ORCA – Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:https://orca.cardiff.ac.uk/id/eprint/129358/

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Rostami-Tabar, Bahman and Ziel, Florian 2022. Anticipating special events in emergency department forecasting. International Journal of Forecasting 38 (3) , pp. 1197-1213. 10.1016/j.ijforecast.2020.01.001

Publishers page: https://doi.org/10.1016/j.ijforecast.2020.01.001

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.

Anticipating special events in Emergency Department forecasting

Abstract

Accurate daily forecast of Emergency Department (ED) attendance helps roster planners in allocating available resources more effectively and potentially influences staffing. Since special events affect human behaviours, they may increase or decrease the demand for ED services. Therefore, it is crucial to model their impact and use them to forecast future attendance to improve roster planning and avoid reactive strategies. In this paper, we propose, for the first time, a forecasting model to generate both point and probabilistic daily forecast of ED attendance. We model the impact of special events on ED attendance by considering real-life ED data. We benchmark the accuracy of our model against three time series techniques and a regression model without considering special events as alternatives. We show that the proposed model outperforms its benchmarks across all horizons for both point and probabilistic forecasts. Results also show that our model is more robust with increasing forecasting horizon. Moreover, we provide evidence on how different types of special events may increase or decrease ED attendance. Our model can easily be adapted for use not only by EDs but also by other health services. It could also be generalised to include more types of special events.

Keywords: Forecasting, Emergency Department, Forecast accuracy, Special events, Health

1. Introduction

Health service increasingly regards patients as consumers of its service and aims to increase customer satisfaction. This is particularly the case in the Emergency Department (ED), which has been viewed as the shop window of the hospital service (Booth et al., 1992). ED allows patients to access nonstop urgent medical services. Poor service delivery can often determine the difference between life and death. Figures indicate the number of people being treated in ED units is constantly changing (Kamali et al., 2018) and often exceeds the capacity the department can suitably manage (Blunt, 2014). This measure of relative occupancy is an indication of the pressure in the department and is naturally associated with longer average waiting times that result in service deterioration (Izady and Worthington, 2012). Increase in the number of attendees to the ED and failure of patient flow through the department are caused by a multitude of factors. Some factors are internal to the ED and health service, such as lack of resources, whereas others are external factors, such as population growth, changes in life style and changing behaviours due to special events. Among the factors that may impact ED attendance, and consequently roster planning, are special events, such as public holidays, festive days or sports events, which play an important role (Cheang et al., 2003). Special events are deterministic events external to the ED that are known in advance. They can be used for modelling purposes to improve the forecast accuracy of ED attendance. Forecasting ED attendance and its workload represents a vital component of ED planning. The ability to forecast the number of attendances on any given day allows for more accurate staffing schedules and can reduce congestion in the ED, which then improves the quality of service for patients, reduces pressure on ED staff, decreases ambulance queues and reduces costs for the health system (Xu and Chan, 2016). Most published research in forecasting ED attendance uses time series techniques that are based purely on past historical observations of ED attendance (Marcilio et al., 2013). Few studies investigate the impact of public holidays on ED attendance (Tai et al., 2007; Sun et al., 2009; Faryar, 2013; Rotstein et al., 1997; Marcilio et al., 2013; Diehl et al., 1981; Alhusain et al., 2017). In these studies, the effect of types of holidays has not been modelled. Moreover, research in this area fails reproducibility principles and lacks strong methodological rigour (Boylan et al., 2015; Boylan, 2016). No analysis has been conducted to accurately generate both point and probabilistic long-term daily attendance in EDs by taking account of the types of special events in addition to existing effects in the historical data.

This paper aims to bridge this gap by first analysing the impact of special events on ED attendance and then proposing a model to forecast ED attendance in the presence of the types of events. In addition to point forecasts, we provide probabilistic forecasts to highlight uncertainties around future ED attendances. This is crucial for roster planners in the ED, helping them to balance resource allocations and consider the risk of over-capacity and under-capacity. The study uses a daily data set of patient attendance for 6 years and 3 months from an emergency department in a major, public sector, acute care, regional general hospital in the UK. Our model can be easily adapted and could, therefore, be useful for other EDs. It takes historical time series of ED attendance and time series of types of events as input and generates point and probabilistic forecasts of daily ED attendance for the required horizons.

