Forecasting interrupted time series

Bahman Rostami-Tabar
Cardiff Business School
Wales CF10 3EU
United Kingdom
Email: rostami-tabarb@cardiff.ac.uk
Corresponding author

Rob J Hyndman
Department of Econometrics & Business Statistics
Clayton VIC 3800
Australia
Email: Rob.Hyndman@monash.edu

27 July 2023

JEL classification: C10,C14,C22

Forecasting interrupted time series

Abstract

A brief summary of our ideas

Keywords: blah, blah

1 Introduction

Time series are sometimes interrupted by unusual events; for example, a natural disaster may occur, or a there may be a temporary policy change. Many forecasters have faced this issue recently with the COVID-19 pandemic, where historical patterns may have been severely disrupted due to lockdowns and other restrictions. In this paper, we consider the problem of forecasting after such an event has occurred.

This is a different problem from change point detection. In the situations we consider, we know that a change has occurred, and we want to forecast the future after the change. Change point detection is about identifying when a change has occurred.

The changes in the series as a result of the disruption may be relatively simple, for example a left shift at the start of the disruption, and another at the end of the disruption. Or they may be more complex, with changes to the seasonal patterns, and changes to the level, which evolve over time. In this paper, we consider a range of models that can be used to handle such changes, and compare their performance on some real data sets.

Time series models usually assume that the future is similar to the past, at least in how the data evolve. But a big event like covid makes the future different from the past, at least in the short term.

2 Handling interruptions when forecasting

We consider several possible ways to handle interruptions when forecasting.

2.1 Use a highly adaptive model

Highly adaptive models can adjust to the interruption as it is happening, and will therefore be able to approximate the data generating process relatively well. For example, an ETS model with large smoothing parameters will be able to adjust to the interruption relatively quickly. This has the advantage of being a very simple solution that is easy to implement and fast to compute.

However, the prediction intervals will be large, because the model will have heavily discounted past data. In fact, the model will largely forget the past data other than the most recent observations, so there is no memory of the seasonal patterns and other dynamics that were present before the interruption. Consequently, the approach works best if there is no assumption that the post-interruption period will be similar to the pre-interruption period.

2.2 Use a dynamic regression model with intervention covariates

A dynamic regression model with intervention covariates can be used to model the interruption explicitly. For example, if the intervention involves a simple level shift, with a reverse level shift at the end of the intervention, we can use a dummy variable to indicate the interruption period, and allow the model to adjust to the interruption. More complicated interventions can be handled by using more covariates.

This has the advantage of retaining the memory of the past, and so the seasonal patterns and other dynamics will be retained. However, the model will assume that the post-interruption period will be similar to the pre-interruption period, and so the prediction intervals may be be too narrow.

Advantages: retains full memory of the past, and allows the change period to be effectively modelled provided you choose the covariates well.

Disadvantages: requires a lot of thought to choose the covariates well. Assumes that the post-pandemic period will be similar to the pre-pandemic period.

2.3 Treat the covid period as missing and use a model that handles missing values

Many time series models will handle periods of missing values. So the problematic observations that occur during the period of disruption can be set to missing, and the model should continue to produce forecasts as if the interruption had not occurred. Of course, the forecasts will not be accurate for the period of disruption, but they can be interpreted as "what might have been". This solution requires a judgement to be made about when the disruption has begun, and when normality resumes.

Because no information is retained during the disruption, the prediction intervals will become large during the disruption, and after the disruption they will remain large until the model has enough data to estimate the forecast distribution more accurately.

2.4 Estimate what might have been and adjust the data

A fourth solution is to estimate what might have been during the period of disruption, and then use the adjusted data to fit a model. The estimates could be made using any convenient method. One approach would be to set the estimates to equal the forecasts made using only pre-interruption data. This then becomes almost equal to the previous solution except that the prediction intervals will be narrower because the estimation uncertainty has not been taken into account.

3 Examples

Let's consider some examples

3.1 Australian tourism

The Australian tourism data set is a monthly time series showing the number of short-term overseas visitors to Australia. The data are available from January 2000 to May 2023, and are shown in Figure 1. As the borders closed in March 2020, the number of visitors to Australia dropped to near zero, and remained there until towards the end of 2021. The borders officially reopened on 21 February 2022, although it seems visitors began to arrive earlier than that. The data are available from the Australian Bureau of Statistics.

