Hierarchical Time Series Forecasting in the Emergency Departments of NHS Wales

Janice Hsu

2023-10-30

Contents

1 Abstract

1.1 Introduction

National Health Service (NHS) Wales employs several time series forecasting techniques. In order to support the processes for planning and decision-making, both strategically (at All Wales level) and operationally (at local health board level). However, using base forecasting to analyse individual time series and the aggregated level can cause inconsistencies. Therefore, hierarchical time series forecasting method is to be used in this study.

1.2 Background

The COVID-19 outbreak created unexpected difficulties of the forecasting for the attendance of Emergency Department in NHS-Wales. Therefore, NHS-Wales is seeking a reliable and consistent forecasting model regardless of the shifting patterns of patients' arrivals during the lockdowns and health policy changes.

1.3 Goal

This project's main goal is to create a forecasting models for emergency departments that can consistently deliver outcomes at the individual hospital, each local health board, and national levels. The research will allow for a comparison of:

1. Forecasting national time series and with local time series combined.
2. Forecasting time series hierarchically.

By doing this, I hope to create a system that synchronises policymakers' projections with operational forecasts, guaranteeing an efficient decision-making process.

1.4 Techniques

The data is obtained from the StatsWales website and it is publicly accessible.

Data source: https://statswales.gov.wales/Catalogue/Health-and-Social-Care/NHS-Hospital-Waiting-Times/emergency-department/emergencyattendances-by-age-sex-site

The data will be split up into the subsequent sections:

I'll use the following forecasting techniques:

- Base forecasting: I use base forecasting mainly for the attendance under the different local health boards, in order to conduct the exploratory analysis and also to determine which model is a better fit for the data.
- Hierarchical Time Series Forecasting: This method models the data in accordance with the underlying hierarchy in the data (national(Wales), local health boards, hospital).

1.5 Differences between base forecasting and hierarchical forecasting

- Base forecasting: It is used to forecast simple time series data with statistical or machine learning methods. To generate future data, base forecasting observes the historical data, and finds its pattern, trend and seasonality to produce future projections.

- Reconciliation Forecasting (Hierarchical Forecasting): It is used when dealing with hierarchical or grouped time series data. Hierarchical forecasting involves generating forecast on different levels of aggregation. It doesn't only create projections for individual time series, but also reconciles the forecasts to guarantee consistency under multiple hierarchical levels. This method exploits a variety of strategies, such as top down, bottom up, to ensure that the individual projections align with their corresponding aggregated levels. This leads the hierarchical forecasting advantageous due to the fact that it can make sure the consistent and consolidated projections at all levels.

1.6   Anticipated Results

1. Determine the most reliable and accurate forecasting method for the attendance number for emergency departments in Wales.
2. Provide evidence for why hierarchical time series methods are superior than other forecasting methods.
3. Offer a scalable, transparent method that can be used by other NHS departments.

2   Data Introduction

The dataset contains variables related to monthly Emergency Attendances in hospitals in Wales, UK. The data was publicly available and retrieved from StatsWales.
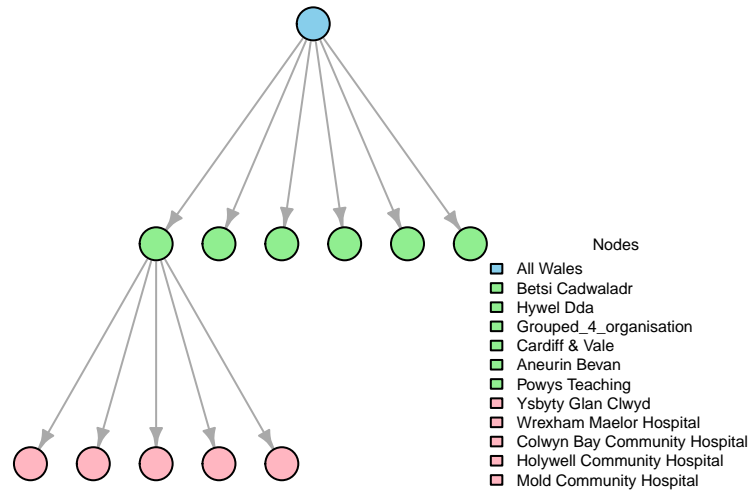
- Data: This column represents the aggregated number of attendances in each emergency department. The data was aggregated according to the other columns.

- YearMonth: This column represents data from 2012 April to 2023 June (135 time points).

- Age_Code: This column provides the patient's age group. There are 17 different age groups: "0 to 4","5 to 17","18 to 24","25 to 29","30 to 34","35 to 39","40 to 44","45 to 49","50 to 54","55 to 59","60 to 64","65 to 69","70 to 74","75 to 79","80 to 84","85" and "Unknown".

- Sex_ItemName_ENG: This column provides patient's gender using 3 categories: "Female", "Male" and "Unknown".

- Hospital_Code: This column represents 42 different hospitals in Wales.

- Hospital_ItemName_ENG: This columns refers to the name of the 42 different hospitals in Wales.

- Hospital_Hierarchy: This column represents the code for the health board that the hospital belongs to.

- Hospital_AltCode1: This column provides an alternate code for the hospital.

- Organisation: This column represents the Local Health Board (LHB), which are responsible for planning and delivering NHS Wales services in their areas, including Emergency Care. In Wales there are 7 distinct LHBs, however two of them, Cwm Taf Morgannwg and Swansea Bay, were previously (before 1 st of April 2019) known as Cwm Taf and Abertawe Bro Morgannwg (therefore this column contains 9 distinct values).

- Organisation_Code: A code for the organisation as well as the LHB.

There are three hierarchies in this dataset: - On the top level, there is all the hospitals in Wales. - On the second level, there are the different LHB, which are "Betsi Cadwaladr","Hywel Dda", "Swansea Bay", "Abertawe Bro Morgannwg", "Cardiff & Vale","Cwm Taf Morgannwg", "Cwm Taf", "Aneurin Bevan" and "Powys Teaching". They will be grouped into 6 LHB later on, I will provide explanation in the subsequent section.

- At the bottom level, there are 42 hospitals.

## 2.1   Introduce the hierarchical structure with graph

(For simplicity, I only put 5 hosptials at the bottom level, however there are 42 in total)

Nodes
- All Wales
- Betsi Cadwaladr
- Hywel Dda
- Grouped_4_organisation
- Cardiff & Vale
- Aneurin Bevan
- Powys Teaching
- Ysbyty Glan Clwyd
- Wrexham Maelor Hospital
- Colwyn Bay Community Hospital
- Holywell Community Hospital
- Mold Community Hospital

