Macroeconomic forecasting with hierarchical time series

Mohamed Affan
(Student ID: 25550926)

ETF5550: Research Project (Semester 2, 2018)
Supervisors:
Dr Anastasios Panagiotelis & Associate Professor George Athanasopoulos

October 29, 2018

Abstract

Panagiotelis, Athanasopoulos, Hyndman, Jiang, and Vahid (2018) found that it is difficult to outperform a random walk with drift for forecasting Australian GDP using seasonally adjusted, real GDP data. Further inspection of Australian GDP and related data revealed that macroeconomic variables such as the components of GDP can be organized into a hierarchical structure.

Therefore, in this project, reconciliation approaches of Bottom-up, WLS, MinT(Sample) and MinT(Shrink) were used in an expanding window experiment, to test whether reconciliation methods improve forecasts, so that they outperform a naïve benchmark.

Using seasonally unadjusted, nominal GDP data, it was concluded that reconciliation approaches do, in fact, improve on forecasts of Australian GDP in most cases and that the naïve benchmark performs poorly.

Contents

1 Introduction

Forecasts of key macroeconomic variables, such as the Gross Domestic Product (GDP), are necessary and important inputs for decision makers both in the private sector and the public sector as it is used in government budget planning, central bank policy making and business decisions. Following studies done on forecasting the GDP and other macroeconomic variables of the United States, Panagiotelis et al. (2018) attempted to use multiple methods to forecast the GDP of Australia. The methods used include univariate benchmarks of naïve forecasts, and the AR model in addition to advanced approaches such as a dynamic factor model, ridge regression, least angle regression and a multivariate Bayesian VAR. They concluded that it is difficult to outperform a random walk with drift for seasonally adjusted, real GDP.

This motivated further inspection of the GDP and related data. According to the Australian Bureau of Statistics (2015), GDP is measured by three approaches in Australia - Production, Expenditure and Income. For each approach, GDP is constructed from disaggregate components. For example, in the Income approach, GDP is the sum of total factor income, taxes less subsidies and the statistical discrepancy calculated for the Income approach. Total factor income is in turn, the sum of the gross operating surplus, gross mixed income and the compensation of employees. Each of these can also be further disaggregated as shown in Figures 2-6. Due to the approaches GDP is measured and the hierarchical structure by which the collection of the related time series can be organized into, forecasting GDP leads naturally to considering forecasting these components and therefore, we enter the world of forecasting hierarchical time series.

It is clear when inspecting the collection of time series, that different time series have different properties - in terms of the trend and seasonality. As a result, the forecasting methods used will have to be able to take into account the varying time series properties and therefore, this paper focuses on using exponential smoothing (ETS) and ARIMA models to generate forecasts at all levels of aggregation independently (referred to as *base* forecasts).

It is common to produce disaggregated forecasts based on disaggregated time series, and it is desirable and often required for the forecasts to add up in the same way as the data. This ensures the forecasts mimic the properties of the real data. This is referred to as "coherent" forecasts. While a simple approach would be to ignore the aggregation constraints and the relationships between series and simply forecast all the series in a collection independently, it is unlikely that the resulting set of forecasts will be coherent unless extremely simple forecasting methods are used (Wickramasuriya, Athanasopoulos, & Hyndman, 2018). Therefore, in order to generate coherent forecasts when ETS and ARIMA models are used, reconciliation approaches have to be used. This paper focuses on the bottom-up approach, weighted least squares (WLS), MinT(Sample), and MinT(Shrink) - from which, the latter three approaches fall under the broader Minimum Trace (MinT) reconciliation approach.

The bottom-up approach is the simplest approach to forecasting hierarchical time series and involves forecasting only the most disaggregated series and adding the resultant forecasts to form forecasts of the various aggregated series. However, Wickramasuriya et al. (2018) noted that the bottom-up approach ignores the relationships between series, and performs poorly, especially on highly disaggregated data which tend to have a low signal-to-noise ratio.

Wickramasuriya et al. (2018) extended previous work in the field by framing the problem of obtaining coherent forecasts and taking into account the relationships between series in terms of finding a set of minimum variance unbiased estimates of future values of all time series across the entire collection, which they referred to as Minimum Trace (MinT) reconciliation. The positive definite covariance matrix of the h-step-ahead base forecast errors can be specified differently to obtain the approaches mentioned above - weighted least squares (WLS), MinT(Sample), and MinT(Shrink).

The weighted least squares (WLS) method was initially suggested by Hyndman, Lee, and Wang (2016), taking account of the variances on the diagonal of the variance-covariance matrix but ignoring the off-diagonal covariance elements. Additionally, as estimates of the variances are not readily available in practice, it was proposed that the variances of the base forecast errors be used. The theoretical justification for this proxy, however, was provided later by Wickramasuriya et al. (2018).

MinT(Sample) involves using the unrestricted sample covariance estimator. While it may be relatively simple to obtain, it may not be a good estimate when the number of series (m) is greater or of the same order as the number of observations (T). This leads to MinT(Shrink) which enables the limitation of MinT(Sample) to be overcome by using a shrinkage parameter introduced by Schäfer and Strimmer (2005).

In this paper, the above mentioned hierarchical time series forecasting/reconciliation approaches were used in an expanding window, to test whether reconciliations methods improve on forecasts, and whether forecasts could be improved to outperform a seasonal random walk with drift or a naïve benchmark. Other reconciliation methods such as the top-down or the middle-out approaches were not considered as Hyndman, Ahmed, Athanasopoulos, and Shang (2011) showed that any top-down method introduces bias into the reconciled forecasts at each disaggregation level even if the base forecasts are unbiased.

The rest of the paper is organized as follows. Section 2 provides a detailed explanation of the methodology used in the paper. Section 3 explains the data used, the issues that arose and the time series properties of the data. Section 4 explains the experiment conducted, while Section 5 provides the main empirical results and the forecast performance was investigate. Section 6 concludes.

2 Methodology

2.1 Reconciliation Methods

Forecasts can be generated for individual time series and it is in fact, common to produce disaggregated forecasts based on disaggregated time series (which will hereby be referred to as *base* forecasts). These can be generated using any method. It is desirable and required for the forecasts to add up in the same way as the data i.e. the forecasts to mimic the properties of the real data. This is referred to as "coherent" forecasts. In order for the forecasts to be coherent, reconciliation methods are used.

The rest of this section would explain the notation and the methodology of the reconciliation approaches used in this project - Bottom-up, Weighted least squares (WLS), MinT(Sample) and MinT(Shrink).

2.1.1 Notation

Let $\mathbf{y}_t$ be an m-vector containing all observations at time t, and $\mathbf{b}_t$ be an n-vector containing the observations at the most disaggregated level only - which results in general matrix representation

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t$$

where $\mathbf{S}$ is a "summing matrix" of order $m \times n$ which aggregates the bottom level series to the series at aggregation levels above.

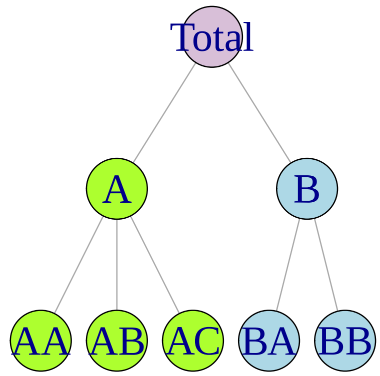


Figure 1: A simple hierarchical time series

A simple example of a small hierarchical time series depicted in the tree diagram is shown in the above Figure - whereby each parent comprises of the sum of its children. Let y_t denote the observations at time t at the most aggregate level 0; $y_{A,t}$ and $y_{B,t}$ the observations at aggregation level 1; and $y_{AA,t}, y_{AB,t}, ..., y_{BB,t}$ the observations at the most disaggregated/lowest level.

In this example, $n = 5$, $m = 8$, $\mathbf{y}_t = [y_t, y_{A,t}, y_{A,t}, y_{AA,t}, y_{AB,t}, y_{AC,t}, y_{BA,t}, y_{BB,t}]'$, $\mathbf{b}_t = [y_{AA,t}, y_{AB,t}, y_{AC,t}, y_{BA,t}, y_{BB,t}]'$, and the summing matrix will be as shown below;

$$S = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ & & \mathbf{I}_n & & \end{bmatrix}$$

where $\mathbf{I}_n$ is an identity matrix of dimension of $n = 5$.

The same general notation is applied to the much larger collection of time series in this research project whereby each aggregation constraint is represented by a row in the summing matrix $\boldsymbol{S}$.

Letting $\hat{\boldsymbol{y}}_T(h)$ be a vector of h-step-ahead base forecasts for each time series in the collection made using data up to time T, and stacking them in the same order as $\boldsymbol{y}_t$, all reconciliation methods can be written in the form of

$$\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_T(h),$$

for some appropriately selected matrix $\boldsymbol{P}$ of order $n \times m$, and $\tilde{\boldsymbol{y}}_T(h)$ is a set of reconciled forecasts which are coherent by construction.

Therefore, the basic idea behind forecast reconciliation methods is to linearly map a given set of base forecasts to a set of reconciled forecasts. The role of $\boldsymbol{P}$ would be to map the base forecasts into bottom level disaggregated forecasts which are then summed by the summing matrix, $\boldsymbol{S}$. (Wickramasuriya et al., 2018)

Let

$$\hat{\boldsymbol{e}}_T(h) = \boldsymbol{y}_{T+h} - \hat{\boldsymbol{y}}_T(h)$$

be defined as the h-step ahead conditionally stationary base forecast errors with $E[\hat{\boldsymbol{e}}_T(h)]|\boldsymbol{\tau}_t] = 0$, where $\boldsymbol{\tau}_t = \boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_T$ denote data observed up to time T. This implies that the base forecasts are unbiased, i.e., $E[\hat{\boldsymbol{y}}_T(h)|\boldsymbol{\tau}_t] = E[\boldsymbol{y}_{T+h}|\boldsymbol{\tau}_t]$. Letting $\hat{\boldsymbol{b}}_T(h)$ be the the bottom level base forecasts where $E[\hat{\boldsymbol{b}}_T(h)|\boldsymbol{\tau}_t] = \boldsymbol{\beta}_T(h)$, implies $E[\hat{\boldsymbol{y}}_T(h)|\boldsymbol{\tau}_t] = \boldsymbol{S}\boldsymbol{\beta}_T(h)$. Therefore, a set of reconciled forecasts will also be unbiased iff $\boldsymbol{SPS} = \boldsymbol{S}$ or equivalently $\boldsymbol{PS} = \boldsymbol{I}_n$. This results in $E[\tilde{\boldsymbol{y}}_T(h)|\boldsymbol{\tau}_t] = \boldsymbol{S}\boldsymbol{\beta}_T(h)$ as from the previous equation. (Wickramasuriya et al., 2018)

2.1.2 Bottom-up

The bottom-up approach as described in Shlifer and Wolff (1979) is another simple approach whereby the the most disaggregated (or the bottom level) series are forecasted, and the results are summed up to form forecasts of the various aggregated series. While an advantage of this approach is that forecasting at the bottom-level of a structure results in no information being lost due to aggregation, a disadvantage is that the bottom-level data can be noisy and challenging to forecast (Hyndman & Athanasopoulos, 2018).

Using the same compact notation above, the bottom-up approach can be mathematically expressed as $\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}\hat{\boldsymbol{b}}_T(h)$, by setting $\boldsymbol{P} = [\boldsymbol{0}_{n\times(m-n)} \vdots \boldsymbol{I}_n]$ where $\boldsymbol{0}_{i\times j}$ is the $i \times j$ null matrix. This extracts bottom level base forecasts from $\hat{\boldsymbol{y}}_T(h)$ and then gets summed by $\boldsymbol{S}$ to return bottom-up forecasts.

2.1.3 Minimum Trace (MinT) Reconciliation

The Minimum Trace reconciliation has three sources of errors - h-step ahead conditionally stationary base forecast errors ($\hat{\boldsymbol{e}}_T(h)$), reconciled forecast errors ($\tilde{\boldsymbol{e}}_T(h)$) and the estimated coherency errors ($\tilde{\boldsymbol{\epsilon}}_t$). Let these be defined as

$$\hat{\boldsymbol{e}}_t(h) = \boldsymbol{y}_{T+h} - \hat{\boldsymbol{y}}_T(h),$$

$$\tilde{\boldsymbol{e}}_t(h) = \boldsymbol{y}_{T+h} - \tilde{\boldsymbol{y}}_T(h),$$

$$\tilde{\boldsymbol{\epsilon}}_h = \hat{\boldsymbol{y}}_T(h) - \tilde{\boldsymbol{y}}_T(h),$$

for t = 1,2,. . . , where $\tilde{\boldsymbol{y}}_T(h)$ are the h-step-ahead reconciled forecasts using information up to and including time t, and $\boldsymbol{y}_{T+h}$ are the observed values of all series at time $t + h$.

Wickramasuriya et al. (2018) explained that for any $\boldsymbol{P}$ subject to the constraint $\boldsymbol{SPS} = \boldsymbol{S}$, the covariance matrix of the h-step ahead reconciled forecast errors is defined as $Var[\boldsymbol{y}_{T+h} - \tilde{\boldsymbol{y}}_T(h)|\boldsymbol{\tau}_t] = \boldsymbol{SPW}_h\boldsymbol{P'S'}$ and $\boldsymbol{W}_h = E[\hat{\boldsymbol{e}}_t(h)\hat{\boldsymbol{e}}'_t(h)|\boldsymbol{\tau}_t]$ is the variance-covariance matrix of the h-step ahead base forecast errors. A value of $\boldsymbol{P}$ that minimizes the trace of h-step ahead reconciled forecast errors would give the best (minimum variance) linear unbiased reconciled forecasts which is referred to as MinT (minimum trace) reconciliation. This optimal $\boldsymbol{P}$ is given by $\boldsymbol{P} = (\boldsymbol{S'W}_h^{-1}S)^{-1}\boldsymbol{S'W}_h^{-1}$ and therefore, the reconciled forecasts from the MinT approach are computed by

$$\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}(\boldsymbol{S'W}_h^{-1}S)^{-1}\boldsymbol{S'W}_h^{-1}\hat{\boldsymbol{y}}_T(h)$$

.

However, as $\boldsymbol{W}_h$ is challenging to estimate, especially for $h > 1$, there are several alternatives for its definition - from which, the relevant methodology will be explained below.

2.1.3.1 Weighted Least Squares

Setting $\boldsymbol{W}_h = k_h diag(\hat{\boldsymbol{W}}_1), \forall h$, where $k_h > 0$ and

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

is the unbiased sample estimator of the in-sample one-step ahead base forecast errors, in which case, MinT was described as a WLS estimator. (Athanasopoulos, Hyndman, Kourentzes, & Petropoulos, 2017; Hyndman et al., 2016)

The WLS estimator involves ignoring correlations across series but allowing for heterogeneity. This eliminates the need to estimate any covariances, and is denoted as WLS_v (i.e. WLS applying variance scaling). (Athanasopoulos et al., 2017)

2.1.3.2 MinT(Sample)

Setting $\boldsymbol{W}_h = k_h \hat{\boldsymbol{W}}_1, \forall h$, where $k_h > 0$, the unrestricted sample covariance estimator for $h = 1$, and was referred to as MinT(Sample). This is relatively simple to obtain, but may not be a good estimate when $m > T$ or when m is of the same order as T, where m is the number of series and T the number of observations (Wickramasuriya et al., 2018). This is not a good estimator in these cases, as it is an infeasible estimator due to the W_h matrix not being invertible.

2.1.3.3 MinT(Shrink)

Setting $\boldsymbol{W}_h = k_h \hat{\boldsymbol{W}}^{\star}{}_{1,D}, \forall h$, where $k_h > 0$, $\hat{\boldsymbol{W}}^{\star}{}_{1,D} = \lambda_D \hat{\boldsymbol{W}}_{1,D} + (1-\lambda_D)\hat{\boldsymbol{W}}_1$ is a shrinkage estimator with a diagonal target, $\hat{\boldsymbol{W}}_{1,D}$ is a diagonal matrix comprising the diagonal entries of $\hat{\boldsymbol{W}}_1$, and λ_D is the shrinkage intensity parameter. Therefore, the off-diagonal elements of $\hat{\boldsymbol{W}}_1$ are shrunk towards zero and diagonal elements (variances) are left unchanged. (Wickramasuriya et al., 2018)

As proposed by Schäfer and Strimmer (2005) the shrinkage intensity parameter (λ_D) is scale and location invariant, and can be defined as

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

where $\hat{r}_{ij}$ is the ijth element of $\hat{\boldsymbol{R}}_1$, the 1-step-ahead sample correlation matrix. The role of the shrinkage intensity parameter is to shrink the 1-step-ahead sample correlation matrix towards an identity matrix. An advantage of this is that matrix will automatically be positive semi definite and invertible - which solves the issue MinT(Sample) faced.

3 Data

3.1 Approaches

According to the Australian Bureau of Statistics (2015), Gross Domestic Product (GDP) is measured by three approaches in Australia - Production, Expenditure and Income. These form the basis by which the GDP is disaggregated and create the hierarchical structure for GDP. Each approach will be highlighted in the following sections. Furthermore, each approach's hierarchy will be plotted, whereby the most aggregated level (or top level) would be displayed in grey, bottom level series in red, aggregations that continue on and shown in following plots in orange and the remaining aggregations at different levels in blue.

3.1.1 Production Approach (GDP(P))

Australian Bureau of Statistics (2015) explains that GDP is based on the concept of value added, which is defined as the unduplicated value of goods and services produced in any given period.

The formula which forms the skeleton of the hierarchy under this approach is;

$$\text{GDP(P)} = \text{Gross value added} + \text{Taxes on products} - \text{Subsidies on products} + \text{Statistical discrepancy (P)}$$

This is further disaggregated and organized according to the hierarchy shown below.

