Analysing Trends and Predictive Modeling of Accident and Emergency Department Attendances and Waiting Times in Scotland

by

Yanfu Ding

Executive Summary

This report presents an analysis of Accident and Emergency (A&E) department attendances and waiting times across Scotland from July 2007 to April 2024. Using Generalised Additive Models (GAMs), the study aimed to predict monthly attendances and the percentage of patients receiving treatments within four hours. The statistical analysis syggested that A&E attendances are affected by seasonal effects, prior attendance levels, and the effects of COVID-19. Seasonal trends showed increased attendances during winter months and a rapid decline in early 2020 due to lockdowns. Moreover, higher attendances in the previous month were found to lead to higher attendances in the subsequent month.

Model predictions suggest a fluctuating trend in future attendances, with a slight improvement in waiting times expected over the next few months overall. However, these numbers are projected to remain below pre-COVID levels, indicating continued challenges for A&E departments in Scotland.

As for potential improvement of the model, the report suggests including factors like age, gender, and socio-economic deprivation in future models to boost their accuracy and predictive capability. Considering these elements will offer a more thorough understanding, helping healthcare systems better manage demand and improve patient outcomes.

Contents

List of Tables

List of Figures

1 Introduction

1.1 Background

The effective management of Accident and Emergency (A&E) departments is a central aspect of global healthcare systems. They are the primary point of access to emergency medical care, providing essential treatments and interventions. The Scottish government has set up a national standard for A&E that 95% of patients should wait no more than 4 hours from arrival to admission, discharge or transfer to A&E since 2007 (PHS 2023). However, the reality has been less optimistic, especially following the outbreak of the COVID-19 pandemic, with this percentage gradually declining. Figure 1 summarises the percentage of patients in Scottish A&E departments who waited no more than four hours each month from July 2007 to April 2024. Around 30% of patients were waiting longer than four hours, which was far from the 5% target. Furthermore, the proportions of patients waiting over 8 hours and 12 hours have also increased. Public Health Scotland (PHS) reports that the percentage of patients waiting more than eight hours increased from 10.8% to 13.4%, and 6% of arrivals had to wait over 12 hours (PHS 2024).

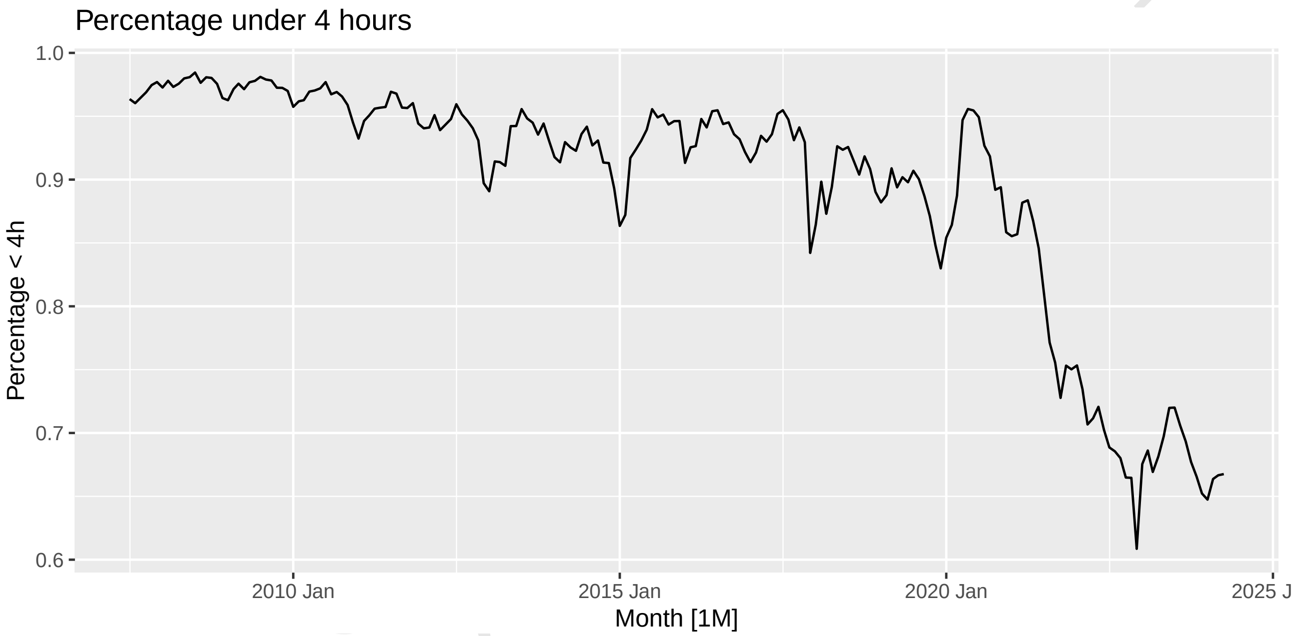

Figure 1: Percentage of A&E attendances completed within four hours in Scotland (2007-2024)

To address this situation, PHS is creating discrete event simulation models to predict demand flow and evaluate the impact on capacity. This report seeks to provide a small contribution towards aiding PHS in comprehending the monthly total demand for A&E attendances and the monthly percentage of long waits (exceeding four hours) over time. Specifically, it will investigate how these outcomes fluctuate over time and offer predictions for a six-month period starting from May 2024.

1.2 Data

Data was sourced from the open dataset provided by PHS, which records monthly A&E activity and waiting times statistics on new and unplanned return attendances at A&E services across Scotland, spanning a period from July 2007 to April 2024. Each row in the dataset represents monthly aggregated statistics for various Health Board of first Treatment (HBT) and treatment locations. Variables of interest in this dataset include:

- Month: The month and year of the recorded data.
- Total Attendances (Episode): The total number of A&E attendances in the given month (by different levels).

- Four-Hour Performance: The percentage of A&E attendances that were admitted, transferred, or discharged within four hours.

For more detailed information, the dataset can be accessed through: https://www.opendata.nhs.scot/dataset/monthly-accident-and-emergency-activity-and-waiting-times/resource/37ba17b1-c323-492c-87d5-e986aae9ab59?inner_span=True.

1.3 Software and Algorithms

The models were fitted using the `mgcv`(Wood 2011; Wood et al. 2016) package in `R` (R Core Team 2024) ,which is designed for Generalised Additive Models (GAMs). For data visualisation and plotting, `ggplot2` (Wickham 2016) was utilised, with additional tools from `cowplot` (Wilke 2024) to arrange multiple plots and `gratia` (Simpson 2024) for GAM-specific plotting. Time series decomposition and analysis were carried out using the `fpp3` (Hyndman 2024) and `tsibble` ((Wang et al. 2020) packages. Data manipulation and aggregation were executed with `dplyr` (Wickham et al. 2023) and `tidyr` (Wickham et al. 2024). Tables were created using `knitr` (Xie 2024, 2015, 2014) and `kableExtra` (Zhu 2024). Detailed session information for R, including the versions of the packages used, is provided in the appendix.

2 Exploratory Data Analysis

2.1 Total Attendances by Month

The dataset was preprocessed by removing quality flags and unnecessary columns, and the remaining data was aggregated to provide a monthly overview of A&E attendances.

2.1.1 Total monthly attendances

Figure 2 presents a plot of the total number of A&E attendances per month. Noteworthy fluctuations in monthly attendances were observed, with pronounced increases during winter months likely attributable to seasonal illnesses. Conversely, a considerable decline in attendances was observed around early 2020, reflecting the impact of the COVID-19 pandemic.

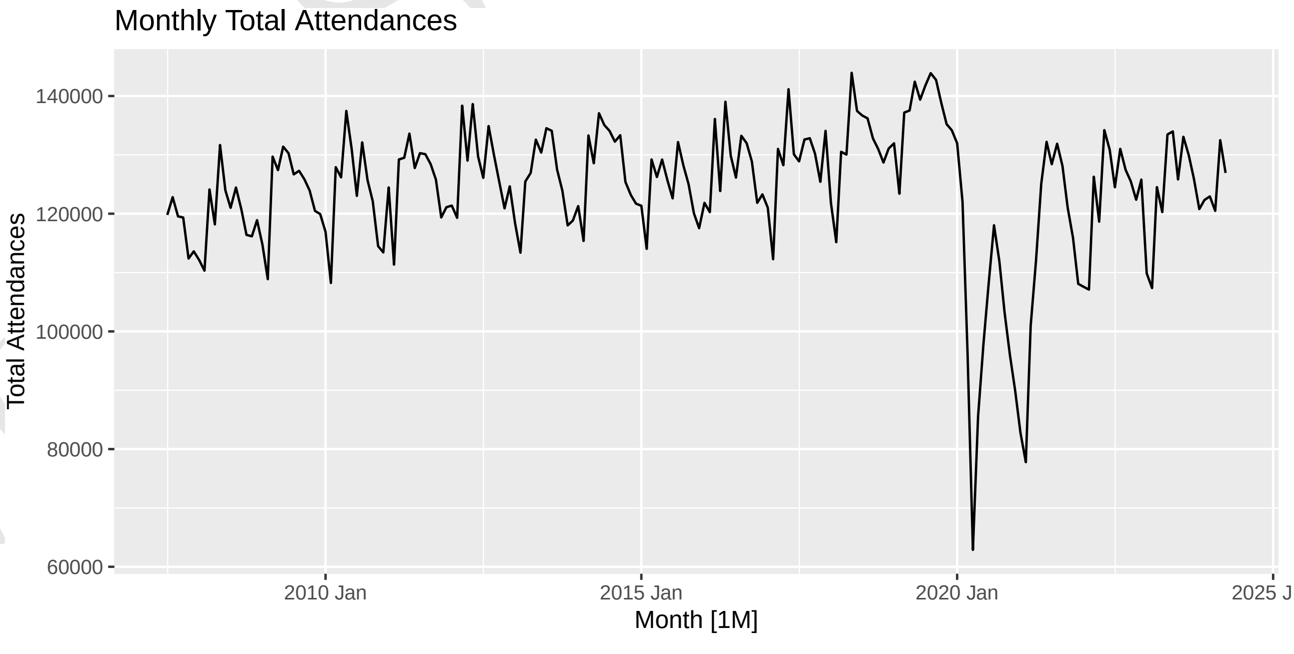

Figure 2: Monthly A&E Attendances in Scotland (2007-2024)

