowledge, no study has explored the accuracy gains and aligned decision making that would result in using cross-temporal reconciliation in the context of GDP forecasting. This research attempts to address this gap by proposing an alternative direct approach to the two step cross-temporal reconciliation approach introduced by Kourentzes and Athanasopoulos (2019).

1.5. Objectives

The main objective of this research is to explore the application of the cross-temporal forecast reconciliation methodology in the context of GDP forecasting. In this regard, we consider an empirical application which focusses on Sri Lankan production approach real GDP to obtain coherent forecasts while improving forecast accuracy. The motivation of this application is to explore the ability of this method to produce coherent forecasts which improve forecast accuracy compared to traditional bottom-up and direct approaches.

The contribution of this study to existing literature is significant in several aspects. First, it extends the cross-temporal forecast reconciliation methodology to macroeconomic forecasting. Further, it will strengthen current forecast models with the addition of this novel approach to GDP forecasting. Moreover, it will produce GDP forecasts which are coherent across all the sub activities as well as across time. This will align the short term quarterly projections with long term annual forecasts and facilitate the exploration of detailed sub activities which are drivers behind the forecasted GDP growth. It provides a better understanding of the current situation. This will facilitate policymakers to identify economic activities which have significant impact and focus on specialised policies to address specific economic activities under consideration. Methodologically, the exploration of the alternative direct approach to the two step cross-temporal reconciliation approach introduced by Kourentzes and Athanasopoulos (2019) will extend the current literature in this area.

1.6. Outline

Section 2 provides a detailed review of the literature on cross-sectional and temporal hierarchical forecasting approaches developed over time. Section 3 elaborates on the current methodology of forecast reconciliation and introduces the direct cross-temporal forecast reconciliation approach developed in this research study. Section 4 focusses on the empirical application of cross-temporal hierarchical forecasting for Sri Lankan GDP. Finally, section 6 summarises the conclusions of this study.

2. Literature review

2.1. Approaches in forecasting hierarchical time series

Earlier approaches in forecasting hierarchical time series mainly focused on selecting a single level of aggregation and then these were combined in a linear manner to generate coherent forecasts for the hierarchical structure. Top-down and bottom-up are two approaches prominent in literature (Syntetos et al. (2016)). The bottom-up approach involves forecasting the most disaggregated bottom-level series at the lowest level in the hierarchy and using simple aggregation to obtain forecasts at higher levels of the hierarchy (Hyndman and Athanasopoulos (2018)). The top-down approach starts with the forecast for the most aggregated top-level and disaggregates the forecast for the lower levels in the hierarchy as needed. The disaggregation can be based on weights derived from historical data as suggested by Gross and Sohl (1990). However, historical proportions do not reflect the dynamic changes in proportions over time. Athanasopoulos, Ahmed, and Hyndman (2009) propose using proportions based on forecasts to overcome this issue. Another less prominent approach uses a combination of bottom-up and top-down approaches. This is referred to as the middle out approach as it chooses an intermediate middle level to forecast and then aggregating bottom-up, as well as disaggregating top-down (Syntetos et al. (2016)).

Relative comparison of top-down and bottom-up methods in different fields in literature is rather inconclusive on the superiority of any method as conclusions depend on the characteristics of the empirical problem considered. Research that favours top-down approaches argue that disaggregate data are error prone and would produce imprecise forecasts due to high volatility and noise and hence top-down will result in better performance as it focuses on forecasting a smooth aggregated series which can reduce specification error (Grunfeld and Griliches (1960)). Research that favours bottom-up argues that information loss is substantial when aggregating series in a top-down approach (Dunn, Williams, and DeChaine (1976)); Weatherford, Kimes, and Scott (2001). Another set of researchers argues that the best approach depends on the correlation among the time series (Fliedner (1999)) or the underlying data generation process (Zotteri, Kalchschmidt, and Caniato (2005); Zotteri and Kalchschmidt (2007)).

The methods discussed so far all have a common limitation. They only consider one aggregation level and do not incorporate information from the entire hierarchical structure. Furthermore, as highlighted by Kourentzes, Barrow, and Petropoulos (2019) overreliance on a single model for all forecasts may increase model selection risk. On the other hand, if forecasts are generated independently for each level in the hierarchy as a simple method to use information from all levels, they may not be coherent and would fail to account for inherent correlation structure.

2.2. Forecast reconciliation methods

To overcome these limitations in traditional methods in forecasting hierarchical time series Hyndman et al. (2011) introduced a forecast reconciliation method. As explained above, if we forecast each of the time series in the hierarchical structure independently, it will not guarantee that the forecast generated will be coherent. In this context, forecast reconciliation can be considered as a process of adjusting forecasts to make them coherent. The basic idea of the methodology introduced by Hyndman et al. (2011) is to first forecast each time series in the hierarchical structure independently, which they term as "base forecasts". Then, to use a regression model to optimally combine and reconcile these forecasts to produce coherent forecasts. The Ordinary Least Squares (OLS) approach introduced in this paper computes reconciliation weights that only depend on the hierarchical structure and they are completely independent of the data. Hyndman et al. (2011) and Athanasopoulos, Ahmed, and Hyndman (2009) show that this method outperforms the commonly used top-down and bottom-up approaches. Extending this concept, Wickramasuriya, Athanasopoulos, and Hyndman (2019) show that reconciled forecast may be improved by using the information on the variance covariance matrix of the reconciled forecast errors. They further strengthen this approach by providing theoretical justification and introduce a new forecast reconciliation method which they refer to as minimum trace (MinT) reconciliation. In this method they produce an optimal forecast reconciliation approach by minimising the mean squared error of the coherent forecasts across the entire collection of time series which are given by the trace of the variance covariance matrix of the reconciled forecast errors under the assumption of unbiasedness.

The focus of all the above methods was limited to a cross-sectional forecast reconciliation setting. Athanasopoulos et al. (2017) extends this reconciliation approach in the time dimension with the introduction of the Temporal Hierarchical Forecasting (THieF) approach. Temporal aggregations can be constructed for any time series by computing non-overlapping temporal aggregates. In this reconciliation approach, the forecasts produced at all aggregation levels are combined to produce temporally reconciled, accurate and robust forecasts. The strength of this concept is based on combining information and borrowing strength from various levels of temporal aggregation of a time series, to generate forecasts. Apart from enabling aligned decision making in different planning horizons Athanasopoulos et al. (2017) show both in simulations and multiple empirical settings that the THieF approach results in improved forecast accuracy in all forecast horizons.

In literature there are only a limited number of attempts which focus on combining temporal and cross-sectional reconciliation. Kourentzes and Athanasopoulos (2019) combine these two concepts, namely the temporal hierarchical forecasting which align different planning horizons and cross-sectional hierarchical forecasting which align the forecast across the cross-sectional structure to produce forecast which are reconciled in both dimensions. This provides greater transparency as forecasts will align in one direction when different viewpoints within the organisation are considered. Apart from this transparency in decision making, Kourentzes and Athanasopoulos (2019) show that this method improves accuracy when forecasting Australian tourism demand. Highlighting the challenge of dimensionality that would result if the cross-temporal reconciliation is performed in one step, they propose an alternative two step procedure.

Another approach to produce cross-temporally reconciled forecast is presented in the work by Spiliotis et al. (2020) where they attempt to apply cross-sectional and temporal hierarchical forecast reconciliation sequentially. Further, they emphasise that multiple temporal aggregation enables to reduce model uncertainty and combining this with cross-sectional hierarchies result in substantial gains in forecast accuracy. However, the sequential nature of this approach does not guarantee coherent forecast across all dimensions.

2.4. Hierarchical forecasting methods for GDP forecasting

National accounting methodologies present three main disaggregation approaches in computing headline GDP. These are namely, expenditure, production, and income approaches. The expenditure approach is a demand side view which uses the national accounting identity that production equals domestic expenditure made on final goods and services. The production approach is a supply oriented decomposition of value added by economic activities. The income approach measures GDP as the sum of factor income flows (European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development, the United Nations, and World Bank (2009)).

In the context of GDP forecasting, the direct approach is dominant in empirical literature. Modelling and forecasting disaggregated subcomponents of GDP based on either the demand side or production side is limited in recent literature. The focus has been on the debate on whether direct GDP forecasting or bottom-up GDP forecasting produce better results. An early contribution in this topic is the study by Fair and Shiller (1990) which compares direct and bottom-up GDP forecasting for the USA. They use a VAR model to forecast aggregated Gross National Product (GNP) directly. Then they use Autoregressive Component (AC) models separately to forecast each of the disaggregated component of GNP and sum up the forecasts based on the GNP identity to produce the final GNP forecast. They find that the disaggregated AC model improves forecasting accuracy compared to the direct approach. Hahn and Skudelny (2008) extend the bottom-up approach to the production side to derive forecasts for Euro area real GDP growth but do not provide a comparison with the direct approach. Barhoumi et al. (2012) produce forecasts for GDP growth in France by aggregating component forecasts from both the supply and demand sides using bridge equation models. They emphasise that disaggregated forecasting produces more background information to build up the story around the forecasts. Moreover, GDP growth seems to be more precisely forecasted using the supply side approach. Heinisch and Scheufele (2018) compare bottom-up and direct GDP forecasting for Germany using an indicator based approach and conclude that the direct approach outperforms the bottom-up approach. Furthermore, the comparison of the performance of the production side disaggregated forecasting to the demand side revealed that using the production approach generates more accurate forecasts. Esteves (2013) studies the question of direct or bottom-up approaches for GDP forecasting using a different perspective. He emphasises that the choice of the approach is not dependent on the forecast performance but the level of analysis that forecasters wish to perform and on their expertise. In particular, the institution that focuses on short term forecasts will opt for a bottom-up approach as they must be able to explain the reasons behind the forecasts and identify current developments to help build the medium term forecasts.

In forecast reconciliation literature, only research that attempts to employ reconciliation methods in the context of GDP forecasting is that of Athanasopoulos et al. (2020). They focus on the application of cross-sectional forecast reconciliation using both income and the expenditure approach national accounting identities for Australian GDP. The study concludes that forecast reconciliation produces coherent forecasts and improves the overall forecast accuracy compared to a bottom-up approach when simple ARIMA models are used to derive the base forecasts.

This review of existing literature in this area indicates that to the best of my knowledge that no study has explored the application of cross-temporal reconciliation in the context of GDP forecasting. When it comes to GDP forecasting, coherent forecasts are of utmost importance to align policy direction. To achieve this objective, coherence in temporal dimension as well as cross-sectional dimension is important. Temporal coherence will ensure that short term policy direction is aligned with long term policy direction. Cross-sectional coherence will enable to identify economic activities which contributed to the forecasts. Therefore, it is valuable to investigate the application of cross-temporal forecast reconciliation for GDP forecasting, and this research aims to address this gap in literature.