3 Exploratory Data Analysis

3.1 Number of patients entering ED under different hospital hierarchy

Number of patients who enter ED
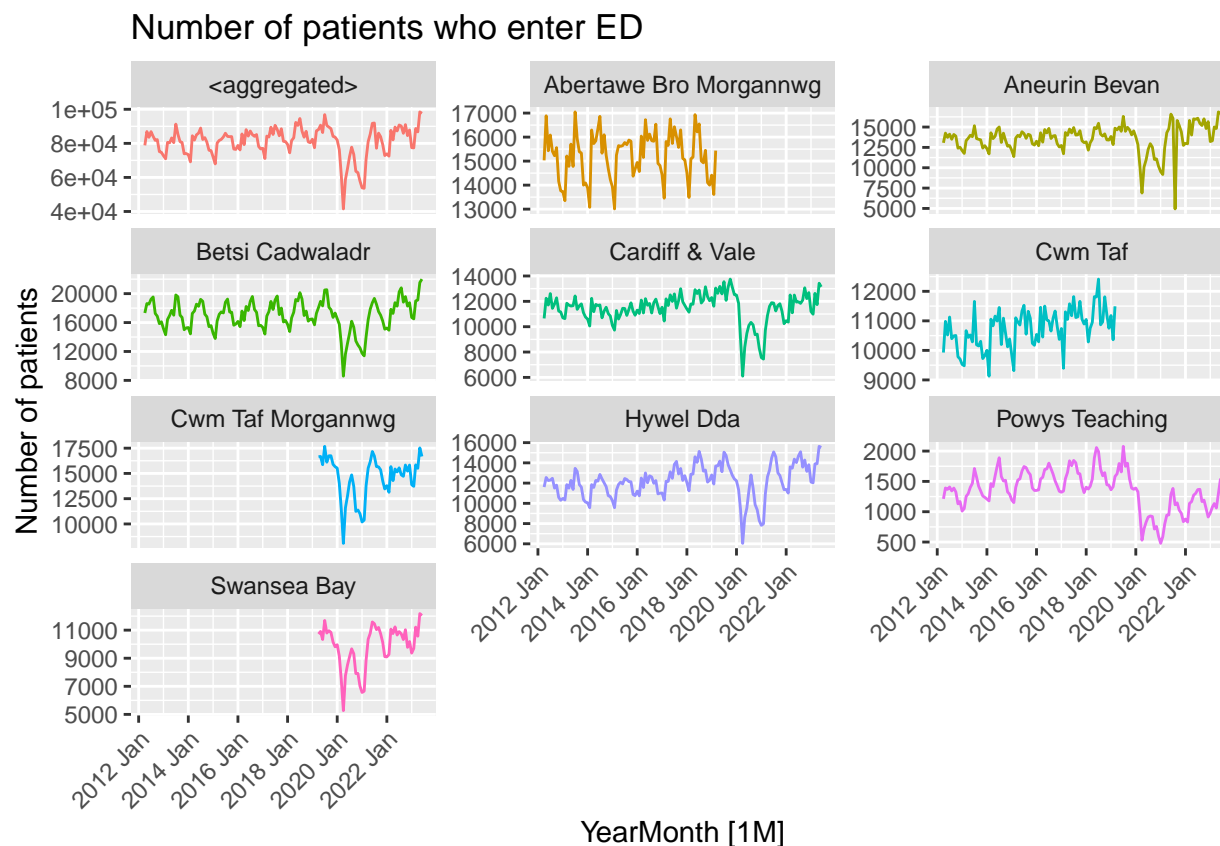
Explanation for the change of local health boards (LHB)

As already mentioned, two LHBs were redefined from the 1st of April 2019 onwards: Cwm Taf ((# hospitals = 27)–> Cwm Taf Morgannwg ((# hospitals = 30)// Abertawe Bro Morgannwg ((# hospitals = 26) –> Swansea Bay ((# hospitals = 31).

Since the number of hospitals included before and after the change in each LHB was different, these 4 organisations were aggregated into one unique group to enable their inclusion in the forecasting analysis (defined as "Grouped_4_organisation"in the code snippets).

3.2 Group the changed Local Health Board together

3.2.1 There are 6 Local Health Boards

```
## [1] "Betsi Cadwaladr"      "Hywel Dda"         "Grouped_4_organisation"
## [4] "Cardiff & Vale"       "Aneurin Bevan"     "Powys Teaching"
```

What are Local Health Board?

The Welsh Government oversees the country's healthcare system, which is different from that of the rest of the UK. In Wales, an institution designated as a "Local Health Board" (LHB) is in charge of providing all NHS healthcare services within a particular geographic region.

Planning and executing NHS services in their respective regions is the responsibility of local health boards. These medical services consist of:

- pharmacy
- dental
- optical, and mental health

They also bear accountability for:

- enhancing the results of physical and mental health
- promoting well being
- decreasing disparities in population health
- contracting outside organizations to provide services in order to address residents' needs

(NHS Wales Health Boards and Trusts | GOV.WALES, 2023)

3.3 Number of patients who enter ED under 6 different local health boards

Number of patients who enter ED
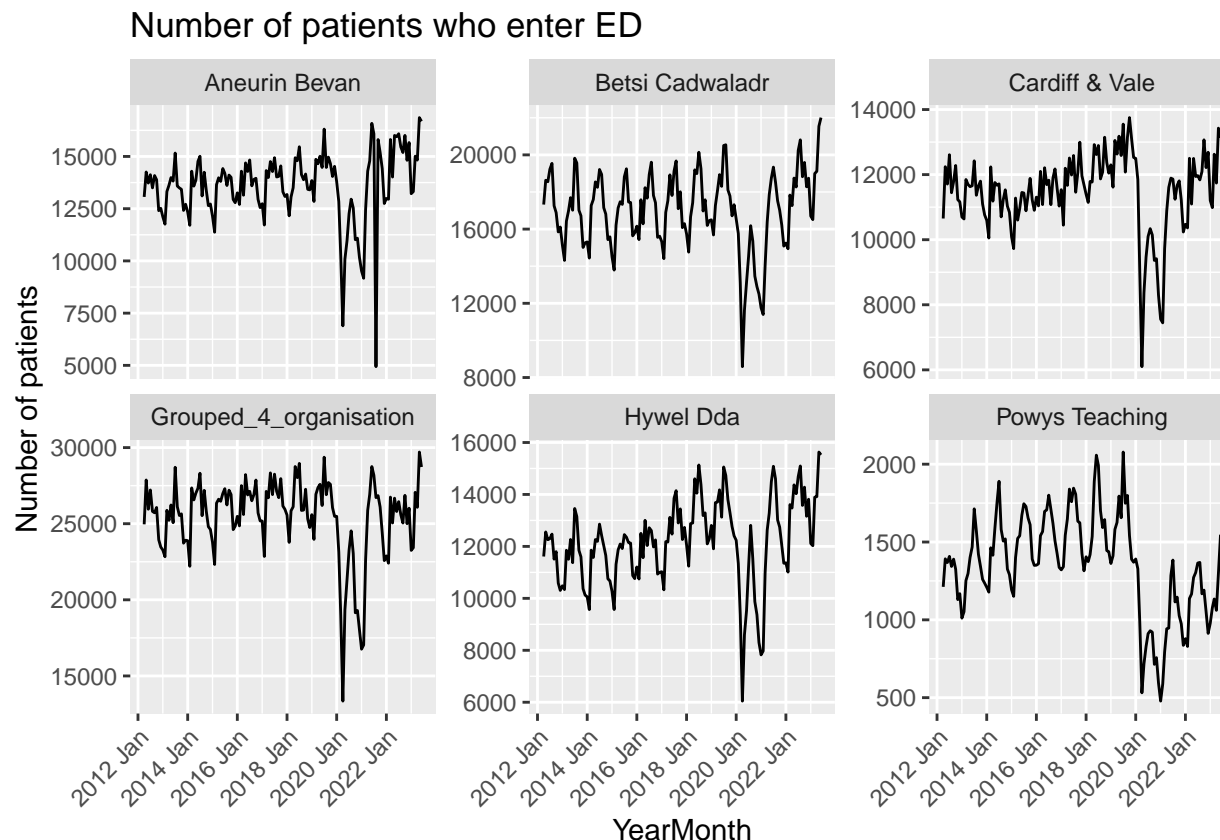
Analysing Health Board Attendance Trends Amidst and Post-COVID-19:

Healthcare systems around the world faced significant difficulties during the COVID-19 pandemic, which caused them to quickly shift into crisis-response mode. Intriguing trends during and after the epidemic era have been revealed by a thorough review of attendance data from local health boards.

- During the pandemic:

Surprisingly, several local health boards experienced a large reduction in attendance throughout the epidemic period. This significant drop in attendance, which may have been brought on by reasons including stringent public health regulations, widespread concern over virus exposure, and maybe changes in healthcare delivery strategies, was an anomalous departure from historical attendance trends.