2.1.2 Seasonal Trend Decomposition, Autocorrelation and Partial Autocorrelation

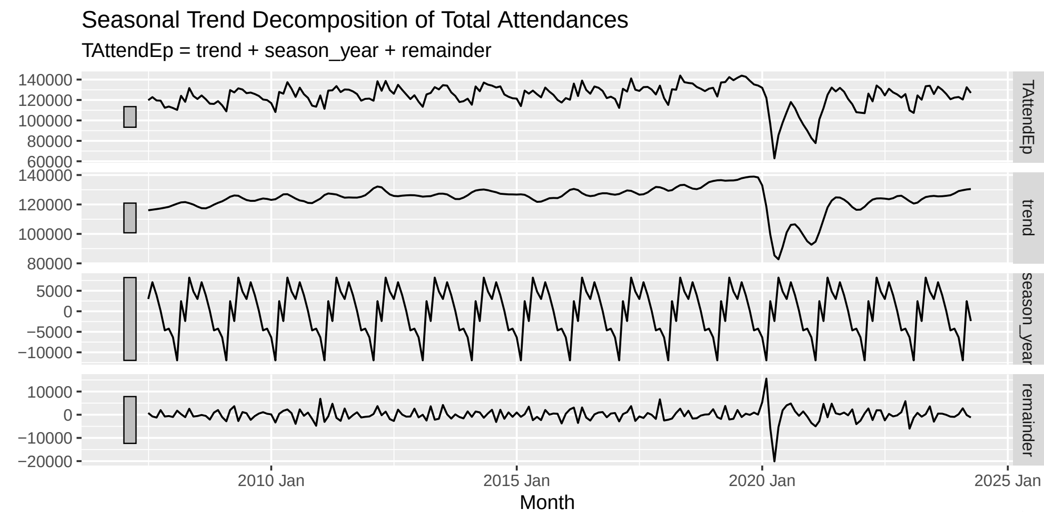

Figure 3: Seasonal Trend Decomposition of Monthly Attendances

Figure 3 presents the seasonal trend decomposition of total attendances, wherein the time-series is partitioned into trend, seasonal and residual components. The trend component demonstrates a predominantly consistent or slightly increasing pattern in attendance figures over time. However, a notable decline was observed in the early years of the pandemic (2020), followed by a gradual recovery. The seasonal term reveals a consistent periodic pattern of seasonal variations approximately every 12 months, and the remainder term suggests a white-noise structure except for the pandemic period. Meanwhile, Figure 4 reveals significant autocorrelation at multiple lags, suggesting that historical attendances have a substantial impact on future values. These findings imply the need for incorporating lag terms in the modelling process.

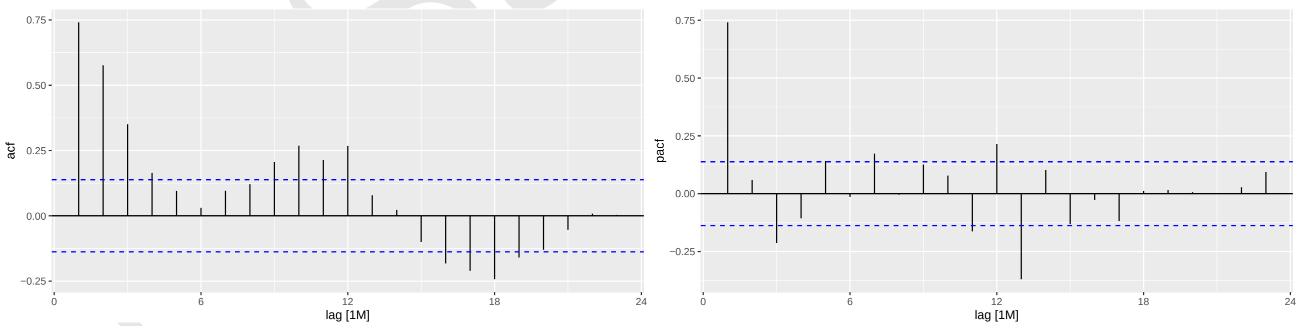

Figure 4: Monthly Attendances: ACF and PACF Plots

2.1.3 Moving Averages and Anomaly Detection

The 12-month moving average plot (Figure 5) smooths short-term fluctuations, emphasising longer-term trends and indicating overall stability with a slight upward trajectory in attendances despite COVID-19 disruptions. Anomalies are detected through STL decomposition (Figure 5), marked by red points, highlight significant deviations, particularly during early 2020.

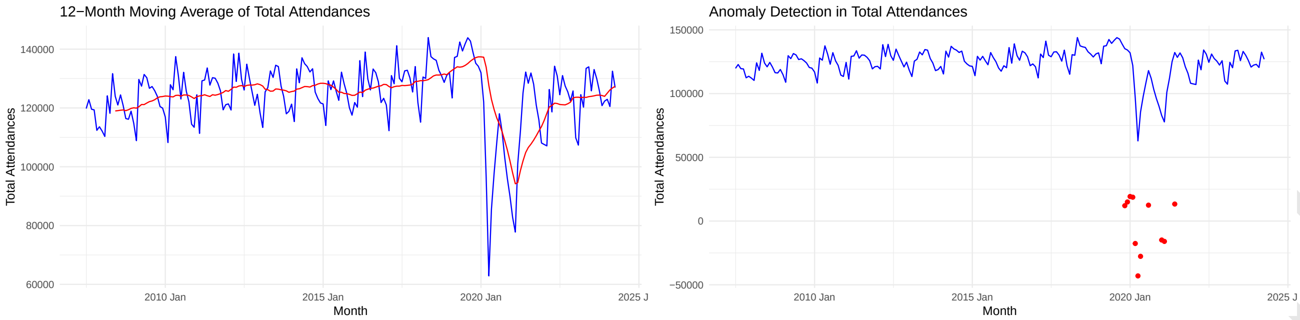

Figure 5: 12-Month Moving Average of Monthly Attendances and Anomaly Detection

2.2 Percentage of Short Waits

It follows from Figure 1 that the percentage of short waits remained high and stable from 2007 to around 2017,, consistently between 90% and 95%, indicating that most patients were seen within the target time frame. However, starting around 2017, there is a gradual decline, which sharply accelerates in early 2020 with the onset of the COVID-19 pandemic.

The most significant drop occurs in early 2020, with the percentage falling rapidly from around 85% to below 60%, reflecting the substantial strain on A&E departments during the pandemic. Despite a slight recovery after that, the values remain significantly lower than pre-2017 levels. Towards the end of the graph, a slight upward trend suggests potential improvements in A&E performance, but the percentage has not yet returned to pre-pandemic levels.

2.3 EDA Summary

- It follows from the STL decomposition that monthly attendances suggest possible seasonal and cyclical effects.
- ACF/PACF suggest the need for autoregressive terms in modelling monthly attendances.
- The effect of COVID-19 pandemic might be considered for modelling monthly attendances and percentage of short waits.
- Monthly attendances and short waits percentages display non-linear effect w.r.t Month.

3 Methods

Generalised Additive Models (GAMs) are applied for modelling both the total attendances and the percentage of short waits. This section gives an overview on GAMs and specify possible models that are used for the analysis.

3.1 GAM

Generalised Additive Models (GAMs) represent an adaptable and robust class of regression models "skillful" at capturing non-linear relationships between response and predictors. Developed by Hastie and Tibshirani in 1987, GAMs enhance traditional linear models by permitting the linear predictor to incorporate smooth functions of the predictor variables (Hastie and Tibshirani 1987).
GAMs are structured as follows:

$$g(E(Y)) = \beta_0 + f_1(X_1) + f_2(X_2) + \ldots + f_p(X_p)$$

where g is a link function, $E(Y)$ is the expected value of the response variable Y, β_0 is the intercept, and $f_1, f_2, \ldots, f_p$ are smooth functions of the predictors $X_1, X_2, \ldots, X_p$. These smooth functions can be non-parametric and are typically estimated using techniques such as splines or local regression (Wood 2017).

One of the significant benefits of GAMs is their flexibility. GAMs are capable of modeling complex, non-linear relationships, which is particularly advantageous in scenarios where the actual relationship between variables is unknown or non-linear (Solonen and Staboulis 2023). Meanwhile, compared to "black-box" machine learning algorithms, they provide much more interpretable insights into the effects of each predictor on the response variable (Radenovic et al. 2022). Additionally, GAMs incorporate smoothing penalties to prevent overfitting (control wiggleness) (Elston and Proe 1995; Feng and Simon 2018). Given these advantages, and considering the non-linear nature of the data, GAMs are employed for modelling to uncover potentially unobserved patterns.

3.2 Models for Total Attendances

This subsection describes three GAMs designed to predict total attendances ($TAttendEp$) in A&E departments, incorporating different variables and modelling techniques. These models incorporate smooth functions to capture trends and seasonality, lag variables to address autocorrelation, and a COVID-19 period indicator to account for pandemic effects (COVID period: 24 months starting from March 2020), all of which uses a Gaussian family.

3.2.1 Model 1: Basic GAM with Lagged Variable

The first model includes a smooth function of the scaled month to account for long-term trends, a cyclic smooth function of the month to capture seasonality, a lag variable, and a COVID-19 period indicator. Mathematically,

$$\mathbb{E}(TAttendEp_t) = \beta_0 + f_1(Time_t) + f_2(nMonth_t) + \beta_1 \cdot Lag1_t + \beta_2 \cdot CovidPeriod_t$$

where:

- $Time$ is a rescaled version of the month variable. Calculated as the numeric representation of the month minus the minimum month value in the dataset. Used to capture long-term trends in the data.
- $f_1(Time_t)$ is a smooth function of Time fitted by a shrinkage version of Cubic splines.
- $f_2(nMonth_t)$ is a cyclic smooth function of the month, fitted by cyclic cubic splines.
- $Lag1_t$ is the lagged total attendances at time $t-1$.
- $CovidPeriod_t$ is a binary indicator for the COVID-19 period (0 for non-COVID period).

The corresponding syntax is as follows:
`TAttendEp ~ s(Time, bs = "cs") + s(nMonth, bs = "cc") + CovidPeriod + Lag1`.

3.2.2 Model 2: GAM with Interaction Term and Multiple Lags

The second model extends the first by adding multiple lag variables to capture autocorrelation more comprehensively and by including a smooth term that varies by the COVID-19 period, i.e.

$$\begin{aligned}\mathbb{E}(TAttendEp_t) = \beta_0 + f_1(Time_t) + f_2(nMonth_t) + f_3(Time_t|CovidPeriod_t)\\ + \beta_1 \cdot Lag1_t + \beta_2 \cdot Lag2_t + \beta_3 \cdot Lag3_t + \beta_4 \cdot CovidPeriod_t\end{aligned}$$

where:

- $f_3(Time_t|CovidPeriod_t)$ represents separate smooths of time modulated by the COVID-19 period, fitted by thin plate regression splines.
- $Lag2_t$ and $Lag3_t$ are the lagged total attendances at time $t-2$ and $t-3$, respectively.