In this paper, our objectives are fourfold: (i) we examine how different types of events affect ED attendance and provide evidence of their usefulness in forecasting daily ED attendance; (ii) we propose a model to accurately forecast daily ED attendance with consideration of different types of events, weekday effects, Auto-regressive effects, long-term trends and date effects; (iii) we provide probabilistic forecasts that quantify uncertainties in future ED attendance; (iv) we benchmark the accuracy of our model against three time series techniques including 1) Naive, 2) AutoRegressive, AR(p) (Box et al., 2015), 3) exponential smoothing state space model (ETS) (Durbin and Koopman, 2012) and a regression model without considering special events as alternatives.

The remainder of the paper is organised as follows: Section 2 describes different categories of events considered in this study. Section 3 provides an overview of the use and development of forecasting in ED and section 4 describes the proposed model and alternative benchmarks. Section 5 describes the data set used and the setup of our evaluation, while section 6 presents the results, followed by concluding remarks in section 7.

2. Special events in forecasting ED attendance

Special events can refer to specific days of the year when people might be away from their professional obligations. This may include public holidays, celebratory dates, school/university holidays, festivals and sports events. The existence of these types of events tends to change peoples behaviours. Therefore, ED attendance may also be influenced by this change, which consequently has an impact on the ED workload. Events shift the place of activities from business areas to entertainment centres; i.e. while some people may stay home, others may prefer outdoor activities, travel, parties and drink. These activities may increase or decrease the number of people requiring ED services, depending on the types of events. The occurrence of events is usually known in advance. It is the deterministic nature of an event that renders it an interesting variable to be considered for forecasting. This can be modelled appropriately from a forecasting perspective. However, events usually occur once, or at least once, a year, so it can be seen as a rare event and rather difficult to establish a model which would be able to credit for such a day. Moreover, ED attendance may have weekly seasonal patterns. Events may violate weekly patterns and shift attendance to another day.

2.1. Event classification

In this study, we consider three different types of events based on their characteristics: (i) fixed-date events; (ii) flexible-date events; (iii) long-date events.

Table 1: Event classification

Fixed-date	Flexible-date	Long-date
New Year Day	Pancake Day	Winter School Holiday
Valentines Day	Mothering Sunday	Spring Half-Term School Holiday
St. Davids Day	Good Friday	Spring School Holiday
St. Patricks Day	Holy Saturday	Summer Half-Term School Holiday
Halloween Day	Easter Sunday	Summer School Holiday
Guy Fawkes Night	Easter Monday	Autumn Half-Term School Holiday
Christmas Day	Early May Bank Holiday	
Boxing Day	Spring Bank Holiday	
Royal Wedding Bank Holiday	Fathers Day	
	August Bank Holiday	
	Remembrance Sunday	
	Black Friday	

The first group contains fixed-date events. These always occur on a fixed date each year. Typical examples are New Years Day (1 January), Christmas Day (25 December) and Boxing Day (26 December). Many national holidays linked to historically important dates in specific countries also fall into this category. The second group contains flexible-date events. These events always occur on the same weekday; however, the occurrence date varies from year to year. Typical examples are Good Friday and Easter Monday. Flexible-date events may also differ depending on geographical location. Summer bank holiday, for example, usually differs between England-Wales and Scotland. The third category of events are long- date events that contain both school and university holidays. Each school holiday term usually lasts for one week but no more than two weeks, with the exception of the Summer School Holiday, which lasts for more than a month. Universities generally have three holiday periods: (i) Winter Recess; (ii) Easter Recess; and (iii) Summer Recess. University holidays may last up to one/two and a half months. School and University holidays are important considerations for modelling ED attendance, especially in student cities. Table 2.1 provides a summary of holidays considered in this study.