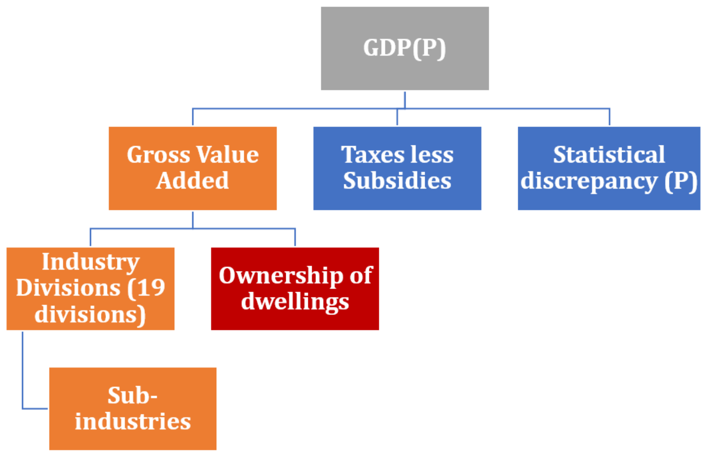


Figure 2: Hierarchy of the Production Approach

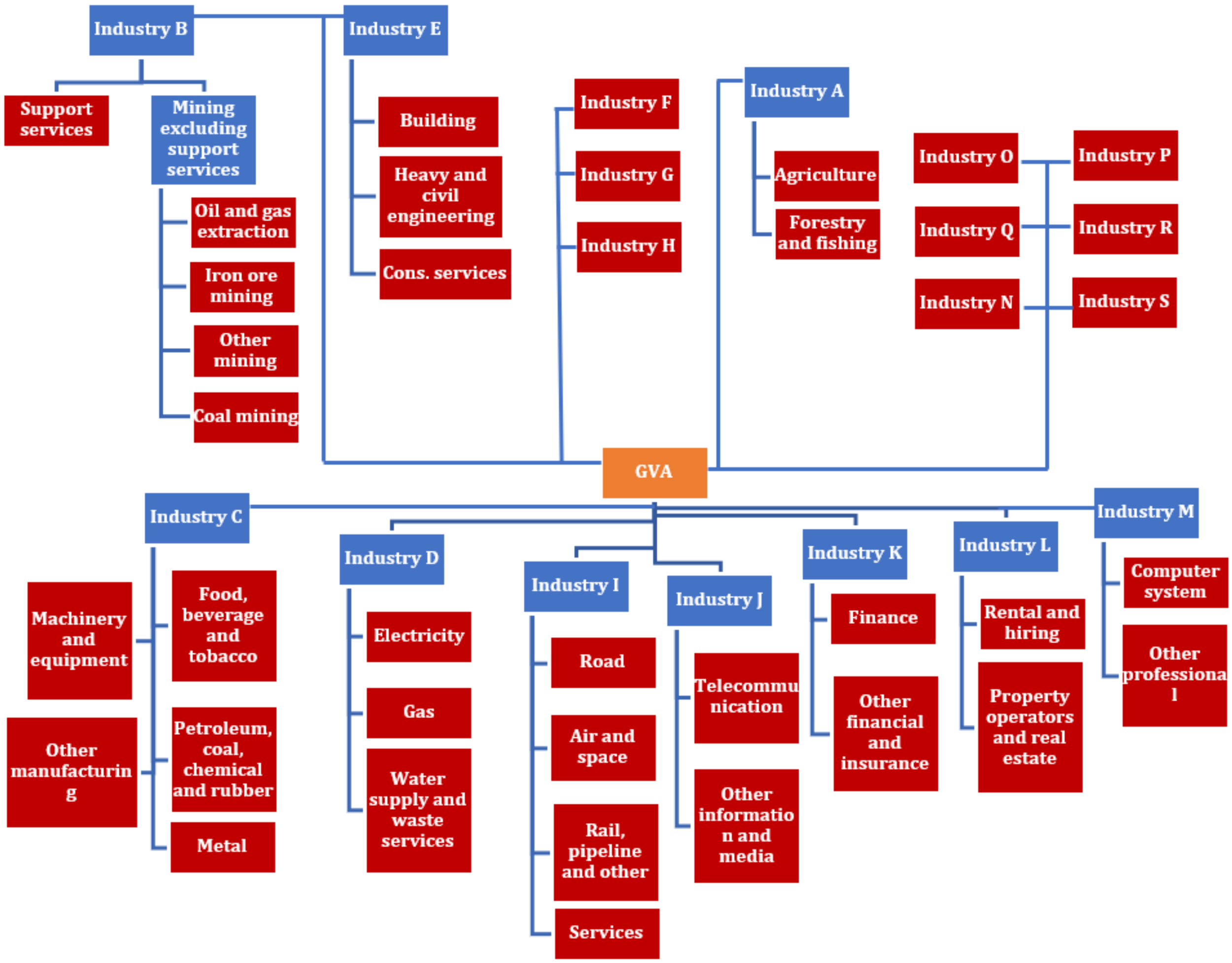


Figure 3: Hierarchy of Gross Value Added under the Production Approach

3.1.2 Expenditure Approach (GDP(E))

GDP can also be derived as the sum of all final expenditures (which comprises of final consumption expenditure and gross fixed capital formation (GFCF), in this context), the changes in inventories of finished goods, work-in-progress and raw materials, and the value of exports of goods and services less the value of imports of goods and services. (Australian Bureau of Statistics, 2015)

The equivalent formula under this approach is;

$$
\begin{aligned}
\text{GDP(E)} &= \text{Gross National Expenditure} + \text{Exports} - \text{Imports} + \text{Statistical discrepancy (E)} \\
&= \text{Domestic Final Demand} + \text{Changes in inventories} + \text{Exports} - \text{Imports} + \text{Statistical discrepancy (E)} \\
&= \text{Final consumption expenditure} + \text{Gross fixed capital formation} + \text{Changes in inventories} \\
&\quad + \text{Exports} - \text{Imports} + \text{Statistical discrepancy (E)}
\end{aligned}
$$

Although the net acquisition of valuables is measured, it is not separately identified in the Australian System of National Accounts (ASNA), and therefore, not included in this hierarchy either.

This is further disaggregated and organized according to the hierarchy shown below.

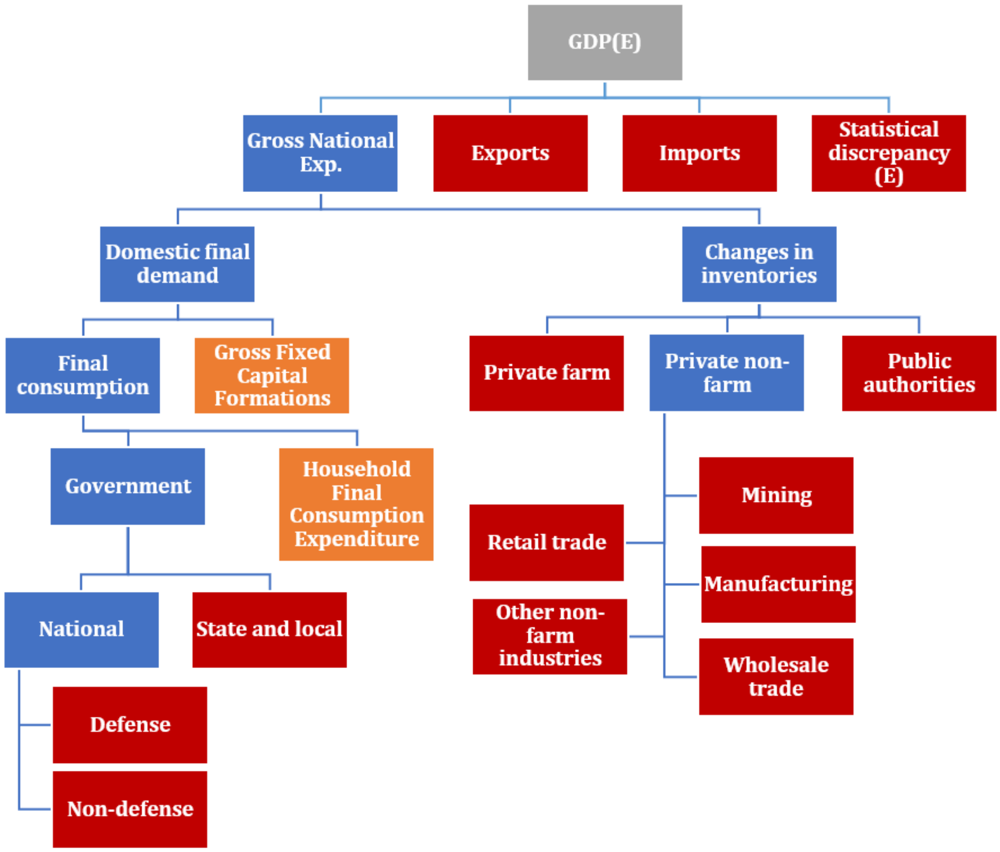


Figure 4: Hierarchy of the Expenditure Approach

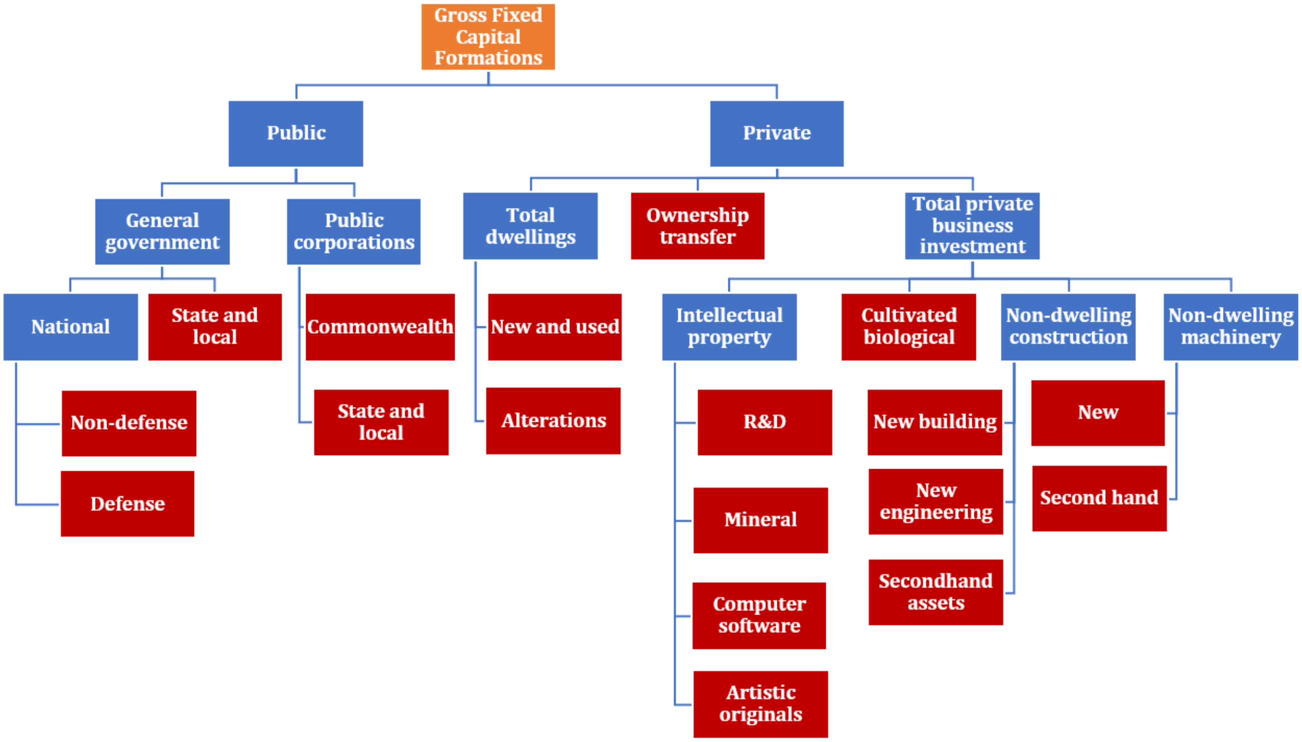


Figure 5: Hierarchy of Gross Fixed Capital Formations under the Expenditure approach

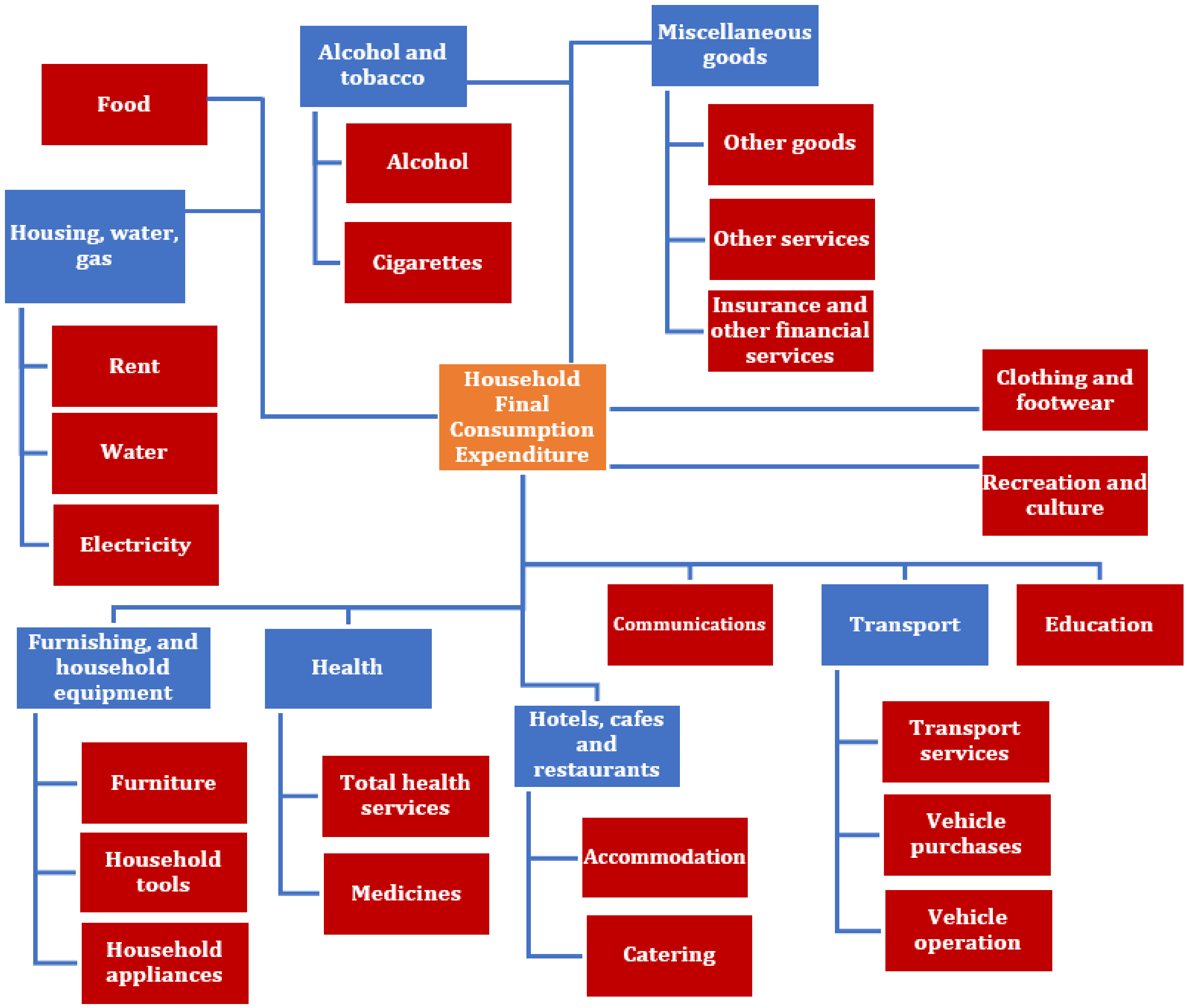


Figure 6: Hierarchy of Household Final Consumption Expenditure (HFCE) under the Expenditure approach

3.1.3 Income Approach (GDP(I))

Australian Bureau of Statistics (2015) also measures GDP by the sum of income flows, and the equivalent formula under this approach is;

$$\begin{aligned} \text{GDP(I)} &= \text{Total Factor Income} + \text{Taxes on production and imports} - \text{Subsidies on production and imports} \\ &+ \text{Statistical discrepancy (I)} \\ &= \text{Compensation of employees} + \text{Gross Operating Surplus} + \text{Gross Mixed Income} \\ &+ \text{Taxes on production and imports} - \text{Subsidies on production and imports} + \text{Statistical discrepancy (I)} \end{aligned}$$

This is further disaggregated and organized according to the hierarchy shown below.

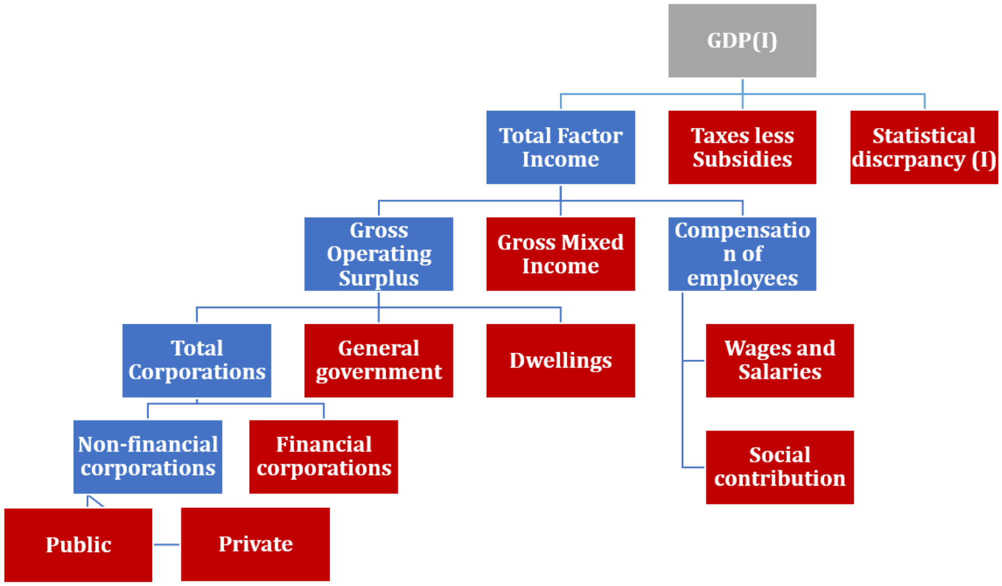


Figure 7: Hierarchy of the Income Approach

3.2 Price and Volume Measures

As the economic transactions are a product of the quantity produced or sold and the unit price, the total monetary value of the transactions are simply just the sum of individual transaction values in a particular period, which is referred to as the current price value. In order to measure the volume growth of production and expenditures (growth free of the effects of price change), current price values are inadequate as they are subject to the effects of changing prices. Although measures of constant prices, and chain volume indexes of Laspeyres, Paasche and Fisher's Ideal Index (referred to as Fisher's Index, and is a geometric mean of the Laspeyres and Paasche) are available to the ABS, it uses chain volume Laspeyres indexes. (Australian Bureau of Statistics, 2015)

The annual chain Laspeyres price indexes is formed by multiplying or compounding consecutive year-to-year indexes using the formula:

$$L_Q^y = \frac{\sum_{i=1}^n P_i^0 Q_i^1}{\sum_{i=1}^n P_i^0 Q_i^0} \times \frac{\sum_{i=1}^n P_i^1 Q_i^2}{\sum_{i=1}^n P_i^1 Q_i^1} \times ... \times \frac{\sum_{i=1}^n P_i^{y-1} Q_i^y}{\sum_{i=1}^n P_i^{y-1} Q_i^{y-1}}$$

Quarterly chain volume indexes are found similarly by using annual weights to derive quarterly Laspeyres chain volume indexes and linking them by using the one-quarter overlap method (Australian Bureau of Statistics, 2015).