The corresponding syntax is:
`TAttendEp ~ s(Time, bs = "cs") + s(nMonth, bs = "cc") + s(Time, by = CovidPeriod) + Lag1 + Lag2 + Lag3 + CovidPeriod.`

3.2.3 Model 3: GAM with Additional Seasonal Components

The third model builds on the second by incorporating additional seasonal components through quarterly dummy variables and refining the smooth functions with specific parameters:

$$\begin{aligned}\mathbb{E}(TAttendEp_t) = \beta_0 + f_1(Time_t) + f_2(nMonth_t) + f_3(Time_t|CovidPeriod_t) \\ + \beta_1 \cdot Lag1_t + \beta_2 \cdot Lag2_t + \beta_3 \cdot Lag3_t + \beta_4 \cdot CovidPeriod_t + \beta_5 \cdot Quarter_t\end{aligned}$$

where:

- f_1, f_2, and f_3 are smooth functions.
- $Quarter_t$ represents the quarterly dummy variables.

The corresponding syntax is:
`TAttendEp ~ s(Time, bs = "cs", k = 20) + s(nMonth, bs = "cc", k = 12) + s(Time, by = CovidPeriod, k = 10) + Lag1 + Lag2 + Lag3 + CovidPeriod + Quarter.`

3.3 Models for Percentage of Short Waits

Assume the percentage of people waiting less than 4 hours follows a Beta distribution. The first model (Model 4) includes smooth functions of the scaled month and the month of the year to capture long-term trends and seasonality, respectively. Additionally, it incorporates separate smooths for the scaled month during the COVID-19 period.

$$\text{logit}(\mathbb{E}(Percentage_t)) = \beta_0 + f_1(Time_t) + f_2(nMonth_t) + f_3(Time_t|CovidPeriod_t)$$

The corresponding syntax is:

```
Percentage ~ s(Time, bs = "cs", k = 50) + s(nMonth, bs = "cc", k = 12) +
             s(Time, by = CovidPeriod, k = 20) + CovidPeriod.
```

Meanwhile, the second model ((Model 5)) simplifies the first by removing separate smooths for the scaled month and the Covid Period factor, focusing primarily on the smooth functions for long-term trends and seasonality:

$$\text{logit}(\mathbb{E}(Percentage_t)) = \beta_0 + f_1(Time_t) + f_2(nMonth_t)$$

The corresponding syntax is:

```
Percentage ~ s(Time, bs = "cs", k = 20) + s(nMonth, bs = "cc", k = 12).
```

4 Results

4.1 Workflow

The project workflow involved a systematic approach to model and predict total attendances and the percentage of short waits. Initially, the dataset was divided into training and test sets, with data from November 2023 to April 2024 used for training, and the remaining used for testing. Multiple model suggested in the previous section were fitted using the training set to capture long-term trends, seasonality, autocorrelation, and the impact of the COVID-19 pandemic. Model comparison was conducted using Root Mean Squared Error (RMSE) and Akaike Information Criterion (AIC) to identify the best-performing model, which was then visually compared through plots of actual, fitted, and predicted values. The selected model underwent validation, including residual analysis. After validation, the model was refitted using the entire dataset. Forecasts for May 2024 to October 2024 were then generated, with prediction intervals calculated. These intervals, along with the forecasted values, were plotted to visually represent expected future trends and their associated uncertainty.

4.2 GAM: Total Attendances

4.2.1 Model Comparsions

Figure 6~8 display the performance of different models in predicting the monthly attendances at A&E departments. Each figure compares the actual attendances (black line) with the fitted values on the training set (blue line) and the predicted values on the test set (green line).

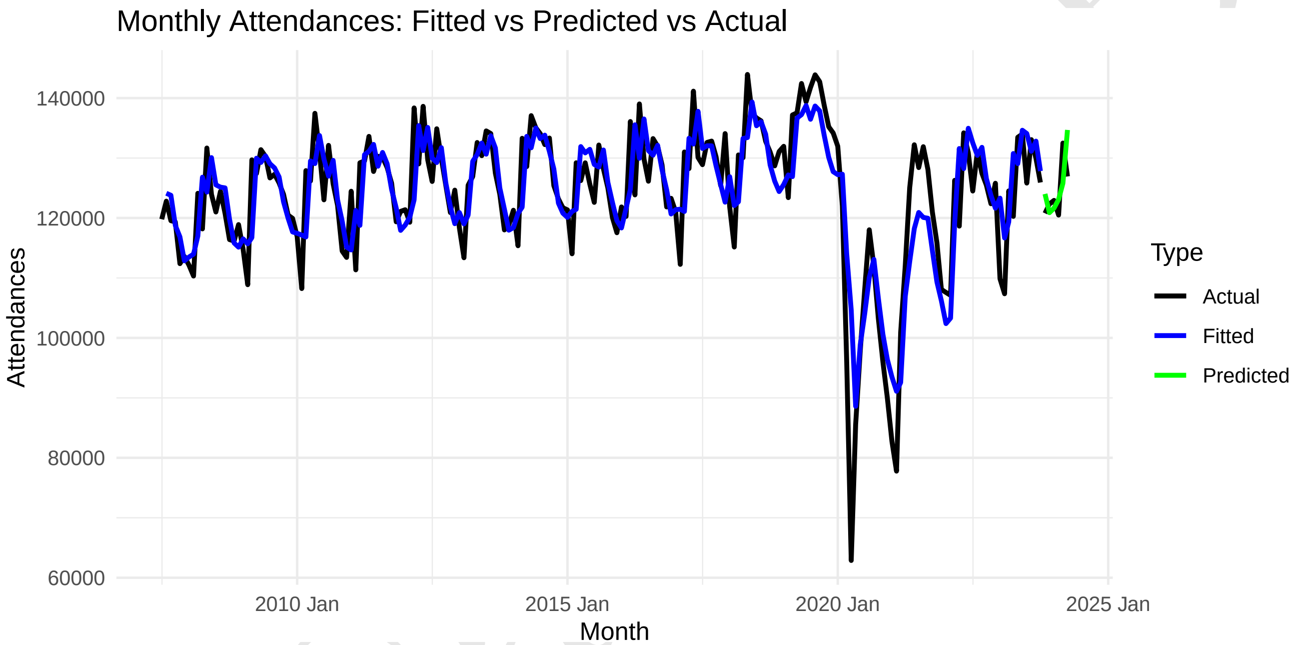

Figure 6: Monthly Attendances: Fitted vs Predicted vs Actual(Model 1)

It follows that Model 3, with its comprehensive approach, offers the best performance in fitting and predicting A&E attendances, as evidenced by the close alignment of fitted and predicted values with the actual data, as well as its lowest RMSE and AIC values in Table 1.

Table 1: Model Performance Metrics (Monthly Attendances)

Model	RMSE	AIC
Model 1	4585.681	4000.049
Model 2	4394.503	3843.654
Model 3	3149.481	3750.068

4.2.2 Model Validations

Table 2 presents the basis dimension (k) checking results for the smooth terms in the model, including each term's effective degrees of freedom (edf) and p-values. The k-index values are close to 1, and the relatively high p-values suggest that the selected basis dimensions are suitable and the model is not overfitting.

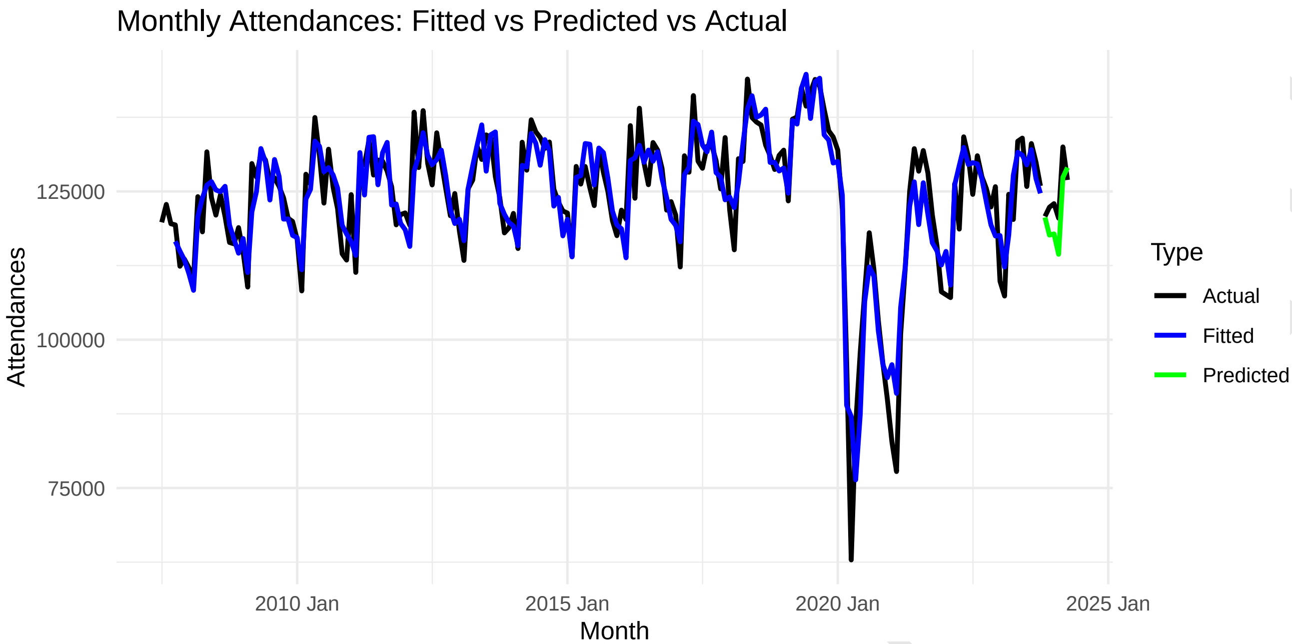

Figure 7: Monthly Attendances: Fitted vs Predicted vs Actual(Model 2)