3. Methodology

3.1. Hierarchical time series

Following a notation similar to Kourentzes and Athanasopoulos (2019), let $\mathbf{y}$ be an n- dimensional vector containing observations of the complete hierarchical structure and $\mathbf{b}$ be an m-dimensional vector of the most disaggregated times series which is often referred to as the bottom-level time series. We can write the aggregation constraints in any hierarchy as,

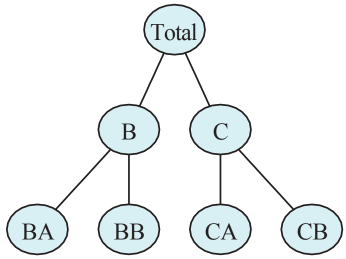
$$\mathbf{y} = \mathbf{S}\mathbf{b} \qquad (1)$$

where $\mathbf{S}$ is the summing matrix of order $n \times m$ which contains the linear aggregation constraints in the hierarchical structure in terms of bottom-level series.

For example, consider a simple two-level hierarchical time series either in the cross-sectional or temporal dimension which is represented in Figure 1. Level 0 is the most aggregated level, level 1 is the first level of disaggregation, and level 2 is the most disaggregated time series.

Let $\boldsymbol{y_{Total}}$ denote the observation of the most aggregated level 0 and $\boldsymbol{y_i}$ the observation corresponding to the node $\mathbf{i}$ of the levels below the top-level. The aggregation constraints

Figure 1: Two-level hierarchical structure



for this hierarchy in terms of the most disaggregated bottom-level time series can be represented by,

$$\begin{aligned} y_{Total} &= y_B + y_C \\ &= y_{BA} + y_{BB} + y_{CA} + y_{CB} \\ y_B &= y_{BA} + y_{BB} \\ y_C &= y_{CA} + y_{CB} \qquad (2) \end{aligned}$$

For this example, n which is the total number of nodes in the hierarchy is 7 and m which is the number of bottom-level series is 4. $\boldsymbol{y} = [y_{Total}, y_B, y_C, y_{BA}, y_{BB}, y_{CA}, y_{CB}]'$ and $\boldsymbol{b} = [\, y_{BA}, y_{BB}, y_{CA}, y_{CB}]'$ and the summing matrix is given by,

$$\boldsymbol{S} = \begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \square & \boldsymbol{I_4} & \square & \square \end{bmatrix}$$

where $\mathbf{I}_4$ is 4 x 4 identity matrix. Each aggregation constraint is represented by a row in the summation Matrix **S**. Thus, the same notation can be applied to represent any complex hierarchical structure.

3.2. Forecast reconciliation

The first step in forecast reconciliation is to generate h-steps ahead base forecasts for the complete hierarchy. Any forecasting method can be used to produce these forecasts, even multivariate models. However, these forecasts almost certainly will not be coherent. In other words, these will not follow hierarchical aggregation constraints other than in the case where a simple model such as naïve is used to generate base forecasts.

Let $\hat{\boldsymbol{y}}_{\boldsymbol{h}}$ be the h-step ahead base forecasts stacked in the same order as data **y**. Then, linear reconciliation methods can be written as,

$$\hat{y}_h = SG\hat{y}_h \qquad (3)$$

An appropriately selected matrix **G** of order m × n linearly maps base forecasts $\hat{\boldsymbol{y}}_{\boldsymbol{h}}$ to bottom-level forecasts. Then **S** sums these up to a set of reconciled forecasts $\hat{\boldsymbol{y}}_{\boldsymbol{h}}$ which are coherent. Thus, **SG** is referred to as the reconciliation matrix.

In traditional methods **G** only uses information from a single level from base forecasts which is a major drawback as highlighted earlier. For example, in the bottom-up approach $\boldsymbol{G} = [\boldsymbol{0}_{\boldsymbol{m}\times(\boldsymbol{n}-\boldsymbol{m})}|\boldsymbol{I}_{\boldsymbol{m}}]$ where $\boldsymbol{0}_{\boldsymbol{m}\times(\boldsymbol{n}-\boldsymbol{m})}$ is a null matrix of order m × (n − m) and $\boldsymbol{I}_{\boldsymbol{m}}$ is an identity matrix of order m × m. Thus, **G** only extracts bottom-level base forecasts from $\hat{\boldsymbol{y}}_{\boldsymbol{h}}$ and then these are summed by **S** to return the bottom-up coherent forecasts for the entire hierarchy.

Hyndman et al. (2011) show that if the base forecasts are unbiased the reconciled forecasts will preserve that unbiasedness if **SGS=S**. This holds for the bottom-up but not for top-down approaches. Therefore, this study will only focus on the bottom-up method for comparison. The identification of appropriate **G** which uses information from all levels within the hierarchy and which is also unbiased is important for the better performance of the forecast reconciliation method.

3.3. Optimal MinT reconciliation

Wickramasuriya, Athanasopoulos, and Hyndman (2019) frame the problem of finding appropriate **G** as an optimisation problem. They show that the variance covariance matrix of the h-step-ahead coherent forecast errors is given by,

$$V_h = var[y - \hat{y}_h] = SGW_hG'S' \qquad (4)$$

Where $\boldsymbol{W}_{\boldsymbol{h}} = E[\hat{\boldsymbol{e}}_{\boldsymbol{h}}\hat{\boldsymbol{e}}_{\boldsymbol{h}}']$ is a positive definite covariance matrix of the base forecast's errors $\hat{\boldsymbol{e}}_{\boldsymbol{h}} = \boldsymbol{y} - \hat{\boldsymbol{y}}_{\boldsymbol{h}}$. Then the error variances of the coherent forecast are on the diagonal of the matrix $\mathbf{V}_{\mathbf{h}}$. Hence, the sum of all the error variances is given by the trace of this matrix. Wickramasuriya, Athanasopoulos, and Hyndman (2019) shows that the form of the matrix **G** that minimises the trace of $\mathbf{V}_{\mathbf{h}}$ subject to **SGS=S** is given by,

$$G = (S'W_hS)^{-1}S'^{W_h^{-1}} \qquad (5)$$

This would give the best (minimum variance) linear unbiased reconciled forecasts and is referred to as MinT (minimum trace) reconciliation. Substituting **G** into Equation 3, reconciled forecasts from the MinT approach are given by,

$$\hat{y}_h = S(S'W_hS)^{-1}S'^{W_h^{-1}}\,\hat{y}_h \qquad (6)$$

The MinT approach has the ability of incorporating the full correlation structure of the hierarchy through $\boldsymbol{W}_{\boldsymbol{h}}$. However, the challenge in this approach is to estimate $\boldsymbol{W}_{\boldsymbol{h}}$ which is the variance covariance matrix of the base forecast which is of the dimension n × n. Thus, several alternative estimators for $\boldsymbol{W}_{\boldsymbol{h}}$ are used in literature.

3.4. OLS reconciliation

Set $\boldsymbol{W_h} = \boldsymbol{k_h I_n}$ where $\boldsymbol{k_h} > 0$ is a proportionality constant and $\boldsymbol{I_n}$ is n × n identity matrix. This will reduce the form of the MinT estimator to the OLS estimator proposed by Hyndman et al. (2011). This simplified assumption has performed well in practice (Hyndman et al., 2011; Athanasopoulos, Ahmed, and Hyndman, 2009). In this approach **G** only depends on **S**. Thus, this method can be used with forecasts generated from any forecasting method, such as judgmental forecasting. However, even though this is easy to apply, it ignores the correlations across series and the scale differences between the levels of the hierarchy due to aggregation.

3.5. Variance scaling

Set $\boldsymbol{W_h} = \boldsymbol{k_h diag(\widehat{W}_1)}$ for all h where $\boldsymbol{k_h} > 0$ and $\boldsymbol{\widehat{W}_1} = \frac{\boldsymbol{1}}{\boldsymbol{T}}\sum_{\boldsymbol{t=1}}^{T} \boldsymbol{\hat{e}_t \hat{e}_t}'$ where $\boldsymbol{\hat{e}_t}$ is the in-sample one-step ahead forecast errors of the base forecasts. This is referred to as a weighted least squares (WLS) estimator as it scales the base forecasts using the variance of in-sample residuals. This will account for heterogeneity within aggregation levels as well as across aggregation levels.

3.6. Structural scaling

Athanasopoulos et al. (2017) proposed to set $\boldsymbol{W_h} = \boldsymbol{k_h \Lambda}$ for all h where $k_h > 0$, $\boldsymbol{\Lambda} = \mathbf{diag(S1)}$ where **1** is a unit vector of dimension n. This is specifically applicable in the context of temporal hierarchies as it assumes that each of the bottom-level base forecasts has equal error variance $\boldsymbol{k_h}$ and are uncorrelated. In this approach error variances of the higher levels are taken as the sum of error variances that contributing to that aggregation level. As the weight scheme only depends on the aggregation structure, this is referred to as structural scaling. In contrast to the OLS approach this only assumes equal forecast error variances at the bottom level of the structure and not across all levels. Furthermore, as this does not require an estimate of variances of forecast errors it can be used with forecasts generated from any forecasting method, such as judgmental forecasting where sample residuals may not be available.

3.7. Sample covariance estimate for MinT

Set $\boldsymbol{W_h} = \boldsymbol{k_h \widehat{W}_1}$ for all h where $\boldsymbol{k_h} > 0$. This assumes $\boldsymbol{W_h}$ to be proportional to unrestricted sample covariance estimator for h=1. This is relatively simple to obtain and provides a good estimate for small hierarchies. However, when the number of bottom-level series (m) is larger compared to the length of the series T, this will not provide reliable results (Wickramasuriya, Athanasopoulos, and Hyndman (2019); Athanasopoulos et al. (2020)).

3.8. Shrinkage covariance estimator for MinT

Set $\boldsymbol{W_h} = k_h \boldsymbol{\widehat{W}^*_{1D}}$ for all h where $\boldsymbol{k_h} > 0$ and $\boldsymbol{\widehat{W}^*_{1D}} = \boldsymbol{\lambda \widehat{W}_{1D}} + \boldsymbol{(1-\lambda)\widehat{W}_1}$. This estimator shrinks the sample covariances to the diagonal target matrix $\boldsymbol{\widehat{W}^*_{1D}}$ which comprises of the diagonal elements of $\boldsymbol{\widehat{W}_1}$. Thus, off diagonal elements of $\boldsymbol{\widehat{W}_1}$ are shrunk towards zero. As proposed by Schäfer and Strimmer (2005) the shrinkage intensity parameter **λ** is set to,

$$\hat{\lambda} = \frac{\sum_{i \neq j} Var(\hat{r}_{ij})}{\sum_{i \neq j} \hat{r}_{ij}} \qquad (7)$$

where $\boldsymbol{\hat{r}_{ij}}$ is the $\mathbf{ij}^{\text{th}}$ element of $\boldsymbol{\widehat{R}_1}$, one-step ahead sample correlation matrix.

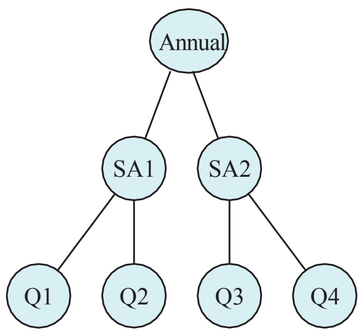
3.9. Cross-sectional forecast reconciliation

A cross-sectional hierarchy can be defined as a collection of time series that follows an aggregation constraint as shown in Figure 1. For example, consider a case where several geographical regions add up to give the total number for the whole country. In this setting the time series within each level and across each level represent different entities. Thus, we must account for heterogeneity within the levels and across the levels. Therefore, when estimating $\boldsymbol{W_h}$ more suitable estimators would be Variance scaling and Shrinkage MinT.