3. Research background: ED forecasting

Multiple studies have analysed historical time series demand in emergency departments. Some studies are descriptive in nature; others use time series analysis or regression to forecast ED attendance. Rotstein et al. (1997), Sun et al. (2009) and Kam et al. (2010) analyse whether calendar seasons have an impact on ED visits. Sun et al. (2009) show that

Monday has the highest attendance in the week; however, Kam et al. (2010) conclude that it is Sunday. They also observe different behaviours in winter and summer. Some studies demonstrate that public holidays increase ED patient volume (Tai et al., 2007; Zheng et al., 2007; Sun et al., 2009; DaGar et al., 2014). Others, however, reveal a decrease in ED attendance during public holidays (Faryar, 2013; Rotstein et al., 1997; Marcilio et al., 2013; Diehl et al., 1981; Alhusain et al., 2017). Baer et al. (2011) also indicate that local weather conditions affect ED attendance. Diehl et al. (1981) have examined the effect of calendar and meteorological factors on ED attendance. They used a stepwise regression model to forecast future attendance. They found that patient flow peaked on Monday and declined steadily during the remainder of the week. They also discovered that fewer visits occurred during autumn and winter than during summer months. Moreover, their study indicates that higher temperatures are associated with more visits and daytime rainfall with fewer.

Forecasting ED attendance has been widely studied in the literature. Wargon (2009) provides a literature review of forecasting patient visits to ED. They state that the most frequently used forecasting approaches are either linear regression models, including calendar variables, or time series models. Champion et al. (2007) used Seasonal Exponential Smoothing (SES) and Autoregressive Integrated Moving Average (ARIMA) of order (0,1,1) to forecast the monthly number of patients at the ED of a hospital in regional Victoria. They used monthly time series from 2000 to 2005 to evaluate forecast accuracy through Root Mean Square Error (RMSE). Results show that SES outperforms ARIMA (0,1,1). Reis and Mandl (2003) forecast both the overall visits and the respiratory-related visits using ARIMA. They found that the mean absolute percentage error (MAPE) of the ARIMA model is 9.37% for overall visits and 27.54% for respiratory visits. Using an hourly data set of ED bed occupancy from July 2005 to June 2006 for three hospitals, Schweigler et al. (2009) produced short-term forecasts of ED bed occupancy. They used three methods for each site: 1) hourly historical average, 2) seasonal autoregressive integrated moving average (ARIMA), and 3) sinusoidal with an autoregression (AR)-structured error term. They evaluated the forecast accuracy of four and twelve hours forecast in advance using root mean squared error (RMSE). They demonstrated that both the sinusoidal model with AR-structured error term and a seasonal ARIMA model robustly predict ED bed occupancy. Kam et al. (2010) used moving average (MA), univariate and multivariate seasonal auto-regressive integrated moving average (SARIMA) models to forecast daily visits at a regional ED. They use weather and calendar information as explanatory variables in building the model. The result suggests that SARIMA model provides more accurate results than MA. They suggest incorporating weather information (temperature and rain) to predict daily volumes. Marcilio et al. (2013) compared the forecast accuracy of the Generalised Linear Model (GLM), Generalised Estimating Equations (GEE), and Seasonal Autoregressive Integrated Moving Average (SARIMA) methods using total daily patient visits to an ED in Sao Paulo, Brazil, from January 1, 2008, to December 31, 2010. They used the final three months of data to measure the forecast accuracy of each model using MAPE. They concluded that GLM and GEE models provide more accurate forecasts than SARIMA model. They also state that forecast performance is better in the short-term horizon (seven days in advance) than the mid-term (30 days in advance). Calegari et al. (2016) have examined the accuracy of the simple seasonal exponential smoothing (SES), seasonal multiplicative Holt-Winters (SMHW), SARIMA, and Multivariate Autoregressive Integrated Moving Average (MSARIMA) to forecast patient volumes in the ED. They considered horizons of one, seven, fourteen, twenty-one, and thirty days. They used a daily data set of twenty-seven months for the analysis. They calculated MAPE for each forecasting horizon using ninety days as test set. They concluded that SARIMA is the most accurate approach for all horizons. Holleman et al. (1996) examined the effect of calendar and weather variables on daily unscheduled patient volume in a walk-in clinic and emergency department. They considered (i) calendar variables, such as season, week of month, day of week, holidays, and federal check delivery days; and (ii) weather variables, such as high temperature and snowfall, as external variables. A daily data set from November 1991 to August 1994 was used for the forecasting experiment. They constructed a linear regression model to predict daily patient volume. They claim that five calendar variables (season, day of week, week of month, relation to federal holiday, and federal check-delivery day) accurately predict daily patient volume based on in-sample evaluation measured by $R^2 = .86$, and p-value = .0001. However, they have not reported the forecast accuracy using genuine test set and reliance on R^2 could be misleading, as a good fit does not guarantee accurate forecasts. Batal et al. (2001)examined the impact of the calendar variables in staff scheduling using a daily patient volume from February 1998 to July 2000. They show that when the calendar variable is the only variable in the model, the model displays an accuracy of 73%. Accuracy measurements for the validation set are not reported in the study. Sun et al. (2009) used SARIMA model to forecast daily patient volume for each patient acuity level. The data set includes daily patient attendances at ED between July 2005 and March 2008. The final six months are used as test and the remainder for training the model. MAPE is used to choose the best-fit model. They fitted separate ARIMA models to the three categories of acuity and overall data. The authors conclude that the predictions had good accuracy. Moreover, they observe that the impact of weather is not significant. They conclude that the ARIMA model is effective for both short term (weekly) and long term (three months) forecast horizons. The authors do not compare SARIMA models with any benchmark method in the study.