- Post pandemic:

After the pandemic's intensity started to lessen, there was a noticeable increase in attendance, however there were notable differences amongst the local health boards. In contrast to most boards, which showed a strong recovery and raised their attendance figures to levels prior to the pandemic, Powys Teaching stood out as an outlier and defied this recovery trend. To fully understand the subtleties affecting these disparate courses and influence upcoming strategic planning, more research is necessary.

- Findings on seasonality:

Initial data analysis reveals the existence of seasonality, which is characterized by repeated swings in attendance across all health boards. This pattern calls for a perceptive investigation into:

Identify Underlying Causes: Researching seasonality's probable triggers, such as public health initiatives, flu seasons, or holiday seasons, in order to comprehend the factors influencing these patterns.

Determine Seasonal influence: Establishing whether the observed seasonality has an equal influence on all health boards or whether there are differences, which may be a sign of regional causes or policy changes.

Planning and Forecasting: By using this understood seasonality to improve forecasting models and ensuring that they are calibrated to account for these seasonal changes, more precise and useful predictive insights are made possible.

3.4 Seasonality of number of attendances

To further explore whether there was any trend and/or seasonality in the data by LHB, the time series was decomposed using the STL method and the extracted components (trend, seasonal and residuals) plotted:

Seasonal plots of ED Attendances (by LHB)
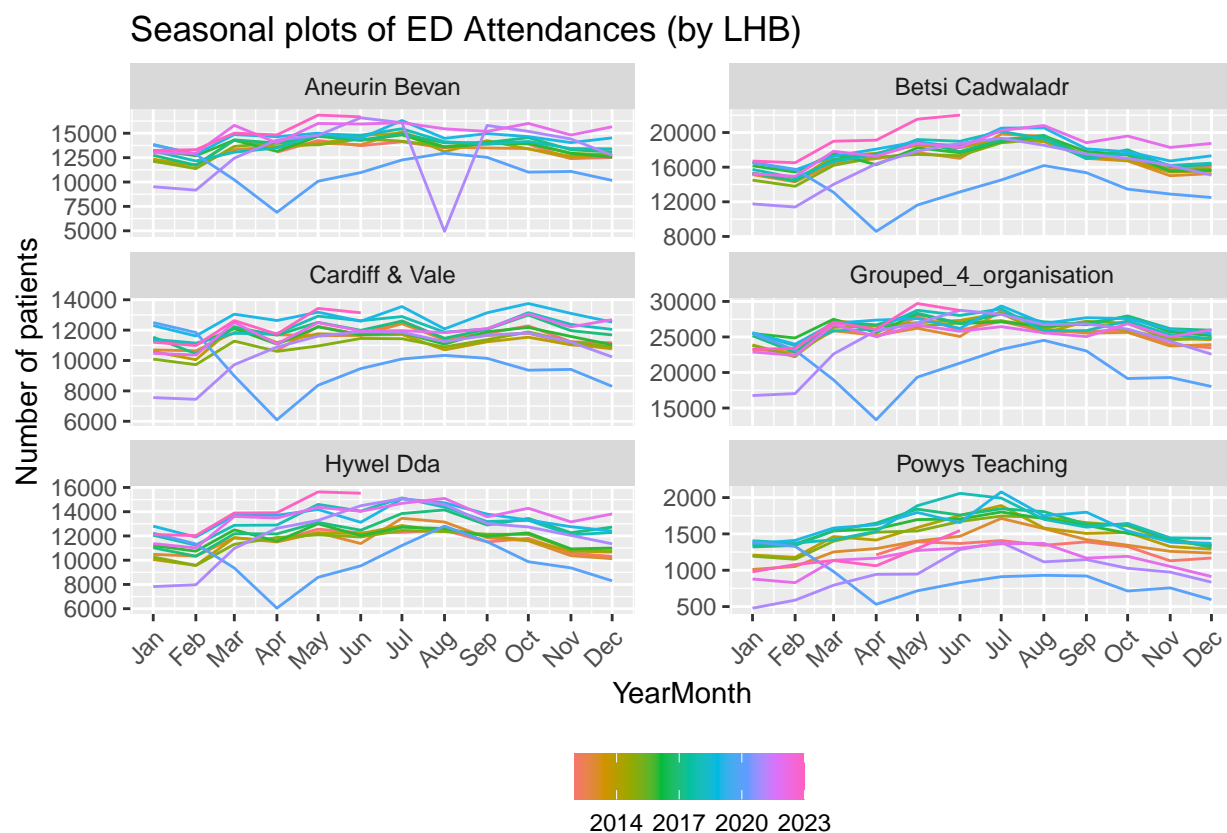
- Findings:

1. Most LHBs have a trend whereby attendance is lower in some months, like April and November, and higher in others, like June and July.

2. In many of the LHBs, the attendance in 2020 differs noticeably from other years. In comparison to similar months in past years, the majority of LHBs appear to have a fall or reduction in ED attendances during several months of 2020. This could be a sign of outside influences influencing access to or use of healthcare because of the pandemic.

3.

- Aneurin Bevan: The middle of the year for 2020 shows a clear decline.
- Cardiff & Vale: Attendances in 2020 appear to be continuously lower than in previous years, particularly in the early months.
- Hywel Dda: Around mid-year, there is a noticeable decline in attendance.
- Betsi Cadwaladr: While the trend for 2020 appears to be largely constant, several months still show a decline in attendance.
- Grouped_4_organisation: Attendance clearly drops off during particular months.
- Powys Teaching: The variations are more pronounced, with obvious dips during particular months.

3.5   Decompose Time Series

3.6   Plotting

STL Decomposition of ED Attendance

Component — Trend — Seasonal — Residual
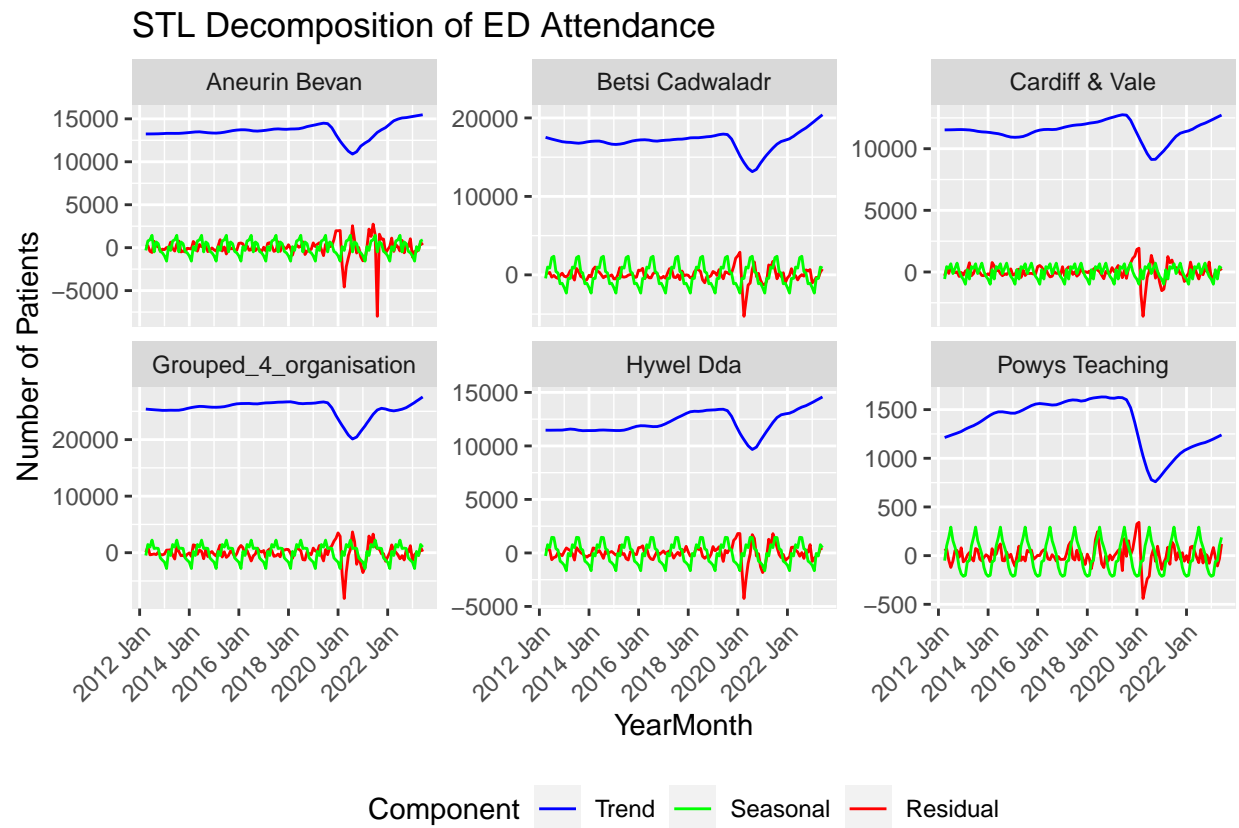
- Findings:

1. Trend: There appears to be an upward trend in the number of ED visits over time for all LHBs, but there are outliers. The most obvious decline occurs around 2020, which may be a sign of the COVID-19 pandemic's effects.

2. Seasonality: The seasonal component illustrates the periodicity in ED attendances by showing predictable patterns that repeat annually. There is a pronounced surge in the number of patient's attendance in the middle of the year (approximately in June or July). This seasonal pattern underlines the recurrent nature of patient admissions.

3. Residuals: The residuals are the data's unexplained variation after the trend and seasonal components have been removed. The majority of LHBs have residuals that oscillate near zero, which shows that the STL decomposition has effectively caught the main trends in the data.

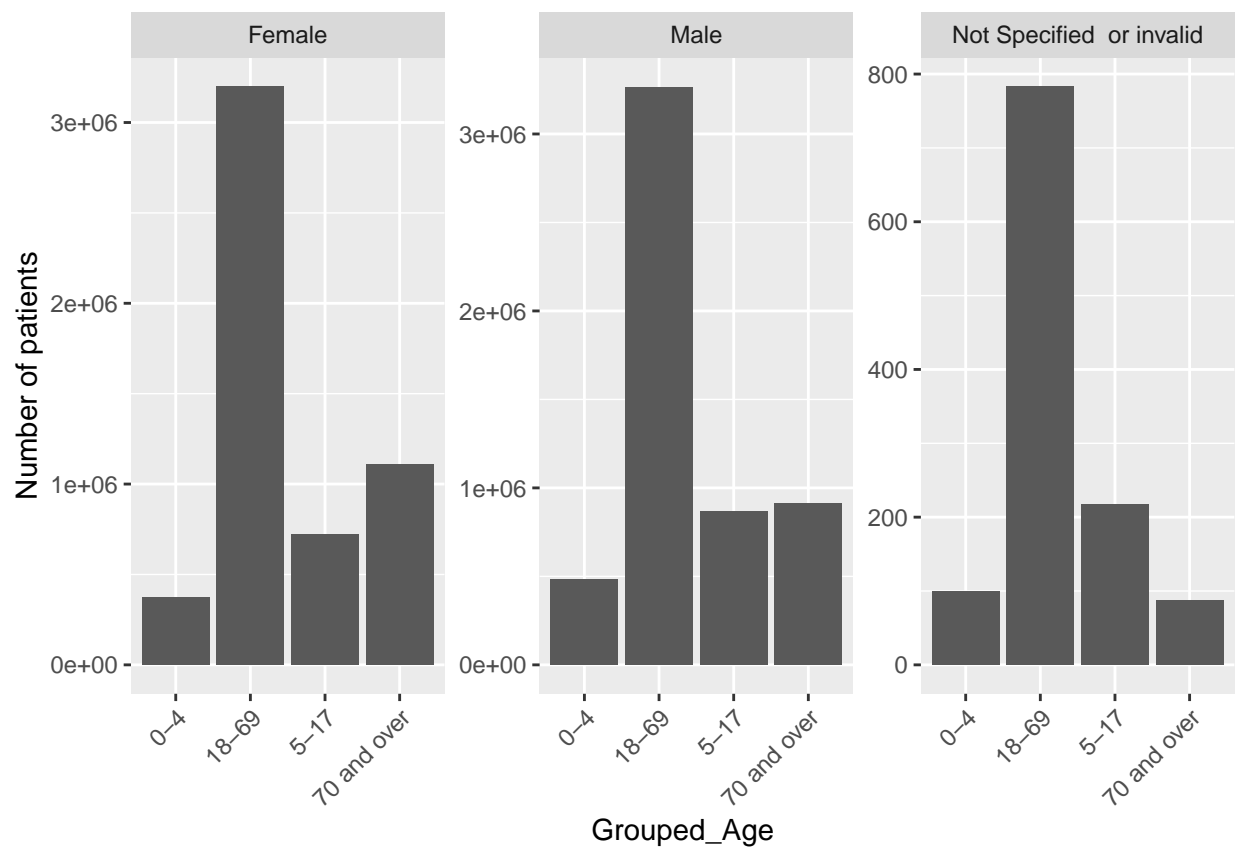## 3.7 Change the Age_Code structure into different groups for simplicity and interpretability

The original age groups

```
##  [1] "0 to 4"   "18 to 24" "25 to 29" "30 to 34" "35 to 39" "40 to 44"
##  [7] "45 to 49" "5 to 17"  "50 to 54" "55 to 59" "60 to 64" "65 to 69"
## [13] "70 to 74" "75 to 79" "80 to 84" "85"          "Unknown"
```

New age group

```
## [1] "0-4"          "18-69"          "5-17"          "70 and over"
```

3.8 Plot Number of Patients in different age groups

- Findings:

It is expected that the age group 18-69 has the most amount of attendance, given the fact that it is the biggest group among all. However, the second largest group is from the oldest age bracket. It is aligned with the general understanding of that elders may need more health care.

4 Forecast with ARIMA

AutoRegressive Integrated Moving Average, or ARIMA, is a widely used forecasting technique for time series data.

- AutoRegressive, or AR:

The link between one observation and several lag observations (prior periods or terms) is the subject of this component. The word "auto" denotes that the variable is being regressed against itself.

- I (Integrated):

This is the original data difference that is used to make the time series stationary, meaning that the data values are independent of time. The mean of stationary data will not change over time, and neither will the variance. It won't display seasonality or patterns.
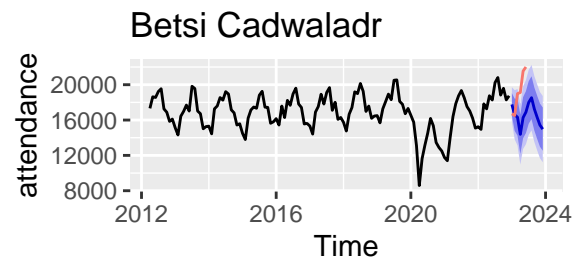
- Moving Average, or MA:

This part uses a combination of earlier error terms to model the time series error. The MA term's intuitive purpose is to model the error.
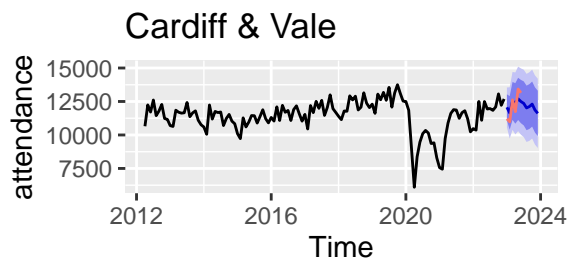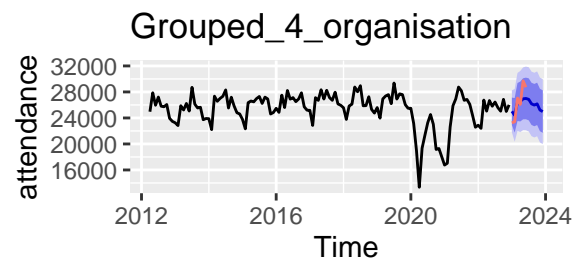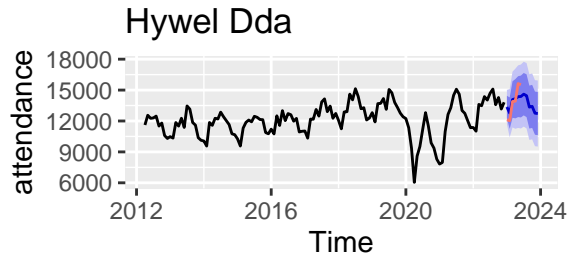
4.1 Plotting using ARIMA
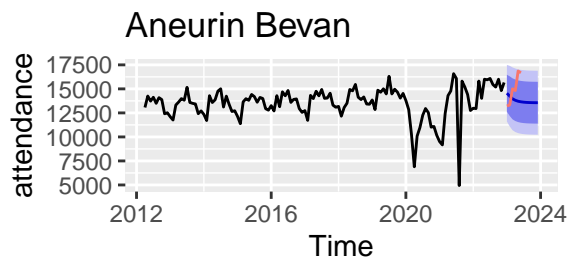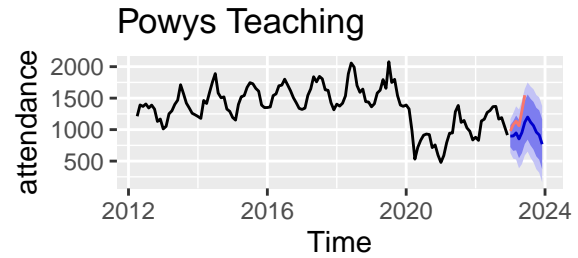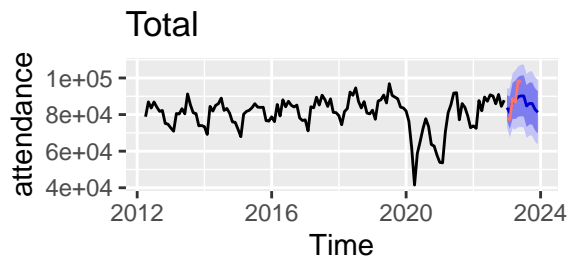
5 Forecasting with ETS

Error, Trend, Seasonality, or ETS, is a widely used model for predicting time series data. A generalization of these techniques, the ETS model includes a variety of exponential smoothing methods.

The elements of the ETS model breaks down as follows:

Error (E):

The error component represents the discrepancies between observed values and those predicted by a model. There are two primary types of error models: additive and multiplicative.

1. **Additive Error**:
   - Assumes that the model's residuals (errors) are added together.
   - Suitable when the magnitude of errors remains relatively constant regardless of the time series' level.

2. **Multiplicative Error**:
   - Assumes that the errors change proportionally to the time series' level.
   - Relevant when the magnitude of errors increases or decreases in proportion to the actual value of the series.

Trend (T):

The trend component captures the long-term progression of a time series.

1. **No Trend**:

- Indicates that the time series remains relatively constant over time and does not exhibit any discernible upward or downward movement.

2. **Additive Trend**:

   - Represents a consistent linear progression in the time series.
   - The series displays a steady increase or decrease at a constant amount over equal intervals of time.

3. **Multiplicative Trend**:

   - Denotes an exponential progression in the series.
   - The data exhibits growth or decline at a rate that increases or decreases proportionally over time.

Seasonality (S):

The seasonality component captures periodic fluctuations in a time series that recur at regular intervals, such as daily, monthly, or yearly.

1. **No Seasonality**:

   - Implies that the time series does not exhibit any periodic or recurring fluctuations. The data remains consistent throughout without any noticeable seasonal patterns.

2. **Additive Seasonality**:

   - Suggests that seasonal patterns are consistent in magnitude throughout the time series.
   - The seasonal effect is added to or subtracted from the series in a constant manner, irrespective of the time series level.

3. **Multiplicative Seasonality**:

   - Indicates that the seasonal fluctuations change in proportion to the time series' level.
   - As the series grows or declines, the impact of seasonality amplifies or diminishes respectively, leading to varying magnitudes of seasonal effects over time.

5.1 Plotting with ets
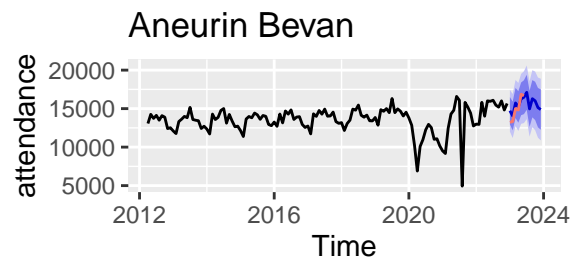
Aneurin Bevan

series — Actual

Betsi Cadwaladr

series — Actual

Cardiff & Vale

series — Actual

Grouped_4_organisation
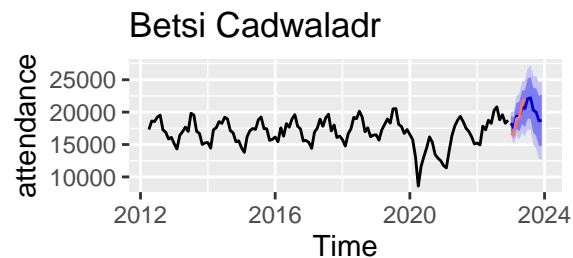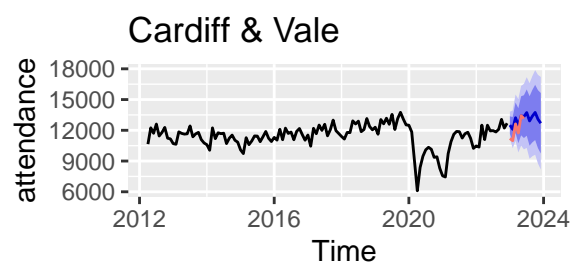
series — Actual
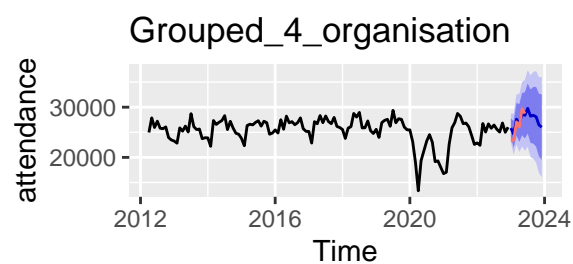
Hywel Dda

series — Actual
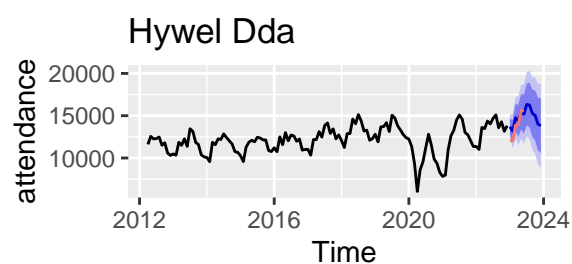
Powys Teaching
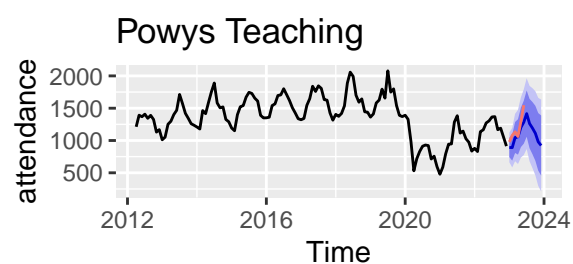
series — Actual

Total

series — Actual

Conformity:

Hywel Dda, Powys Teaching, Cardiff & Vale, Grouped_4_Organisation, and Total (All-Wales): For these

health boards, the ARIMA model effectively predicted the downturn that was observed in the latter portions of 2023 and successfully reflected the actual data, indicating a cogent direction. This demonstrates ARIMA's ability to effectively model and anticipate time-series data for these specific entities, despite the unpredictably fluctuating conditions brought on by the epidemic.