3.10. Temporal forecast reconciliation

The concept of temporal forecast reconciliation was introduced by Athanasopoulos et al. (2017). A temporal hierarchy can be developed for any time series by creating non overlapping temporal aggregates which do not introduce non-integer seasonality. If m is the highest frequency observed per year of a series, then each of the temporal aggregates created should be a factor of m. For example, if a series is observed in quarterly frequency then a temporal hierarchy can be constructed as shown in Figure 2, where the bottom level comprises of four quarterly observations (Q1, Q2, Q3, Q4) which adds up to the two semi-annual series (SA1, SA2) in the intermediate level which adds up to the total annual at the top level.

Figure 2: Temporal hierarchy for quarterly data



In contrast to cross-sectional forecast reconciliation the forecast horizon at each aggregation level will differ and it will depend on the specific aggregation level. For example, if we consider 4 quarters ahead forecasts, then the forecast horizon will be 4 when we consider the quarterly series, while it will be 1 and 2 for annual and semi-annual frequencies, respectively. In general, if h^* is the maximum required forecast horizon at the most disaggregated level and m is the highest frequency observed per year, then we would require $h = \lceil h^*/m \rceil$ forecasts at the most aggregated level. Then for each aggregation level k, we must generate $M_k h$ step ahead forecasts conditional on $\lfloor T/k \rfloor$ observations, where M_k is the number of observations per year for the k^{th} aggregation level and T is the length of the time series based on the highest frequency.

In this setting as forecasts for each level are created by one series, it is reasonable to assume homogeneous forecast errors within each level. Therefore, when estimating $\boldsymbol{W_h}$ assumptions behind the structural scaling estimator are justifiable in this situation.

3.11. Cross-temporal forecast reconciliation

In order to construct a cross-temporal hierarchy, a cross-sectional hierarchy needs to be combined with a respective temporal hierarchy. To illustrate this let us consider the simple cross-sectional hierarchy with one levels shown in Figure 3, where the two series B and C add up to the total and the temporal hierarchy for quarterly data shown in Figure 2.

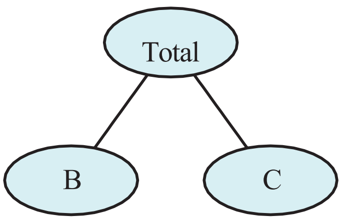
To develop a cross-temporal hierarchy we must consider the temporal aggregation at each of the cross-sectional nodes as shown in Figure 4. In this cross-temporal hierarchy, there are m = 8 bottom-level series, which comprise of four quarterly series at each of the two cross-sectional nodes. Further, with seven temporal aggregates at each cross-sectional node and with three cross-sectional nodes, there are $n = 7 \times 3 = 21$ nodes in the total cross-temporal hierarchy.

To create the cross-temporal Summation matrix (**S**) we need to combine the cross-sectional summation matrix ($\boldsymbol{S_C}$) and the temporal summation matrix ($\boldsymbol{S_T}$). In this regard, each of the elements in cross-sectional $\boldsymbol{S_C}$ need to be replaced with temporal $\boldsymbol{S_T}$.

Mathematically this is given by the Kronecker product of $\boldsymbol{S_C}$ with $\boldsymbol{S_T}$;

$$S = S_C \otimes S_T \qquad (8)$$

Figure 3: Simple cross-sectional hierarchy



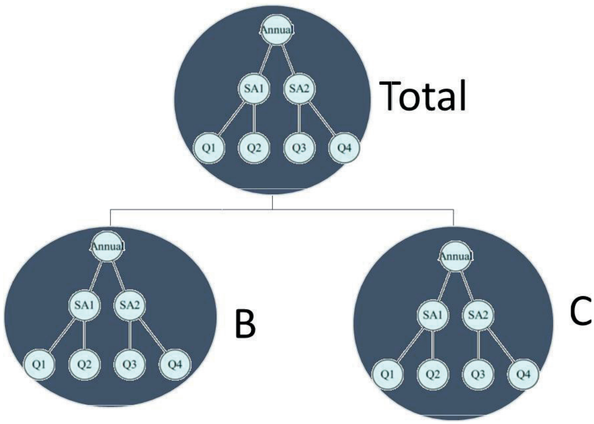
For example, the cross-sectional summation matrix ($\boldsymbol{S_C}$) corresponding to the hierarchy in Figure 3 in terms of the two bottom-level series B and C is,

$$S_C = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}_{3\times 2}$$

The temporal summation matrix for the temporal hierarchy in Figure 2 in terms of the four quarterly observations in the bottom-level is given by,

$$S_T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ \square & I_4 & \square & \square \end{bmatrix}_{7\times 4}$$

Figure 4: A cross-temporal hierarchy with quarterly data



Thus, the corresponding cross-temporal summation matrix for the cross-temporal hierarchy in Figure 4 would be,

$$\mathbf{S} = \mathbf{S_C} \otimes \mathbf{S_T}$$
$$= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ & I_4 & & & & & I_4 & \\ 1 & 1 & 1 & 1 & & & & \\ 1 & 1 & 0 & 0 & & 0_{7\times4} & & \\ 0 & 0 & 1 & 1 & & & & \\ & I_4 & & & & & & \\ & & & & 1 & 1 & 1 & 1 \\ & 0_{7\times4} & & & 1 & 1 & 0 & 0 \\ & & & & 0 & 0 & 1 & 1 \\ & & & & & & I_4 & \end{bmatrix}_{21\times8}$$

If we stack all the series in the cross-temporal hierarchy in vector **y** and all the bottom-level series in vector **b** then aggregation constraints in any cross-temporal hierarchy can be also represented by Equation 1. In the case of the above example. We let,

$$\boldsymbol{y} = [y_{Total_A}, y_{Total_{SA1}}, y_{Total_{SA2}}, y_{Total_{Q1}}, y_{Total_{Q2}}, y_{Total_{Q3}}, y_{Total_{Q4}}, y_{B_A}, y_{B_{SA1}}, y_{B_{SA2}},$$
$$y_{B_{Q1}}, y_{B_{Q2}}, y_{B_{Q3}}, y_{B_{Q4}}, y_{C_A}, y_{C_{SA1}}, y_{C_{SA2}}, y_{C_{Q1}}, y_{C_{Q2}}, y_{C_{Q3}}, y_{C_{Q4}}]'_{21\times1}$$
$$\boldsymbol{b} = [y_{B_{Q1}}, y_{B_{Q2}}, y_{B_{Q3}}, y_{B_{Q4}}, y_{C_{Q1}}, y_{C_{Q2}}, y_{C_{Q3}}, y_{C_{Q4}}]'_{8\times1}$$

Using the reconciliation matrix **SG** with **G** specified in Equation 5, optimal MinT reconciliation for the cross-temporal hierarchies can be computed directly using the same procedure as explained in Section 3.3. However, estimating $\boldsymbol{W_h}$ will be more challenging compared to considering cross-sectional and temporal dimensions separately as its dimension will become very large very quickly.

OLS reconciliation and structural scaling estimates of $\boldsymbol{W_h}$ can be directly applied to cross- temporal hierarchy with the developed $\mathbf{S}$ matrix as it does not require an estimate of forecast error variance. If we consider the cross-temporal hierarchy given in Figure 4 as it has 21 nodes, OLS reconciliation estimator of $\boldsymbol{W_h = k_h I_{21}}$, where I_{21} is 21 × 21 identity matrix. This is referred to as OLS in the results that follow. The structural scaling estimator for $\boldsymbol{W_h}$ with the assumption that equal forecast error variance at the bottom-level series is given by,

$$W_h = k_h diag(8, 4, 4, 2, 2, 2, 2, 4, 2, 2, 1, 1, 1, 1, 4, 2, 2, 1, 1, 1, 1)$$

The 8 at the top of the diagonal matrix represents that 8 bottom-level series are used to construct the top-level annual series. This is referred to as Struc in results to follow. Even though the assumptions behind these estimators are highly restrictive they are the only estimators that are applicable when in-sample forecast error variances are not available (e.g. with judgmental forecasts).

The variance scaling estimator of $\boldsymbol{W_h}$ for the cross-temporal hierarchies can be computed in a similar way as explained in Section 3.5 with in-sample residuals of the base forecasts stacked in the same way as the data. For example, the resulting estimator for the cross-temporal hierarchy in Figure 4 is given by,

$$\boldsymbol{W_h} = k_h diag(\sigma^2_{Total_A}, \sigma^2_{Total_{SA1}}, \sigma^2_{Total_{SA2}}, \sigma^2_{Total_{Q1}}, \sigma^2_{Total_{Q2}}, \sigma^2_{Total_{Q3}}, \sigma^2_{Total_{Q4}}, \sigma^2_{B_A}, \sigma^2_{B_{SA1}},$$
$$\sigma^2_{B_{SA2}}, \sigma^2_{B_{Q1}}, \sigma^2_{B_{Q2}}, \sigma^2_{B_{Q3}}, \sigma^2_{B_{Q4}}, \sigma^2_{C_A}, \sigma^2_{C_{SA1}}, \sigma^2_{C_{SA2}}, \sigma^2_{C_{Q1}}, \sigma^2_{C_{Q2}}, \sigma^2_{C_{Q3}}, \sigma^2_{C_{Q4}})$$

where σ^2_i is the estimated variance of the in-sample residuals corresponding to each time series. This is referred to as VAR in the results that follow. The variance scaling estimator using the diagonal of the sample covariance matrix requires fewer error variances to be estimated as compared to sample covariance estimate for MinT. However, the sample available to estimate each variance is limited to $\lfloor T/m \rfloor$. This will create stability problems with time series with limited history. Therefore, an alternative variance scaling estimator was also considered similar to the series variance scaling estimator introduced by Athanasopoulos et al. (2017). This assumes a common variance within the same temporal aggregation level in each of the cross-sectional nodes. This assumption is not unreasonable as the base forecast errors within the same aggregation level are for the same series in that particular frequency (i.e., semi-annual or quarterly). For example, the resulting estimator $\boldsymbol{W_h}$ for the cross-temporal hierarchy of the Figure 4 is given by,

$$\boldsymbol{W_h} = k_h diag(\sigma^2_{Total_A}, \sigma^2_{Total_{SA}}, \sigma^2_{Total_{SA}}, \sigma^2_{Total_Q}, \sigma^2_{Total_Q}, \sigma^2_{Total_Q}, \sigma^2_{Total_Q}, \sigma^2_{B_A}, \sigma^2_{B_{SA}},$$
$$\sigma^2_{B_{SA}}, \sigma^2_{B_Q}, \sigma^2_{B_Q}, \sigma^2_{B_Q}, \sigma^2_{B_Q}, \sigma^2_{C_A}, \sigma^2_{C_{SA}}, \sigma^2_{C_{SA}}, \sigma^2_{C_Q}, \sigma^2_{C_Q}, \sigma^2_{C_Q}, \sigma^2_{C_Q})$$

i.e., four quarterly forecast error variances for each year for each series will be replaced by one common quarterly forecast error variance, and two semi-annual forecast error variances for each year for each series will be replaced by one common semi-annual forecast error variance. This is referred to as SVAR in the results that follow. The shrinkage MinT estimator for the cross-temporal hierarchy can be computed as explained in Section 3.8. For example, the diagonal target matrix $\boldsymbol{\widehat{W}_{1D}}$ which comprises of diagonal elements of in-sample one-step ahead forecasts residual matrix $\boldsymbol{\widehat{W}_1}$ for the cross-temporal hierarchy of Figure 4 is given by,

$$\widehat{\boldsymbol{W}}_{\boldsymbol{1D}} = diag(\sigma^2_{Total_A}, \sigma^2_{Total_{SA1}}, \sigma^2_{Total_{SA2}}, \sigma^2_{Total_{Q1}}, \sigma^2_{Total_{Q2}}, \sigma^2_{Total_{Q3}}, \sigma^2_{Total_{Q4}}, \sigma^2_{B_A}, \sigma^2_{B_{SA1}},$$

$$\sigma^2_{B_{SA2}}, \sigma^2_{B_{Q1}}, \sigma^2_{B_{Q2}}, \sigma^2_{B_{Q3}}, \sigma^2_{B_{Q4}}, \sigma^2_{C_A}, \sigma^2_{C_{SA1}}, \sigma^2_{C_{SA2}}, \sigma^2_{C_{Q1}}, \sigma^2_{C_{Q2}}, \sigma^2_{C_{Q3}}, \sigma^2_{C_{Q4}})$$