3.3 Issues with the data

Australian Bureau of Statistics (2015) benchmarks its quarterly chain volume estimates to their annual counterparts. Unlike constant price estimates, after benchmarking (and re-referencing), they do not usually add up (i.e. the data is not *coherent*). As the coherency of the data is an important condition of the methodology used in this paper, this paper focuses on current price estimates.

A second issue that arose is that only chain volume estimates are provided for quarterly data under the production approach. This is due to information necessary to compile comprehensive current price estimates not being available quarterly (Australian Bureau of Statistics, 2018). As maximizing the number of observations is important to improve forecast accuracy, quarterly data is used in this paper, and therefore, only the expenditure and income approaches are available in current prices.

While each measure should conceptually or theoretically produce the same estimate of GDP, since the three measures are compiled independently using different data sources, they result in different estimates of GDP. The final GDP figure (or the headline measure) is a simple average of the three separate measures compiled in chain volume terms. The current price estimate of GDP is obtained by reflating the average chain volume estimate by the implicit price deflator derived from the expenditure-based estimates (Australian Bureau of Statistics, 2018). Therefore, a statistical discrepancy for each measure is calculated and shown, and due to the coherency condition, the statistical discrepancy is also included and forecasted.

The variables used, series IDs assigned by the ABS, and a brief description of each variable was recorded in Tables 9-12. All data series are quarterly data, and the data sets starts from the earliest possible observation from all data series - 1984 Q4.

It should be noted that under both approaches, all data sets were checked for coherency, and it was found that all data sets do indeed add up except for insignificant rounding issues. This was tested by the formula $\boldsymbol{y}_t - \boldsymbol{S}\boldsymbol{b}_t$ where $\boldsymbol{y}_t$ is an m-vector containing all observations from 1984 Q4 to 2018 Q1, $\boldsymbol{b}_t$ is an n-vector containing the observations at the most disaggregated level only, and $\boldsymbol{S}$ the respective summing matrix - as described in the previous section.

Some of the extreme cases found under the income approach occurred in Q2 1989, Q3 1989 and Q2 1990 at differences of \$6 million, \$5 million and \$-5 million respectively (shown below in the upper panel of Figure 8). In all three cases, the differences were the result of the coherency issues in the total Gross operating surplus which in turn, got carried forward to the GDP at level 0. Similar issues were found for the expenditure approach and has been illustrated in the lower panel of Figure 8 below. However, due to all issues being rounding errors, they were considered insignificant.

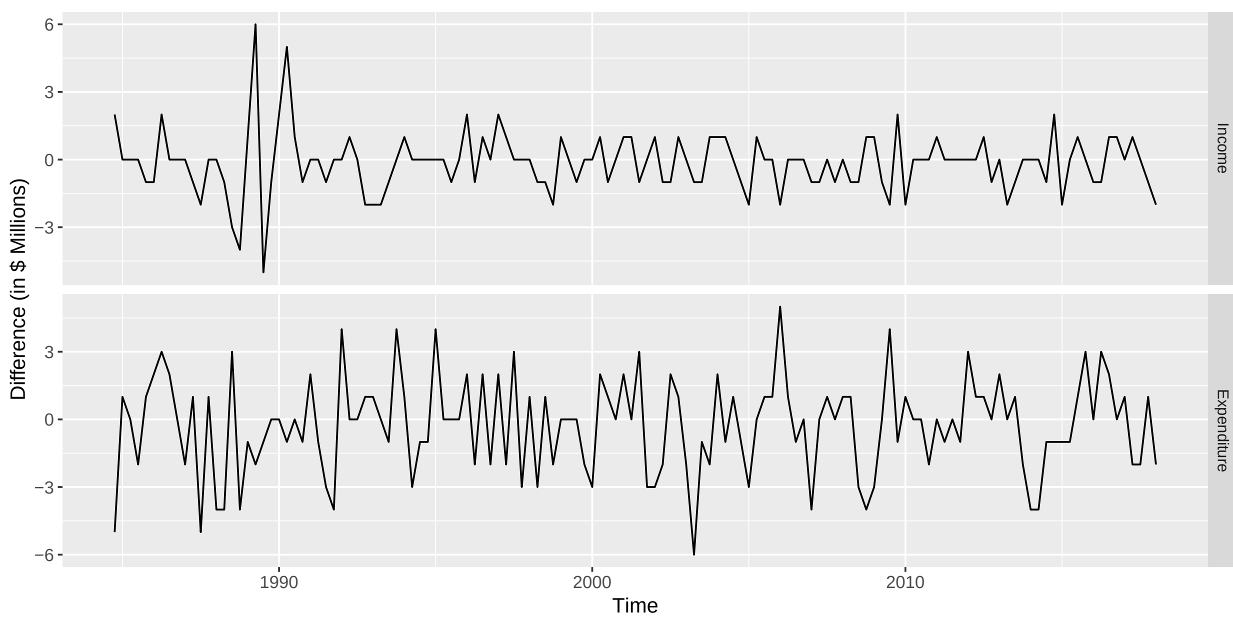


Figure 8: Coherency check for Income approach (upper panel) and Expenditure approach (lower panel). Found by taking the difference between GDP and the sum of the bottom level series.

3.4 Time series properties

When examining the time series properties, it can be seen that different series in the hierarchy have different properties - in terms of both the trend and seasonality. For example, while the GDP (at the most aggregate level) has a strong, positive trend, gross fixed capital formations show that the trend has plateaued. Additionally, some of the most disaggregated series such as New engineering construction had a varying trend, and secondhand non-dwelling machinery has a decreasing trend. Furthermore, changes in inventories can be seen to have minimal trend, with a strong seasonality component and oscillating around a mean of zero. This is illustrated in the figure below. Therefore, the base forecasts generated for each series would also have to reflect these varying time series properties. Plots of all series will be illustrated in the Appendix.

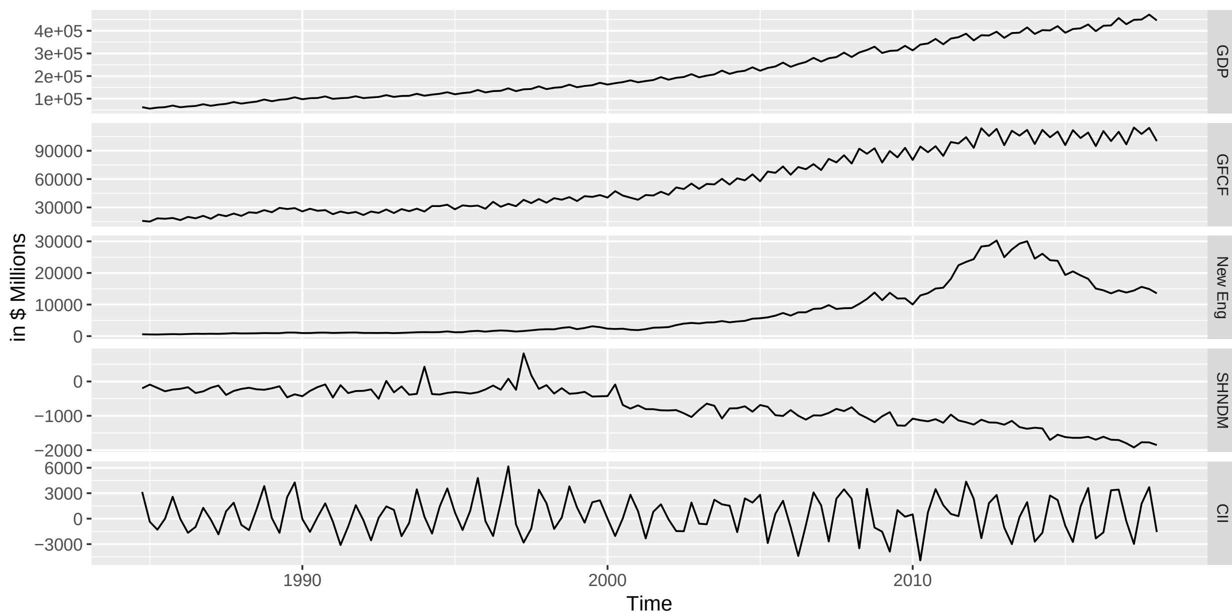


Figure 9: Gross Domestic Product (GDP), Gross Fixed Capital Formations (GFCF), New engineering costs (New Eng), Secondhand non-dwelling machinery (SHNDM), and Changes in Inventories (CII)

4 Experiment

4.1 Forecasting Methods

Due to the number of series in the expenditure approach being 81 series and the income approach being an additional 16 series, automatic algorithms were required for the forecasts to be made. Exponential smoothing (ETS) and AutoRegressive Integrated Moving Average (ARIMA) models allow the varying time series properties described to be modeled accurately in addition to having the framework for automatic algorithms defined. Therefore, base forecasts for all time series would be generated using ETS and ARIMA models, fitted using the default settings in the automated algorithms, `ets` and `auto.arima` respectively, of Hyndman and Khandakar (2008) and implemented in the `forecast` package for R by Hyndman et al. (2018).

Exponential smoothing models (ETS) are allowed to have different combinations of errors, trend and seasonality. Trend is allowed to be either none, additive or additive damped and seasonality is allowed to be either none, additive or multiplicative. For each combination of trend and seasonality, two models exist - one with additive errors and another with multiplicative errors - which in turn, generate different prediction intervals. (Hyndman & Athanasopoulos, 2018)

The automated algorithm, `ets`, attempts to find the best model by minimizing the Corrected Akaike's Information Criterion (AICc) from all the possible ETS models (excluding the models that cause numerical difficulties and the numerically unstable models in the specific scenarios).

As explained by Hyndman and Athanasopoulos (2018), AutoRegressive Integrated Moving Average (ARIMA) models allow both autoregressive terms (past values of the variable) and moving average terms (past forecast errors in a regression-like model). Seasonal ARIMA models (as the ones applicable in this case) takes into account non-seasonal components and seasonal components of a series and can be summarized and written in the form of ARIMA (p, d, q) $(P, D, Q)_m$ whereby;

- p = order of the non-seasonal autoregressive component
- d = degree of first differencing involved
- q = order of the non-seasonal moving average component
- P = order of the seasonal autoregressive component
- D = degree of seasonal differencing involved
- Q = order of the seasonal moving average component
- m = number of observations per year or frequency of the data

The automated algorithm, `auto.arima`, attempts to find the best model after ensuring stationarity using repeated KPSS tests, and by finding an appropriate number of non-seasonal and seasonal AR (p and P respectively) and MA (q and Q respectively) components by minimizing the AICc. With the default settings, it uses approximations and a stepwise search to traverse the model space to speed up the process.

These forecasts would then be reconciled using the alternative reconciliation approaches explained.

4.2 Expanding window

The experiment set up is an expanding window whereby 1 to 4 step ahead (1 year ahead) forecasts would be obtained for each training set. The smallest training set would start in 1984 Q4 and end in 1994 Q3 (comprises of 40 observations). The training set would then be expanded by one quarter and the experiment would be repeated.

In each window, base ETS and ARIMA forecasts, in addition to reconciled forecasts for each using the Bottom-up, Weighted least squares, MinT (Sample) and MinT(Shrink) approaches were estimated. Furthermore, the benchmark of a Seasonal Random Walk with Drift and its reconciliations were estimated as well.

The forecast accuracy measures used were Root Mean (Median) Squared Error (RMSE) and the scale independent Mean (Median) Absolute Scaled Error (MASE). Root Mean Squared Error can be defined as

$$RMSE = \sqrt{\text{mean}(e_t^2)},$$

whereby e_t is a forecast error and the difference between an observed value and its forecast (mathematically expressed as $e_{T+h} = y_{T+h} - \hat{y}_{T+h|T}$). Although widely used, as it is based only on e_t, it is a scale dependent measure and cannot be used to make comparisons between series that involve different units or vastly different scales. (Hyndman & Athanasopoulos, 2018)

This leads to the second forecast accuracy measure of Mean Absolute Scaled Error (MASE) for seasonal time series, proposed by Hyndman and Koehler (2006), which defines the forecast error relative to a seasonal naïve method in the following manner;

$$q_t = \frac{e_t}{\frac{1}{T-m}\sum_{t=m+1}^{T}|y_t - y_{t-m}|};$$
$$\text{MASE} = \text{mean}(|q_t|),$$

where m is the frequency of data (in this case, 4 as it is quarterly data).

MASE is independent of the scale of the data as both the numerator and denominator both involve values on the scale of the original data. MASE is less than one if it arises for a better forecast the average seasonal naïve forecast computed on the training data, and greater than one if the forecast is worse than the average seasonal naïve forecast.

It should be noted that the Root Median Squared Error and Median Absolute Scaled Error is the median equivalent of these error measures.

5 Results

The experiment was examined by computing the forecast accuracy measures across all windows for each series, forecasting method, reconciliation approach and forecast horizon for both the income and expenditure approach. Under each approach, the most aggregate level (GDP) as well as an average Mean (Median) Absolute Scaled Error (MASE) was computed and examined as well.

5.1 Income approach

The following table shows the Root Mean Squared Error (RMSE) for GDP under the income approach in comparison to the benchmark errors. Negative values indicate that the methods improve on the benchmark forecasts, and the best method is the method with the lowest errors.

When examining these errors, it can be seen that the best performing methods are WLS ARIMA for the first forecast horizon (when h=1), followed by WLS ETS in the second forecast horizon, and Bottom-up ARIMA in the following horizons.

Table 1: Root Mean Squared Errors for the most aggregate level (GDP)

F-method	**R-method**	**1**	**2**	**3**	**4**
ARIMA	MinT(Shrink)	-54.664	-31.692	-13.108	-5.054
ARIMA	MinT(Sample)	-53.621	-30.31	-13.709	-8.608
ARIMA	WLS	**-56.682**	-36.089	-19.253	-11.557
ARIMA	Bottom-up	-54.187	-36.172	**-19.896**	**-13.299**
ARIMA	Base	-56.15	-33.809	-15.214	-6.948
ETS	MinT(Shrink)	-54.55	-35.094	-16.967	-4.56
ETS	MinT(Sample)	-51.141	-34.322	-17.115	-4.53
ETS	WLS	-55.212	**-36.676**	-19.446	-8.179
ETS	Bottom-up	-52.32	-33.642	-17.752	-6.229
ETS	Base	-56.094	-34.83	-14.345	-1.515

However, upon closer inspection, it can be seen that there are likely outliers as illustrated in the following figure, which depicts the absolute errors of base ARIMA and ETS forecasts alongside the benchmark forecasts. The outliers are most noticeable around the Financial Crisis of 2007 and late 2016 to early 2017. Hence, the Root Median Squared errors were examined as well.

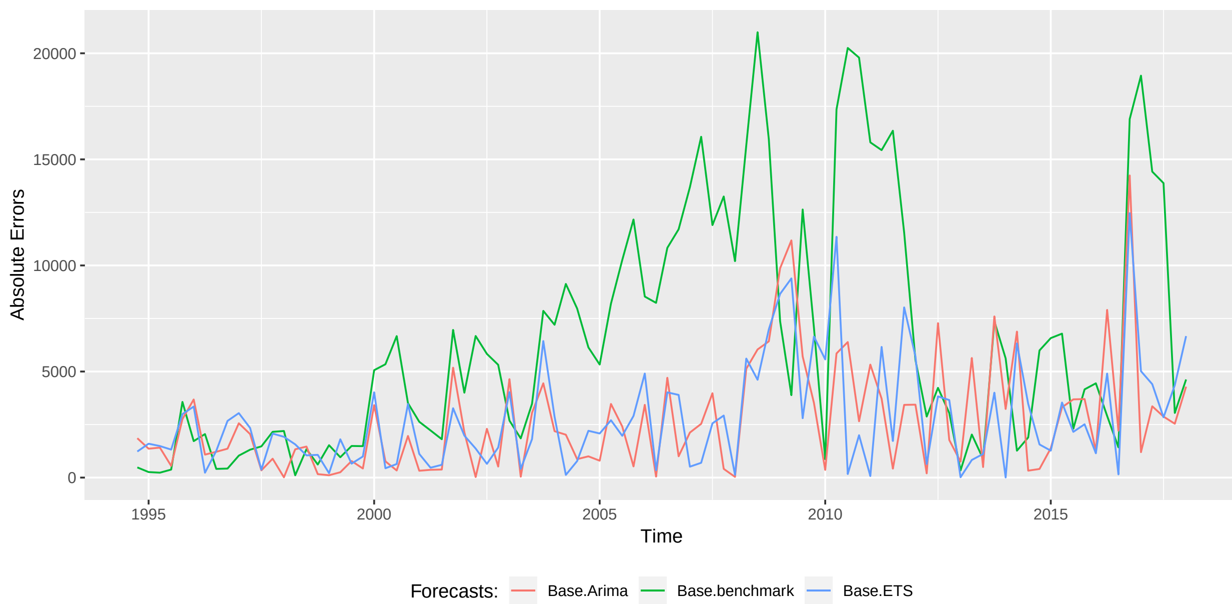


Figure 10: Absolute errors of Base and Benchmark forecasts

When the Root Median Squared Errors were examined, MinT(Sample) ARIMA brings the biggest improvement over the Benchmark forecasts in the first forecast horizon, followed by WLS ETS, MinT(Sample) ARIMA and MinT(Shrink) ARIMA in the following forecast horizons.

Table 2: Root Median Squared Errors for the most aggregate level (GDP)

F-method	**R-method**	**1**	**2**	**3**	**4**
ARIMA	MinT(Shrink)	-61.255	-48.739	-32.64	**-46.26**
ARIMA	MinT(Sample)	**-64.956**	-45.167	**-37.31**	-28.983
ARIMA	WLS	-58.121	-49.827	-28.237	-27.545
ARIMA	Bottom-up	-53.431	-46.385	-18.781	-7.182
ARIMA	Base	-59.788	-46.734	-31.291	-40.235
ETS	MinT(Shrink)	-53.943	-49.147	-36.308	-35.787
ETS	MinT(Sample)	-53.131	-46.582	-36.105	-38.57
ETS	WLS	-54.979	**-53.325**	-25.029	-16.786
ETS	Bottom-up	-56.609	-45.439	-21.589	9.038
ETS	Base	-59.987	-50.261	-34.127	-37.72

Additionally, an average of the whole hierarchy was taken as well. The following table depicts the Median Absolute Scaled Errors averaged over the whole hierarchy for each forecast horizon. The best models is the one with the lowest value. The best methods, in this case, were WLS ETS for first and second forecast horizons, followed by WLS ARIMA for the last two forecast horizons. It should be noted that these results do not change when an average of Mean Absolute Scaled Errors are taken either.