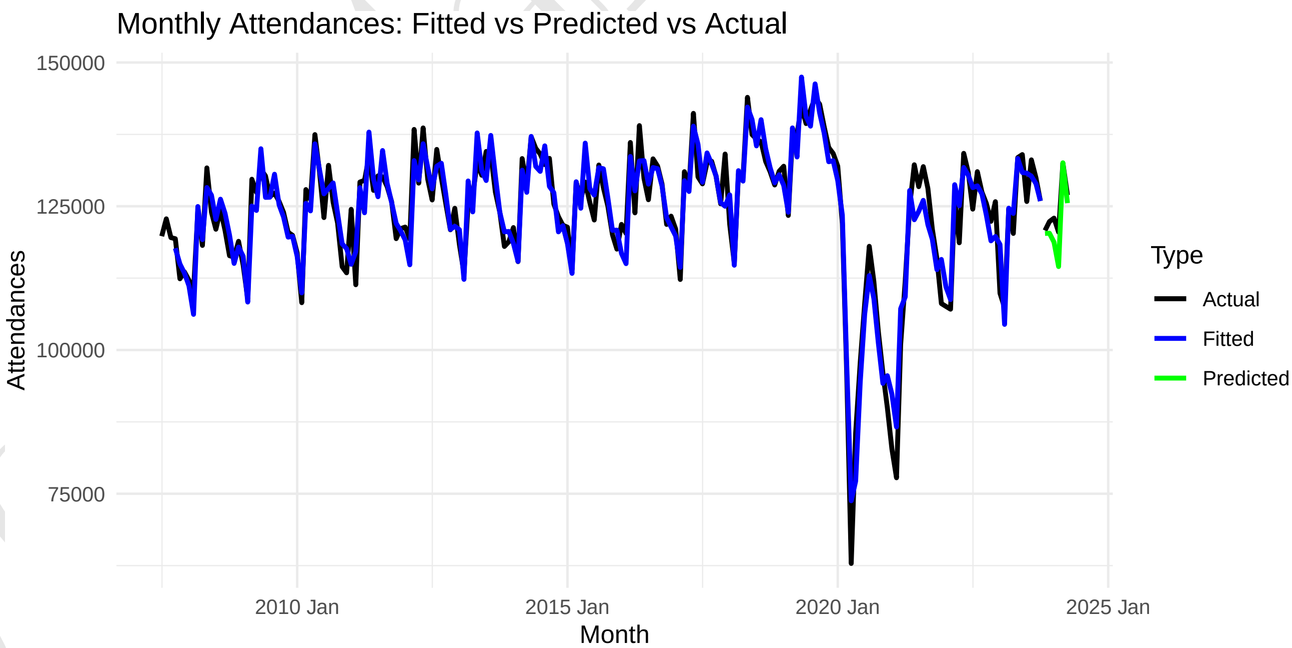

Figure 8: Monthly Attendances: Fitted vs Predicted vs Actual(Model 3)

Table 2: Basis Dimension (k) Checking Results (Monthly Attendances)

	k'	edf	k-index	p-value
s(Time)	19	4.2729454	1.050267	0.7525
s(nMonth)	10	7.8818451	1.039751	0.7275
s(Time):CovidPeriod0	9	6.2727375	1.050267	0.7650
s(Time):CovidPeriod1	9	0.9556506	1.050267	0.7725

Figure 9 shows four diagnostic plots for checking the fit of the model. It follows from the Q-Q plot that the residuals generally adhere to the 45-degree reference line; the residuals versus linear predictor plot does not exhibit any discernible patterns; the histogram of residuals shows that the residuals are centered around zero and exhibit a roughly symmetric distribution; the response versus fitted values plot demonstrates a close alignment between the fitted values and the actual response values. These results indicate that the model is satisfactory.

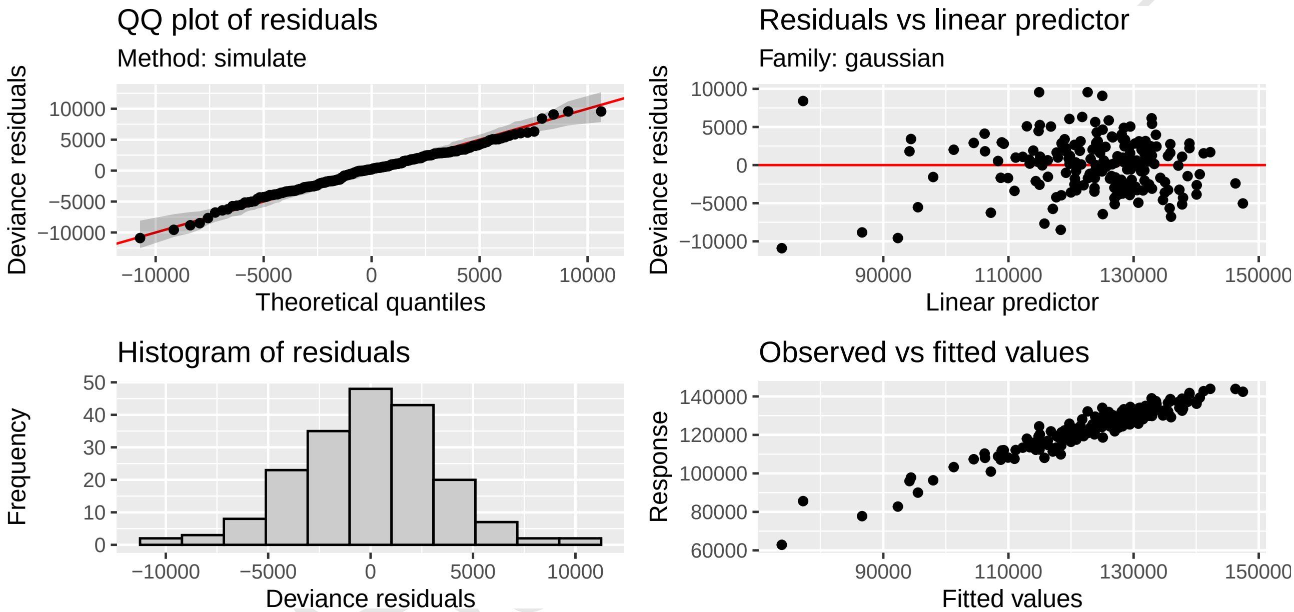

Figure 9: Total Attendances: GAM Checking

4.2.3 Model Summary and Total Attendances Forecast

The following results are the summary of Model 3 fitted using the whole datasets, as well as the forecasting information.

Table 3 shows the summary of parametric coefficients of the model. The coefficient for `Lag1` is 0.55 ($p < 0.0001$), implying that an increase in attendances last month leads to an increase in this month's attendances. Conversely, the coefficients for Lag2 and Lag3 are -0.16 ($p = 0.0287$) and -0.19 ($p = 0.0008$), respectively, indicating significant negative effects from two and three months ago. Meanwhile, the COVID-19 period is represented by the `CovidPeriod1` coefficient, which is -1.267443×10^5 ($p < 0.0001$). This large negative value indicates a significant decrease in attendances during the COVID-19 period, possibly because of lockdowns, excessive burden of the medical system, and people avoiding hospitals, afraid of being infected. Quarterly effects are also included in the model, suggesting a significant decrease in attendances in the second quarter compared to the first quarter, and a significant increase in attendances in the fourth quarter compared to the first quarter.

Table 4 records the summary of smooth terms of the model, indicating the smooths are significantly different from a flat line. However, it follows from Figure10, which illustrates the smooth terms from the model used to analyse total attendances, that some smooths have relatively flat lines with wide confidence intervals (top-left and bottom-left), while `s(nMonth)` captures clear seasonality effect of the month of the year on total attendances.

Note that the adjusted R-squared value is 0.897, implying that approximately 89.7% of the variance in total attendances is explained by the model, demonstrating a high level of fit. The deviance explained is 90.9%, further confirming that the model accounts for a significant portion of the variability in the data.

Table 3: Attendance: Summary of Parametric Coefficients

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.101989e+05	8.054633e+03	13.681430	0.0000000
Lag1	5.504848e-01	6.331110e-02	8.694915	0.0000000
Lag2	-1.636928e-01	7.421560e-02	-2.205639	0.0287217
Lag3	-1.869872e-01	5.467850e-02	-3.419758	0.0007811
CovidPeriod1	-1.267474e+05	1.069816e+04	-11.847594	0.0000000
Quarter2	-2.627879e+04	1.132875e+04	-2.319656	0.0215227
Quarter3	-1.254707e+04	1.107949e+04	-1.132459	0.2590022
Quarter4	4.058136e+03	1.298406e+03	3.125474	0.0020810

Table 4: Attendance: Summary of Smooth Terms

	edf	Ref.df	F	p-value
s(Time)	7.5279809	9.0000000	7.894421	0
s(nMonth)	7.8799915	8.0000000	40.519792	0
s(Time):CovidPeriod0	0.9290833	0.9290808	75.337481	0
s(Time):CovidPeriod1	0.9289926	0.9290328	132.915989	0

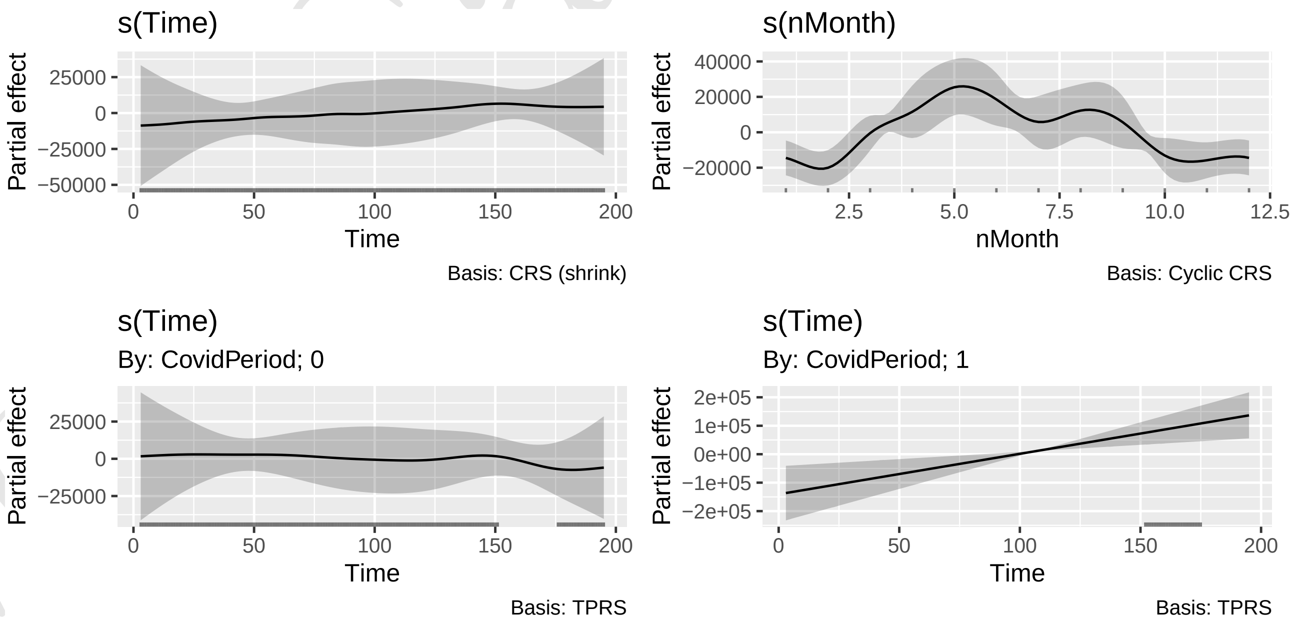

Figure 10: Smooth Terms of the GAM Model for Total Attendances

Finally, Figure 11 and Table 5 provide insights into the trends and forecast of monthly attendances in the A&E departments from May 2024 to October 2024, with 95% prediction intervals specified.

From the forecast results, it can be observed that the number of attendances will not significantly increase compared to the previous months and will fluctuate within a certain range. The number of attendances in May 2024 will increase compared to that in April, but there will be a decrease of about 5% in June and July. This will be followed by a new peak in August, and then a decrease of 5%-6% in the subsequent two months.