The literature review reveals some limitations in forecasting ED attendance which will be summarised as follows: (i) the effect of types of events is not modelled and has not been used for forecasting purposes as studies are mainly descriptive and, if used for forecasting, do not differentiate the types of events; (ii) some studies in this area lack a rigorous experimental design, i.e. there is no benchmark method nor is forecast accuracy reported; (iii) most studies are not reproducible as it is almost impossible to reproduce forecasting models and results; (iv) studies are limited in terms of the length of historical data used for training purposes and forecast performance evaluation; and (v) no study considers probabilistic forecasts that yield more information about the uncertainty of future forecasts, providing the entire picture of future ED attendance. Considering these limitations from the literature, we propose a forecasting model to fill these gaps.

4. Proposed model

We consider a high-dimensional time series model where the components mentioned might be zero after applying a certain shrinkage estimator to fit the model. There are several components that allow possible interpretations, including: a) weekly effects; b) date based effects (= annual effects); c) event effects; d) long-term trend effects; e) autoregressive effects; and f) weekly autoregressive effects.

The weekly effects a) cover deterministic weekly variations, especially potential weekend effects. Date based effects b) cover annual effects that always occur on a certain date or over a period of dates. Thus, these effects repeat every year, although they are not strictly periodic, as a calendar year has either 365 or 366 days. These effects include automatically fixed-date event effects. Other event effects, such as flexible-date and long-date are covered by c). The long-term trend component d) handles long term structural changes, e.g. induced from a different life style or a change in the population of the catchment area. The autoregressive components e) and f) cover autoregressive effects. Whereas the first covers standard linear relationships to the past of the time series, the latter covers periodically autoregressive effects. These can be understood as periodically time-varying autoregressive effects with a period of seven, which corresponds to a week. Thus, the autoregressive effect on a Sunday might be different to this one on a Monday.

We denote Y_t as the number of patients in attendance on day t. To define the model, we require some dummies, which are presented as follows:

- DoW_t^k for the k-th day-of-the-week dummy. i.e. DoW_t^1 is 1 if t is on a Monday and 0 otherwise, up to DoW_t^7 is 1 if t is on a Sunday and 0 otherwise.

- DoY_t^k for the k-th day-of-the-year dummy. i.e. DoY_t^k is 1 if t the k-th day of the year if it has 365 days and 0 otherwise. If a t is in a leap year (which has 366 days) the 60th day (the 29th February) is ignored from the day counting. It takes the same value itself as the 28th of February and both days will have a joint dummy variable.