Contradictions:

Aneurin Bevan and Betsi Cadwaladr: By contrast, the prediction statistics for Aneurin Bevan and Betsi Cadwaladr show a clear departure from the actual numbers. It is necessary to assess the ARIMA model's applicability and accuracy with regard to these specific datasets because there is a discrepancy between the predicted value and the real-time data.
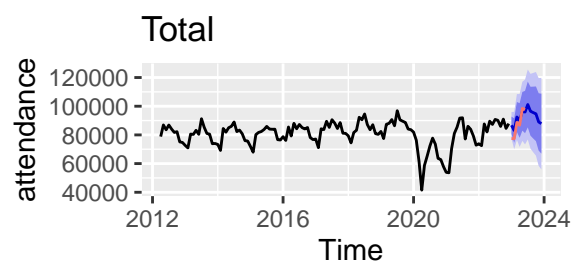
Comparison:

Upon first glance, it appears that the ETS model offers a relatively close alignment with the real data, demonstrating a strong predictive power. When compared to the ARIMA model, the projections closely mimic the actual data patterns rather than merely following them, which suggests a possibility for greater accuracy.

Findings: Consistent Decline in Predictions

Downtrend in 2023: Regardless of the model used, all forecasts uniformly point to a decline in trends during the second half of 2023. This might be explained by the historical data's apparent cyclical nature. In other words, the historical data reveals a pattern in which peaks frequently appear in the middle of the year and then descend into troughs at year's end and the beginning of the next year. This recurrent pattern may have a significant impact on the expected downturn in 2023.

In light of the above, it is crucial to approach model selection with a nuanced comprehension of the underlying data and impacting factors, even though the ETS model shows a potentially higher prediction accuracy within the given circumstances.

An accuracy analysis was then carried out upon for both ARIMA and ETS models to further assess the quality of their predictions.

6 Accuracy assessment for ARIMA and ETS

6.1 Calculating Accuracy Metrics for ARIMA and ETS

6.1.1 Displaying Accuracy Metrics for each Column

LHB	ARIMA	ETS
Aneurin Bevan.MAE	1748.634	688.417
Aneurin Bevan.RMSE	1992.543	801.644
Aneurin Bevan.MAPE	0.113	0.048
Betsi Cadwaladr.MAE	3157.257	915.735
Betsi Cadwaladr.RMSE	3756.403	1050.122
Betsi Cadwaladr.MAPE	0.156	0.049
Cardiff & Vale.MAE	600.429	654.811
Cardiff & Vale.RMSE	648.376	799.030
Cardiff & Vale.MAPE	0.050	0.057
Grouped_4_organisation.MAE	1386.277	1089.509
Grouped_4_organisation.RMSE	1598.559	1315.071
Grouped_4_organisation.MAPE	0.052	0.043
Hywel Dda.MAE	804.597	763.085
Hywel Dda.RMSE	923.853	905.503
Hywel Dda.MAPE	0.058	0.059
Powys Teaching.MAE	238.265	120.999
Powys Teaching.RMSE	264.667	146.796
Powys Teaching.MAPE	0.192	0.098
Total.MAE	4781.817	4297.460
Total.RMSE	5601.875	5014.551
Total.MAPE	0.054	0.052

- With Cardiff & Vale being a significant exception, most places displayed reduced error metrics when modeled with ETS, signifying a higher predictive accuracy as compared to the ARIMA model.

- According to the interest expressed by the NHS executives, I would like to provide some insights for the error metrics according to the quality of the forecast for the non-zero count data:

1. **Root Mean Squared Error (RMSE)**

    - **Description**: Measures the square root of the average squared differences between forecast and actual values.
    - **Pros**: It gives more weight to large errors.
    - **Cons**: It can be influenced significantly by outliers. RMSE might be inflated, if the count data is prone to significant spikes or declines, which should not be the problem for this data except for the Covid-19 era.

2. **Mean Absolute Percentage Error (MAPE)**

    - **Description**: Measures the average of the absolute percentage errors.
    - **Pros**: It is a relative metric, and is scale-independent and easy to interpret.
    - **Cons**: MAPE can overemphasize relative errors on small counts.

3. **Mean Absolute Error (MAE)**

    - **Description**: Measures the average of the absolute differences between forecast and actual values.
    - **Pros**: It's less sensitive to outliers than RMSE. It provides a straightforward average error size.
    - **Cons**: It does not emphasize large errors.

However, RMSE and MAE should not be used for the hierarchical time series data.

And here is why:

- **1. Consistency Between Levels:** Granularity Errors are not consistently distributed or similar across all hierarchical levels due to the variance in data granularity between levels, and there will be an aggregation dilemma. Due to the variation in magnitude and volume at various hierarchical levels, direct aggregation of RMSE or MAE from bottom levels to top levels might produce inconsistent or biased judgments of forecast accuracy.

- **2. Problems with reconciliation:** It is possible for forecasts to not add up coherently in a hierarchical structure when they are independently generated for several levels. Since they assess mistakes without taking into account hierarchical dependencies, RMSE and MAE do not by default address this reconciliation. And the alignment issue will therefore be caused. The use of standard error measurements at various hierarchical levels may produce false conclusions regarding the precision and dependability of the forecasts in the absence of a cogent reconciliation process.

- **3. Metric incomparability:** Impact of a unit error is not constant across levels; for instance, a 100-unit error may be insignificant at a higher aggregate level but significant at a lower one. Additionally, due to the divergent scales of the errors at various levels, it is difficult to use a single statistic, such as RMSE or MAE, to fairly compare performance across all hierarchical levels.

7 Reconciliation

Aggregate data: to aggregate data in a hierarchical structure, I set Aggregated_Organisation which is the local health board to be the parent attribute, and Hospital_ItemName_ENG as a child attribute. Additionally, I also group the patients' age by 0-4yrs, 5-17yrs, 18-69yrs, and 70 and over. The aggregate key is the patients' attendance.

cleaned_data: This is the new dataset I created. It only contains the hospitals which has data for the entire time series (From 2012 Apr to 2023 Jun), for the purpose of conducting the reconciliation. In other words, if the hospital does not have the whole data or contains NAs across the entire period of time, it will be unfeasible to conduct the reconciliation (the function can still run, however, it will return an empty cell for the whole time period).

7.1 Time series cross validation

Time series cross validation is also called rolling-forecast origin and walk-forward validation. It splits the data into training and test datasets. The training data here consists the data before 2022 Jun, and the test data consists the remaining which is from 2022 July to 2023 Jun (1 year). Just as its other name "rolling-forecast origin", the iteration process performs with a model trained on the initial training data and will be used to make a forecast for the next period. And the actual observation will be added to the training set for the next step.

The number of months in the training data is:

```
## [1] 123
```

Reasons to conduct the cross validation on the entire time period

The time series cross validation is conducted on the cleaned_data which has the entire time period due to the fact that I would also like to analyse the effect resulted by Covid-19. Even though Covid-19 was an outburst and seemed to be a sudden accident, the emergency department has to get prepared if there is another outbreak like this.

There are 13 training data when I set the .init = 135(the total time points) -12(test data). Stretch_tsibble will automatically generate a column called .id and there are 13 different ids.

Reasons to filter the data

I chose to analyse the data only for the aggregated level due to the computational constraints. In other words, R consistently terminated itself when I use the entire data to run the accuracy test. The accuracy test was conducted by taking the forecast for 12 time points and compare it with the test data which contains the actual data.