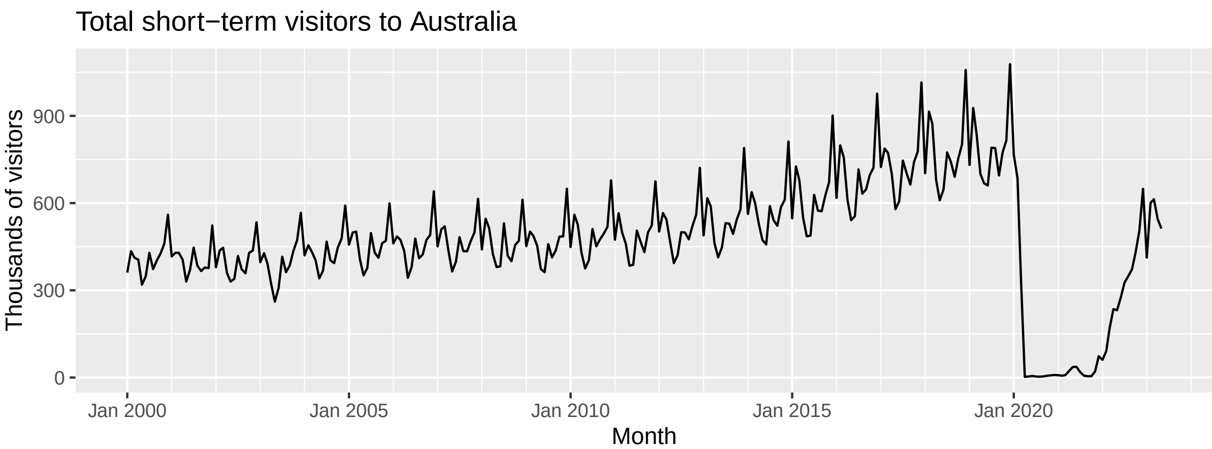


Figure 1: ***?(caption)***

We will apply each of the approaches above to this data set, and compare their performance.