Table 3: Average Median Absolute Scale Error for the whole hierarchy

F-method	R-method	1	2	3	4
ARIMA	MinT(Shrink)	0.5216	0.6935	0.8432	0.9503
ARIMA	MinT(Sample)	0.5424	0.7274	0.8563	0.9496
ARIMA	WLS	0.4456	0.5838	**0.6913**	**0.7775**
ARIMA	Bottom-up	0.4353	0.5694	0.7019	0.8329
ARIMA	Base	0.4524	0.574	0.7145	0.7865
Benchmark	MinT(Shrink)	0.7668	0.7925	0.8145	0.835
Benchmark	MinT(Sample)	0.7668	0.7926	0.8145	0.835
Benchmark	WLS	0.7668	0.7926	0.8145	0.835
Benchmark	Bottom-up	0.7669	0.7926	0.8146	0.835
Benchmark	Base	0.7669	0.7926	0.8145	0.835
ETS	MinT(Shrink)	0.5336	0.7065	0.8388	0.9201
ETS	MinT(Sample)	0.5824	0.7252	0.8475	0.9897
ETS	WLS	**0.4013**	**0.5471**	0.7049	0.7826
ETS	Bottom-up	0.4226	0.5541	0.7411	0.8122
ETS	Base	0.4325	0.5517	0.7188	0.7871

5.2 Expenditure approach

The following table shows the Root Mean Squared Error (RMSE) for GDP under the expenditure approach in comparison to the benchmark errors. When examining these errors, it can be seen that the best performing methods are WLS ARIMA for the first forecast horizon (when h=1), followed by MinT(Shrink) ETS in the following horizons.

Table 4: Root Mean Squared Errors for the most aggregate level (GDP)

F-method	R-method	1	2	3	4
ARIMA	MinT(Shrink)	-53.39	-31.053	-15.2	-7.419
ARIMA	WLS	**-56.586**	-35.245	-17.871	-9.462
ARIMA	Bottom-up	-42.849	-24.491	-10.228	-3.289
ARIMA	Base	-56.15	-33.809	-15.214	-6.948
ETS	MinT(Shrink)	-55.96	**-36.589**	**-22.988**	**-13.956**
ETS	WLS	-55.888	-34.869	-14.311	-1.521
ETS	Bottom-up	-50.299	-30.958	-18.608	-9.715
ETS	Base	-56.094	-34.83	-14.345	-1.515

Due to the outliers as shown above, the Root Median Squared Errors were checked for the expenditure approach as well, as shown in the following table. The best methods, in this case, were the Base ETS forecasts in the first forecast horizon, followed by MinT(Shrink) ETS in the second horizon, Base ETS and Base ARIMA in the third and fourth forecast horizon respectively.

Table 5: Root Median Squared Errors for the most aggregate level (GDP)

F-method	R-method	1	2	3	4
ARIMA	MinT(Shrink)	-53.545	-37.487	-16.631	-16.612
ARIMA	WLS	-59.698	-50.631	-29.415	-26.72
ARIMA	Bottom-up	-47.114	-35.196	-17.567	-11.292
ARIMA	Base	-59.788	-46.734	-31.291	**-40.235**
ETS	MinT(Shrink)	-59.098	**-56.528**	-22.475	-18.973
ETS	WLS	-59.195	-50.261	-29.845	-37.719
ETS	Bottom-up	-56.092	-50.671	-24.217	-10.765
ETS	Base	**-59.987**	-50.261	**-34.127**	-37.72

When the average Median Absolute Scaled Error for the whole hierarchy was checked, the best models were Bottom-up ETS in the first and second horizon, followed by WLS ARIMA in third and fourth forecast horizon. Similar to the findings from the income approach, these results do not change when the average of Mean Absolute Scaled Errors are taken either.

Table 6: Average Median Absolute Scale Error for the whole hierarchy

F-method	R-method	1	2	3	4
ARIMA	MinT(Shrink)	0.8302	0.9901	1.1111	1.2128
ARIMA	WLS	0.5872	0.7282	**0.8422**	**0.949**
ARIMA	Bottom-up	0.6045	0.754	0.8598	0.9593
ARIMA	Base	0.6076	0.7495	0.8643	0.9583
Benchmark	MinT(Shrink)	0.9512	0.9675	0.9847	1
Benchmark	WLS	0.9509	0.9672	0.9846	1.0001
Benchmark	Bottom-up	0.9509	0.9672	0.9845	1
Benchmark	Base	0.951	0.9673	0.9846	1.0001
ETS	MinT(Shrink)	0.7729	0.9183	1.0995	1.2383
ETS	WLS	0.7218	0.9408	1.092	1.2224
ETS	Bottom-up	**0.5854**	**0.7262**	0.8562	0.9709
ETS	Base	0.5871	0.7384	0.8708	0.9831

5.3 Summary of the methods

The following tables show the best forecast method for each error measure and forecast horizon, under each GDP approach.

Table 7: Best forecast method for each error measure under each forecast horizon for the income approach

	1	2	3	4
GDP(I): MeanASE	Base ARIMA	WLS ETS	WLS ARIMA	Base ARIMA
GDP(I): MedianASE	MinT(Shrink) ARIMA	WLS ETS	MinT(Shrink) ETS	MinT(Shrink) ETS
GDP(I): RMeanSE	WLS ARIMA	WLS ETS	BU ARIMA	BU ARIMA
GDP(I): RMedianSE	MinT(Sample) ARIMA	WLS ETS	MinT(Sample) ARIMA	MinT(Shrink) ARIMA
Hierarchy MeanASE	WLS ETS	WLS ETS	WLS ARIMA	WLS ARIMA
Hierarchy MedianASE	WLS ETS	WLS ETS	WLS ARIMA	WLS ARIMA

Table 8: Best forecast method for each error measure under each forecast horizon for the expenditure approach

	1	**2**	**3**	**4**
GDP(E): MeanASE	WLS ARIMA	MinT(Shrink) ETS	MinT(Shrink) ETS	MinT(Shrink) ETS
GDP(E): MedianASE	WLS ARIMA	MinT(Shrink) ETS	Base ARIMA	WLS ETS
GDP(E): RMeanSE	WLS ARIMA	MinT(Shrink) ETS	MinT(Shrink) ETS	MinT(Shrink) ETS
GDP(E): RMedianSE	Base ETS	MinT(Shrink) ETS	Base ETS	Base ARIMA
Hierarchy MeanASE	BU ETS	BU ETS	WLS ARIMA	WLS ARIMA
Hierarchy MedianASE	BU ETS	BU ETS	WLS ARIMA	WLS ARIMA

Tables showing the best forecast method for an average Mean (Median) Absolute Scaled Errors, averaged across all forecast horizons at the most aggregate level and for the whole hierarchy will also be shown in the appendix.

5.4 Other discussions

The two GDP approaches of Income and Expenditure were also compared. When comparing across all three measures described above, it appears as though the income approach generally outperforms the expenditure approach. The only two instances the expenditure approach was more accurate were GDP's Root Mean Squared Error measure's third and fourth forecast horizons.

Furthermore, as seen from Figure 10 and the multiple error tables, the base benchmark forecasts do not perform well for the data that is being used in this project - seasonally unadjusted in current prices. Therefore, it may be more appropriate to compare the reconciled forecasts rather than the benchmark forecasts. The following plots show reconciled forecasts of ARIMA and ETS compared to their respective base forecasts.

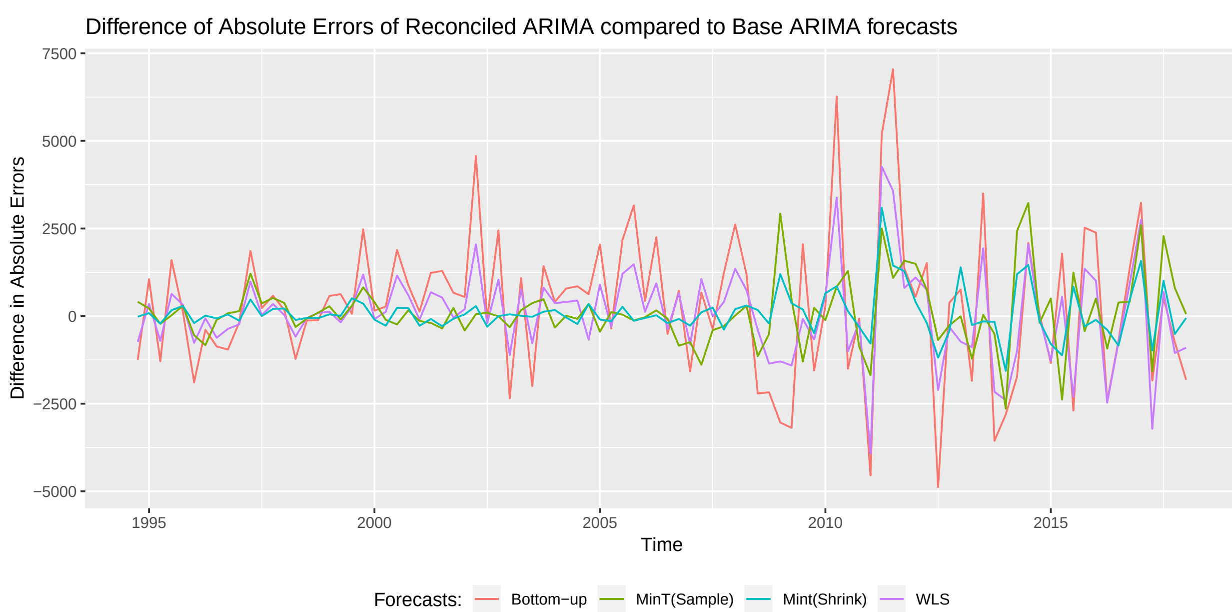


Figure 11: Difference of Absolute Errors of Reconciled ARIMA compared to Base ARIMA forecasts (Income approach)

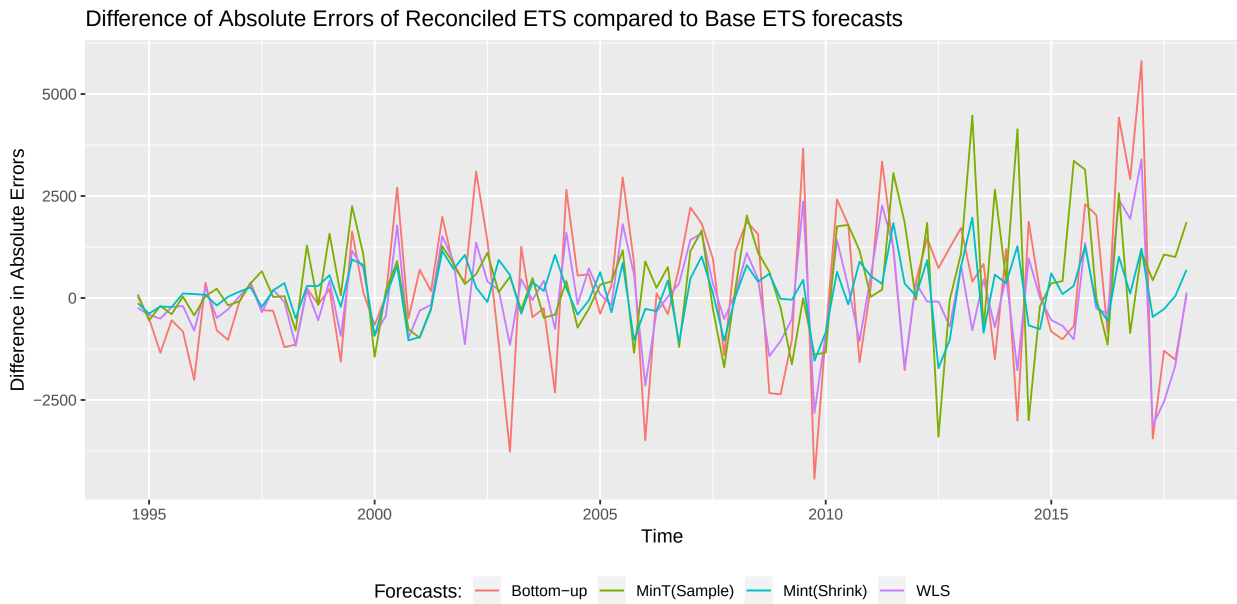


Figure 12: Difference of Absolute Errors of Reconciled ETS compared to Base ETS forecasts (Income approach)

The above figure shows the differences in the absolute errors between reconciled forecasts and base forecasts for the income approach. Negative values indicate that the reconciled forecasts are performing better than the base forecasts and vice versa. From these graphs, it can be seen that from all reconciliation approaches, WLS seems to be the best performing method. A similar story can be seen for the expenditure approach from the graphs provided in the Appendix.

Moreover, further analysis into why benchmark forecasts were not performing as well as expected showed that it is likely due to the benchmark method used in this project being unable to adequately pick up the exponential or non-linear trend that seems to have occurred from early 2000's to 2010. This can be seen in Figures 13-14 illustrated below.

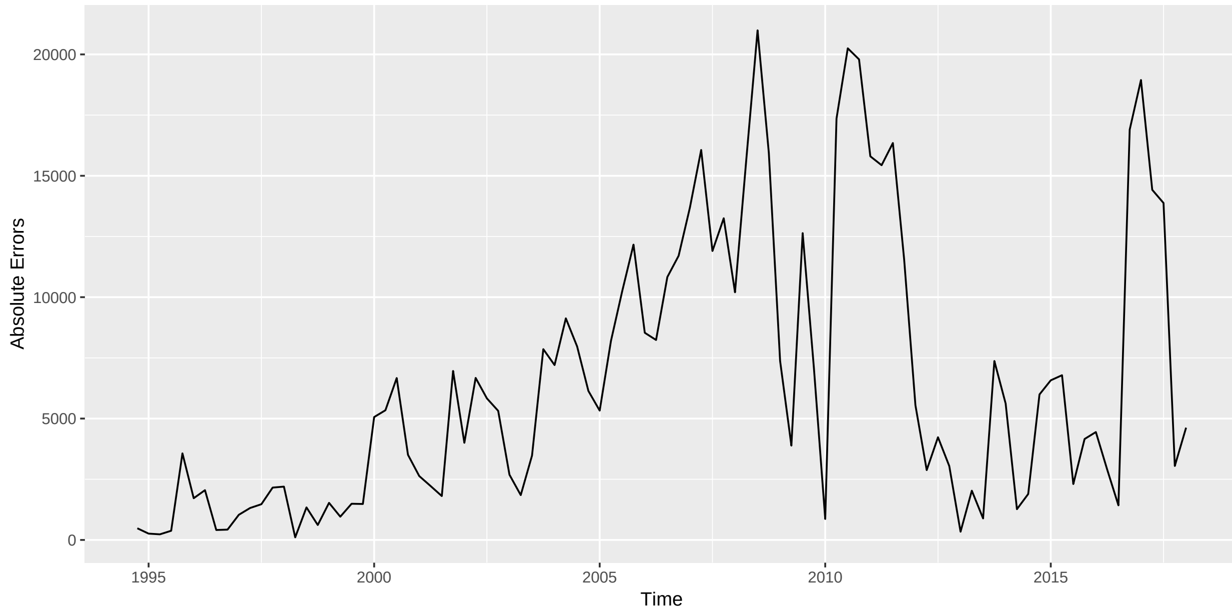


Figure 13: Absolute errors of benchmark foecasts

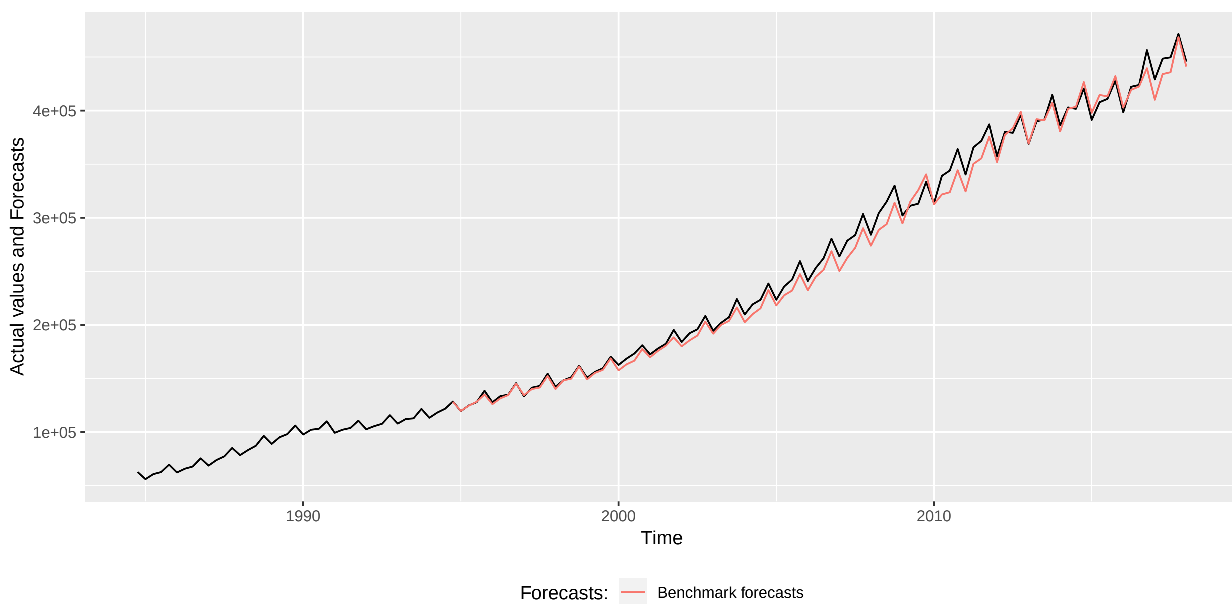


Figure 14: Benchmarks forecasts plotted alongside the GDP

6 Conclusion

To conclude, based on the experiment that was conducted in this project on seasonally unadjusted, nominal GDP, reconciliation approaches improve on GDP forecasts and related data in most cases. It should be noted even in the cases that it does not improve on base forecasts, reconciled forecasts have the advantage of providing coherent forecasts which other methods and approaches cannot provide. Additionally, it was found that while reconciliation approaches did beat the benchmark forecasts, the selected benchmark method did not perform very well for the data.