4 models for generating the forecast

I conducted the forecast with 4 models, naive model, ets, tscount and a combination of all three models called "comb".

Accuracy metrics for 4 different models

```
##         model  Avg_MASE Avg_RMSSE
## 1 naive_model 10.817713  4747.813
## 2   ets_model  7.960357  2873.544
## 3     tscount  9.134870  4519.123
## 4        comb  8.693122  3670.712
```

Interpretation for the accuracy metrics

The error metrics MASE and RMSSE usually take naive model as a benchmark which means these metrics compare others with naive model. Therefore, I also applied this concept. The result reveals that 'ets', 'tscount' and the combination of naive, 'ets' and 'tscount' all outperform the naive model. Given this insight, I excluded naive model for the subsequent reconciliation process.

7.1.1   Time series cross validation, set training data before Covid

Suggestions of setting the training data before Covid

Some professors and NHS-Wales executives suggested me to do the forecast before the Covid-19 outbreak. Due to the stretch_tsibble function, I needed to set the training data one year later in order to obtain desired forecast period (2020 Feb to 2023 Jun), which I believe that this is a bug that needs to be solved in the stretch_tsibble function, I will consult the developer about it later on. Check data_gts_fc2 for the forecasting period in the column of YearMonth.

The number of months in the training data is:

```
## [1] 106
```

Reasons to set the .init =94

This comes from the number of months(106) minus 12 (which I want the number of training data to be).

```
##         model  Avg_MASE    Avg_RMSSE
## 1 naive_model  25.27004   10544.862
## 2   ets_model  19.84085    9689.341
## 3     tscount 391.71459 2818486.809
## 4        comb 135.99048  939580.334
```
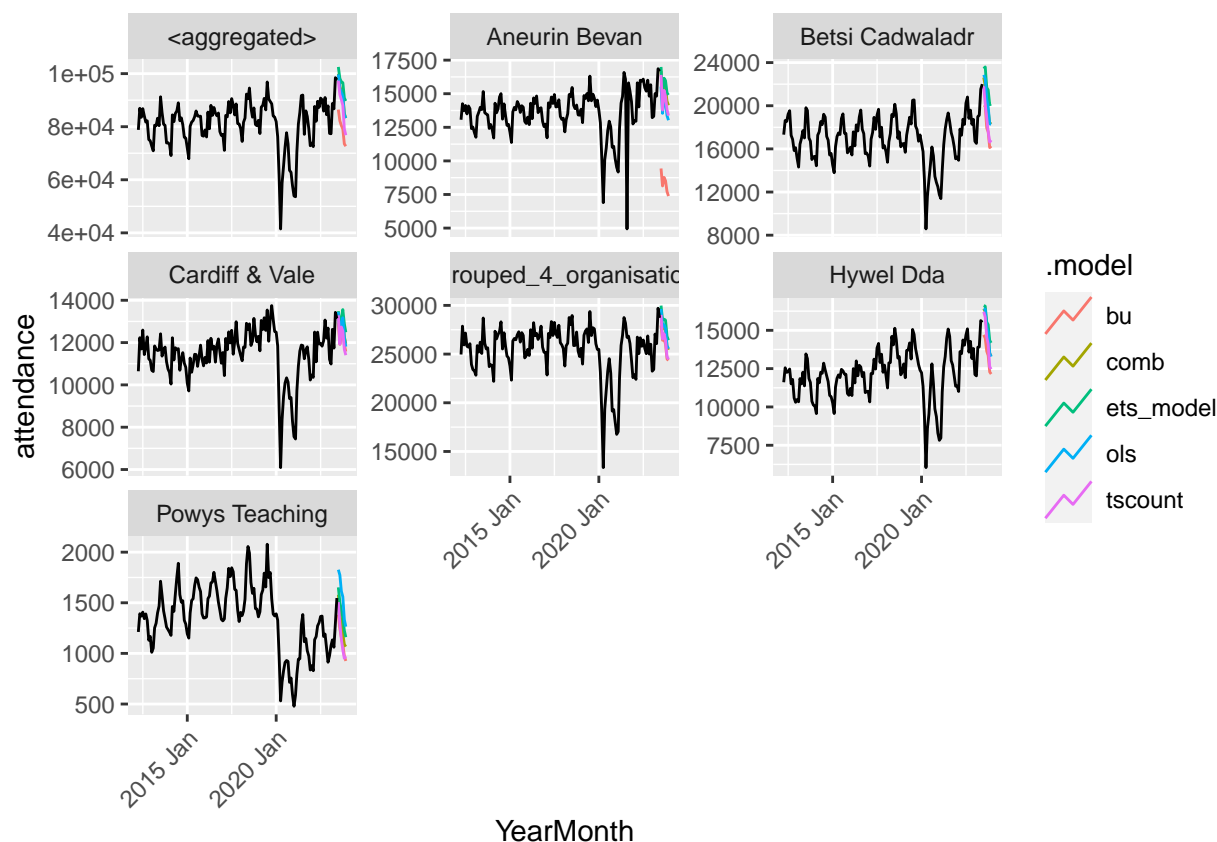
Interpretation for the result

According to the MASE and RMSSE, tscount performs the worst, only the ets model performs better than the naive model. However, all of them did not perform as well as the previous analysis. Therefore, I would continue my analysis with the training data from 2012 Apr to 2022 Jun for the subsequent analysis. Nevertheless, the result is to my surprise, as I first thought that without the Covid-19 period data, the variability would be less and could instead provide a more precise prediction, however, the result for the accuracy suggested otherwise. I reckon the lower accuracy is due to that splitting the training data before covid-19 will result in the need of forecasting for the this period, and it then leads to a lower accuracy because Covid-19 period is basically impossible to be forecasted.

I also thought about excluding the Covid-19 period from the data, however, first of all, it is hard to define when the Covid-19 period ended as the data shows that it ended around 2022 Jan, however, Australia's Chief Medical Officer declared that Covid-19 is no longer a Communicable Disease Incident of National Significance (CDINS) until 20th October 2023(Australian Government Department of Health and Aged Care, 2023). Secondly, the trend and the pattern are ignored if excluding the Covid-19 period. Even though there was a significant drop during the period, there was overall still an upward trend after Covid-19 which we can see from the plot that the attendance raised back up. Therefore, simply excluding the data during the Covid-19 era is not a good idea in my opinion.

## 7.2 Forecast with the training data from 2012 Apr to 2022 Jun

Forecast horizon

The goal of this project is to forecast the 6 months ahead of the emergency department attendance; hence, the forecast horizon equals to 6.

Interpretation for the reconciliation result

I mentioned the overall trend and seasonality in the previous analysis. According to the plot above, it aligns with the result from the exploratory analysis. To give more details for the reconciliation, bu refers to "bottom-up" forecasting, it is an approach that reconciles from the most detailed/bottom level. Additionally, ols stands for ordinary least squares, it is a type of linear regression analysis. Both of them are the reconciliation methods for the hierarchical and grouped time series data. Other methods such as comb, ets and tscount are simply used for the base forecasting.