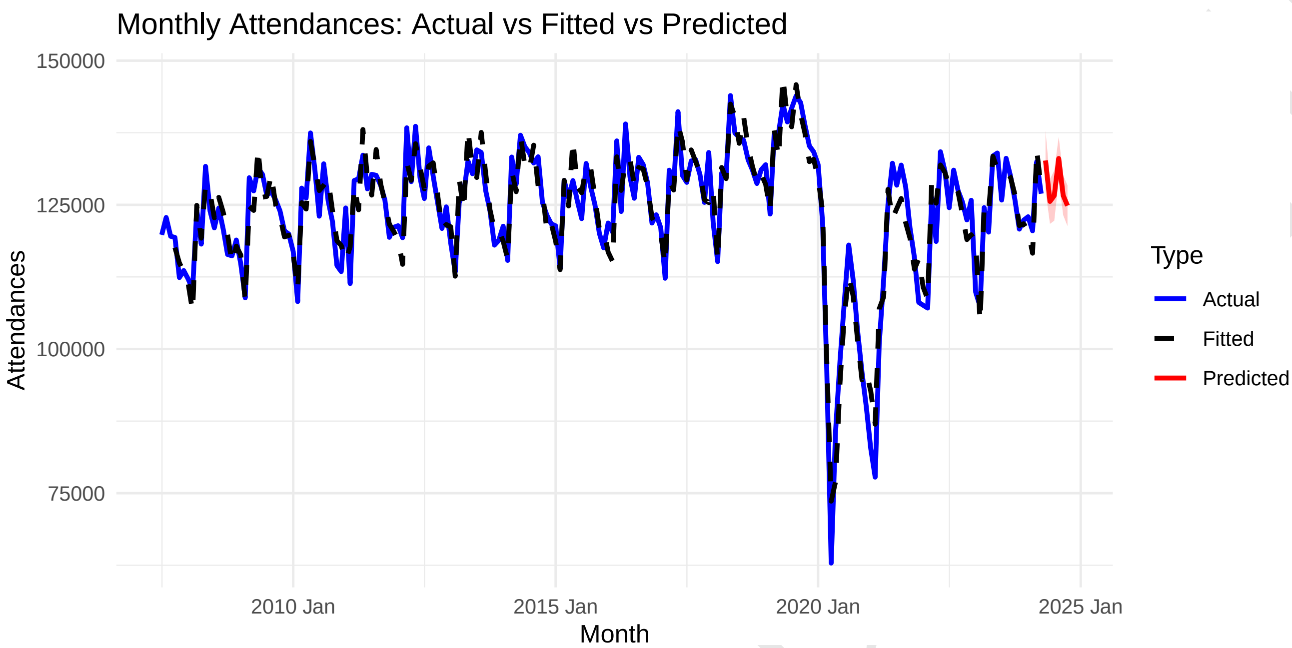

Figure 11: Monthly Attendances: Actual vs Fitted vs Predicted.

Notes: This figure illustrates the observed (Actual) attendances from Jan 2007 to Apr 2024, the model's fitted values over the same period, and the forecasted attendances from May 2024 to October 2024 with 95% prediction intervals.

Table 5: Forecast of Total Attendances: May 2024 to Oct 2024

Month	Forecast	95% Lower Bound	95% Upper Bound
2024 May	132691.4	127530.5	137852.4
2024 Jun	125585.1	121707.6	129462.7
2024 Jul	126602.9	122244.6	130961.2
2024 Aug	133039.8	129258.8	136820.9
2024 Sep	126638.4	123203.1	130073.7
2024 Oct	124858.7	121317.8	128399.7

4.3 GAM: Percentage of Short waits

4.3.1 Model Comparsions

Figure 12-13 display the performance of different models in predicting the monthly percentage of short waits. Each figure compares the actual percentage (black line) with the fitted values on the training set (blue line) and the predicted values on the test set (red line). It follows that Model 5, with a simpler structure, offers a better performance in fitting and predicting percentage of short waits, as evidenced by the close alignment of fitted and predicted values with the actual data, as well as its lower RMSE in Table 6.

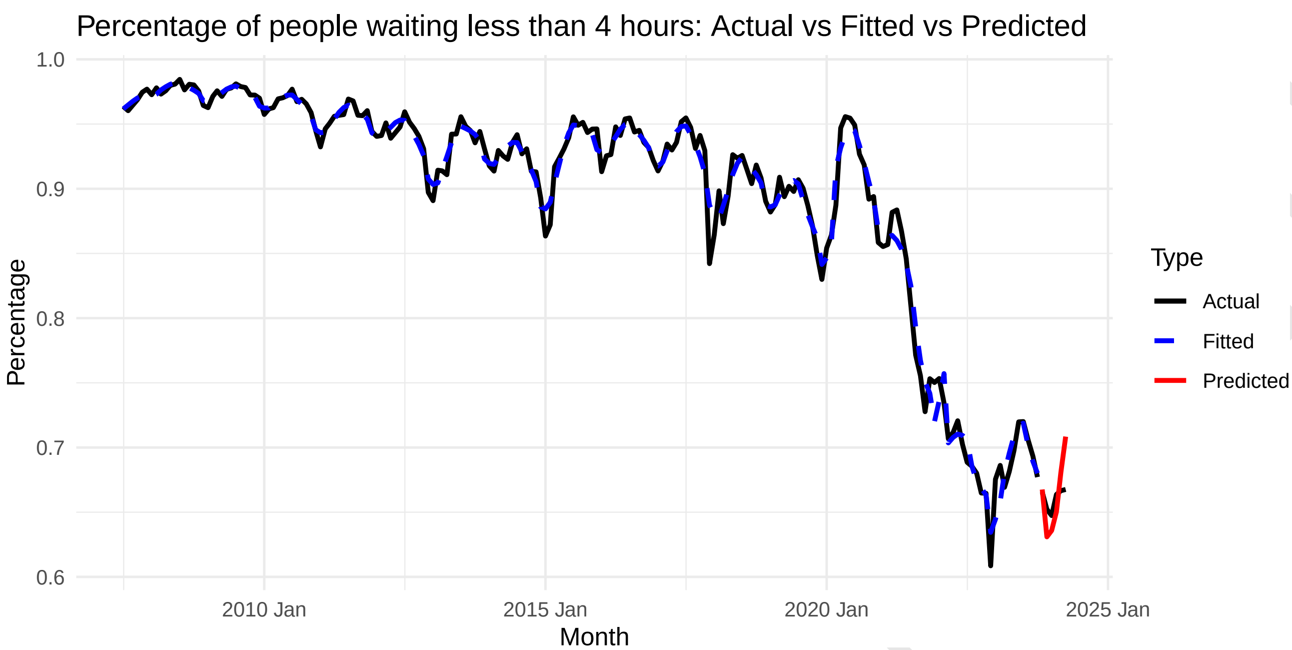

Figure 12: Percentage of people waiting less than 4 hours: Actual vs Fitted vs Predicted (Model 4)

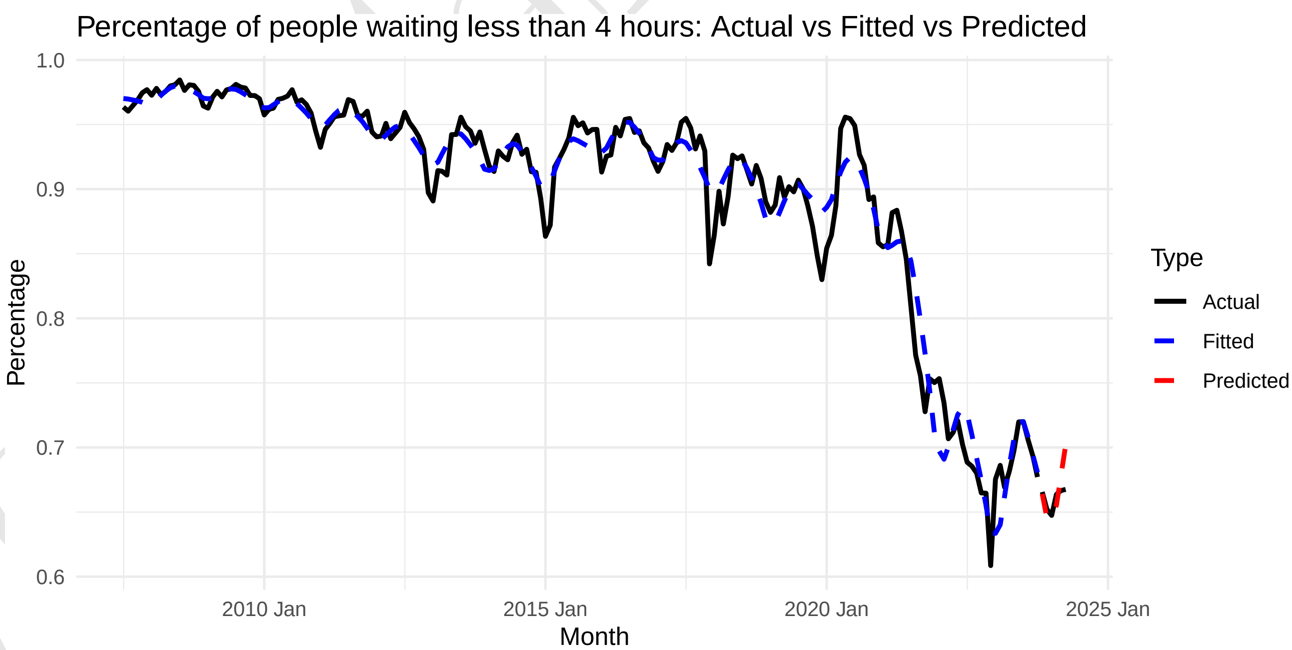

Figure 13: Percentage of people waiting less than 4 hours: Actual vs Fitted vs Predicted(Model 5)

Table 6: Model Performance Metrics (Percentage)

Model	RMSE
Model 4	0.0209768
Model 5	0.0154077

4.3.2 Model Validations

Table 7 presents the basis dimension (k) checking results for the smooth terms in the percentage model, including each term's edf and p-values. While `s(nMonth)` has suitable basis dimensions, the smooth term `s(Time)` is "significant", indicating k is not large enough. However, this is not a significant issue since, unlike predicting attendances, this model does not aim to model autocorrelation. In this case, a large k is not required.

As for GAM checking (Figure 14), the model displays similar goodness of fit shown in the model validations section of total attendances. The Q-Q plot indicates that the residuals largely follow the 45-degree reference line. In the residuals versus linear predictor plot, no noticeable patterns are observed. The histogram of residuals reveals that they are centered around zero and show an approximately symmetric distribution. The response versus fitted values plot shows a strong alignment between the fitted values and the actual response values. These findings suggest that the model performs well.