One possible avenue for future research include reconciliation of data with slack, which does not require a strict equality in terms of data coherency. This would allow components with no economic significance such as the statistical discrepancy to be dropped. Additionally, reconciliation with non-linear constraints (as opposed to linear constraints used in this project) would allow reconciliation to be done on seasonally adjusted, real GDP instead of being constrained to using seasonally unadjusted, nominal GDP.

Furthermore, an additional avenue for future research is to combine both the income and expenditure approaches, to test whether it could provide additional improvements in forecasts. Along the same lines, the hierarchical levels could also be limited based on gauging the signal-to-noise ratio.

Temporal hierarchies could also be used within the current hierarchy of GDP, whereby more frequent data could be allowed to add up to the aggregate for each series. This would involve using monthly data to add up to quarterly data for each series.

Appendix

Table 9: Variables, Series IDs and their descriptions for Income Approach

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

Table 10: Variables, Series IDs and their descriptions for Expenditure Approach

Variable	**Series ID**	**Description**
Gdpe	A2302467A	GDP(E)
Sde	A2302566J	Statistical Discrepancy(E)
Exp	A2302564C	Exports of goods and services
Imp	A2302565F	Imports of goods and services
Gne	A2302563A	Gross national exp.
GneDfdFceGvtNatDef	A2302523J	Gen. gov. - National; Final consumption exp. - Defence
GneDfdFceGvtNatNdf	A2302524K	Gen. gov. - National; Final consumption exp. - Non-defence
GneDfdFceGvtNat	A2302525L	Gen. gov. - National; Final consumption exp.
GneDfdFceGvtSnl	A2302526R	Gen. gov. - State and local; Final consumption exp,
GneDfdFceGvt	A2302527T	Gen. gov.; Final consumption exp.
GneDfdFce	A2302529W	All sectors; Final consumption exp.
GneDfdGfcPvtTdwNnu	A2302543T	Pvt.; Gross fixed capital formation (GFCF)
GneDfdGfcPvtTdwAna	A2302544V	Pvt.; GFCF - Dwellings - Alterations and additions
GneDfdGfcPvtTdw	A2302545W	Pvt.; GFCF - Dwellings - Total
GneDfdGfcPvtOtc	A2302546X	Pvt.; GFCF - Ownership transfer costs
GneDfdGfcPvtPbiNdcNbd	A2302533L	Pvt. GFCF - Non-dwelling construction - New building
GneDfdGfcPvtPbiNdcNec	A2302534R	Pvt.; GFCF - Non-dwelling construction - New engineering construction
GneDfdGfcPvtPbiNdcSha	A2302535T	Pvt.; GFCF - Non-dwelling construction - Net purchase of second hand assets
GneDfdGfcPvtPbiNdc	A2302536V	Pvt.; GFCF - Non-dwelling construction - Total
GneDfdGfcPvtPbiNdmNew	A2302530F	Pvt.; GFCF - Machinery and equipment - New
GneDfdGfcPvtPbiNdmSha	A2302531J	Pvt.; GFCF - Machinery and equipment - Net purchase of second hand assets
GneDfdGfcPvtPbiNdm	A2302532K	Pvt.; GFCF - Machinery and equipment - Total
GneDfdGfcPvtPbiCbr	A2716219R	Pvt.; GFCF - Cultivated biological resources
GneDfdGfcPvtPbiIprRnd	A2716221A	Pvt.; GFCF - Intellectual property products - Research and development
GneDfdGfcPvtPbiIprMnp	A2302539A	Pvt.; GFCF - Intellectual property products - Mineral and petroleum exploration
GneDfdGfcPvtPbiIprCom	A2302538X	Pvt.; GFCF - Intellectual property products - Computer software
GneDfdGfcPvtPbiIprArt	A2302540K	Pvt.; GFCF - Intellectual property products - Artistic originals
GneDfdGfcPvtPbiIpr	A2716220X	Pvt.; GFCF - Intellectual property products Total
GneDfdGfcPvtPbi	A2302542R	Pvt.; GFCF - Total private business investment
GneDfdGfcPvt	A2302547A	Pvt.; GFCF
GneDfdGfcPubPcpCmw	A2302548C	Plc. corporations - Commonwealth; GFCF
GneDfdGfcPubPcpSnl	A2302549F	Plc. corporations - State and local; GFCF
GneDfdGfcPubPcp	A2302550R	Plc. corporations; GFCF Total
GneDfdGfcPubGvtNatDef	A2302551T	Gen. gov. - National; GFCF - Defence
GneDfdGfcPubGvtNatNdf	A2302552V	Gen. gov. - National ; GFCF - Non-defence
GneDfdGfcPubGvtNat	A2302553W	Gen. gov. - National ; GFCF Total
GneDfdGfcPubGvtSnl	A2302554X	Gen. gov. - State and local; GFCF
GneDfdGfcPubGvt	A2302555A	Gen. gov.; GFCF
GneDfdGfcPub	A2302556C	Plc.; GFCF
GneDfdGfc	A2302557F	All sectors; GFCF

Table 11: Variables, Series IDs and their descriptions for Changes in Inventories

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

Table 12: Variables, Series IDs and their descriptions for Household Final Consumption Expenditure

Variable	Series ID	Description
GneDfdHfc	A2302254W	Household Final Consumption Expenditure
GneDfdFceHfcFud	A2302237V	Food
GneDfdFceHfcAbt	A3605816F	Alcoholic beverages and tobacco
GneDfdFceHfcAbtCig	A2302238W	Cigarettes and tobacco
GneDfdFceHfcAbtAlc	A2302239X	Alcoholic beverages
GneDfdFceHfcCnf	A2302240J	Clothing and footwear
GneDfdFceHfcHwe	A3605680F	Housing, water, electricity, gas and other fuels
GneDfdFceHfcHweRnt	A3605681J	Actual and imputed rent for housing
GneDfdFceHfcHweWsc	A3605682K	Water and sewerage charges
GneDfdFceHfcHweEgf	A2302242L	Electricity, gas and other fuel
GneDfdFceHfcFhe	A2302243R	Furnishings and household equipment
GneDfdFceHfcFheFnt	A3605683L	Furniture, floor coverings and household goods
GneDfdFceHfcFheApp	A3605684R	Household appliances
GneDfdFceHfcFheTls	A3605685T	Household tools
GneDfdFceHfcHlt	A2302244T	Health
GneDfdFceHfcHltMed	A3605686V	Medicines, medical aids and therapeutic appliances
GneDfdFceHfcHltHsv	A3605687W	Total health services
GneDfdFceHfcTpt	A3605688X	Transport
GneDfdFceHfcTptPvh	A2302245V	Purchase of vehicles
GneDfdFceHfcTptOvh	A2302246W	Operation of vehicles
GneDfdFceHfcTptTsv	A2302247X	Transport services
GneDfdFceHfcCom	A2302248A	Communications
GneDfdFceHfcRnc	A2302249C	Recreation and culture
GneDfdFceHfcEdc	A2302250L	Education services
GneDfdFceHfcHcr	A2302251R	Hotels, cafes and restaurants
GneDfdFceHfcHcrCsv	A3605694V	Catering services
GneDfdFceHfcHcrAsv	A3605695W	Accommodation services
GneDfdFceHfcMis	A3605696X	Miscellaneous goods and services
GneDfdFceHfcMisOgd	A3605697A	Other goods
GneDfdFceHfcMisIfs	A2302252T	Insurance and other financial services
GneDfdFceHfcMisOsv	A3606485T	Other services

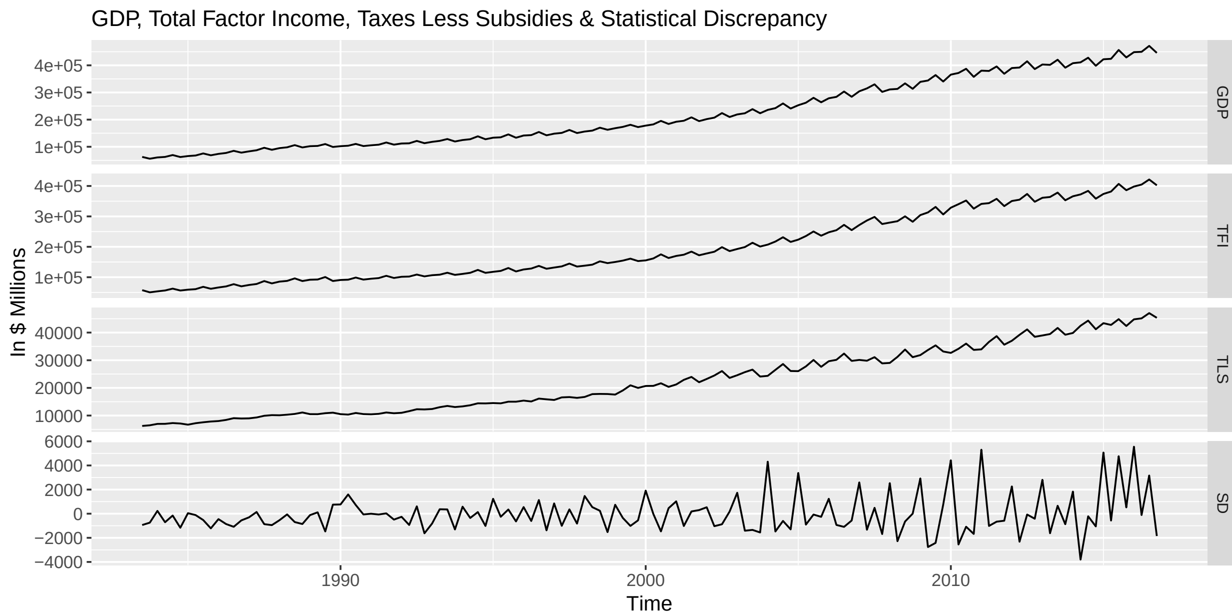


Figure 15: GDP, Total Factor Income, Taxes Less Subsidies and Statistical Discrepancy

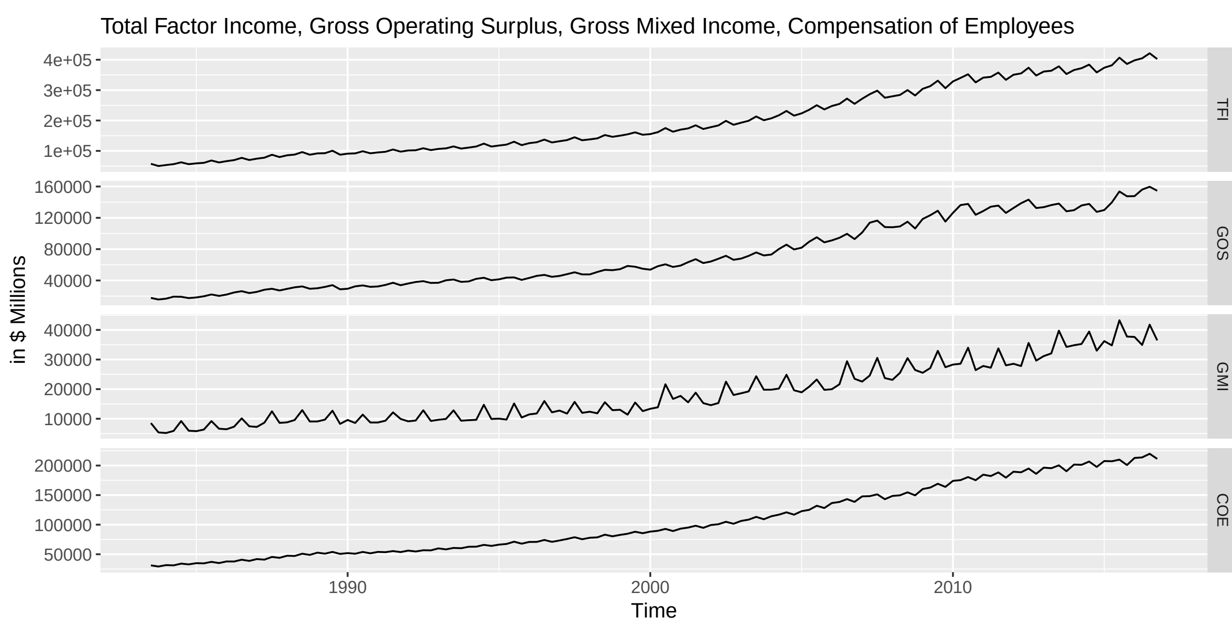


Figure 16: Total Factor Income, Gross Operating Surplus, Gross Mixed Income, Compensation of Employees

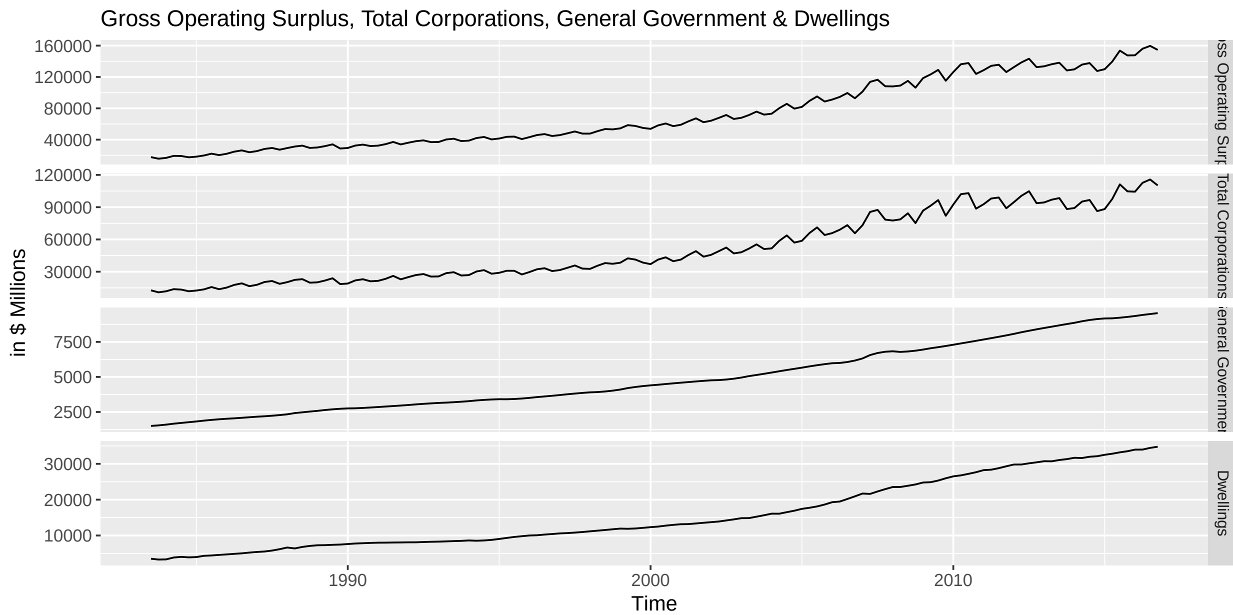


Figure 17: Gross Operating Surplus, Total Corporations, General Government and Dwellings

Figure 18: Compensation of Employees, Wages and salaries, Employers' Social Contributions