A closer look
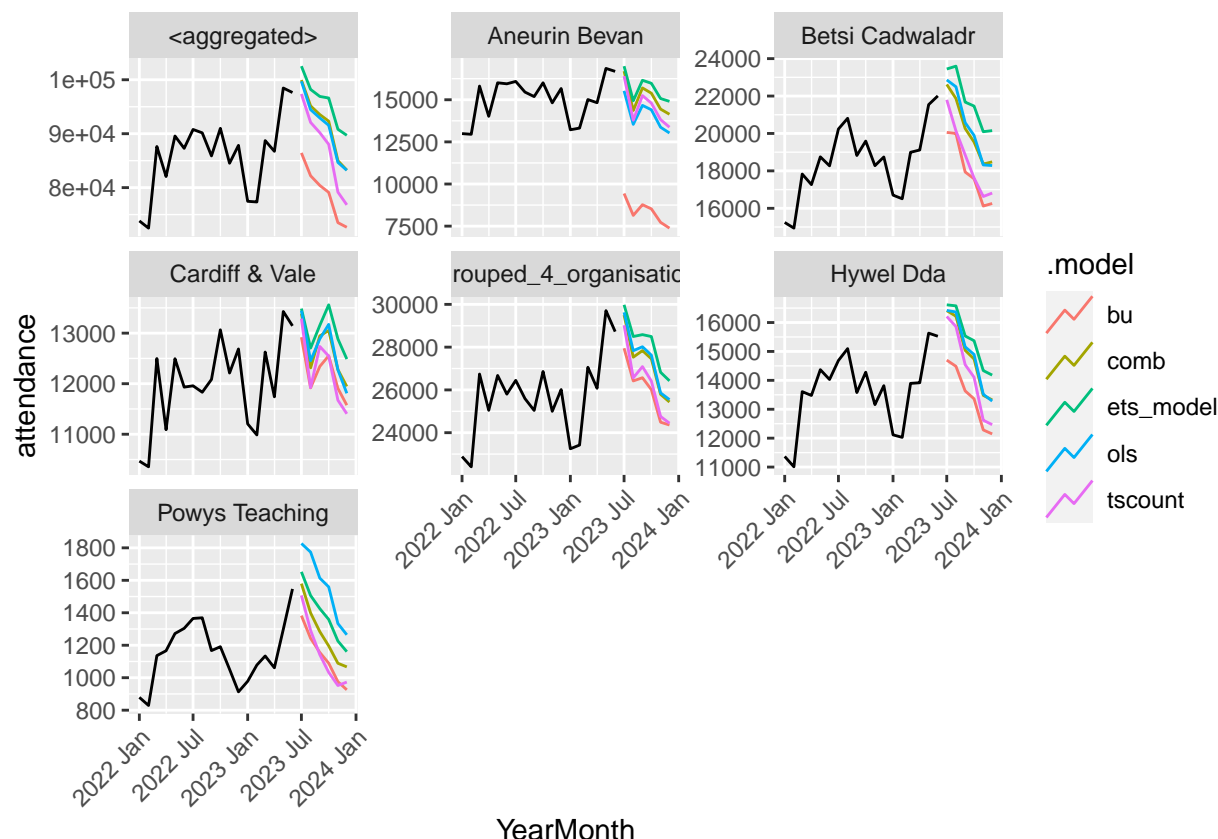
Downward trend

The base forecasting and the reconciliation forecasting both show a downward trend for the future 6 months. After the discussion with the NHS-Wales executive, we concluded that this might be due to the seasonal effect that there always seems to have a peak in the middle of the year and a trough in the end of the year. The other reason is so called the holiday effect, which points out that patients are reluctant to visit the hospital during Christmas or New Year as they would like to spend more time with family and friends during the special season. However, this might also be contradictory as during the holidays in the year, it is usually that emergency department is the only department that is open. Therefore, we might need to include some global variables into this, and create a multivariate regression to discover what is actually happening during the later half of the year.

8   Comments on this project and what can be done in the future?

I find this project very interesting as it is very close to our daily lives. Emergency and accidents happen all the time. My goal for this project is to provide some forecast and information to give the hospitals, local health boards and NHS-Wales some insights. So the health care system can be more prepared for the coming future. Additionally, they can also be more insightful when allocating the budget and adjusting the health

care policy for the future purpose. In addition to this, NHS-Wales can use the forecast model and code I provide to produce the forecast for more than 6 months ahead if they would like to make the use of it.

Furthermore, this is my first time conducting the hierarchical forecasting, and I really learnt a lot from it, such as the skill to cope with hierarchical time series or grouped time series data. I believe with the additional skill I equipped from this project and unit, I can be more qualified to conduct the forecasting analysis in the future. It will be really useful in the data analytics field.

However, from the analysis, even though the models show a fairly good accuracy. It might be more insightful if there is another model such as probabilistic model which can take parameters such as holidays or any other outbreaks (for example, flu season) to create a multivariate model. Additional information is that for some specific days, only Emergency Department is open, and therefore, there should be more people visiting. However, we cannot observe the pattern from the given data. Therefore, another thing that can be done is to include a more detailed data, for instance, to include global variable.

9   References

Australian Government Department of Health and Aged Care. (2023, October 20). AHPPC statement – End of COVID-19 emergency response. https://www.health.gov.au/news/ahppc-statement-end-of-covid-19-emergency-response

Csardi G, Nepusz T (2006). "The igraph software package for complex network research." *InterJournal*, *Complex Systems*, 1695. https://igraph.org.

Garrett Grolemund, Hadley Wickham (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1-25. URL https://www.jstatsoft.org/v40/i03/.

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Hamner B, Frasco M (2018). *Metrics: Evaluation Metrics for Machine Learning.* R package version 0.1.4, https://CRAN.R-project.org/package=Metrics.

Hyndman R (2023). *fpp3: Data for "Forecasting: Principles and Practice" (3rd Edition).* R package version 0.5, https://CRAN.R-project.org/package=fpp3.

Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropoulos F, Razbash S, Wang E, Yasmeen F (2023). *forecast: Forecasting functions for time series and linear models.* R package version 8.21.1, https://pkg.robjhyndman.com/forecast/.

Hyndman R, Lee A, Wang E, Wickramasuriya S (2021). *hts: Hierarchical and Grouped Time Series.* R package version 6.0.2, https://CRAN.R-project.org/package=hts.

Liboschik T, Fokianos K, Fried R (2017). "tscount: An R Package for Analysis of Count Time Series Following Generalized Linear Models." *Journal of Statistical Software*, *82*(5), 1-51. doi:10.18637/jss.v082.i05 https://doi.org/10.18637/jss.v082.i05.

NHS Wales health boards and trusts | GOV.WALES. (2023, February 3). GOV.WALES. https://www.gov.wales/nhs-wales-health-boards-and-trusts

O'Hara-Wild M, Hyndman R, Wang E (2023). *fable: Forecasting Models for Tidy Time Series.* R package version 0.3.3, https://CRAN.R-project.org/package=fable.

O'Hara-Wild M, Hyndman R, Wang E (2023). *feasts: Feature Extraction and Statistics for Time Series.* R package version 0.3.1, https://CRAN.R-project.org/package=feasts.

Pedersen T (2022). *ggraph: An Implementation of Grammar of Graphics for Graphs and Networks.* R package version 2.1.0, https://CRAN.R-project.org/package=ggraph.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Wang, E, D Cook, and RJ Hyndman (2020). A new tidy data structure to support exploration and modeling of temporal data, Journal of Computational and Graphical Statistics, 29:3, 466-478, doi:10.1080/10618600.2019.1695624.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." *Journal of Open Source Software*, *4*(43), 1686. doi:10.21105/joss.01686 https://doi.org/10.21105/joss.01686.

Wickham H, François R, Henry L, Müller K, Vaughan D (2023). *dplyr: A Grammar of Data Manipulation.* R package version 1.1.3, https://CRAN.R-project.org/package=dplyr.

Wickham H, Henry L (2023). *purrr: Functional Programming Tools.* R package version 1.0.2, https://CRAN.R-project.org/package=purrr.

Wickham H, Vaughan D, Girlich M (2023). *tidyr: Tidy Messy Data.* R package version 1.3.0, https://CRAN.R-project.org/package=tidyr.

Zeileis A, Grothendieck G (2005). "zoo: S3 Infrastructure for Regular and Irregular Time Series." *Journal of Statistical Software*, *14*(6), 1-27. doi:10.18637/jss.v014.i06 https://doi.org/10.18637/jss.v014.i06.