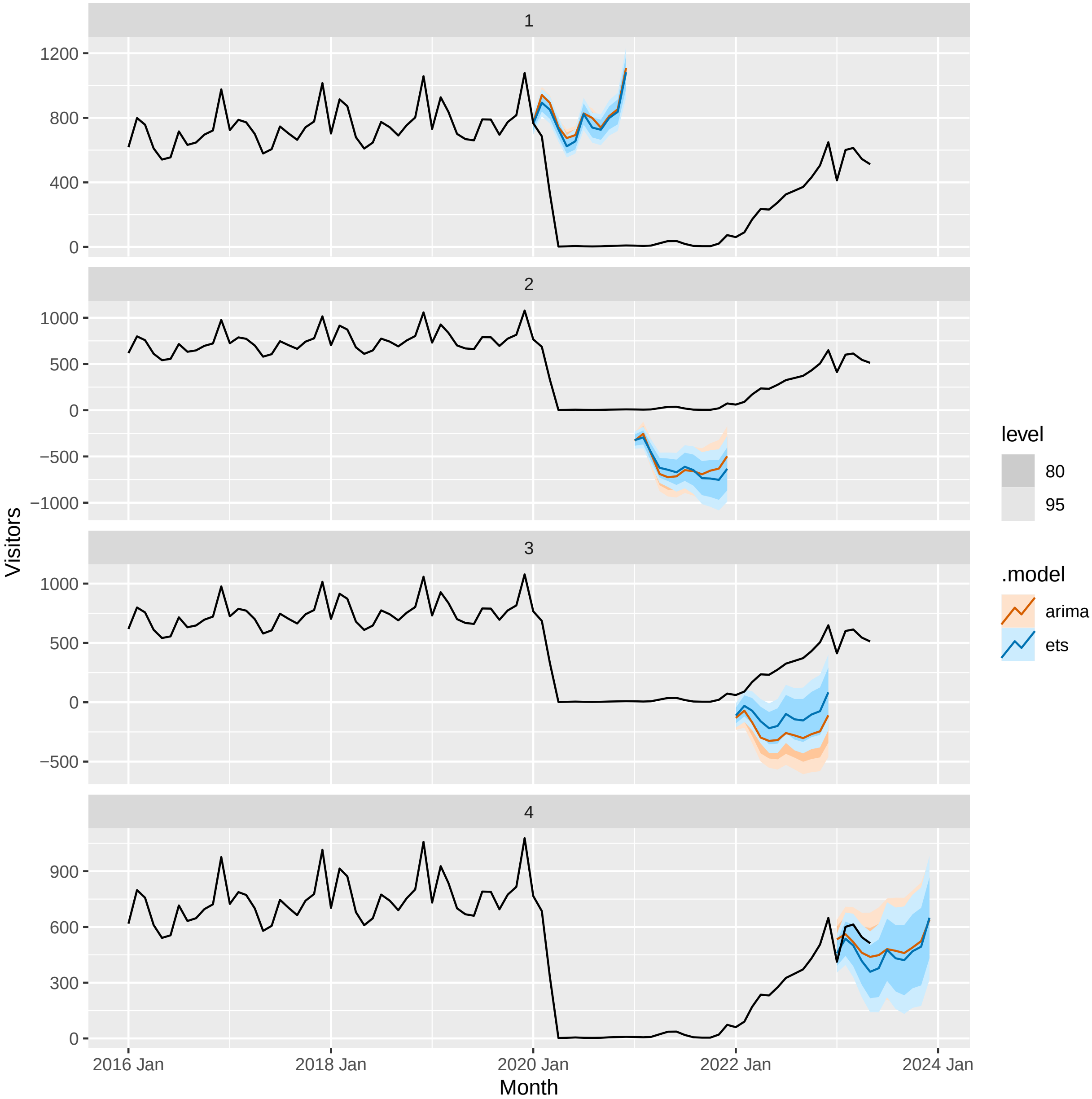


Figure 2: *?(caption)*

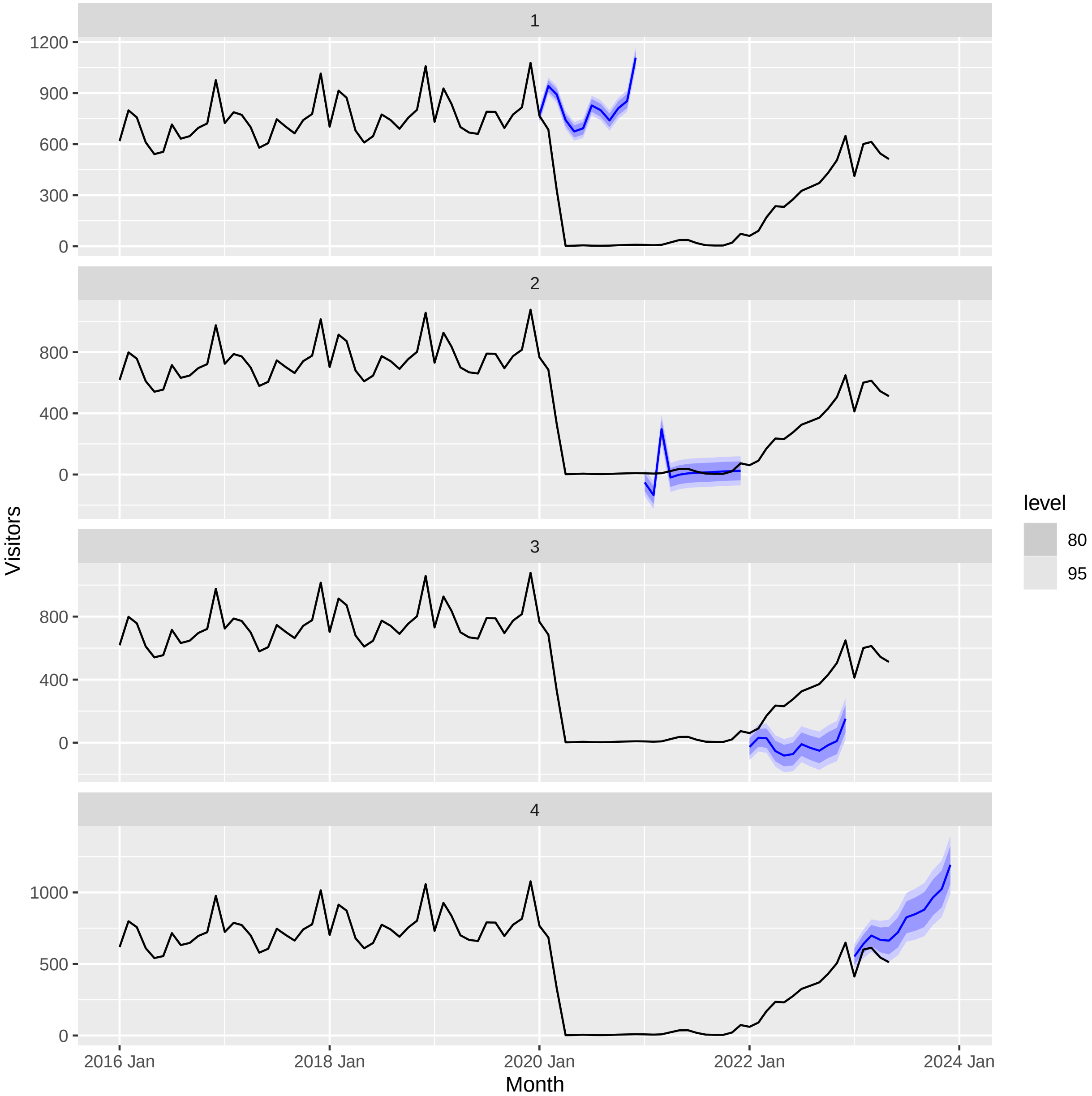