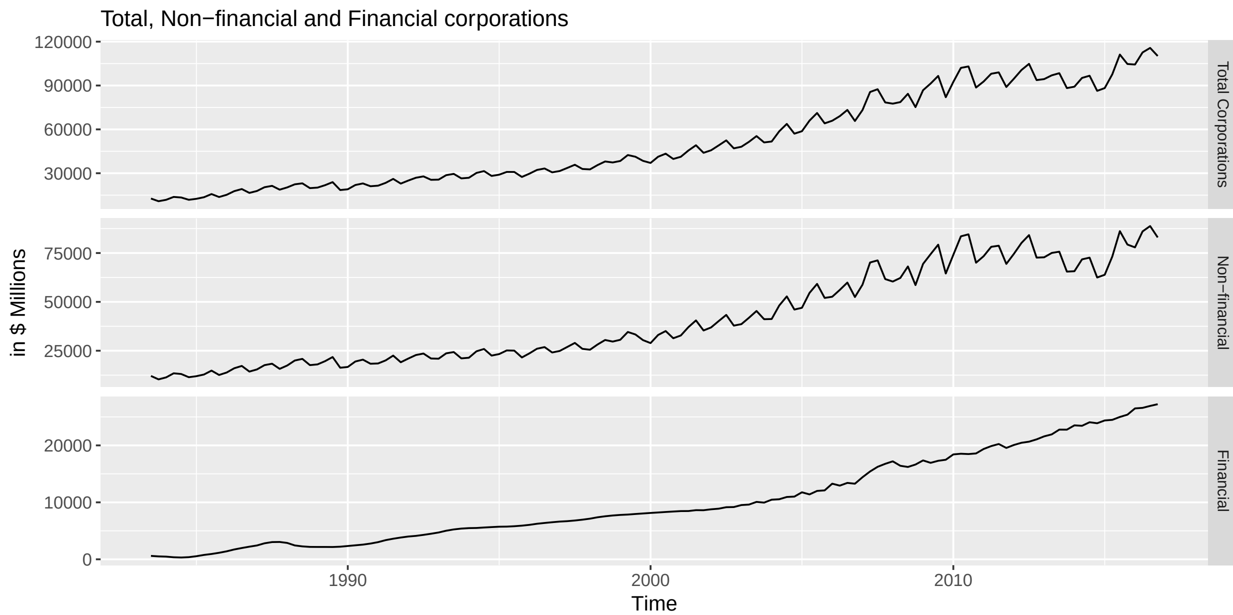


Figure 19: Total, Non-financial and Financial corporations

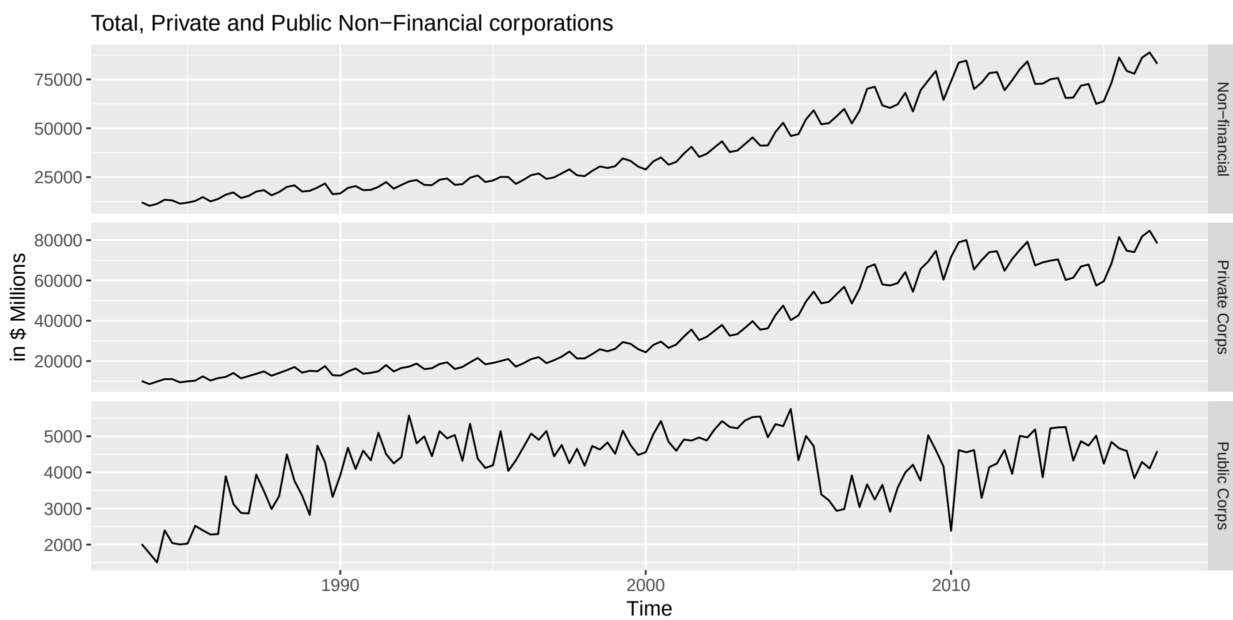
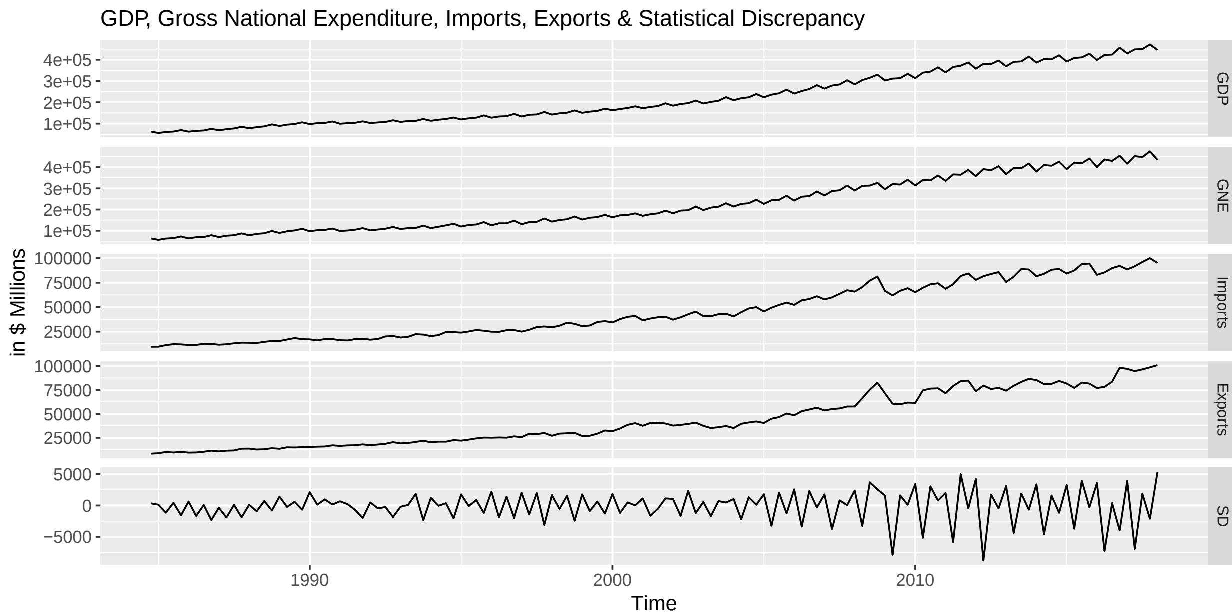
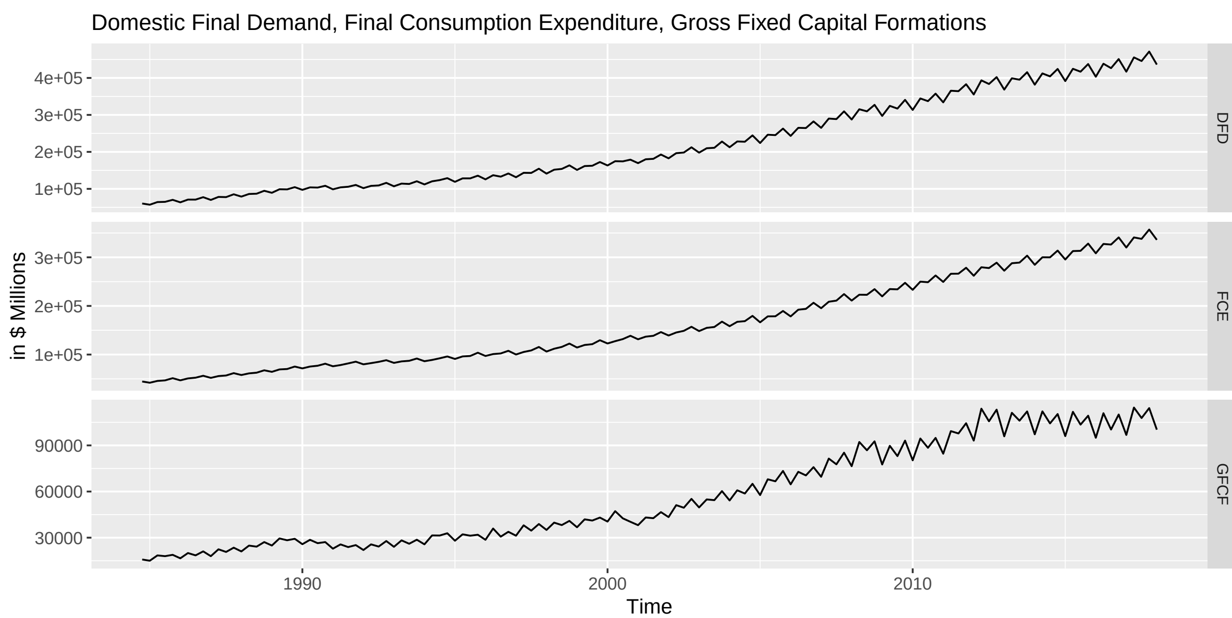
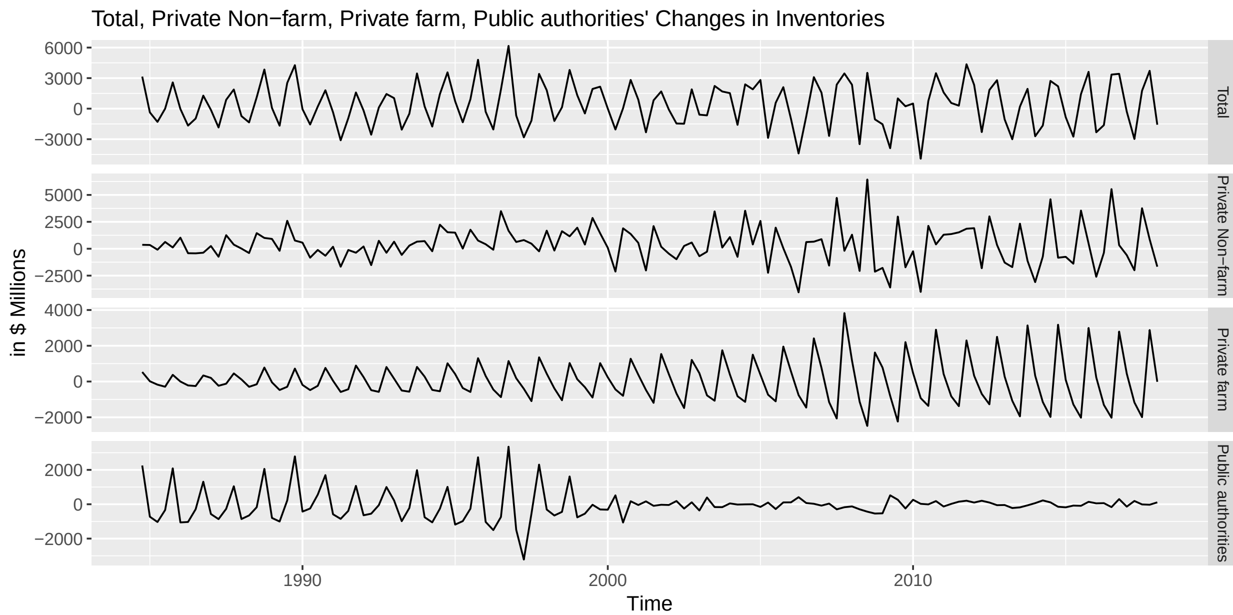
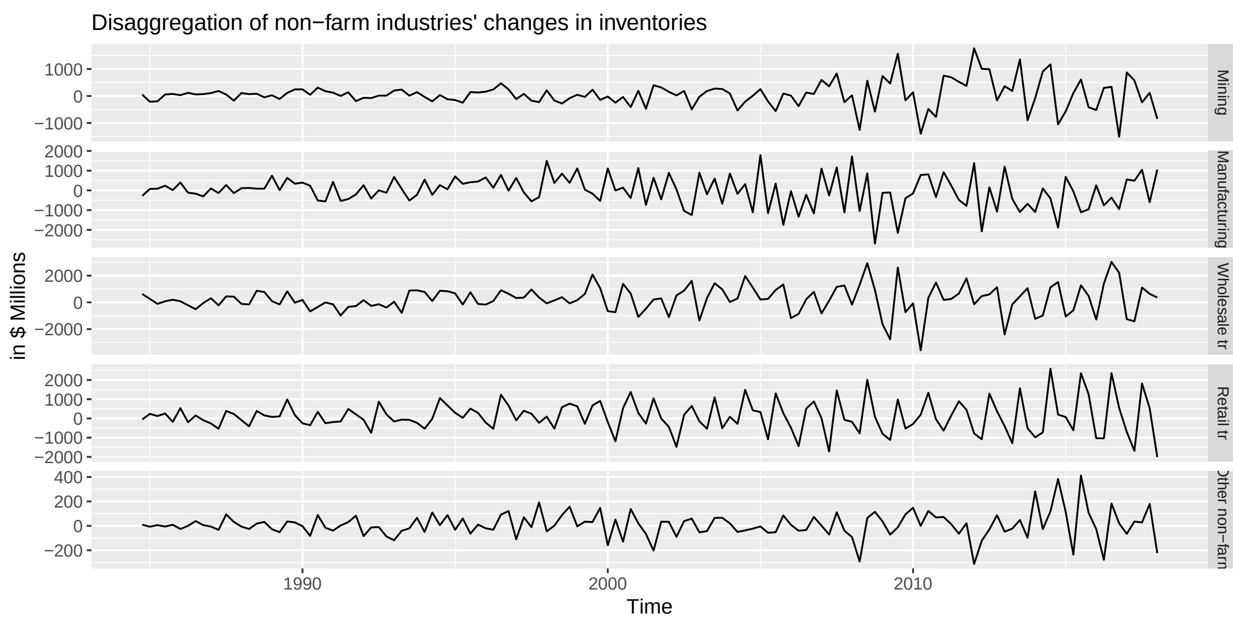


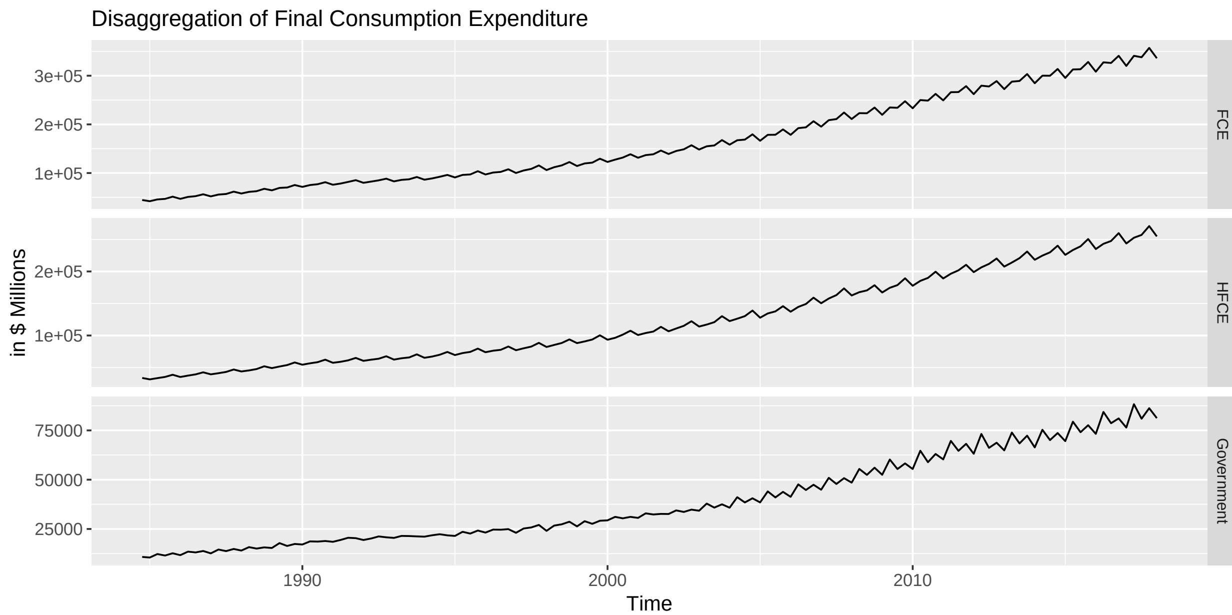
Figure 20: Total, Private and Public Non-Financial corporations

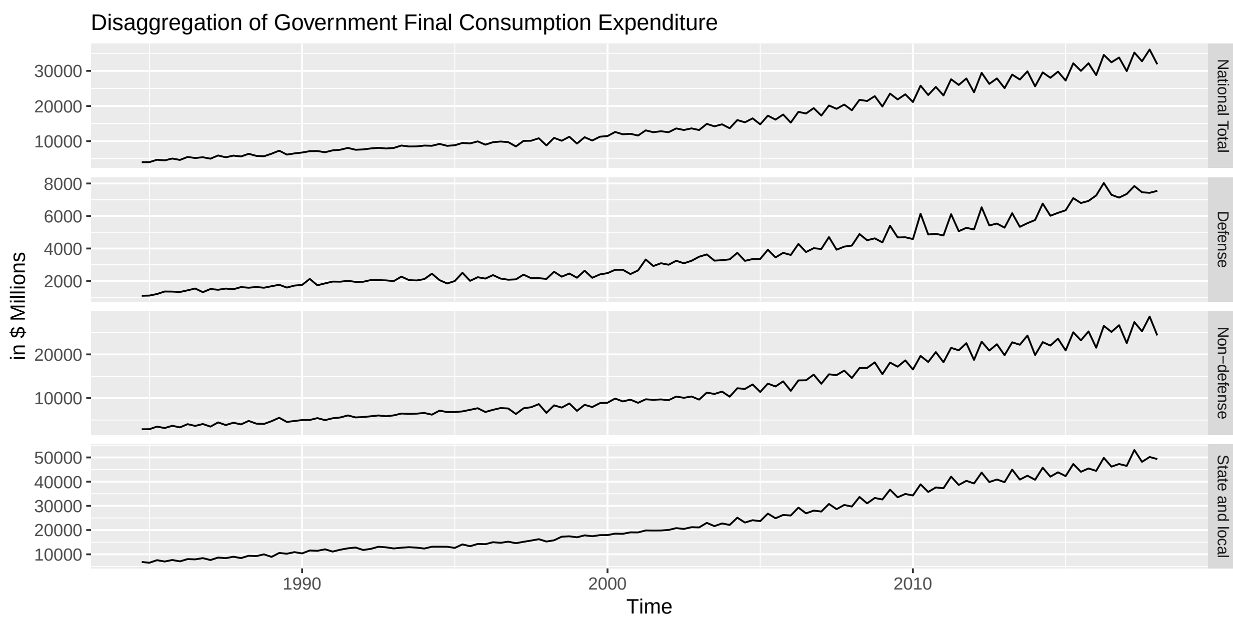

GDP, Gross National Expenditure, Imports, Exports & Statistical Discrepancy
in $ Millions
Time
GDP
GNE
Imports
Exports
SD
1990
2000
2010


Domestic Final Demand, Final Consumption Expenditure, Gross Fixed Capital Formations
in $ Millions
Time
DFD
FCE
GFCF
1990
2000
2010

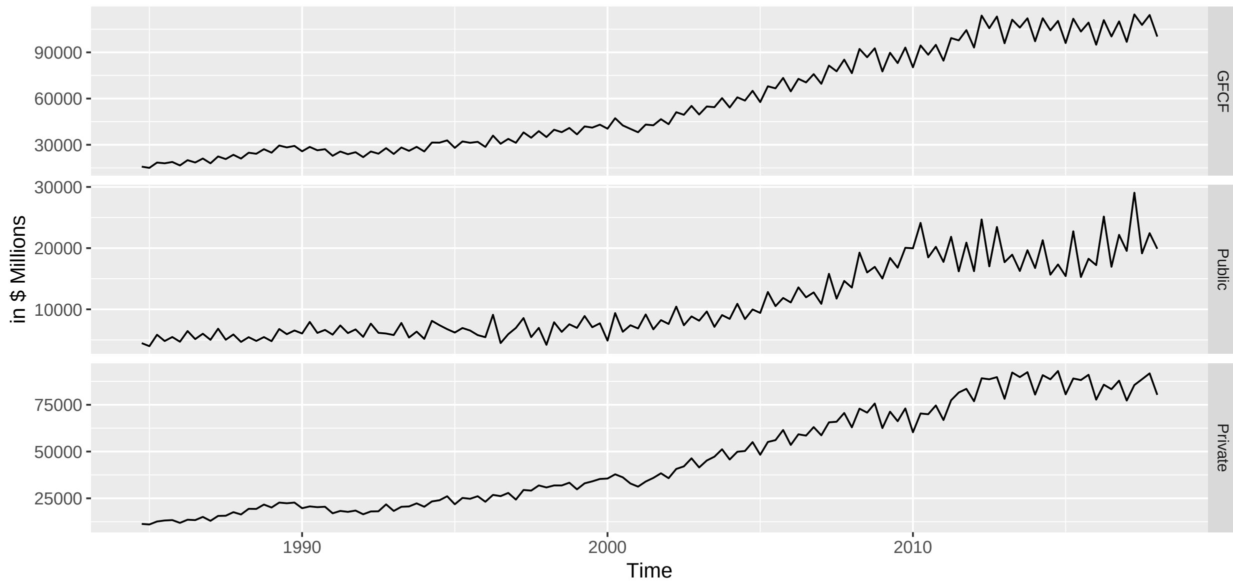

Total, Private Non−farm, Private farm, Public authorities' Changes in Inventories
in $ Millions
Total
Private Non−farm
Private farm
Public authorities
1990
2000
2010
Time