- E_t^k is 0 if t is at the k-th event or in the k-th event period. We consider 12 single events, which are all flexible-date events, and 6 long event periods, which are school and University holidays (see Table 2.1). Additionally, we take special care of the impact of Spring School Holidays. This holiday period is usually around Easter - Easter is after the school holidays, before the school holidays, or in the middle of the school holidays. Hence, we separate it into three sub-periods: a dummy for the period before Good Friday, a dummy for after Easter Monday, and a dummy for the four day period from Good Friday to Easter Monday.

- BS_t^k the k-th B-spline basis of degree 12 with 21 knots and the full support (including the in-sample data and the forecasting horizon), see e.g. Ziel and Liu (2016); Ziel (2019). The choice of the degree and the number of knots seems somehow arbitrary. In fact, the choice with a degree of, for example 13, would lead to very similar results. However, the important aspect of the long-term trend basis functions is that their support covers more than a year to avoid interactions with annual effects or short-term effects.

For the 3 dummy/basis functions DoW_t^k, DoY_t^k and BS_t^k, we also define the cumulative version across the basis functions, so $\mathrm{DoW}_t^{\mathrm{cum},k} = \sum_{j\leq k} \mathrm{DoW}_t^j$, $\mathrm{DoY}_t^{\mathrm{cum},k} = \sum_{j\leq k} \mathrm{DoY}_t^j$ and $\mathrm{BS}_t^{\mathrm{cum},k} = \sum_{j\leq k} \mathrm{BS}_t^j$.

The sum across DoW and DoY add up to 1 that induces collinearity in the model. However, the estimation procedure automatically take care of that by setting the latter parameters to 0 (strictly they are ignored in the optimization).

The idea behind the cumulative basis function is to model the aforementioned effects in a different way, due to another parametrisation. For example, DoY_t^k models the day-of-the-year effect in an absolute manner. Thus, if t is the k-th day of the year, then there is an absolute effect of this day, such as a shock that is active for this particular day. This is highly relevant for special events. In contrast, $\text{DoY}_t^{\text{cum},k}$ models the change and not the absolute effect of the specific day of the year. Therefore, if this dummy is active, then the effect persists for the remaining basis function period; thus,a year for $\text{DoY}_t^{\text{cum},k}$. This is particularly important to model special events effects that last multiple days, like summer or winter holidays which, potentially, last multiple weeks.

Following the description of the elements of the model, the proposed model is given by

$$
\begin{aligned}
Y_t = \beta_0 &+ \underbrace{\sum_{k=1}^{7} \beta_k \text{DoW}_t^k + \beta_{7+k} \text{DoW}_t^{\text{cum},k}}_{\text{Weekday effects}} + \underbrace{\sum_{k=1}^{365} \beta_{14+k} \text{DoY}_t^k + \beta_{379+k} \text{DoY}_t^{\text{cum},k}}_{\text{Fixed-date effects}} \\
&+ \underbrace{\sum_{k=1}^{22} \beta_{744+k} \text{E}_t^k}_{\text{Flexible-date and long-date effects}} + \underbrace{\sum_{k=1}^{22} \beta_{766+k} \text{BS}_t^k + \beta_{788+k} \text{BS}_t^{\text{cum},k}}_{\text{Long-term trend effects}} \\
&+ \underbrace{\sum_{k=1}^{28} \beta_{810+k} Y_{t-k}}_{\text{Autoregressive effects}} + \underbrace{\sum_{k=1}^{28} \sum_{j=1}^{7} \beta_{838+28(i-1)+k} \text{DoW}_t^j Y_{t-k}}_{\text{Weekly autoregressive effects}} + \varepsilon_t
\end{aligned} \tag{1}
$$

Due to autoregressive effects the model has a recursive structure. This specific structure will be taken into account for the forecasting procedure, explained in Section 5.3.