The shrinkage intensity parameter λ was estimated using the method proposed by Schäfer and Strimmer (2005) which is implemented in the SHIP package (Jelizarow and Guillemot (2015)) for R (R Core Team (2020)). This is referred to as Shrk in the results that follow.

The sample covariance estimator for MinT was not considered for cross-temporal hierarchy. Even though it is straight forward to apply, estimates are highly unstable with the increasing dimensionality.

4. Results and discussion

GDP is the total value of goods and services produced within the boundaries of a country in a particular period. The System of National Accounts (SNA) (European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations, and World Bank (2009)) presents an internationally agreed standard set of recommendations on how to compile measures of economic activity including GDP. As defined in this framework, "GDP is derived from the concept of value added. Gross value added (GVA) is the difference between output and intermediate consumption. GDP is the sum of gross value added of all resident producer units plus that part (possibly the total) of taxes on products, less subsidies on products, that is not included in the valuation of output." Based on this methodology there are three approaches of computing GDP, which are: the Production approach, Income approach, and Expenditure approach. These approaches compute GDP as an aggregate of various economic variables. This forms a natural cross-sectional hierarchy. Thus, using a hierarchical approach to forecasting will enable us to improve forecasting accuracy, preserve coherency of the forecasts as well as provide aligned information on the contributors of the forecasts generated.

4.1. Sri Lankan GDP

Sri Lankan National Accounts are currently compiled by the Department of Census and Statistics (DCS) in compliance with guidelines given in SNA 2008 using 2010 as the base year. This case study uses the production approach of GDP by economic activity at constant prices from 2010-Q1 to 2019-Q4. I restrict my attention to the Production approach which is also known as the Output approach as it presents the supply-side decomposition of value added by economic activities. It allows the tracking of the overall performance of the whole economy. This approach provides data for the analysis of the productivity of each economic activity and changes in the structure of the economy. Furthermore, it allows policy makers to analyse the performance of specific economic activities against the industry averages (Viet (2009))

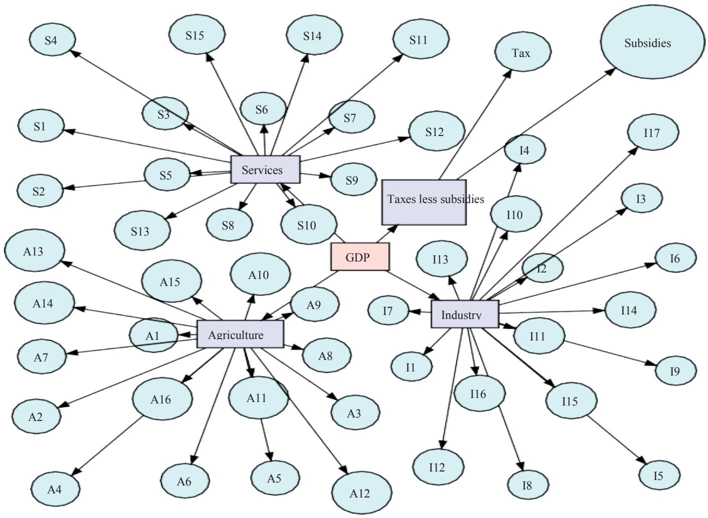
GDP is defined by the production approach as the sum of the GVA at basic prices of all resident producers plus taxes on products payable less subsidies on products receivable (European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations, and World Bank (2009)).

$$\textbf{GDP = GVA at basic prices + all taxes on products − all subsidies on products} \quad (9)$$

The GVA is an aggregated value added based on value added generated by economic activities which are classified according to Sri Lanka Standard Industry Classification based on International Standard Industry classification - Rev.4.

The most detailed dissemination table provides 48 economic activities which are categorised into 3 main streams: 16 activities related to Agriculture, forestry and fishing activities, 17 activities related to Industry activities and 15 activities relating to Services activities.

Figure 5: Hierarchical structure of the income approach for GDP



Note: The Pink cell contains GDP which the most aggregated series purple cells contain intermediate-level series and blue cells contain bottom-level series.

Figure 5 shows the full hierarchical structure capturing all components aggregated to form GDP using the production approach. This hierarchy has three levels. The most aggregated top-level of the hierarchy, which is level 0, comprises of the GDP. Level 1 comprises of GVA generated by three main activities and the component tax less subsidies. The bottom level has 50 series. Thus, in total this hierarchy has n = 55 series. These are summarised in Table 1.

Figure 6 shows some of the time series in the production approach. The top panel shows the most aggregated time series which is the total GDP as well as, Level 1 series namely: Agriculture, Industry and Services activities along with the component of taxes less subsidies on products (TaxLessSubsi). The bottom panel shows some selected series in the most disaggregated bottom level. Each series shows diverse dynamics with some series showing prominent seasonality while others simply showing a trend. This highlights the need to account for the different dynamics observed to produce a better model for forecasting each series.

Table 1: Number of time series per level of hierarchy

Hierarchy	Number of series
Level 0 (top-level)	1
Level 1	4
Level 4(bottom-level)	50
Total	55

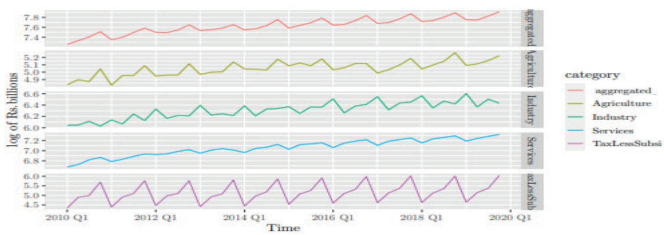
Figure 7 plots the same hierarchy as in Figure 6 but now in the annual frequency. As expected, series are now much smoother with a prominent trend, as seasonality is filtered out. Therefore, different temporal aggregation levels capture different features of the times series. Thus, these features could be extracted to improve forecast accuracy with temporal reconciliation. The cross-sectional reconciliation will enable to extract diverse dynamics of each of the series within the hierarchy. Moreover, using cross-temporal reconciliation will enable to extract these diverse signals from both cross-sectional and temporal dimensions to improve overall forecast accuracy.

4.2. Empirical application methodology

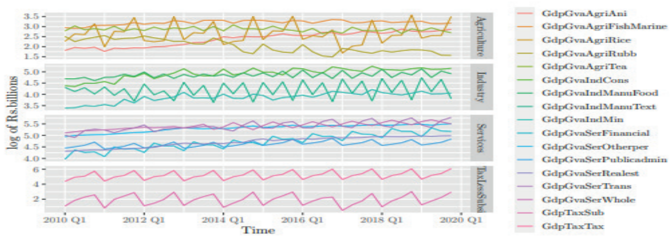
The data are quarterly from 2010-Q1 to 2019-Q4. As an only limited history is available, the last 8 quarters (2 years) will be considered as the test set to evaluate the forecast accuracy of competing approaches and to identify the potential of cross-temporal reconciliation to improve forecast accuracy.

The cross-temporal structure is not currently supported in an R package. Thus, I expand on the base implementations of cross-sectional hierarchical structure facilitated in the fpp3 package (Hyndman, Athanasopoulos, and O'Hara-Wild (2020)) for R (R Core Team (2020)). The code developed for this can be shared if requested.

Figure 6: Time plots for series from different levels of production approach hierarchy in quarterly frequency

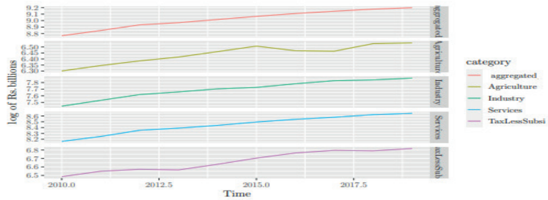


(a) *Top-level and level 1 series*

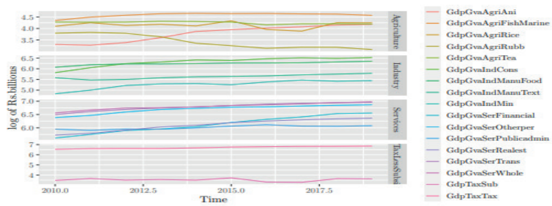


(b) *Selected bottom-level series in each main activity*

annual frequency



(a) *Top-level and level 1 series*



(b) *Selected bottom-level series in each main activity*

4.3. Forecasting models

The first step in forecast reconciliation is to obtain base forecasts for all the series in the hierarchy. The cross-sectional aggregation structure comprises 55 series and with 3 temporal aggregation levels. Thus, the cross-temporal hierarchy has $55 \times 3 = 165$ different series. To develop base forecasts for each of these I consider, two classes of forecasting models namely ExponenTial Smoothing (ETS) and AutoRegressive Integrated Moving Average (ARIMA) models as implemented in the ARIMA and ETS functions in the fable package (Hyndman, Athanasopoulos, and O'Hara-Wild (2020)) for R (R Core Team (2020)). The appropriate ETS and ARIMA models are chosen by minimising the Akaike Information Criterion (AIC) corrected for small sample sizes.

ETS models are commonly used in empirical research as they perform well with limited data and are relatively simple to build (Kourentzes and Athanasopoulos (2019)). ETS captures time series as the total of four fundamental components of a time series which are level, trend, seasonality, and the error process, where these components are combined additively or multiplicatively. Forecasts produced using exponential smoothing methods are weighted averages of past observations, with the weights decaying exponentially as the observations get older. ARIMA models aim to describe the autocorrelations in the data as opposed to ETS models which are based on a description of the trend and seasonality in the data. The Autoregressive component of the ARIMA model captures the habitual elements in the time series by regressing the variable of interest using a linear combination of past values of the variable after the series is difference as required to make it stationary. The moving average component regresses the variable of interest using a linear combination of past forecast errors of the stationary time series to smooth out the inherent noise in the data (Hyndman and Athanasopoulos (2018)).