Disaggregation of non−farm industries' changes in inventories
in $ Millions
Mining
Manufacturing
Wholesale tr
Retail tr
1990
2000
2010
Time

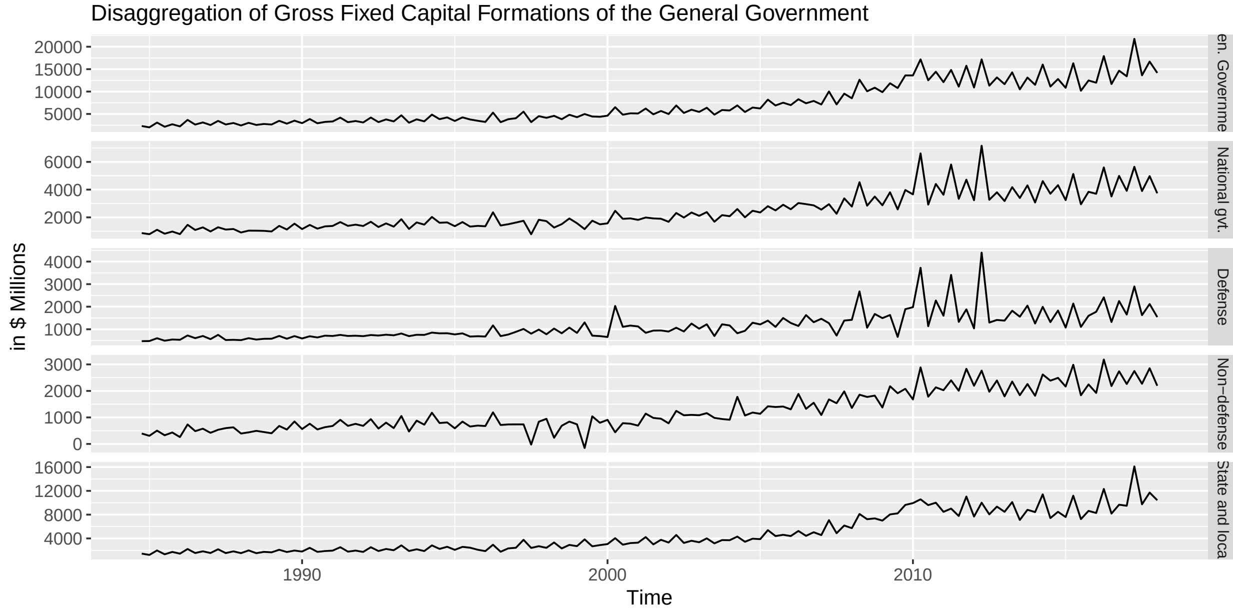

Disaggregation of Final Consumption Expenditure
in $ Millions
3e+05
2e+05
1e+05
FCE
2e+05
1e+05
HFCE
75000
50000
25000
Government
1990
2000
2010
Time


Disaggregation of Government Final Consumption Expenditure
in $ Millions
30000
20000
10000
National Total
8000
6000
4000
2000
Defense
20000
10000
Non−defense
50000
40000
30000
20000
10000
State and local
1990
2000
2010
Time

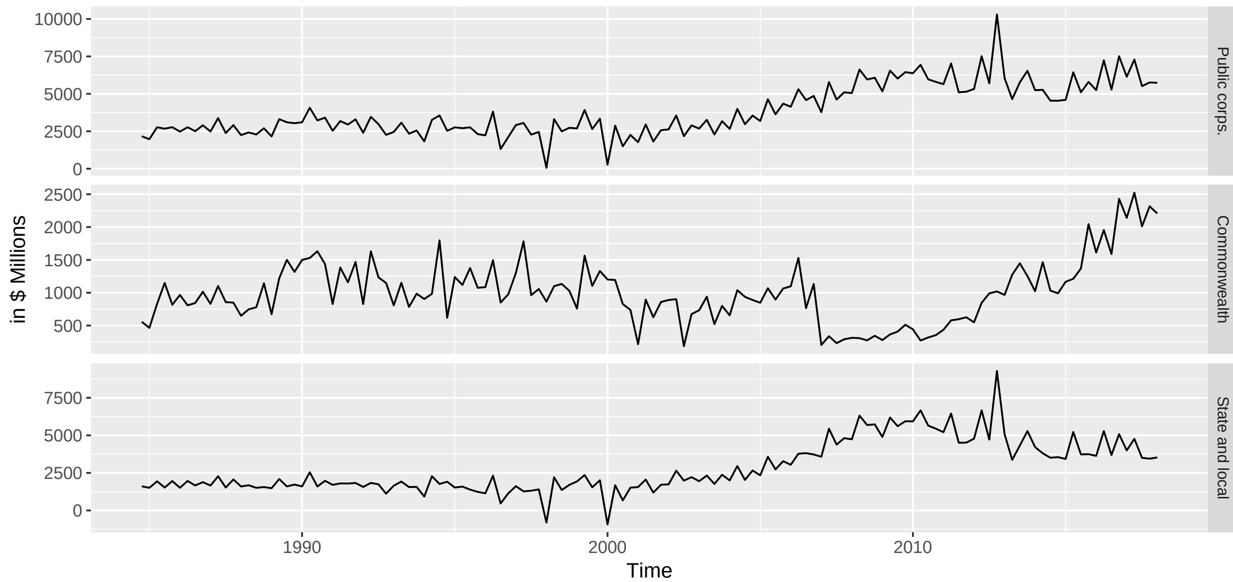
Disaggregation of Gross Fixed Capital Formations



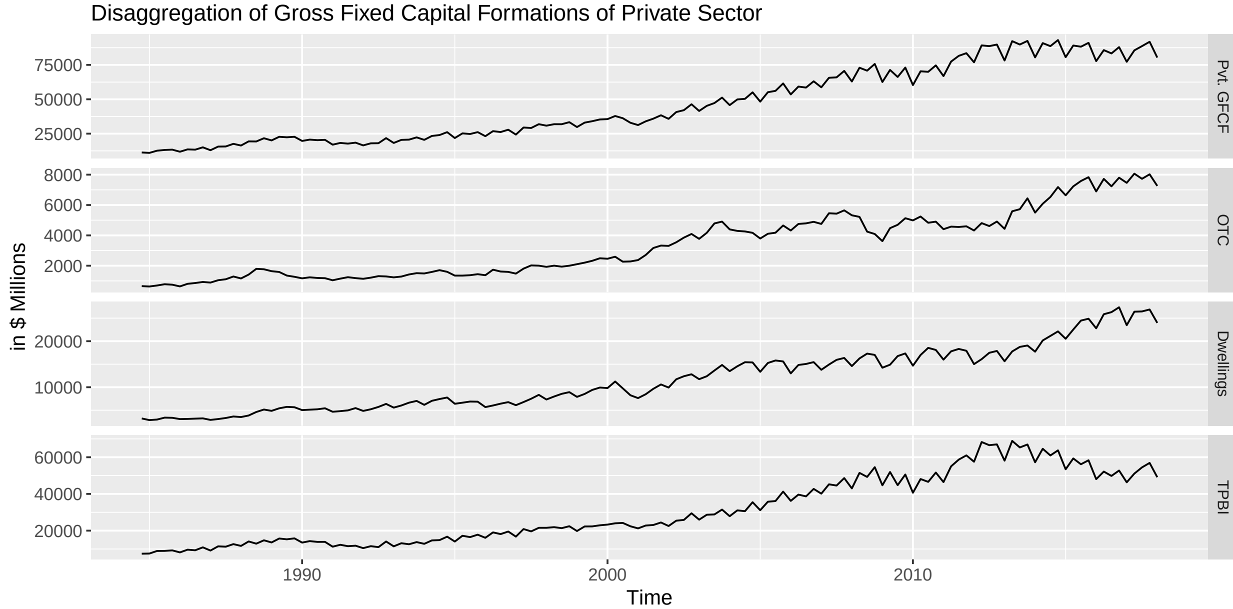
Disaggregation of Gross Fixed Capital Formations of the General Government

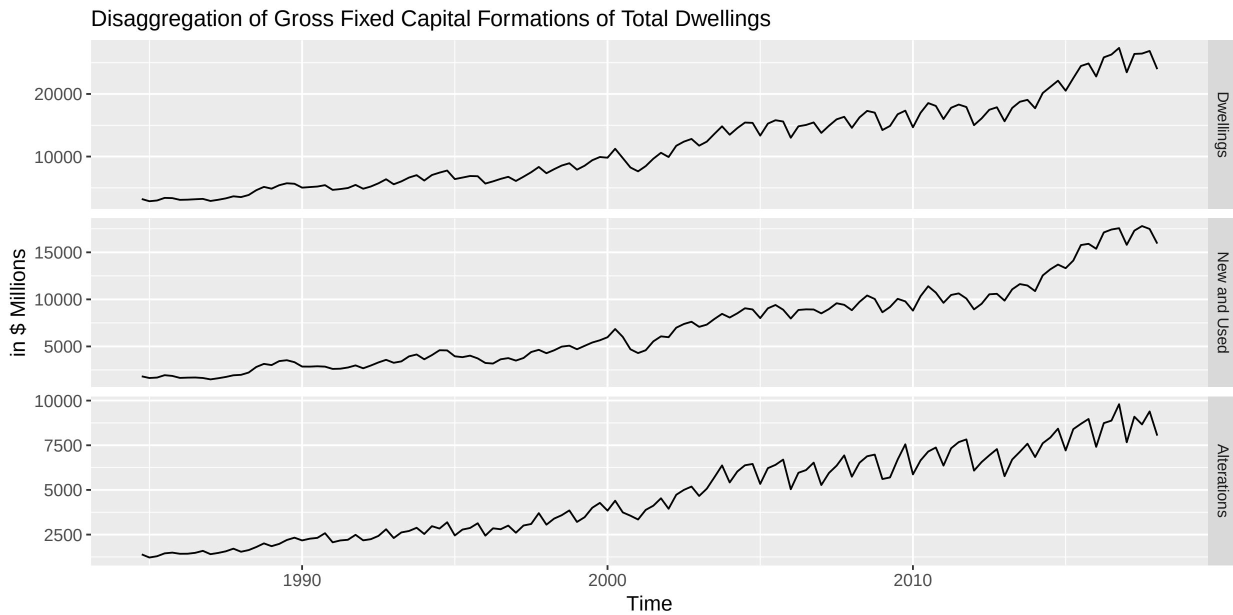


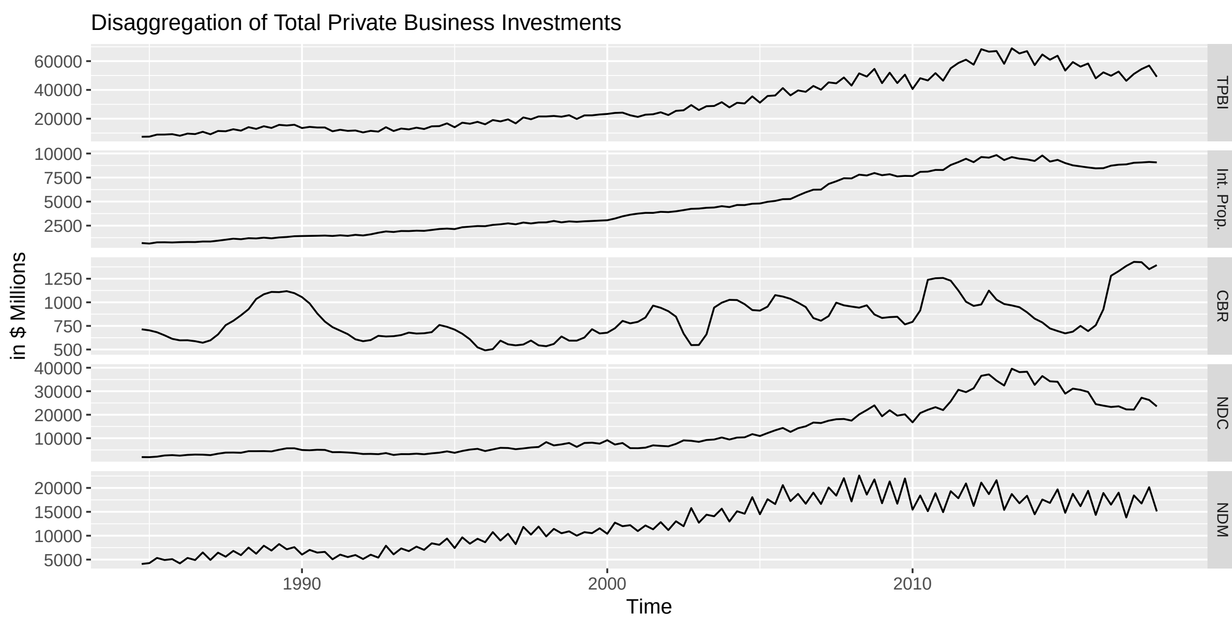
Disaggregation of Gross Fixed Capital Formations of Public Corporations

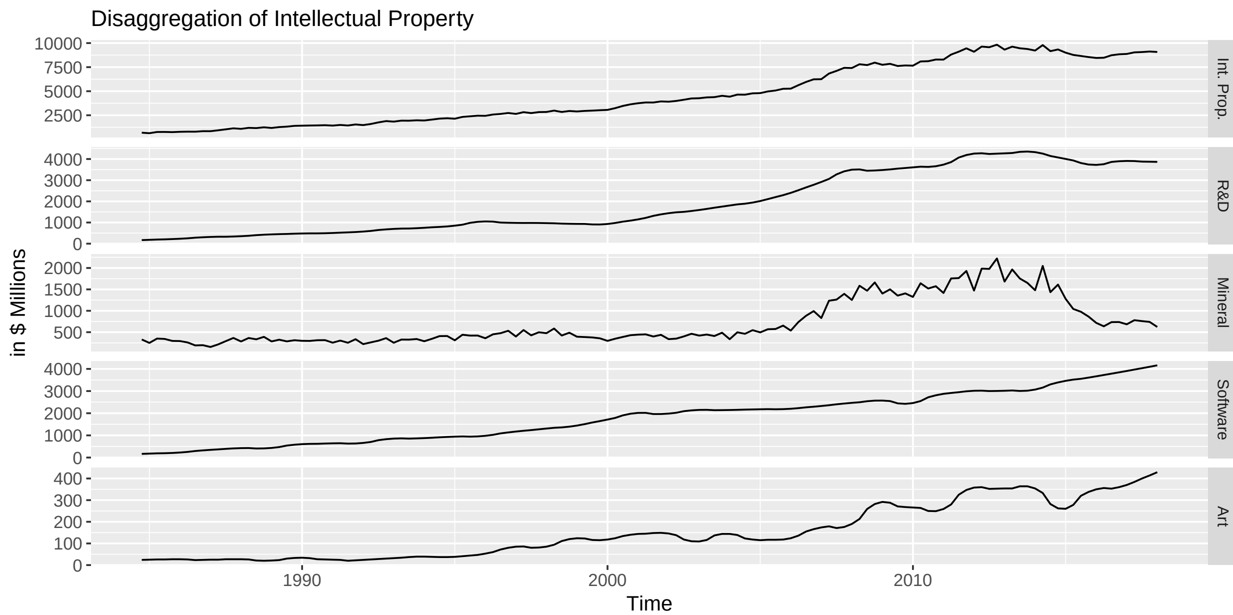


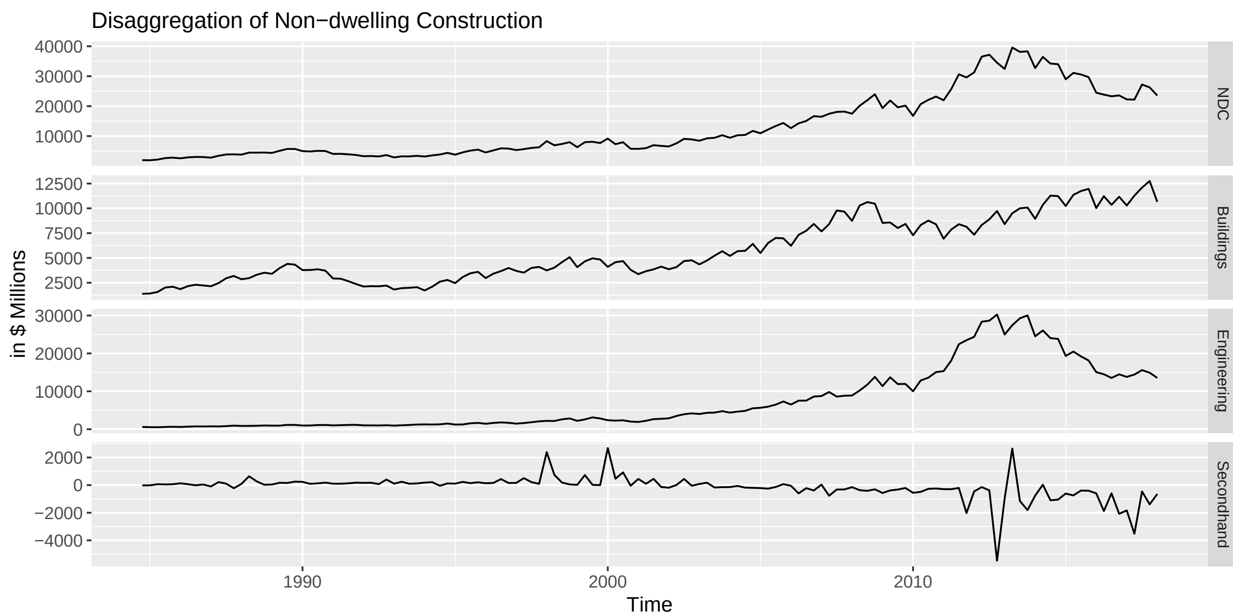
Disaggregation of Gross Fixed Capital Formations of Private Sector




Disaggregation of Gross Fixed Capital Formations of Total Dwellings
in $ Millions
Dwellings
New and Used
Alterations
Time
1990
2000
2010


Disaggregation of Total Private Business Investments
in $ Millions
TPBI
Int. Prop.
CBR
NDC
NDM
Time
1990
2000
2010