Figure 3: *?(caption)*

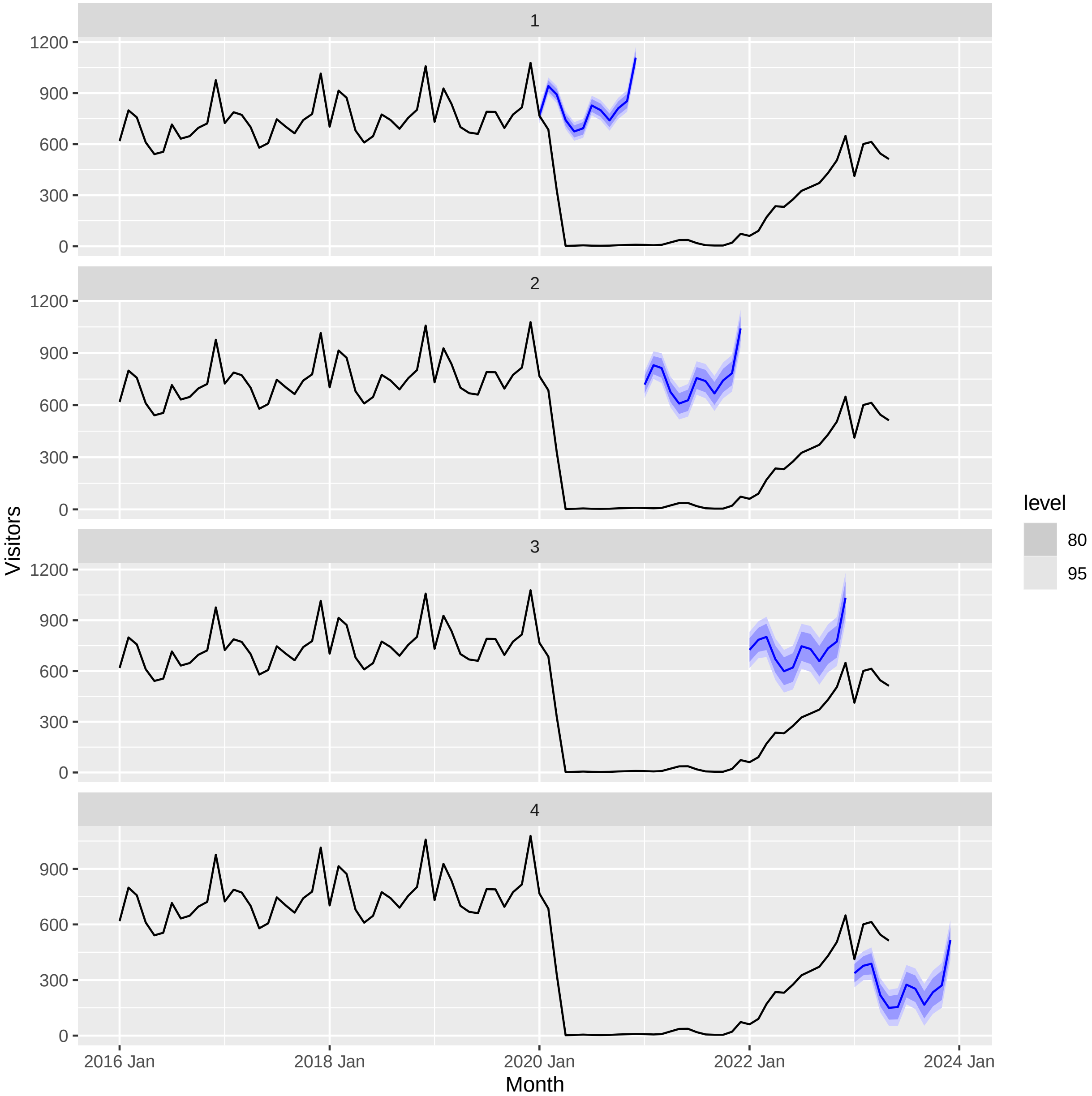