Table 7: Basis Dimension (k) Checking Results(Percentages)

	k'	edf	k-index	p-value
s(Time)	19	15.202572	0.4162130	0.0000
s(nMonth)	10	4.343284	0.9607421	0.2825

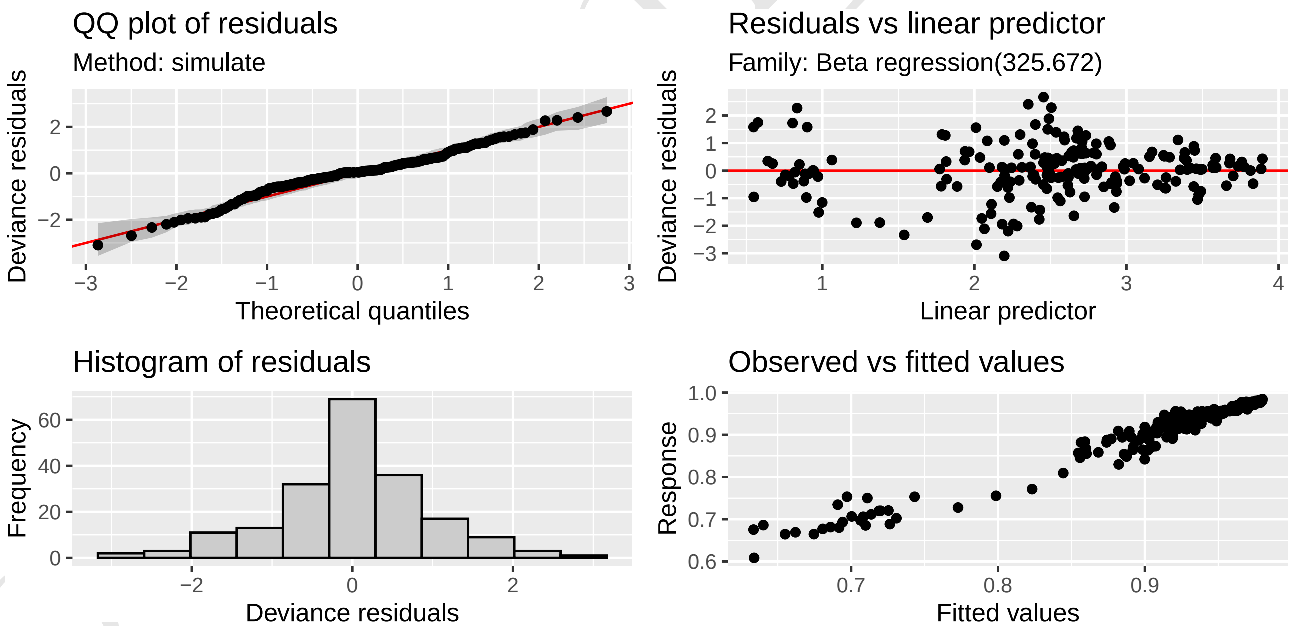

Figure 14: Percentages: GAM Checking

4.3.3 Model Summary and Short Waits Percentage Forecast

The following results are the summary of Model 5 fitted using the whole datasets, as well as the forecasting information.

For the smooth terms, the edf for `s(Time)` is 15.201, with a Chi-square value of 4153.36 ($p < 2e\text{-}16$). This suggests that the relationship between Time and the response variable is highly flexible and significant, highlighting the importance of long-term trends in understanding the percentage of short waits. Similarly, the smooth term s(nMonth) has an edf of 4.343 and a Chi-square value of 98.56 ($p < 2e{-}16$). This indicates significant seasonal patterns in the data, showing that the percentage of short waits varies considerably across different months of the year. The above two results can also be viewed straightforwardly through Figure 15. Note that the summary of parametric term is not shown here, as the only one is the intercept.

The model fit is satisfactory, as indicated by an adjusted R-squared value of 0.958, which means that 95.8% of the variability in the response variable is accounted for by the model. Furthermore, the deviance explained is 96.1%, which reinforces the model's strength in capturing the underlying patterns.

Table 8: Percentages: Summary of Smooth Terms

	edf	Ref.df	Chi.sq	p-value
s(Time)	12.569469	14	4560.1458	0
s(nMonth)	4.237677	10	96.4945	0

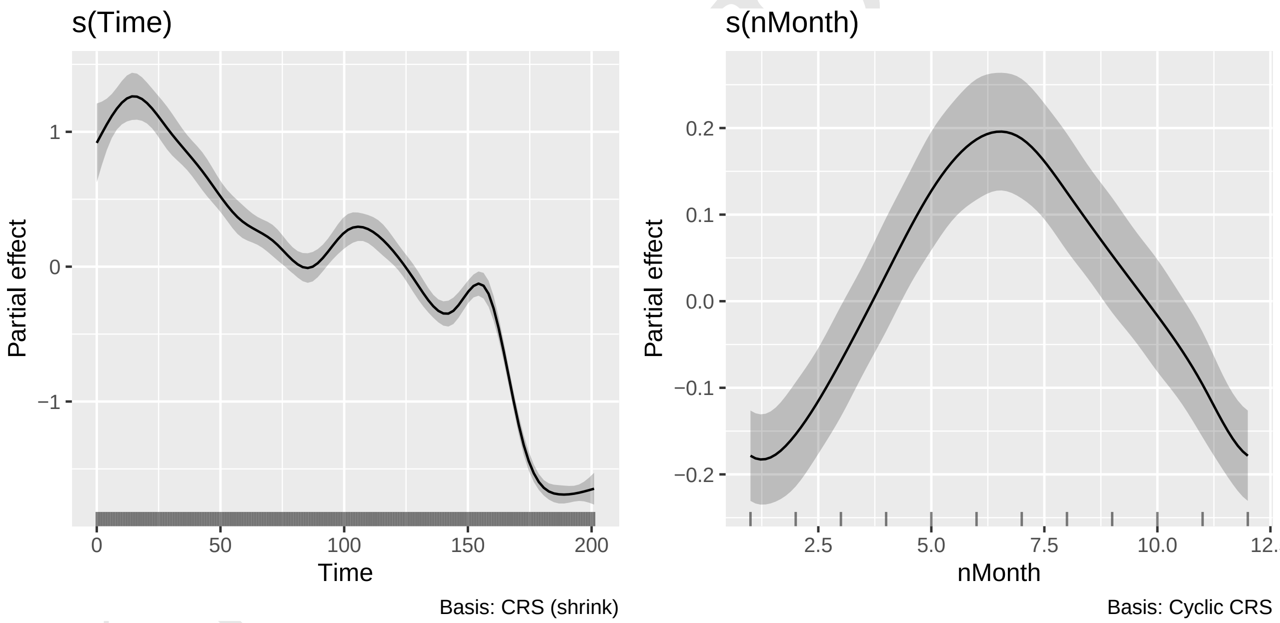

Figure 15: Smooth functions of predictors for Percentages

Finally, Figure 16 and Table 9 provide insights into the trends and forecast of the percentage of people waiting less than 4 hours in the A&E departments from May 2024 to October 2024, with 95% prediction intervals specified. The forecast results indicate that the percentages will fluctuate within a certain range. The percentage in May 2024 is expected to be around 71.4%, with a slight increase in June and July, reaching approximately 72.9% in July. A new peak is predicted for August at around 71.7%, followed by a decrease in September and October, dropping to approximately 68.9% in October.

Notes:

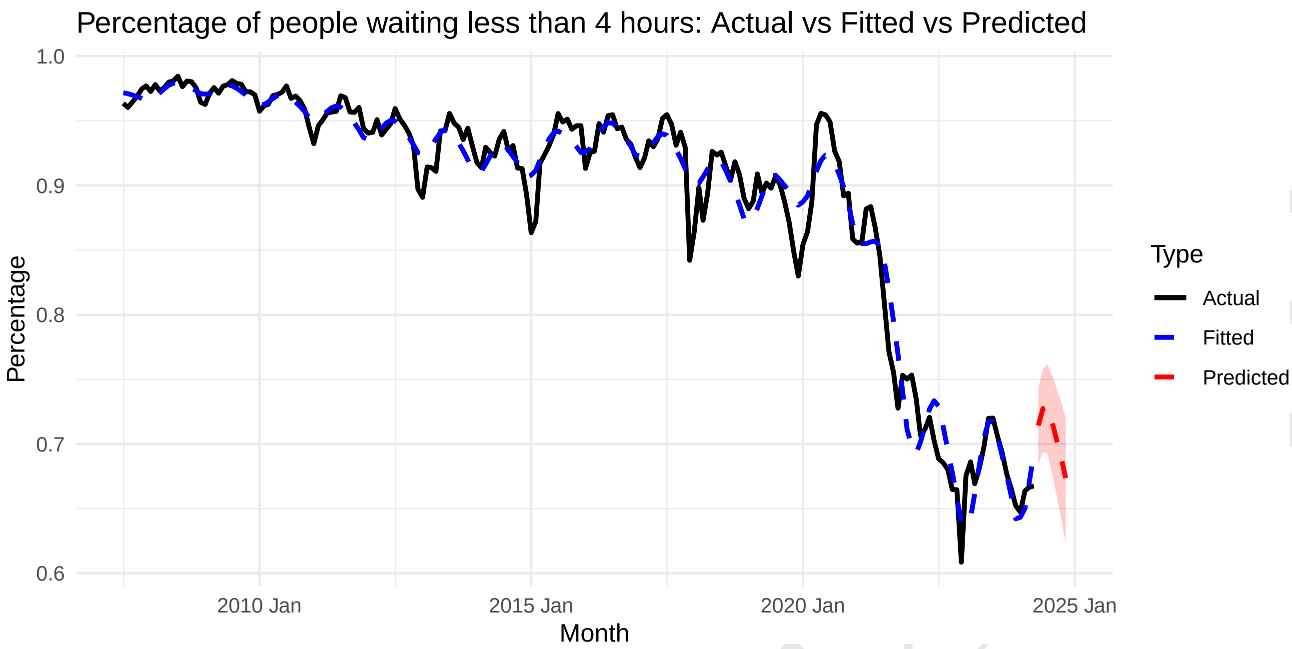

Figure 16: Monthly Percentages: Actual vs Fitted vs Predicted.
Notes: This figure illustrates the observed (Actual) percentages from Jan 2007 to Apr 2024, the model's fitted values over the same period, and the forecasted percentages from May 2024 to October 2024 with 95% prediction intervals.