For this application ETS forecasts were on average more accurate than the ARIMA forecasts (Refer AppendixA.2 TableA.4 to TableA.6) and using ARIMA models had minimal impact on conclusion of this study. Thus, I will only present the results obtained using ETS models. The results obtained for ARIMA models are given in AppendixA.3 TableA.7 to TableA.9.

Apart from these univariate models, other sophisticated multivariate models such as VAR models or indicator based regression type models can also be used for specific series to generate these base forecasts as its completely flexible and independent of the reconciliation methodology which is an advantage of forecast reconciliation.

The base forecasts do not adhere to the aggregation constraints in the cross-temporal hierarchy, and they also do not consider information available in other temporal or cross- sectional aggregation levels. Hence, cross-temporal coherent forecasts are generated reconciling the base forecasts as per the reconciliation Equation 3. The cross-temporal summation matrix was compiled according to the process explained in Section 3.11. The cross-sectional GDP hierarchy as summarised in Table1 has n = 55 series in total with m = 50 bottom-level series. Thus, the cross-sectional summation matrix is of order 55 × 50. As the series are observed in quarterly frequency the corresponding temporal summation matrix will be a matrix of order 7 × 4. Therefore, the corresponding cross- temporal summation matrix complied by taking the kronecker product of cross-sectional and temporal summation matrices will be a large matrix of order 385 × 200 with m = 200 bottom-level series.

The first set of cross-temporally coherent forecasts were generated using the bottom-up method which only use the information from the bottom-level of the hierarchy. This is referred to as BU in the results to follow and provides the natural benchmark to assess the benefit of generating forecasts at all aggregation levels (Athanasopoulos et al. (2017)). Three sets of alternative reconciled forecasts were also generated using OLS reconciliation (OLS), Structural scaling (Struc) and the Series Variance scaling (SVAR). The Variance scaling estimator and Shrinkage covariance estimator were also used but due to limited length of the series, forecasts error variances estimated for certain series were close to zero and it created problems in using these approaches. Reconciled forecasts were also computed using only cross-sectional reconciliation to compare the accuracy gain of using cross-temporal reconciliation.

4.4. Forecast accuracy evaluation

The forecast accuracy was evaluated using Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Several accuracy measures are considered in this empirical application to calibrate the results and to evaluate whether forecast performance depends on the accuracy measure used. For a particular series $\mathbf{j}$ in a particular aggregation level, for h-steps ahead forecast:

$$RMSE_j = \sqrt{\frac{1}{h}\sum_{i=1}^{h}(y_{ij} - \hat{y}_{ij})^2} \qquad (9)$$

$$MAE_j = \frac{1}{h}\sum_{i=1}^{h}\left|y_{ij} - \hat{y}_{ij}\right| \qquad (10)$$

$$MAPE_j = \frac{1}{h}\sum_{i=1}^{h}\left|\frac{100(y_{ij} - \hat{y}_{ij})}{y_{ij}}\right| \qquad (11)$$

where $\boldsymbol{y_{ij}}$ and $\widehat{\boldsymbol{y}}_{\boldsymbol{ij}}$ are actual and forecast values for the series $\mathbf{j}$ in the period $\mathbf{i}$. RMSE and MAE are the most commonly used accuracy measures, but they have the disadvantage of being scale dependent. However, they are useful in evaluating different methods applied to the same data set. MAPE has an advantage of being independent of scale and frequently used to compare the forecast accuracy of different data sets (Hyndman and Athanasopoulos (2018)). There are also certain issues in MAPE such as being unidentified or infinite if $\boldsymbol{y_{ij}}$ is zero or close to zero, assuming a meaningful zero and imposing a heavier penalty on positive errors than on negative errors (Hyndman and Koehler (2006)). However, in this application $\boldsymbol{y_{ij}}$ has a meaningful zero and it is not close to zero. Further, over estimation of growth may be more harmful than under estimation so imposing heavier penalty on positive errors can be justifiable. The summary accuracy measures in the tables that follow are the arithmetic mean of these accuracy measures calculated for each of the time series in the dimension considered.

It is common in the forecasting literature to express the accuracy measures in terms of a skill score (Wheatcroft (2019)), which is defined as,

$$skill\ score = \frac{A_f - A_r}{A_p - A_r} \qquad (12)$$

where A_p is the value of the accuracy measure if the outcome is known perfectly and A_f and A_r are the values of the accuracy measure using the method of interest and reference method, respectively. A_p is zero for the forecast accuracy measures considered in this application and incoherent base forecasts are taken as the reference forecasting method. Skill score can be interpreted as the proportional increase in accuracy of the forecasting method of interest compared to base forecasts. Thus, if the skill score is positive, it represents an improvement in forecasting accuracy over the base forecasts while negative values represent a deterioration. The summary measures in the tables that follow are the skill scores calculated based on arithmetic mean of the accuracy measures in the dimensions considered.

4.5. Results

Table 2 summarises the skill scores calculated based on average MAPE of the all cross- sectional series in the temporal dimension considered, where MAPE was computed based on forecasts up to and including the forecast horizon h. The results are presented for the complete hierarchy, bottom-level series, and top-level series (i.e., GDP) separately. Furthermore, results are presented for each temporal aggregation level (i.e., annual, semi-annual, and quarterly) separately together with an average measure across all temporal aggregation levels. The incoherent base forecasts were taken as the reference method. The Table 2 summaries the resulting skill scores of coherent forecasts obtained from the classical method bottom-up and the reconciliation methods. It should be noted that for cross- sectional reconciliation VAR referred to Variance scaling (Refer Section 3.5) and in cross- temporal reconciliation SVAR refers to Series variance scaling (Refer Section 3.11). The measures are summarised for cross-sectional and cross-temporal reconciliation separately to evaluate the accuracy gains of using cross-temporal reconciliation. The colored cells show the best performing method in each row (i.e., the temporal aggregation level). The darker the shade, the higher the improvement across the temporal aggregation levels. Skill scores calculated based on MAE and RMSE are given in Appendix A.1 Table A.2 and Table A.3. The conclusion based on these measures was also similar to that of MAPE.

The first panel refers to results summarised over all series, the second panel refers to top-level GDP series, and the last panel refers to the bottom-level. Reported figures are skill scores computed based on average MAPE over the entire test set of h=1 to 8. A positive (negative) entry shows a relative improvement (loss) over the base incoherent forecasts. Skill scores are summarized for cross-sectional and cross-temporal reconciliation separately to evaluate the accuracy gains of using cross-temporal reconciliation. The coloured cells show the best performing method in each row (i.e. the temporal aggregation level). The darker the shade, the higher the improvement across the temporal aggregation levels. Skill scores calculated based on MAE and RMSE are given in Appendix A.1 TableA.2 and Table A.3. The conclusion based on these measures was also similar to that of MAPE.

Table 2: Skill scores for point forecasts from alternative methods (with reference to incoherent base forecasts) using MAPE for Sri Lankan production approach

All-levels									
		Cross-Sectional			Cross-Temporal				
Temporal level	h	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	0.00	-0.42	-0.11	-0.01	0.13	-0.24	0.08	**0.17**
Semi-annual	4	0.02	-1.86	-1.24	0.02	0.11	-0.44	-0.13	**0.16**
Quarterly	8	0.01	-2.00	-0.99	0.00	0.01	-0.72	-0.37	**0.04**
Average		0.01	-1.47	-0.80	0.00	0.08	-0.48	-0.15	**0.12**
Top-level									
		Cross-Sectional				Cross-Temporal			
Temporal level	h	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	-0.05	0.00	-0.02	-0.04	0.88	0.57	0.76	**0.94**
Semi-annual	4	0.72	-0.02	0.37	0.52	0.75	0.18	0.54	**0.89**
Quarterly	8	-0.40	0.01	0.16	0.17	-0.40	-1.55	-0.54	**0.28**
Average		0.16	-0.01	0.12	0.15	0.71	0.23	0.56	**0.86**
Bottom-level									
		Cross-Sectional				Cross-Temporal			
Temporal level	h	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	0.00	-0.44	-0.12	-0.01	0.10	-0.27	0.05	**0.15**
Semi-annual	4	0.00	-1.95	-1.30	0.01	0.09	-0.47	-0.15	**0.14**
Quarterly	8	0.00	-2.06	-1.03	-0.01	0.00	-0.74	-0.39	**0.03**
Average		0.00	-1.53	-0.85	0.00	0.06	-0.51	-0.18	**0.11**

First, we compare cross-sectional reconciliation with cross-temporal reconciliation. It can be clearly seen from the Table 2, that in general using cross-temporal reconciliation has improved the forecast accuracy in all the cross- sectional and temporal levels considered irrespective of the reconciliation method. It shows that extracting and sharing information from different temporal aggregation levels to supplement the signals extracted from the cross-sectional hierarchy improves the forecast accuracy of all the reconciliation methods considered. As shown in Figure 5 and Figure 6 the seasonal component of the series dominates at quarterly frequency, possibly concealing the trend when it comes to model selection and estimation. At the annual level trends become dominant but estimation efficiency will be low due to limited sample size. Therefore, using temporal aggregation with cross-sectional aggregation will extract seasonal information and estimation efficiency to annual level and extract the trend information from annual level to the quarterly level. Thus, cross-temporal reconciliation gives a better view of the data in different angles which allows to bring in more information and improve the overall forecast accuracy.

The strength of cross-temporal reconciliation is not limited to accuracy gains. Another gain is the cross-temporally coherent forecasts which align the decision making and provide transparency within the organization. The short term view will align with the long term view while the disaggregated activity level forecasts will align with the country level GDP forecasts. This will facilitate consistent, transparent and align policy implementation.

It is interesting to note that, although all cross-temporal reconciliation alternatives perform better than cross-sectional reconciliation, cross-temporal SVAR reconciliation forecasts are consistently the most accurate in every scenario considered. Further, all the skill scores of SVAR forecasts are positive indicating that these are more accurate compared to the incoherent base forecasts, which is taken as the reference method. In general, OLS and Struc have failed to perform better than the conventional bottom-up method and even worse than the incoherent base forecasts for this application. However, it is noteworthy to highlight that accuracy gains are positive for OLS and Struc if performance is evaluated based on MAE and RMSE (refer AppendixA.1 TableA.2 and TableA.3). According to Table 2 the accuracy of OLS and Struc based on MAPE are worse at the bottom-level series, this was not evident in the skill scores based on RMSE and MAE. Therefore, there is some indication that OLS and Struc are performing relatively poorly at some low base series in the bottom level which result in overall loss in accuracy when evaluated based on MAPE.

The results are also disaggregated to top-level and bottom-level of the hierarchy for further investigation. These results are presented in the 2nd and the 3rd panels of Table 2. The accuracy gains of the top-level are higher than the bottom level. This indicates that additional information received at the top-level from the bottom level is arguably more influential than the additional information received at the bottom level.