Disaggregation of Intellectual Property
in $ Millions
Int. Prop.
R&D
Mineral
Software
Art
Time
1990
2000
2010


Disaggregation of Non−dwelling Construction
in $ Millions
NDC
Buildings
Engineering
Secondhand
Time
1990
2000
2010

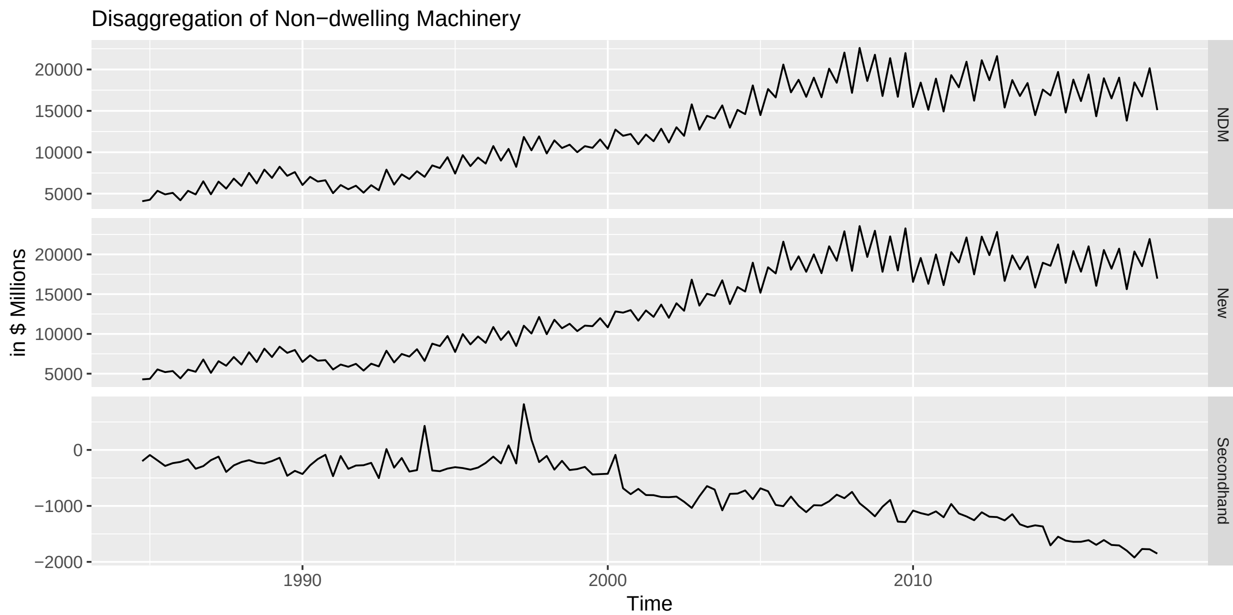


Table 13: Income Approach: Mean Absolute Scaled Error for the most aggregate level (GDP)

F-method	**R-method**	**1**	**2**	**3**	**4**
ARIMA	MinT(Shrink)	-57.259	-38.917	-26.34	-21.538
ARIMA	MinT(Sample)	-56.889	-37.339	-24.763	-21.909
ARIMA	WLS	-56.767	-41.042	**-28.718**	-20.906
ARIMA	Bottom-up	-53.035	-39.537	-23.469	-14.693
ARIMA	Base	**-58.487**	-40.205	-27.974	**-22.293**
ETS	MinT(Shrink)	-53.698	-39.66	-25.095	-19.871
ETS	MinT(Sample)	-50.129	-38.408	-24.46	-19.353
ETS	WLS	-55.711	**-42.709**	-24.884	-15.49
ETS	Bottom-up	-53.53	-38.9	-18.909	-4.577
ETS	Base	-56.002	-40.963	-22.3	-17.379

Table 14: Income Approach: Average Mean Absolute Scaled Error for the whole hierarchy

F-method	R-method	1	2	3	4
ARIMA	MinT(Shrink)	0.6896	0.9286	1.1065	1.2104
ARIMA	MinT(Sample)	0.7208	0.9669	1.149	1.2373
ARIMA	WLS	0.5604	0.7389	**0.8967**	**0.9841**
ARIMA	Bottom-up	0.5633	0.7386	0.9033	0.9917
ARIMA	Base	0.5666	0.7425	0.8984	0.9909
Benchmark	MinT(Shrink)	0.9649	0.9912	1.0133	1.0312
Benchmark	MinT(Sample)	0.9649	0.9912	1.0133	1.0312
Benchmark	WLS	0.9649	0.9912	1.0133	1.0312
Benchmark	Bottom-up	0.9649	0.9912	1.0133	1.0312
Benchmark	Base	0.9649	0.9912	1.0133	1.0312
ETS	MinT(Shrink)	0.681	0.9116	1.0823	1.1675
ETS	MinT(Sample)	0.7187	0.9409	1.0968	1.1965
ETS	WLS	**0.5427**	**0.7307**	0.9074	0.9994
ETS	Bottom-up	0.5596	0.7409	0.9197	1.0147
ETS	Base	0.5574	0.7397	0.9121	1.0071

Table 15: Income Approach: Median Absolute Scaled Error for the most aggregate level (GDP)

F-method	R-method	1	2	3	4
ARIMA	MinT(Shrink)	**-59.135**	-42.425	-26.978	-28.696
ARIMA	MinT(Sample)	-56.352	-39.99	**-28.593**	-20.425
ARIMA	WLS	-55.573	-43.031	-23.83	-27.469
ARIMA	Bottom-up	-44.125	-44.841	-17.142	2.612
ARIMA	Base	-53.672	-39.362	-30.985	-22.435
ETS	MinT(Shrink)	-49.57	-36.143	-31.2	**-30.46**
ETS	MinT(Sample)	-43.868	-38.107	-26.89	-27.193
ETS	WLS	-50.307	**-45.425**	-28.35	-7.387
ETS	Bottom-up	-51.649	-43.935	-14.321	24.603
ETS	Base	-51.269	-43.079	-25.475	-21.834

Table 16: Expenditure Approach: Mean Absolute Scaled Error for the most aggregate level (GDP)

F-method	R-method	1	2	3	4
ARIMA	MinT(Shrink)	-53.39	-31.053	-15.2	-7.419
ARIMA	WLS	**-56.586**	-35.245	-17.871	-9.462
ARIMA	Bottom-up	-42.849	-24.491	-10.228	-3.289
ARIMA	Base	-56.15	-33.809	-15.214	-6.948
ETS	MinT(Shrink)	-55.96	**-36.589**	**-22.988**	**-13.956**
ETS	WLS	-55.888	-34.869	-14.311	-1.521
ETS	Bottom-up	-50.299	-30.958	-18.608	-9.715
ETS	Base	-56.094	-34.83	-14.345	-1.515

Table 17: Expenditure Approach: Average Mean Absolute Scaled Error for the whole hierarchy

F-method	R-method	1	2	3	4
ARIMA	MinT(Shrink)	1.0722	1.2452	1.4049	1.5462
ARIMA	WLS	0.7762	0.9428	**1.0921**	**1.214**
ARIMA	Bottom-up	0.7944	0.959	1.1022	1.2246
ARIMA	Base	0.7923	0.958	1.1069	1.2281
Benchmark	MinT(Shrink)	1.195	1.2237	1.2494	1.2702
Benchmark	WLS	1.195	1.2237	1.2494	1.2703
Benchmark	Bottom-up	1.195	1.2237	1.2494	1.2702
Benchmark	Base	1.195	1.2237	1.2494	1.2703
ETS	MinT(Shrink)	0.9867	1.192	1.4235	1.5623
ETS	WLS	0.9591	1.2092	1.4528	1.6098
ETS	Bottom-up	**0.765**	**0.933**	1.0993	1.2253
ETS	Base	0.7705	0.9437	1.1123	1.2443

Table 18: Expenditure Approach: Median Absolute Scaled Error for the most aggregate level (GDP)

F-method	R-method	1	2	3	4
ARIMA	MinT(Shrink)	-53.545	-37.487	-16.631	-16.612
ARIMA	WLS	-59.698	-50.631	-29.415	-26.72
ARIMA	Bottom-up	-47.114	-35.196	-17.567	-11.292
ARIMA	Base	-59.788	-46.734	-31.291	**-40.235**
ETS	MinT(Shrink)	-59.098	**-56.528**	-22.475	-18.973
ETS	WLS	-59.195	-50.261	-29.845	-37.719
ETS	Bottom-up	-56.092	-50.671	-24.217	-10.765
ETS	Base	**-59.987**	-50.261	**-34.127**	-37.72

Table 19: Income approach: Average Mean Absolute Scaled Error for the most aggregate level (GDP)

R-method	ARIMA	Benchmark	ETS
MinT(Shrink)	0.4513	0.7023	0.4613
MinT(Sample)	0.4568	0.7023	0.4716
WLS	0.4453	0.7023	0.4608
Bottom-up	0.4748	0.7023	0.5014
Base	**0.4427**	0.7023	0.4645

Table 20: Income approach: Average Mean Absolute Scaled Error for the whole hierarchy

R-method	ARIMA	Benchmark	ETS
MinT(Shrink)	0.9814	0.9998	0.9584
MinT(Sample)	1.0162	0.9998	0.9861
WLS	0.7931	0.9998	**0.793**
Bottom-up	0.7973	0.9998	0.8066
Base	0.7977	0.9998	0.802

Table 21: Income approach: Average Median Absolute Scaled Error for the most aggregate level (GDP)

R-method	ARIMA	Benchmark	ETS
MinT(Shrink)	**0.3075**	0.5503	0.3340
MinT(Sample)	0.3357	0.5504	0.3495
WLS	0.3258	0.5503	0.3444
Bottom-up	0.3733	0.5503	0.3888
Base	0.3165	0.5503	0.3381

Table 22: Income approach: Average Median Absolute Scaled Error for the whole hierarchy

R-method	ARIMA	Benchmark	ETS
MinT(Shrink)	0.7216	0.8015	0.7247
MinT(Sample)	0.7346	0.8015	0.7592
WLS	0.5971	0.8015	**0.5849**
Bottom-up	0.6018	0.8015	0.6073
Base	0.6079	0.8015	0.6035

Table 23: Expenditure approach: Average Mean Absolute Scaled Error for the most aggregate level (GDP)

R-method	ARIMA	Benchmark	ETS
MinT(Shrink)	0.4927	0.7023	0.4619
WLS	0.448	0.7023	0.4658
Bottom-up	0.5436	0.7023	0.4934
Base	**0.4427**	0.7023	0.4645

Table 24: Expenditure approach: Average Mean Absolute Scaled Error for the whole hierarchy

R-method	ARIMA	Benchmark	ETS
MinT(Shrink)	1.3150	1.2342	1.2885
WLS	1.0043	1.2342	1.3048
Bottom-up	1.0181	1.2342	**1.0036**
Base	1.0194	1.2342	1.0155

Table 25: Expenditure approach: Average Median Absolute Scaled Error for the most aggregate level (GDP)

R-method	ARIMA	Benchmark	ETS
MinT(Shrink)	0.3887	0.5503	0.3635
WLS	0.3291	0.5503	0.3381
Bottom-up	0.4447	0.5502	0.3741
Base	**0.3165**	0.5503	0.3381

Table 26: Expenditure approach: Average Median Absolute Scaled Error for the whole hierarchy

R-method	ARIMA	Benchmark	ETS
MinT(Shrink)	1.0141	0.9742	0.9787
WLS	**0.7504**	0.9738	0.9549
Bottom-up	0.7679	0.9737	0.7540
Base	0.7659	0.9738	0.7660

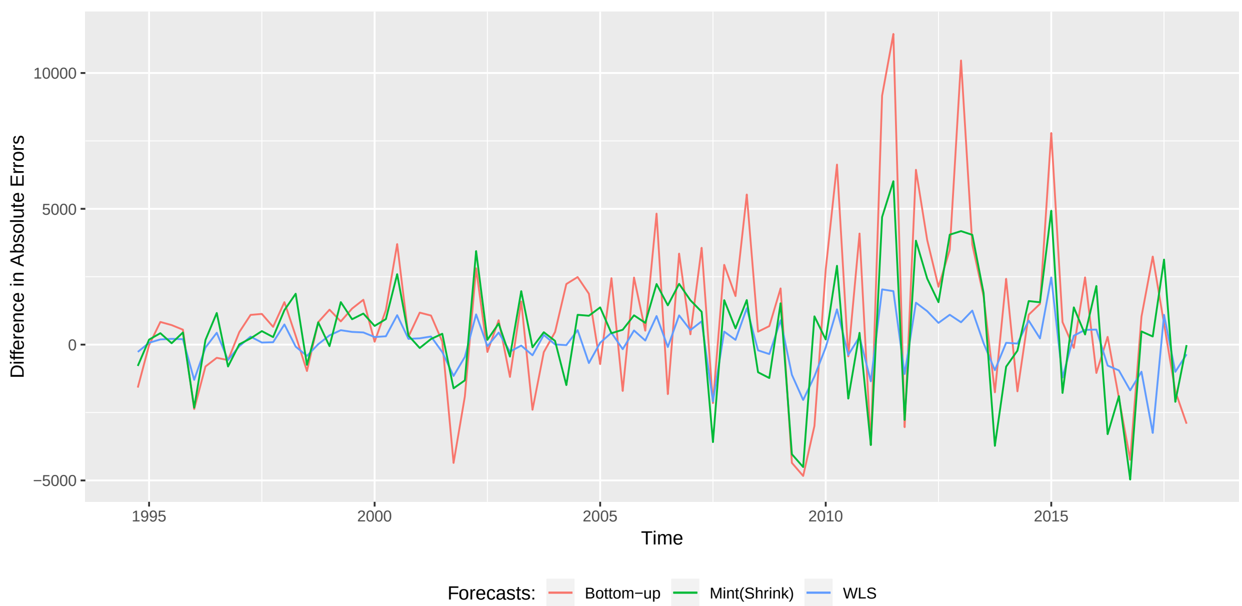


Figure 21: Difference of Absolute Errors of Reconciled ARIMA compared to Base ARIMA forecasts (Expenditure approach)

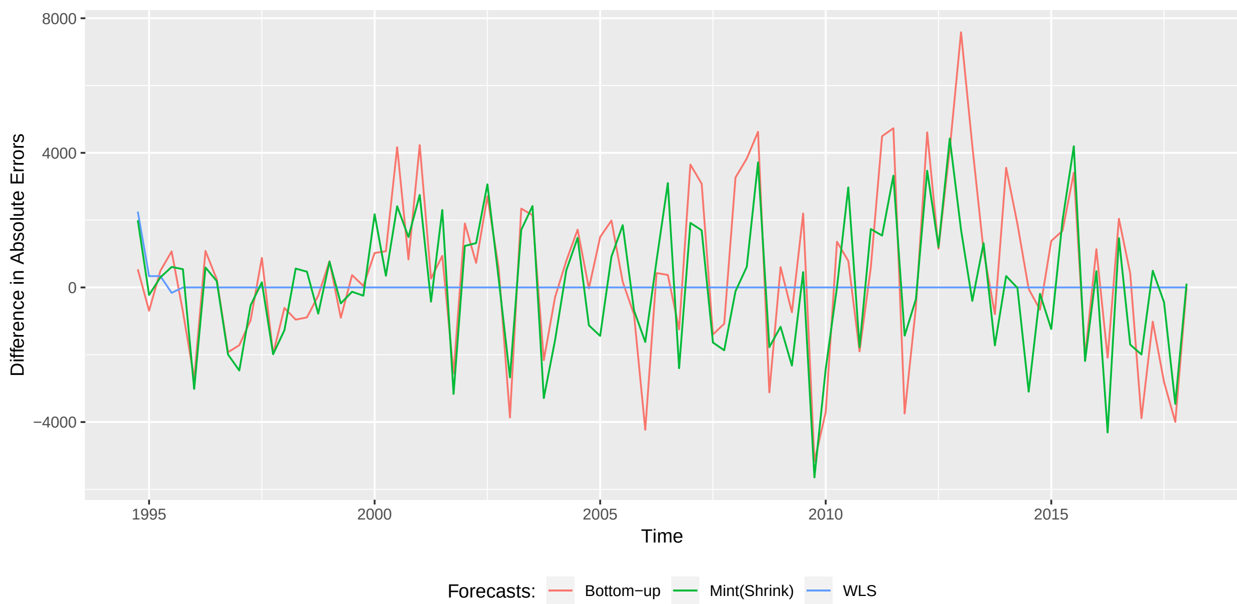


Figure 22: Difference of Absolute Errors of Reconciled ETS compared to Base ETS forecasts (Expenditure approach)

References

Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Petropoulos, F. (2017). Forecasting with temporal hierarchies, *262*(1), 60–74. doi:10.1016/j.ejor.2017.02.046

Australian Bureau of Statistics. (2015). *Australian System of National Accounts: Concepts, Sources and Methods (Catalogue: 5216.0).* ABS. Retrieved from http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/5216.02015?OpenDocument

Australian Bureau of Statistics. (2018). 5206.0 - Australian National Accounts: National Income, Expenditure and Product: Explanatory Notes. c=AU; o=Commonwealth of Australia; ou=Australian Bureau of Statistics. Retrieved from http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/5206.0Explanatory Notes1Mar 2018?OpenDocument

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2nd ed.). OTexts. Retrieved from https://otexts.org/fpp2/

Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, *27*(3). doi:10.18637/jss.v027.i03

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*. doi:10.1016/j.ijforecast.2006.03.001

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics and Data Analysis*, *55*(9), 2579–2589. doi:10.1016/j.csda.2011.03.006

Hyndman, R. J., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., … Zhou, Z. (2018). Package 'forecast': Forecasting Functions for Time Series and Linear Models. Retrieved from https://cran.r-project.org/web/packages/forecast/index.html

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

Hyndman, R. J., Lee, A. J., Wang, E., & Wickramasuriya, S. L. (2018). hts: Hierarchical and Grouped Time Series. Retrieved from https://cran.r-project.org/web/packages/hts/index.html

Panagiotelis, A., Athanasopoulos, G., Hyndman, R. J., Jiang, B., & Vahid, F. (2018). *Macroeconomic forecasting for Australia using a large number of predictors*. Monash University, Department of Econometrics; Business Statistics. Retrieved from http://business.monash.edu/econometrics-and-business-statistics/research/publications

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Schäfer, J., & Strimmer, K. (2005). A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics. *Statistical Applications in Genetics and Molecular Biology*, *4*(1).

Shlifer, E., & Wolff, R. W. (1979). Aggregation and Proration in Forecasting. *Management Science*, *25*(6), 594–603.

Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J. (2018). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association*, (just-accepted), 1–45. doi:10.1080/01621459.2018.1448825