Figure 4: ***?(caption)***

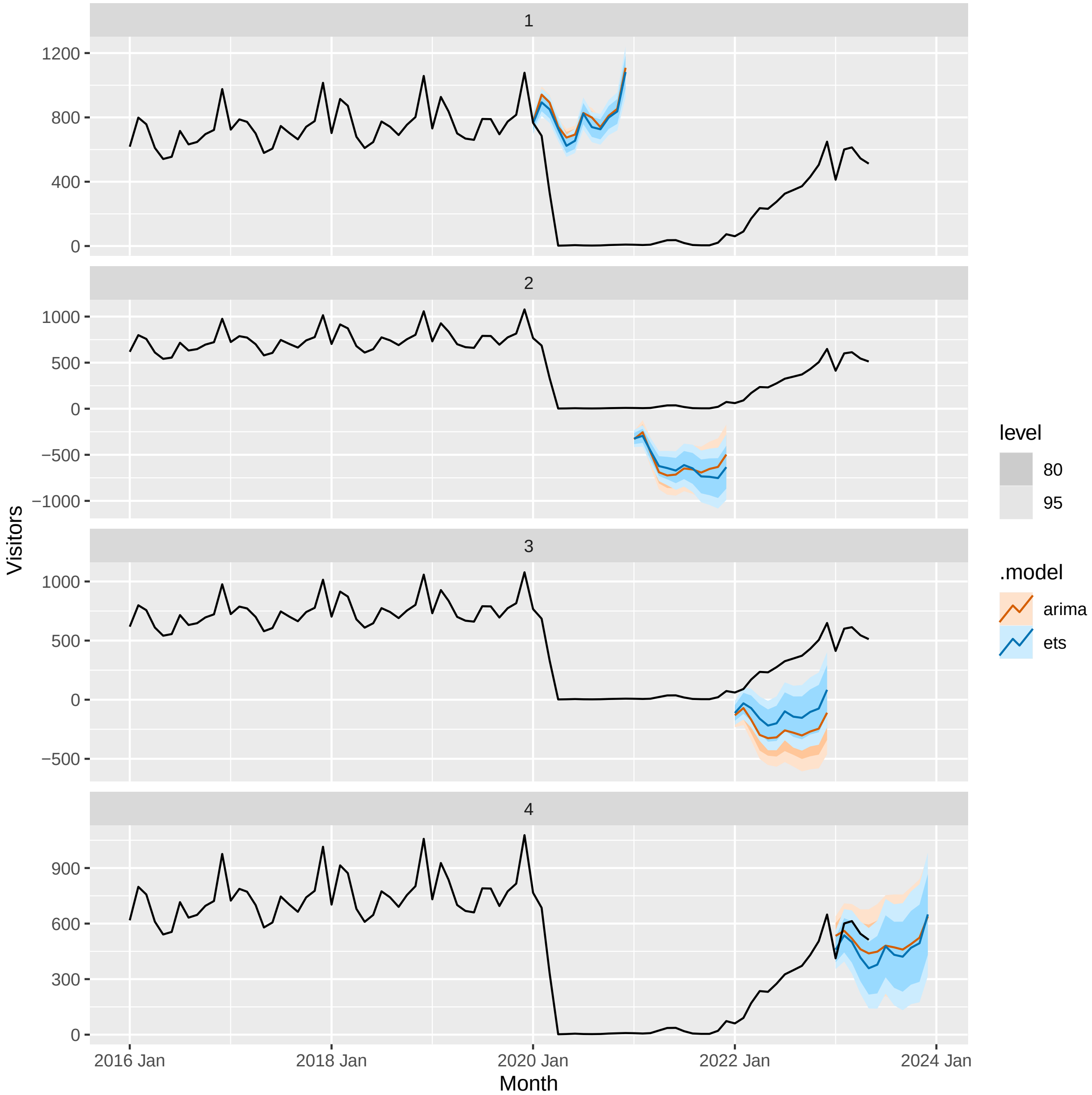
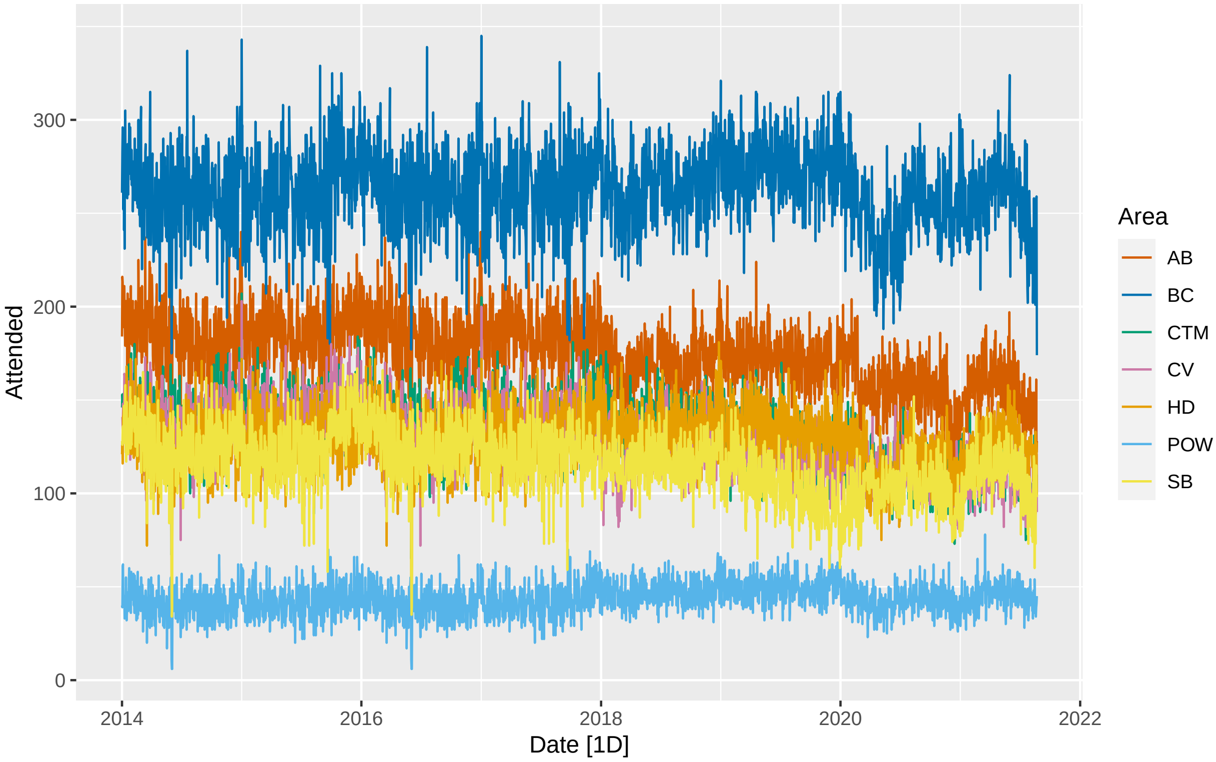