Table 9: Forecast of Monthly Percentages: May 2024 to Oct 2024

Month	Forecast	95% Lower Bound	95% Upper Bound
2024 May	0.7143187	0.6836207	0.7431572
2024 Jun	0.7273497	0.6939749	0.7583521
2024 Jul	0.7285565	0.6925448	0.7618009
2024 Aug	0.7172151	0.6778216	0.7535432
2024 Sep	0.7033778	0.6603328	0.7430902
2024 Oct	0.6897451	0.6429528	0.7329520

5 Discussion

Models analysing monthly attendances and the percentage of patients waiting less than four hours provide valuable insights. Both models highlight seasonal variations, showing certain times of the year with consistently higher or lower attendances and short wait percentages, possibly influenced by factors like seasonal illnesses, holidays, and weather conditions. Hospitals can use this data to schedule elective surgeries during periods of low demand and prepare for peak seasons. Additionally, the models show a significant drop in attendances and an increase in waiting times during the COVID-19 pandemic, reflecting the strain on healthcare systems and changes in patient behavior. The lag terms in the attendance model reveal that previous attendance levels significantly influence current figures, aiding in forecasting and resource management. Moreover, a declining trend in short wait percentages over the long term suggests increased pressure on medical services, potentially due to a growing population, aging demographics, or healthcare system inefficiencies.

However, it is evident that the two established models employed for predicting attendances and the percentage of short waits have considered only temporal factors and the impact of the COVID-19 period. Consequently, the scope of the models' explanatory power and their predictive accuracy are somewhat limited. In reality, numerous factors such as age, deprivation, type of A&E department, spatial distribution (including different hospitals, various health boards of first treatment (HBT), regional population sizes, numbers of registered GPs, holidays, and the convenience of transportation near hospitals) significantly influence attendances and the proportion of short waits.

- Regarding age, older individuals (those over 30) are more likely to visit A&E, whereas young adults aged 18-24 have the lowest attendance rates (see Figure 17). This discrepancy is likely due to variations in physical health and health awareness across different age groups. For instance, young people typically do not seek A&E services for minor injuries or ailments, often managing them independently. Moreover, younger individuals generally possess a higher capacity for self-recovery, resulting in fewer attendances compared to older age groups.

- Concerning department type, it is the emergency departments that heavily utilise medical resources, thus experiencing significantly higher attendances than minor injuries units or other departments (see Figure 18).

- With respect to deprivation, it is observed that individuals from more deprived areas are more inclined to attend A&E departments (see Figure 19), primarily due to factors such as limited access to primary care and health disparities .

- As for HBTs, variations in attendance numbers and trends are noted across different health boards. For example, HBT=31, corresponding to NHS Greater Glasgow & Clyde, registers the highest attendances (see Figure 20 and (see Figure 21)), which is understandable given the area's high population density and urban nature.

These examples illustrate additional factors that may influence attendances and short waits. Therefore, future models should consider incorporating such factors, potentially by including them within a single model, modelling according to different categories, or integrating random effects such as random slopes, random intercepts, or even spatiotemporal random effects. These avenues present opportunities for further exploration.

Due to time constraints, this report did not incorporate the critical factors mentioned above. Part of the challenge also arises because the PHS public database contains a significant amount of missing data regarding age, sex, and deprivation. Although these aggregated data can be used to analyse different categories of attendances, the missing data could lead to considerable inaccuracies in attendance predictions. At present, the author has not identified an appropriate method for handling missing data or for using such data in predictions, but this remains a promising area for future research.

6 Conclusion

This report analyses the trends and patterns in A&E department attendances and waiting times across Scotland from 2007 to 2024. GAMs were applied to forecast total monthly attendances and the percentage of patients seen within four hours over the months. The findings demonstrate a notable decline in the four-hour performance metric since 2017, which was further compounded by the COVID-19 pandemic starting in 2020. The analysis demonstrated that attendance levels fluctuate significantly due to seasonal influences, with higher numbers observed during winter months due to seasonal illnesses. Prior attendance levels also have a notable impact, indicating that higher numbers of patients attending in previous months tend to result in higher numbers of attendances in subsequent months. The pandemic has had a profound impact, with attendances declining markedly during the peak periods of the pandemic. However, as the situation improved, there was a gradual recovery in attendances. The models predict that there will be continued fluctuations in attendance over the coming months, with slight overall improvements in waiting times. However, these metrics are projected to remain below pre-pandemic levels, which demonstrates ongoing difficulties in A&E departments.

It is important to note that the proposed models only consider the time effect and the effect of the pandemic. While various factors were acknowledged as influential, they have not been incorporated into the models due to data limitations. To improve the models' accuracy and prediction potential, future studies should address the problems with missing data and include other variables including sex, age, and deprivation. A more thorough knowledge can be attained by taking these factors into account, which will help healthcare systems better manage demand and enhance patient outcomes.

7 References

Elston, DA, and MF Proe. 1995. "Smoothing Regression Coefficients in an Overspecified Regression Model with Interrelated Explanatory Variables." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 44 (3): 395–406.

Feng, Jean, and Noah Simon. 2018. "Gradient-Based Regularization Parameter Selection for Problems with Nonsmooth Penalty Functions." *Journal of Computational and Graphical Statistics* 27 (2): 426–35.

Hastie, Trevor, and Robert Tibshirani. 1987. "Generalized Additive Models: Some Applications." *Journal of the American Statistical Association* 82 (398): 371–86.

Hyndman, Rob. 2024. *Fpp3: Data for "Forecasting: Principles and Practice" (3rd Edition)*. https://CRAN.R-project.org/package=fpp3.

PHS. 2023. *A&E Activity and Waiting Times - Month Ending 31 March 2023 - A&E Activity and Waiting Times - Publications - Public Health Scotland — Publichealthscotland.scot*. Https://publichealthscotland.scot/publications/ae-activity-and-waiting-times/ae-activity-and-waiting-times-month-ending-31-march-2023.

PHS. 2024. *Accident and Emergency - Urgent and Unscheduled Care - Acute and Emergency Services - Our Areas of Work - Public Health Scotland — Publichealthscotland.scot*. Https://publichealthscotland.scot/our-areas-of-work/acute-and-emergency-services/urgent-and-unscheduled-care/accident-and-emergency/#section-2-2.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

Radenovic, Filip, Abhimanyu Dubey, and Dhruv Mahajan. 2022. "Neural Basis Models for Interpretability." *Advances in Neural Information Processing Systems* 35: 8414–26.

Simpson, Gavin L. 2024. *gratia: Graceful ggplot-Based Graphics and Other Functions for GAMs Fitted Using mgcv*. https://gavinsimpson.github.io/gratia/.

Solonen, Antti, and Stratos Staboulis. 2023. "On Bayesian Generalized Additive Models." *arXiv Preprint arXiv:2303.02626*.

Wang, Earo, Dianne Cook, and Rob J Hyndman. 2020. "A New Tidy Data Structure to Support Exploration and Modeling of Temporal Data." *Journal of Computational and Graphical Statistics* 29 (3): 466–78. https://doi.org/10.1080/10618600.2019.1695624.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Davis Vaughan, and Maximilian Girlich. 2024. *Tidyr: Tidy Messy Data*. https://CRAN.R-project.org/package=tidyr.

Wilke, Claus O. 2024. *Cowplot: Streamlined Plot Theme and Plot Annotations for 'Ggplot2'*. https://CRAN.R-project.org/package=cowplot.

Wood, S. N. 2017. *Generalized Additive Models: An Introduction with r*. 2nd ed. Chapman; Hall/CRC.

Wood, S. N. 2011. "Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models." *Journal of the Royal Statistical Society (B)* 73 (1): 3–36.

Wood, S. N., N., Pya, and B. S"afken. 2016. "Smoothing Parameter and Model Selection for General Smooth Models (with Discussion)." *Journal of the American Statistical Association* 111: 1548–75.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.

Xie, Yihui. 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Chapman; Hall/CRC. https://yihui.org/knitr/.

Xie, Yihui. 2024. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.