5. Conclusion

This study investigates a direct cross-temporal hierarchical forecasting approach specifically in a macroeconomic setting relating to the forecasting of GDP. The main aim was to produce a set of forecasts which are cross-temporally coherent so that it will facilitate aligned policy decisions directed towards a one number forecast. This study proposes a direct approach in combining cross-sectional reconciliation and temporal reconciliation to get the maximum information available in the hierarchical structure as an alternative method to the two-step approach introduced by Kourentzes and Athanasopoulos (2019).

The results of the empirical applications revealed that cross-temporal reconciliation can further improve the forecasting accuracy obtained through cross-sectional reconciliation. This can be attributable to the valuable information provided by temporal hierarchies within the cross-sectional structure. As highlighted by Athanasopoulos et al. (2017) the source of forecast improvement in using temporal hierarchies is that it can strengthen the signal to noise ratio and reduced outlier effect at the aggregated lower frequencies of the time series, while mitigating loss of information and estimation efficiency at higher frequencies. Accuracy gains are greater for the top-level single series GDP compared to the bottom-level series. In addition, gains received at the lower frequencies are higher than the gains received at the higher frequencies.

Evaluation of alternative reconciliation methods revealed that cross-temporal SVAR, which is the series variance scaling reconciliation method yields the highest improvement in forecast accuracy in forecasting the Sri Lankan GDP.

Cross-temporal reconciliation aligns decisions within an organization towards one number. Temporal reconciliation aligns short term forecasts with more strategic long term forecasts. Cross-sectional reconciliation will align the view of the decision makers at different levels within the hierarchy. This is possible as reconciliation methods are model free, so judgmental forecasts produced at strategic levels can also be combined with data driven forecasts at more operational bottom-level in a transparent data driven method. It should be highlighted that this feature is not available with the classical bottom-up method. Furthermore, this will facilitate the alignment of the overall policy direction. This is very important specifically in GDP forecasting as policy decisions should be taken to direct the country towards one direction. To achieve this objective, short-term forecasts should align with long term forecasts. In addition, forecasts of the disaggregated economic activities should also align with the overall GDP forecasts.

In developing cross-sectional forecasts within the GDP hierarchy, reconciliation methods provide the benefit of using different models for different scenarios as the concept is independent of models used. This gives the opportunity to combine different specialised or in other words judgmental forecasts for certain economic activities with data driven sophisticated forecasting models. This is an advantage as for some disaggregated economic activities and at lower frequencies, availability of data or indicator variables will be limited to develop multivariate models. This ability to reconcile different views in a transparent method to enhance efficiency in managerial decision making is the main outcome of this cross-temporal reconciliation approach. In addition, the concept of forecast reconciliation involves forecasting GDP through disaggregated economic activities. This has an additional benefit over direct GDP forecasting which is commonly used in GDP forecasting literature as it has the ability of identifying economic activities which contributed to the overall projected GDP growth. Thus, policymakers can identify any issues at the bottom-levels and design specialised policies to address them.

References

Athanasopoulos, G, RA Ahmed, and RJ Hyndman (2009). "Hierarchical forecasts for Australian domestic tourism." *International Journal of Forecasting,* vol. **25** no.1, pp.146–166.

Athanasopoulos, G, P Gamakumara, A Panagiotelis, RJ Hyndman, and M Affan (2020). "Hierarchical forecasting". *Macroeconomic Forecasting in the Era of Big Data*, edited by Peter Fuleky. Springer, pp.689–719.

Athanasopoulos, G, RJ Hyndman, N Kourentzes, and F Petropoulos (2017). "Forecasting with temporal hierarchies." *European Journal of Operational Research.* Vol. **262**, no.1, pp. 60–74.

Australian Bureau of Statistics (2012). *Australian System of National Accounts: Concepts, Sources and Methods*. Vol. Cat 5216.0.

Barhoumi, K, O Darné, L Ferrara, and B Pluyaud (2012). "Monthly GDP forecasting using bridge models: Application for the French economy." *Bulletin of Economic Research,* vol. **64**, s53–s70.

Dunn, D, W Williams, and T DeChaine (1976). "Aggregate versus subaggregate models in local area forecasting." *Journal of the American Statistical Association,* vol. **71** no.353, pp. 68–71.

Esteves, PS (2013). "Direct vs bottom–up approach when forecasting GDP: Reconciling literature results with institutional practice." *Economic Modelling,* vol. **33**, pp. 416–420.

European Commission, International Monetary Fund, Organisation for Economic Co- operation and Development, United Nations, and World Bank (2009). *System of national accounts 2008*.

Fair, RC and RJ Shiller (1990). "Comparing information in forecasts from econometric models." *The American Economic Review*, pp. 375–389.

Fliedner, G (1999). "An investigation of aggregate variable time series forecast strategies with specific sub aggregate time series statistical correlation." *Computers & operations research,* vol. **26,** no.10-11, pp. 1133–1149.

Gross, CW and JE Sohl (1990). "Disaggregation methods to expedite product line forecasting." *Journal of forecasting,* vol. **9,** no.3, pp. 233–254.

Grunfeld, Y and Z Griliches (1960). "Is aggregation necessarily bad? ",*The Review of Economics and Statistics*, pp. 1–13.

Hahn, E and F Skudelny (2008). "Early estimates of euro area real GDP growth: a bottom up approach from the production side." European Central Bank.

Heinisch, K and R Scheufele (2018). "Bottom-up or direct? Forecasting German GDP in a data-rich environment." *Empirical Economics,* Vol. **54**, no.2, pp.705–745.

Higgins, PC (2014). "GDPNow: A Model for GDP'Nowcasting." Federal Reserve Bank of Atlanta.

Hyndman, R, G Athanasopoulos, and M O'Hara-Wild (2020). *fpp3:Forecasting: Principles and Practice(3rd Edition).* R Package version 0.3.

Hyndman, RJ, RA Ahmed, G Athanasopoulos, and HL Shang (2011). "Optimal combination forecasts for hierarchical time series. "*Computational statistics & data analysis,* vol. **55**, no.9, pp. 2579– 2589.

Hyndman, RJ and G Athanasopoulos (2018). *Forecasting: principles and practice*. OTexts.

Hyndman, RJ and AB Koehler (2006). "Another look at measures of forecast accuracy." *International journal of forecasting,* vol. **22**, no.4, pp. 679–688.

Hyndman, RJ, AJ Lee, and E Wang (2016). "Fast computation of reconciled forecasts for hierarchical and grouped time series." *Computational statistics & data analysis,* vol. **97**, pp. 16–32.

Jelizarow, M and V Guillemot (2015). *SHIP:SHrinkage covariance Incorporating Prior knowledge*. R Package version 1.0.2.

Kourentzes, N and G Athanasopoulos (2019). :"Cross-temporal coherent forecasts for Australian tourism." *Annals of Tourism Research,* vol.**75**, pp.393–409.

Kourentzes, N, D Barrow, and F Petropoulos (2019). "Another look at forecast selection and combination: Evidence from forecast pooling." *International Journal of Production Economics,* vol. **209**, pp.226–235.

Park, M and M Nassar (2014). "Variational Bayesian inference for forecasting hierarchical time series." In: *ICML Workshop*. Citeseer.

Pennings, CL and J van Dalen (2017). "Integrated hierarchical forecasting." *European Journal of Operational Research,* vol. **263**, no.2, pp. 412–418.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Schäfer, J and K Strimmer (2005). "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics." *Statistical applications in genetics and molecular biology, vol.* **4**, no.1.

Spiliotis, E, F Petropoulos, N Kourentzes, and V Assimakopoulos (2020). "Cross-temporal aggregation: Improving the forecast accuracy of hierarchical electricity consumption." *Applied Energy*, vol. **261**, no. 114339.

Syntetos, AA, Z Babai, JE Boylan, S Kolassa, and K Nikolopoulos (2016). "Supply chain forecasting: Theory, practice, their gap and the future." *European Journal of Operational Research,* vol. **252**, no.1, pp. 1–26.

Viet, VQ (2009). "GDP by production approach: A general introduction with emphasis on an integrated economic data collection framework." *Published as United Nations' training material for Statistical Capacity Development in China and Other Developing Countries in Asia*, pp. 1–137.

Weatherford, LR, SE Kimes, and DA Scott (2001). "Forecasting for hotel revenue management: Testing aggregation against disaggregation." *Cornell hotel and restaurant administration quarterly,* vol.**42**, no.4, pp. 53–64.

Wheatcroft, E (2019). Interpreting the skill score form of forecast performance metrics. *International Journal of Forecasting,* vol. **35**, no.2, pp.573–579.

Wickramasuriya, SL, G Athanasopoulos, and RJ Hyndman (2019). Optimal forecast recon- ciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association,* vol. **114**, no.526, pp. 804–819.

Zotteri, G and M Kalchschmidt (2007). "A model for selecting the appropriate level of aggregation in forecasting processes." *International Journal of Production Economics,* vol. 108, no. 1- 2, pp.74–83.

Zotteri, G, M Kalchschmidt, and F Caniato (2005). The impact of aggregation level on forecasting performance. *International Journal of Production Economics,* vol. **93**, pp.479–491.