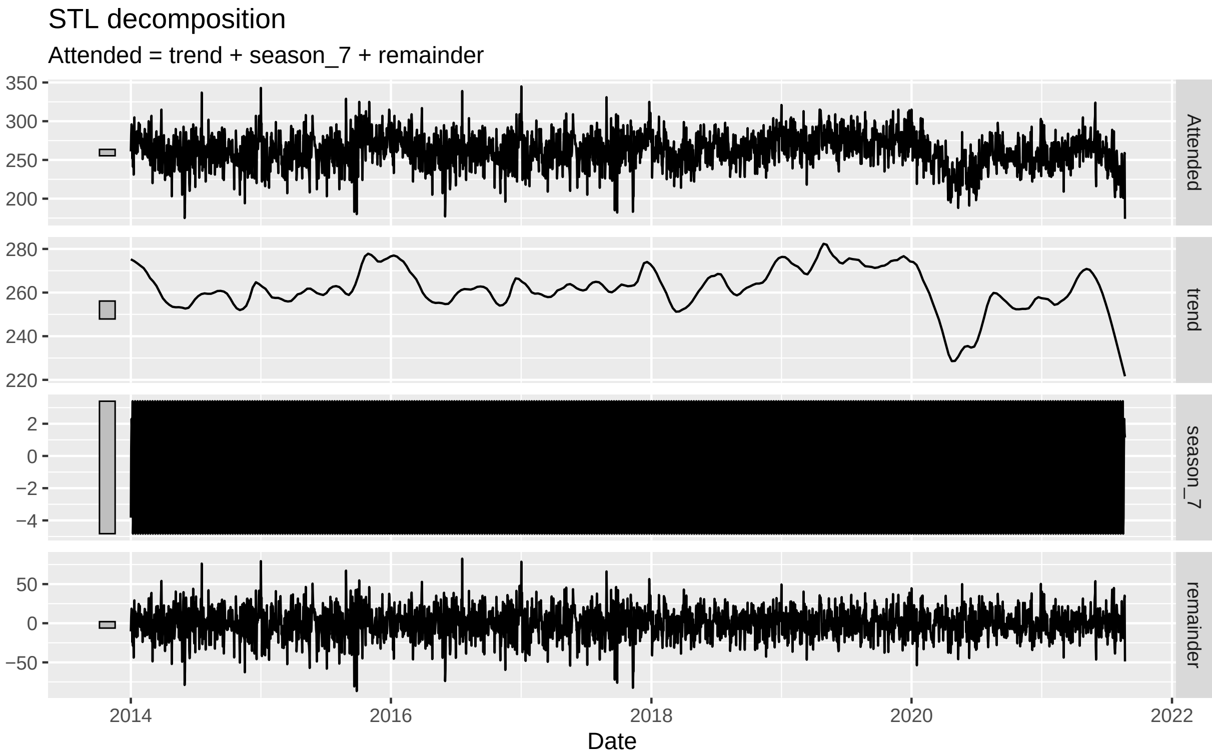
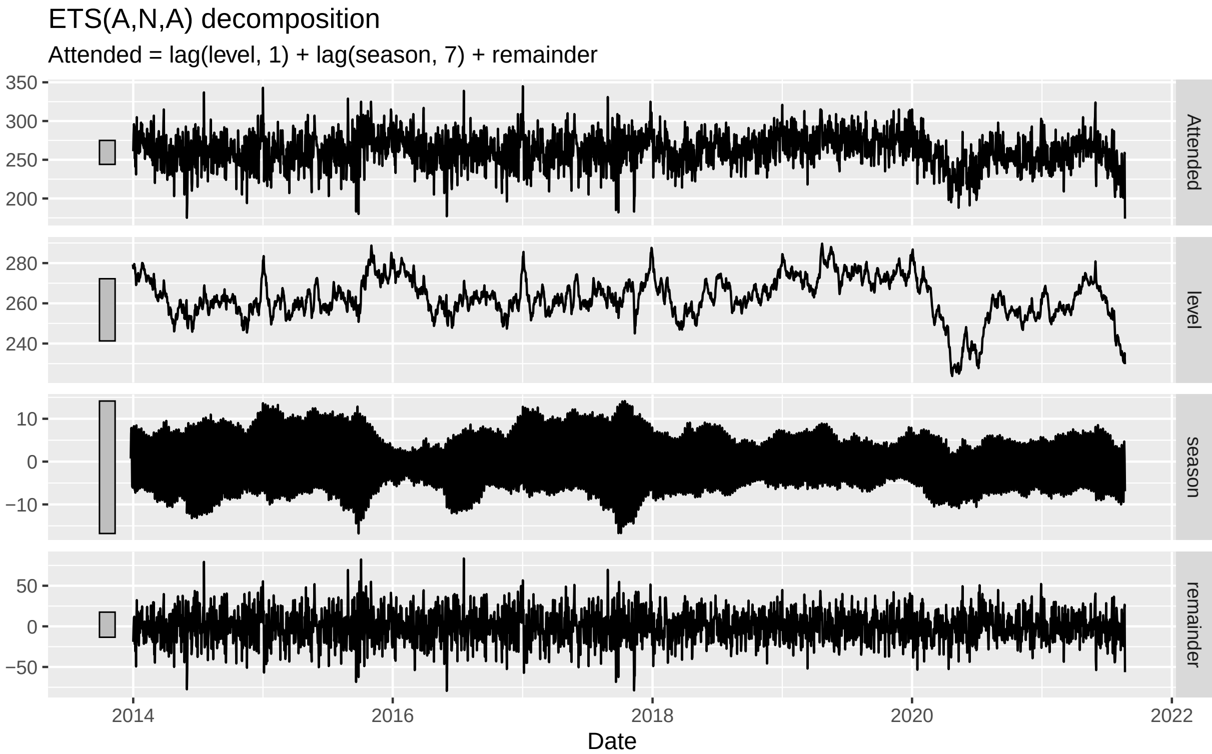