Appendices

A Figures mentioned in the Discussion part

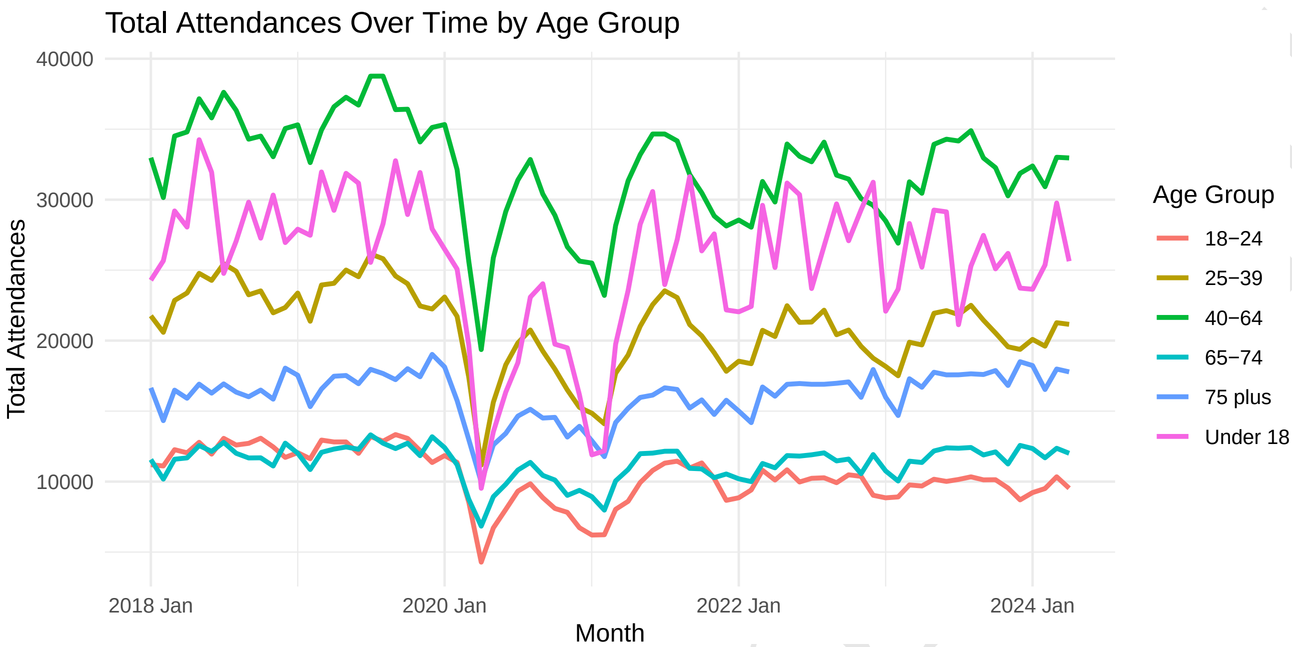

Figure 17: Total Attendances Over Time by Age Group

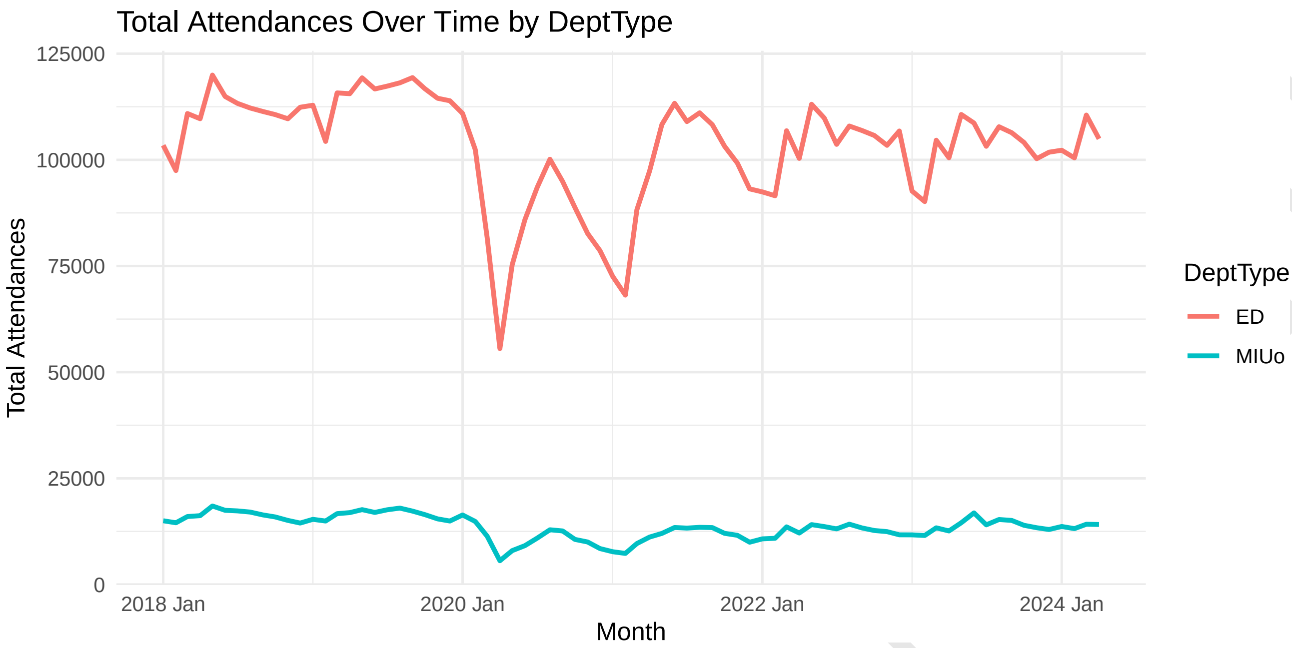

Figure 18: Total Attendances Over Time by DeptType

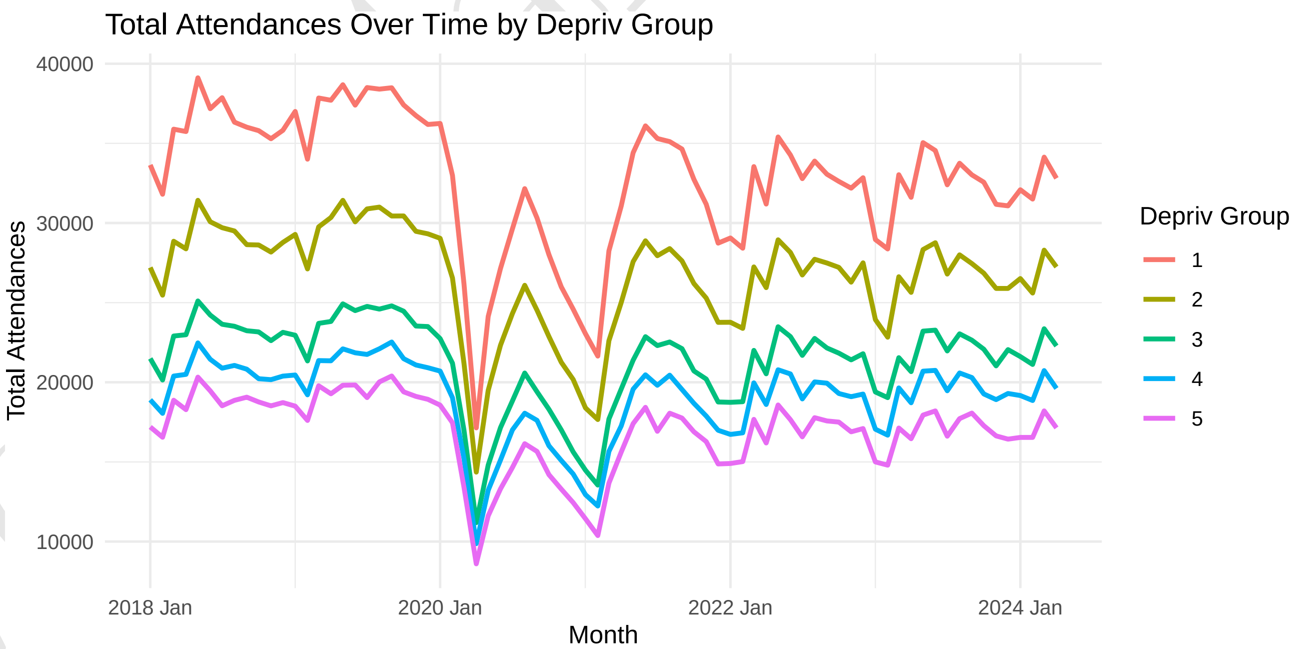

Figure 19: Total Attendances Over Time by Depriv Group

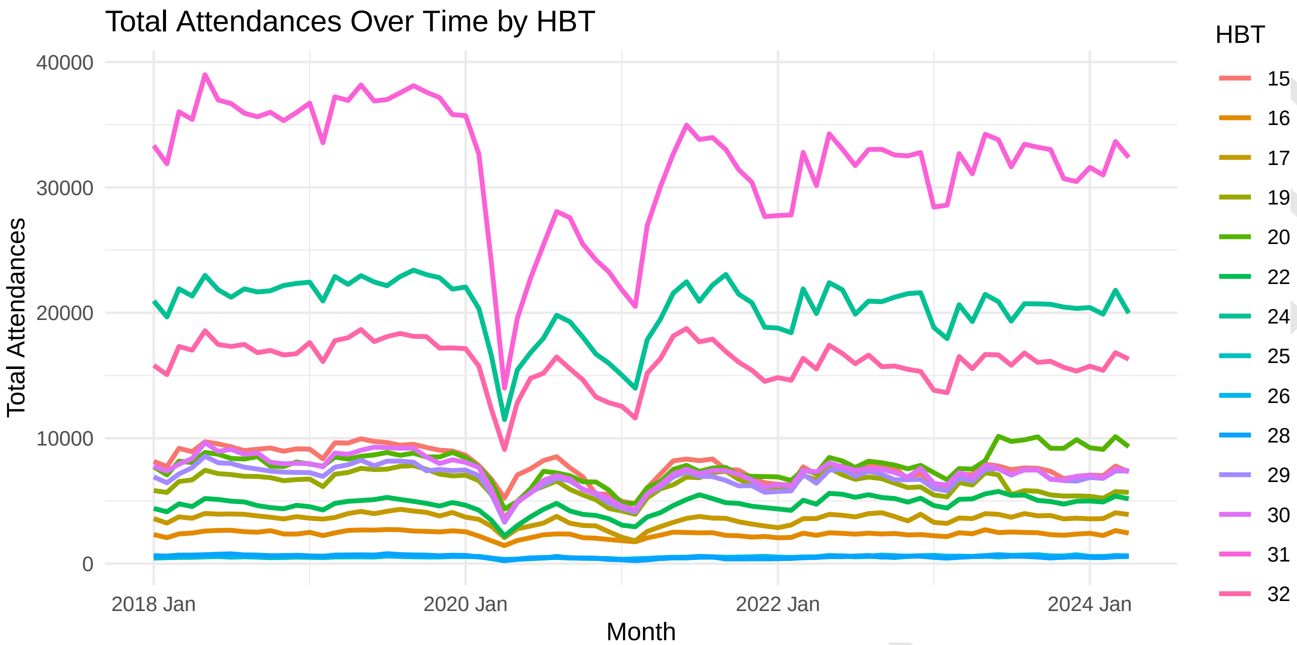

Figure 20: Total Attendances Over Time by HBT

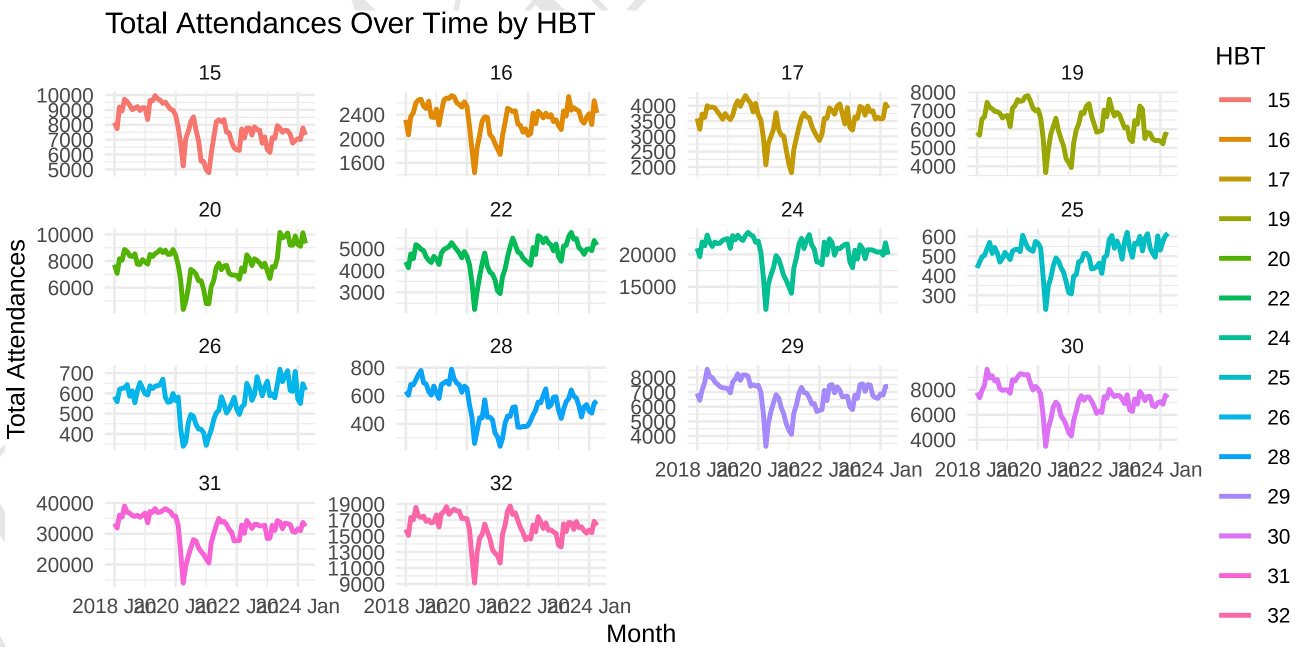

Figure 21: Total Attendances Over Time by HBT (Separate)