Appendices

Table A.1: Detailed Economic activities in Production Approach

Variables	Detailed economic activities	Main Activity
GdpGvaAgriCereal	Growing of Cereals (except rice)	Agriculture, Forestry and Fishing
GdpGvaAgriRice	Growing of Rice	Agriculture, Forestry and Fishing
GdpGvaAgriVege	Growing of Vegetables	Agriculture, Forestry and Fishing
GdpGvaAgriSugar	Growing of Sugar cane, tobacco and other non-perennial crops	Agriculture, Forestry and Fishing
GdpGvaAgrFruits	Growing of fruits	Agriculture, Forestry and Fishing
GdpGvaAgriOle	Growing of Oleaginous Fruits (Coconut, king coconut, Oil palm)	Agriculture, Forestry and Fishing
GdpGvaAgriTea	Growing of Tea (Green leaves)	Agriculture, Forestry and Fishing
GdpGvaAgriBeve	Growing of other beverage crops (Coffee, Cocoa	Agriculture, Forestry and Fishing
GdpGvaAgriSpice	Growing of spices, aromatic, drug and pharmaceutical crops	Agriculture, Forestry and Fishing
GdpGvaAgriRubb	Growing of rubber	Agriculture, Forestry and Fishing
GdpGvaAgriPere	Growing of other perennial crops	Agriculture, Forestry and Fishing
GdpGvaAgriAni	Animal Production	Agriculture, Forestry and Fishing
GdpGvaAgriPlant	Plant propagation and agricultural supporting activities	Agriculture, Forestry and Fishing
GdpGvaAgriForest	Forestry and Logging	Agriculture, Forestry and Fishing
GdpGvaAgriFishMarine	Marine fishing and Marine Aquaculture	Agriculture, Forestry and Fishing
GdpGvaAgriFishInland	Fresh water fishing and Fresh water Aquaculture	Agriculture, Forestry and Fishing
GdpGvaIndMin	Mining and quarrying	Industry
GdpGvaIndManuFood	Manufacture of food, beverages and Tobacco products	Industry
GdpGvaIndManuText	Manufacture of textiles, wearing apparel and leather related products	Industry
GdpGvaIndManuWood	Manufacture of wood and of products of wood and cork, except furniture	Industry
GdpGvaIndManuPaper	Manufacture of paper products, printing and reproduction of media products	Industry
GdpGvaIndManuCoke	Manufacture of coke and refined petroleum products	Industry
GdpGvaIndManuChemi	Manufacture of chemical products and basic pharmaceutical products	Industry
GdpGvaIndManuRubb	Manufacture of rubber and plastic products	Industry
GdpGvaIndManuNonmet	Manufacture of other non- metallic mineral products	Industry
GdpGvaIndManuMetal	Manufacture of basic metals and fabricated metal products	Industry
GdpGvaIndManuMachin	Manufacture of machinery and equipment	Industry
GdpGvaIndManuFurni	Manufacture of furniture	Industry
GdpGvaIndManuOther	Other manufacturing, and Repair and installation of machinery and equipment	Industry
GdpGvaIndElectri	Electricity, gas, steam and air conditioning supply	Industry
GdpGvaIndWater	Water collection, treatment and supply	Industry
GdpGvaIndSewerage	Sewerage, Waste, treatment and disposal activities	Industry
GdpGvaIndCons	Construction	Industry
GdpGvaSerWhole	Wholesale and retail trade	Services
GdpGvaSerTrans	Transportation of goods and passengers including Warehousing	Services
GdpGvaSerPostal	Postal and courier activities	Services
GdpGvaSerAccom	Accommodation, Food and beverage service activities	Services
GdpGvaSerProgram	Programming and broadcasting activities and audio video productions	Services
GdpGvaSerTele	Telecommunication	Services
GdpGvaSerIT	IT programming consultancy and related activities	Services
GdpGvaSerFinancial	Financial Service activities and auxiliary financial services	Services
GdpGvaSerInsurance	Insurance, reinsurance and pension funding	Services
GdpGvaSerRealest	Real estate activities, Including Ownership of dwelling	Services
GdpGvaSerProfess	Professional services	Services
GdpGvaSerPublicadmin	Public administration and defense; compulsory social security	Services
GdpGvaSerEdu	Education	Services
GdpGvaSerHealth	Human health activities, Residential care and social work activities	Services
GdpGvaSerOtherper	Other personal service activities	Services

Table A.2: ***Skill scores for point forecasts from alternative methods (with reference to incoherent base forecasts) using RMSE for Sri Lankan production approach GDP***

All-levels

		Cross-Sectional				Cross-Temporal			
Temporal level	h	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	-0.03	0.01	0.00	-0.02	0.53	0.46	0.57	**0.61**
Semi-annual	4	0.34	-0.08	0.16	0.19	0.36	0.21	0.38	**0.45**
Quarterly	8	-0.02	-0.08	0.04	0.13	-0.02	-0.15	0.07	**0.15**
Average		0.07	-0.02	0.05	0.05	0.43	0.34	0.47	**0.52**

Top-level

		Cross-Sectional				Cross-Temporal			
Temporal level	h	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	-0.05	0.00	-0.02	-0.04	0.86	0.57	0.77	**0.94**
Semi-annual	4	0.74	0.00	0.38	0.52	0.73	0.25	0.59	**0.88**
Quarterly	8	-0.55	0.01	0.08	0.08	-0.55	-1.33	-0.46	**0.11**
Average		0.10	0.00	0.07	0.09	0.78	0.44	0.68	**0.89**

Bottom-level

		Cross-Sectional				Cross-Temporal			
Temporal level	h	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	0.00	0.02	0.02	0.00	0.12	**0.31**	0.30	0.20
Semi-annual	4	0.00	-0.22	-0.09	-0.10	0.02	0.11	**0.14**	0.09
Quarterly	8	0.00	-0.16	-0.08	0.06	0.00	0.04	**0.09**	**0.09**
Average		0.00	-0.07	-0.03	-0.02	0.07	0.21	**0.22**	0.15

The first panel refers to results summarised over all series, the second panel refers to top-level GDP series, and the last panel refers to the bottom level. Reported figures are skill scores computed based on average MAPE over the entire test set of h=1 to 8. A positive (negative) entry shows a relative improvement (loss) over the base incoherent forecasts.

Table A.3: *Skill scores for point forecasts from alternative methods (with reference to incoherent base forecasts) using MAE for Sri Lankan production approach GDP*

All-levels									
Temporal level	h	Cross-Sectional				Cross-Temporal			
		BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	-0.03	0.01	0.00	-0.02	0.55	0.46	0.58	**0.62**
Semi-annual	4	0.34	-0.07	0.21	0.18	0.37	0.20	0.38	**0.45**
Quarterly	8	0.03	-0.10	0.13	0.04	0.03	-0.16	0.09	**0.17**
Average		0.07	-0.02	0.06	0.04	0.46	0.34	0.48	**0.54**
Top-level									
Temporal level	h	Cross-Sectional				Cross-Temporal			
		BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	-0.05	0.00	-0.02	-0.04	0.88	0.57	0.76	**0.94**
Semi-annual	4	0.73	-0.01	0.37	0.52	0.76	0.18	0.55	**0.89**
Quarterly	8	-0.38	0.01	0.17	0.18	-0.38	-1.56	-0.55	**0.29**
Average		-0.62	-0.80	-0.67	-0.65	0.72	0.24	0.56	**0.86**
Bottom-level									
Temporal level	h	Cross-Sectional				Cross-Temporal			
		BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	0.00	0.03	0.02	0.01	0.13	**0.32**	0.31	0.21
Semi-annual	4	0.00	-0.21	-0.08	-0.07	0.01	**0.16**	**0.16**	0.10
Quarterly	8	0.00	-0.19	-0.10	0.04	0.00	0.06	**0.12**	0.09
Average		0.00	-0.07	-0.02	-0.01	0.08	0.24	**0.24**	0.16

The first panel refers to results summarised over all series, the second panel refers to top-level GDP series, and the last panel refers to the bottom level. Reported figures are skill scores computed based on average MAPE over the entire test set of h=1 to 8. A positive (negative) entry shows a relative improvement (loss) over the base incoherent forecasts.

Table A.4: *Average MAPE for Sri Lankan production approach GDP hierarchy*

			Cross-Sectional				Cross-Temporal			
Temporal level	h	Base	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
All-levels										
ETS										
Annual	2	6.89	6.86	9.78	7.63	6.95	6.00	8.52	6.33	5.69
Semi-annual	4	7.64	7.51	21.87	17.08	7.50	6.82	10.97	8.61	6.42
Quarterly	8	7.83	7.77	23.52	15.61	7.83	7.77	13.50	10.73	7.51
Average		7.45	7.38	18.39	13.44	7.42	6.86	11.00	8.56	**6.54**
ARIMA										
Annual	2	9.95	10.14	20.49	17.64	9.54	6.25	15.18	11.08	6.20
Semi-annual	4	7.94	7.80	13.57	10.47	8.07	6.85	16.78	12.13	6.91
Quarterly	8	7.89	7.72	18.85	14.94	7.93	7.72	17.65	13.06	7.82
Average		8.59	8.56	17.64	14.35	8.51	**6.94**	16.54	12.09	6.98
Top-level										
ETS										
Annual	2	4.28	4.49	4.29	4.35	4.46	0.51	1.85	1.03	0.25
Semi-annual	4	2.25	0.62	2.29	1.42	1.09	0.57	1.86	1.03	0.25
Quarterly	8	0.73	1.02	0.73	0.62	0.61	1.02	1.87	1.12	0.53
Average		2.42	2.05	2.44	2.13	2.05	0.70	1.86	1.06	**0.34**
ARIMA										
Annual	2	1.88	5.90	1.95	3.24	3.19	0.61	0.15	0.24	0.71
Semi-annual	4	2.14	1.32	2.09	1.80	1.67	0.61	0.31	0.35	0.70
Quarterly	8	2.62	0.80	2.55	1.91	1.75	0.80	0.71	0.67	0.82
Average		2.21	2.67	2.20	2.32	2.20	0.67	**0.39**	0.42	0.75
Bottom-levels										
ETS										
Annual	2	7.19	7.19	10.36	8.02	7.27	6.48	9.15	6.82	6.13
Semi-annual	4	8.06	8.06	23.75	18.58	8.02	7.36	11.82	9.31	6.91
Quarterly	8	8.37	8.37	25.64	16.99	8.43	8.37	14.58	11.62	8.08
Average		7.87	7.87	19.92	14.53	7.90	7.40	11.85	9.25	**7.04**
ARIMA										
Annual	2	10.57	10.57	22.19	19.10	10.13	6.75	16.48	12.05	6.64
Semi-annual	4	8.36	8.36	14.56	11.23	8.59	7.41	18.22	13.20	7.41
Quarterly	8	8.34	8.34	20.41	16.18	8.48	8.34	19.16	14.20	8.39
Average		9.09	9.09	19.05	15.50	9.07	7.50	17.95	13.15	**7.48**

The 1st and 2nd panels refers to results summarised over all series for ETS and ARIMA models respectively, the 3rd and 4th panel refers to results summarised over the top-level GDP series for ETS and ARIMA models respectively, and the 5th and 6th panels refers to the bottom level results summary for ETS and ARIMA models respectively.Reported figures are average MAPE of the all cross-sectional series in the level considered, where MAPE was computed based on forecasts up to and including the forecast horizon h.Bold figures are the lowest error in the panel.