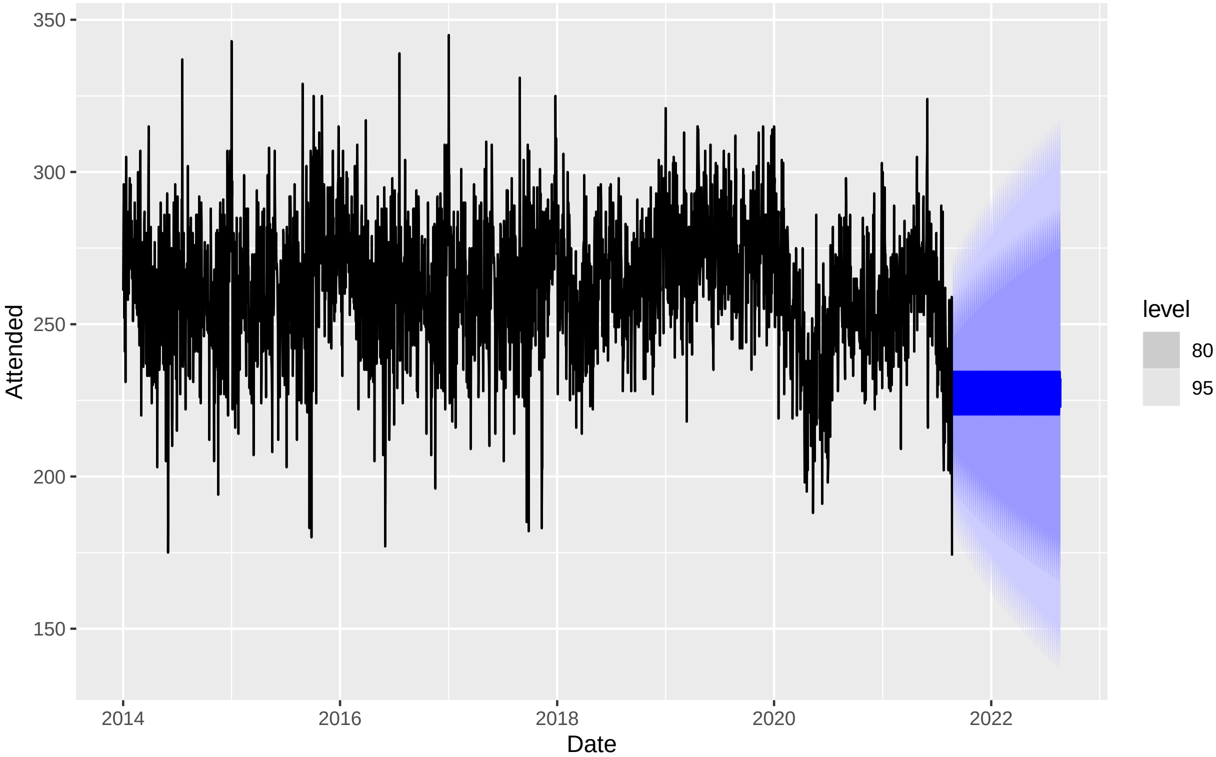
Figure 5: *?(caption)*

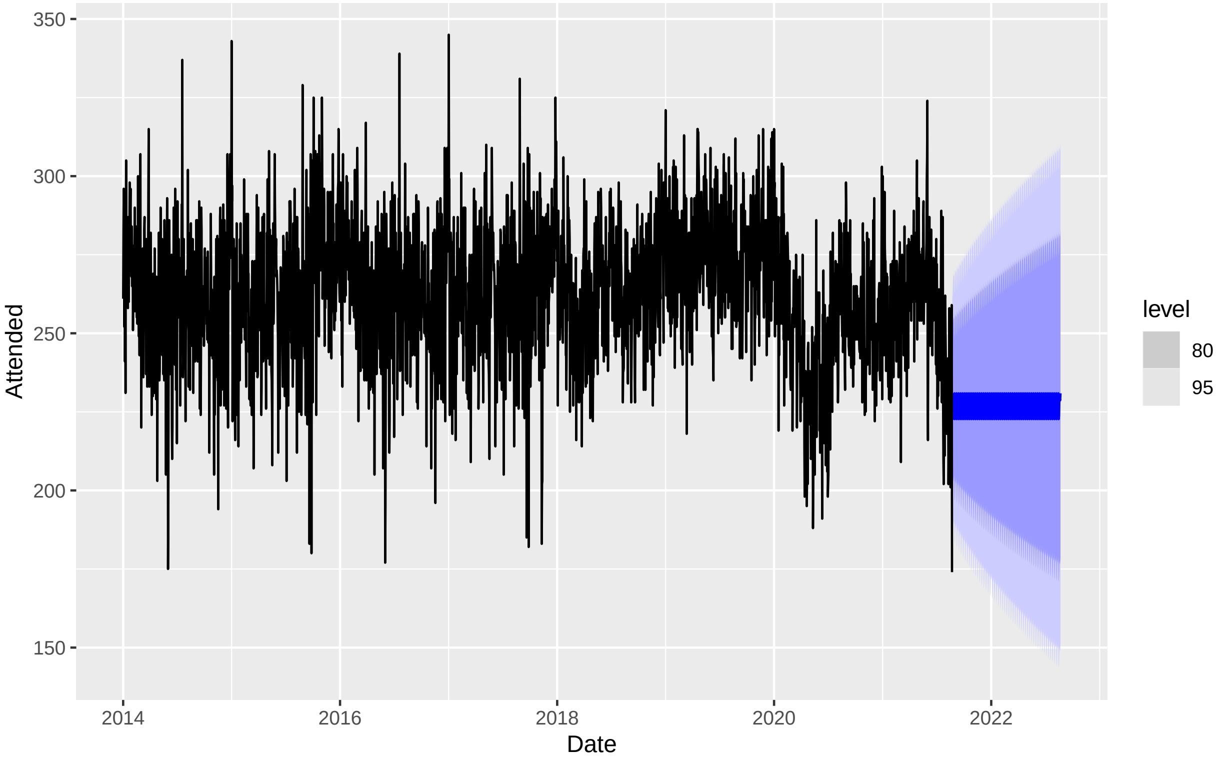
3.2 Example: Ambulance attendances

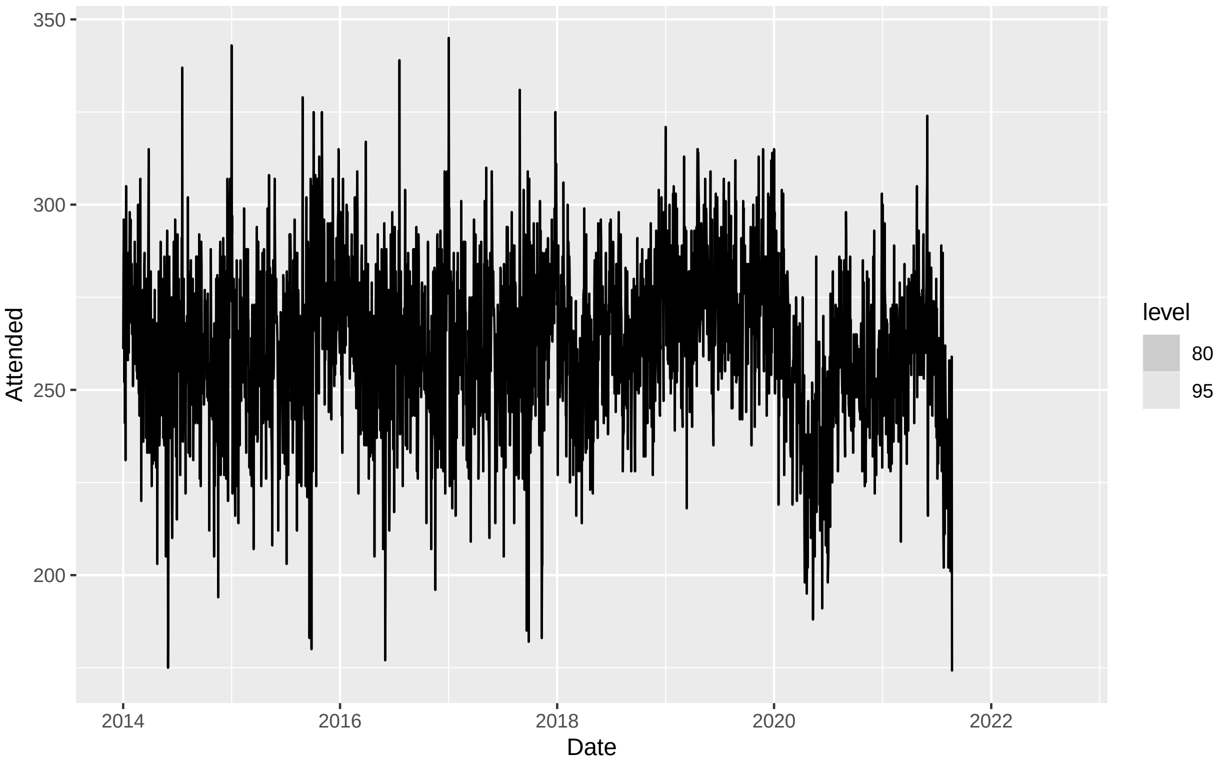





ETS(A,N,A) decomposition
Attended = lag(level, 1) + lag(season, 7) + remainder
350
300
250
200
280
260
240
10
0
−10
50
0
−50
Attended
level
season
remainder
2014
2016
2018
2020
2022
Date


350
300
250
200
150
Attended
2014
2016
2018
2020
2022
Date
level
80
95


350
300
250
200
150
Attended
2014
2016
2018
2020
2022
Date
level
80
95


350
300
250
200
Attended
2014
2016
2018
2020
2022
Date
level
80
95