Table A.5: *Average RMSE for Sri Lankan production approach GDP hierarchy*

Temporal level	h	Base	Cross-Sectional				Cross-Temporal			
All-levels										
ETS										
Temporal level	h	Base	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	25.62	26.29	25.42	25.66	26.15	12.02	13.73	10.90	9.87
Semi-annual	4	10.37	6.85	11.15	8.67	8.35	6.68	8.17	6.43	5.75
Quarterly	8	4.17	4.25	4.48	4.01	3.64	4.25	4.78	3.88	3.55
Average		13.39	12.46	13.69	12.78	12.71	7.65	8.89	7.07	**6.39**
ARIMA										
Temporal level	h	Base	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	22.62	33.42	22.33	22.74	24.34	10.17	13.15	10.23	11.54
Semi-annual	4	10.44	8.60	10.63	9.70	9.64	5.76	7.29	5.83	6.43
Quarterly	8	5.21	3.65	5.45	4.67	4.74	3.65	4.35	3.65	3.91
Average		12.76	15.23	12.80	12.37	12.91	**6.53**	8.26	6.57	7.29
Top-level										
ETS										
Temporal level	h	Base	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	434.14	454.23	434.81	440.83	450.63	59.39	184.76	101.34	27.63
Semi-annual	4	122.82	31.81	122.31	75.99	58.35	33.00	92.38	50.89	14.59
Quarterly	8	21.03	32.50	20.76	19.35	19.45	32.50	48.93	30.71	18.64
Average		192.66	172.85	192.62	178.72	176.14	41.63	108.69	60.98	**20.29**
ARIMA										
Temporal level	h	Base	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	198.75	591.06	204.50	329.39	322.34	63.43	14.39	28.67	78.80
Semi-annual	4	118.06	72.14	115.69	99.93	92.43	31.94	16.47	19.17	41.14
Quarterly	8	71.52	25.07	70.02	53.62	49.53	25.07	19.59	19.63	26.88
Average		129.45	229.42	130.07	160.98	154.77	40.15	**16.82**	22.49	48.94
Bottom-levels										
ETS										
Temporal level	h	Base	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	10.67	10.67	10.44	10.48	10.64	9.38	7.40	7.48	8.51
Semi-annual	4	5.27	5.27	6.41	5.76	5.79	5.16	4.67	4.52	4.78
Quarterly	8	2.95	2.95	3.44	3.19	2.78	2.95	2.84	2.70	2.69
Average		6.30	6.30	6.76	6.48	6.40	5.83	4.97	**4.90**	5.33
ARIMA										
Temporal level	h	Base	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	13.08	13.08	13.01	11.78	13.12	7.78	9.77	8.43	8.36
Semi-annual	4	6.10	6.10	6.36	6.18	6.35	4.45	5.37	4.73	4.73
Quarterly	8	2.65	2.65	2.97	2.80	2.98	2.65	3.04	2.74	2.79
Average		7.28	7.28	7.45	6.92	7.48	**4.96**	6.06	5.30	5.29

The 1st and 2nd panels refers to results summarised over all series for ETS and ARIMA models respectively, the 3rd and 4th panel refers to results summarised over the top-level GDP series for ETS and ARIMA models respectively, and the 5th and 6th panels refers to the bottom level results summary for ETS and ARIMA models respectively.Reported figures are average RMSE of the all cross-sectional series in the level considered, where RMSE was computed based on forecasts up to and including the forecast horizon h.Bold figures are the lowest error in the panel.

Table A.6: ***Average MAE for Sri Lankan production approach GDP hierarchy***

			Cross-Sectional				Cross-Temporal			
All-levels										
ETS										
Temporal level	h	Base	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	24.68	25.34	24.43	24.68	25.17	11.04	13.25	10.42	9.39
Semi-annual	4	9.31	6.12	9.98	7.39	7.64	5.87	7.41	5.82	5.08
Quarterly	8	3.56	3.45	3.90	3.09	3.42	3.45	4.13	3.25	2.95
Average		12.52	11.64	12.77	11.72	12.07	6.78	8.26	6.50	**5.81**
ARIMA										
Temporal level	h	Base	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	21.63	32.48	21.31	21.83	23.57	9.47	12.75	9.65	10.90
Semi-annual	4	9.41	7.65	9.62	8.72	8.66	4.99	6.70	5.18	5.75
Quarterly	8	4.36	2.94	4.65	3.90	3.92	2.94	3.68	3.03	3.23
Average		11.80	14.36	11.86	11.48	12.05	**5.80**	7.71	5.95	6.63
Top-level										
ETS										
Temporal level	h	Base	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	419.87	440.61	420.56	426.79	436.90	50.31	181.44	100.79	24.67
Semi-annual	4	111.23	29.59	112.71	69.86	53.49	26.95	90.72	50.39	12.33
Quarterly	8	17.69	24.34	17.60	14.75	14.55	24.34	45.36	27.36	12.59
Average		101.87	164.85	183.62	170.46	168.31	28.79	77.76	44.43	**14.24**
ARIMA										
Temporal level	h	Base	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	184.68	578.53	191.31	317.83	313.19	59.93	14.36	23.49	70.22
Semi-annual	4	105.67	65.16	103.24	89.20	82.66	29.97	14.80	17.23	35.11
Quarterly	8	64.52	19.36	62.93	47.18	43.11	19.36	17.32	16.38	20.51
Average		118.29	221.01	119.16	151.40	146.32	36.42	**15.50**	19.03	41.95
Bottom-levels										
ETS										
Temporal level	h	Base	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	10.25	10.25	9.99	10.04	10.17	8.90	7.02	7.12	8.10
Semi-annual	4	4.74	4.74	5.74	5.14	5.09	4.69	3.99	3.99	4.27
Quarterly	8	2.52	2.52	3.00	2.78	2.41	2.52	2.37	2.23	2.29
Average		5.84	5.84	6.24	5.98	5.89	5.37	4.46	**4.45**	4.89
ARIMA										
Temporal level	h	Base	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	12.59	12.59	12.49	11.30	12.69	7.20	9.40	8.03	7.86
Semi-annual	4	5.59	5.59	5.87	5.69	5.84	3.86	4.88	4.29	4.26
Quarterly	8	2.19	2.19	2.53	2.36	2.49	2.19	2.58	2.30	2.34
Average		6.79	6.79	6.96	6.45	7.01	**4.42**	5.62	4.87	4.82

The 1st and 2nd panels refers to results summarised over all series for ETS and ARIMA models respectively, the 3rd and 4th panel refers to results summarised over the top-level GDP series for ETS and ARIMA models respectively, and the 5th and 6th panels refers to the bottom level results summary for ETS and ARIMA models respectively.Reported figures are average MAE of the all cross-sectional series in the level considered, where MAE was computed based on forecasts up to and including the forecast horizon h. Bold figures are the lowest error in the panel.

Table A.7: Skill scores for point forecasts from alternative methods (with reference to incoherent base forecasts) using MAPE for Sri Lankan production approach GDP with ARIMA models

		All-levels							
		Cross-Sectional			Cross-Temporal				
Temporal level	h	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	-0.02	-1.06	-0.77	0.04	0.37	-0.53	-0.11	**0.38**
Semi-annual	4	0.02	-0.71	-0.32	-0.02	**0.14**	-1.11	-0.53	0.13
Quarterly	8	0.02	-1.39	-0.89	-0.01	**0.02**	-1.24	-0.66	0.01
Average		0.00	-1.05	-0.67	0.01	**0.19**	-0.92	-0.41	**0.19**
		Top-level							
		Cross-Sectional			Cross-Temporal				
Temporal level	h	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	-2.14	-0.04	-0.72	-0.70	0.68	**0.92**	0.87	0.62
Semi-annual	4	0.38	0.02	0.16	0.22	0.71	**0.85**	0.84	0.67
Quarterly	8	0.69	0.03	0.27	0.33	0.69	**0.73**	**0.74**	0.69
Average		-0.21	0.01	-0.05	0.00	0.70	**0.82**	0.81	0.66
		Bottom-level							
		Cross-Sectional			Cross-Temporal				
Temporal level	h	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	0.00	-1.10	-0.81	0.04	0.36	-0.56	-0.14	**0.37**
Semi-annual	4	0.00	-0.74	-0.34	-0.03	**0.11**	-1.18	-0.58	**0.11**
Quarterly	8	0.00	-1.45	-0.94	-0.02	**0.00**	-1.30	-0.70	-0.01
Average		0.00	-1.10	-0.71	0.00	0.17	-0.98	-0.45	**0.18**

The first panel refers to results summarised over all series, the second panel refers to top-level GDP series, and the last panel refers to the bottom level. Reported figures are skill scores computed based on average MAPE over the entire test set of h=1 to 8. A positive (negative) entry shows a relative improvement (loss) over the base incoherent forecasts.

Table A.8: Skill scores for point forecasts from alternative methods (with reference to incoherent base forecasts) using RMSE for Sri Lankan production approach GDP with ARIMA models

All-levels									
		Cross-Sectional				Cross-Temporal			
Temporal level	h	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	-0.48	0.01	-0.01	-0.08	**0.55**	0.42	**0.55**	0.49
Semi-annual	4	0.18	-0.02	0.07	0.08	**0.45**	0.30	**0.44**	0.38
Quarterly	8	0.30	-0.05	0.10	0.09	**0.30**	0.17	**0.30**	0.25
Average		-0.19	0.00	0.03	-0.01	**0.49**	0.35	**0.49**	0.43

Top-level									
		Cross-Sectional				Cross-Temporal			
Temporal level	h	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	-1.97	-0.03	-0.66	-0.62	0.68	**0.93**	0.86	0.60
Semi-annual	4	0.39	0.02	0.15	0.22	0.73	**0.86**	0.84	0.65
Quarterly	8	0.65	0.02	0.25	0.31	0.65	**0.73**	0.73	0.62
Average		-0.77	0.00	-0.24	-0.20	0.69	**0.87**	0.83	0.62

Bottom-level									
		Cross-Sectional				Cross-Temporal			
Temporal level	h	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	0.00	0.01	0.10	0.00	**0.41**	0.25	0.36	0.36
Semi-annual	4	0.00	-0.04	-0.01	-0.04	**0.27**	0.12	0.22	0.22
Quarterly	8	0.00	-0.12	-0.06	-0.12	**0.00**	-0.15	-0.03	-0.05
Average		0.00	-0.02	0.05	-0.03	**0.32**	0.17	0.27	0.27

The first panel refers to results summarised over all series, the second panel refers to top-level GDP series, and the last panel refers to the bottom level. Reported figures are skill scores computed based on average MAPE over the entire test set of h=1 to 8. A positive (negative) entry shows a relative improvement (loss) over the base incoherent forecasts.

Table A.9: Skill scores for point forecasts from alternative methods (with reference to incoherent base forecasts) using MAE for Sri Lankan production approach GDP with ARIMA models

All-levels

		Cross-Sectional				Cross-Temporal			
Temporal level	h	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	-0.50	0.01	-0.01	-0.09	**0.56**	0.41	0.55	0.50
Semi-annual	4	0.19	-0.02	0.07	0.08	**0.47**	0.29	0.45	0.39
Quarterly	8	0.33	-0.07	0.11	0.10	**0.33**	0.16	0.31	0.26
Average		-0.22	-0.01	0.03	-0.02	**0.51**	0.35	0.50	0.44

Top-level

		Cross-Sectional				Cross-Temporal			
Temporal level	h	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	-2.13	-0.04	-0.72	-0.70	0.68	**0.92**	0.87	0.62
Semi-annual	4	0.38	0.02	0.16	0.22	0.72	**0.86**	0.84	0.67
Quarterly	8	0.70	0.02	0.27	0.33	0.70	**0.73**	0.75	0.68
Average		-0.87	-0.01	-0.28	-0.24	0.69	**0.87**	0.84	0.65

Bottom-level

		Cross-Sectional				Cross-Temporal			
Temporal level	h	BU	OLS	Struc	VAR	BU	OLS	Struc	SVAR
Annual	2	0.00	0.01	0.10	-0.01	**0.43**	0.25	0.36	0.38
Semi-annual	4	0.00	-0.05	-0.02	-0.04	**0.31**	0.13	0.23	0.24
Quarterly	8	0.00	-0.16	-0.08	-0.14	**0.00**	-0.18	-0.05	-0.07
Average		0.00	-0.03	0.05	-0.03	**0.35**	0.17	0.28	0.29

The first panel refers to results summarised over all series, the second panel refers to top-level GDP series, and the last panel refers to the bottom level. Reported figures are skill scores computed based on average MAPE over the entire test set of h=1 to 8. A positive (negative) entry shows a relative improvement (loss) over the base incoherent forecasts.