4.1. Parameter estimation

The model has, in total, 1035 parameters which potentially describe overall behaviour. Not all of these parameters, however, are active in the final model used for forecasting. A parameter is selected if it is significantly different from zero. This selection procedure is conducted by an estimation algorithm. We consider a shrinkage based least square approach, namely the elastic net, for the estimation (Hastie et al., 2015). To define the elastic net, we note that the model (1) can be represented as a standard linear model given by representation

$$Y_t = \boldsymbol{\beta}' \boldsymbol{X}_t + \varepsilon_t$$

with parameter vector $\boldsymbol{\beta}' = (\beta_0, \beta_1, \ldots, \beta_{1034})'$. Then, we consider the observable period $\boldsymbol{Y} = Y_1, \ldots, Y_T$ with the corresponding regression matrix $\boldsymbol{\mathcal{X}} = (\boldsymbol{X}_1', \ldots, \boldsymbol{X}_T')'$. Moreover, we scale both $\boldsymbol{Y}$ and $\boldsymbol{\mathcal{X}}$ such that each row has mean zero and standard deviation one, denoted by $\widetilde{\boldsymbol{Y}}$ and $\widetilde{\boldsymbol{\mathcal{X}}}$. The symbol $'$ is used for transpose.

$$\widehat{\widehat{\boldsymbol{\beta}}} = \arg\min_{\boldsymbol{\beta}} \|\widetilde{\boldsymbol{Y}} - \boldsymbol{\beta}' \widetilde{\boldsymbol{\mathcal{X}}}\|_2 + \lambda \left(\alpha \|\boldsymbol{\beta}\|_1 + \frac{1-\alpha}{2} \|\boldsymbol{\beta}\|_2 \right)$$

is the scaled elastic net estimator which depends on two tuning parameters $\lambda \geq 0$ and $\alpha \geq 0$. Note that we can obtain the unscaled lasso estimator $\widehat{\boldsymbol{\beta}}$ simply by rescaling $\widehat{\widehat{\boldsymbol{\beta}}}$.

We choose $\alpha = 0.5$ as it yields accurate forecasting results in a forecasting study (Uniejewski et al., 2016). Furthermore, it reduces the computational cost of finding an optimal α. As $\alpha = 1$ gives the popular lasso estimator, and $\alpha = 0$ the ridge estimator, the choice $\alpha = .5$ seems to be an intuitive compromise. Note that the elastic net for $\alpha < 1$ retains the sparsity property as the lasso does. Therefore, the more irrelevant a parameter is, the more it shrinks towards zero, including the special case of exactly zero. The shrinking effect depends on the tuning parameter λ. The larger λ the more parameters will be zero. For $\lambda = 0$, we receive the OLS solution as a special case where all parameters are included in the model.

We optimise the tuning parameter λ by minimising the Hannan-Quinn-Information criterion (HQC) on an exponential grid. The HQC led to overall robust results in a large forecasting study by Ziel and Weron (2018) across several data sets in the energy forecasting context. The HQC is a special choice of the generalised information criterion (GIC)

$$\text{GIC}(\kappa; \boldsymbol{\beta}) = T \log(\text{RSS}(\boldsymbol{\beta})/T) + \kappa K(\boldsymbol{\beta})$$

with RSS as residuals sums of squares, κ as penalty parameter and $K(\boldsymbol{\beta})$ as the number of non-zero parameters in the considered model. The penalty parameter κ has to be chosen by the modeller. The HQC follows from the choice $\kappa = 2\log(\log(T))$. It can be regarded as a compromise between the popular Akaike information criterion (AIC) followed by $\kappa = 2$ and the conservative Bayesian information criterion (BIC) with $\kappa = \log(T)$. We also examined the tuning of the λ parameter via block-wise cross-validation by minimising Mean Squared Error (MSE) but the results did not improve compared to the information criterion based approach although it is computationally more demanding.

5. Empirical evaluation

In this section, we provide an overview of the available time series data to examine the empirical validation of our model. We also discuss the experimental design employed in this study, including a description of benchmark techniques, probabilistic forecasts setup and forecast accuracy evaluation.

5.1. Data

Data used in the study comprised counts of patients arrival times at one of the largest ED units in the United Kingdom between January 2010 and March 2016, extracted from the ED administrative database of the hospital. We aggregated the patients arrival times to obtain daily attendance between 01 January 2010 and 31 March 2016, which is used for empirical evaluation in this study. A sample of three years of data is given in Figure 1.

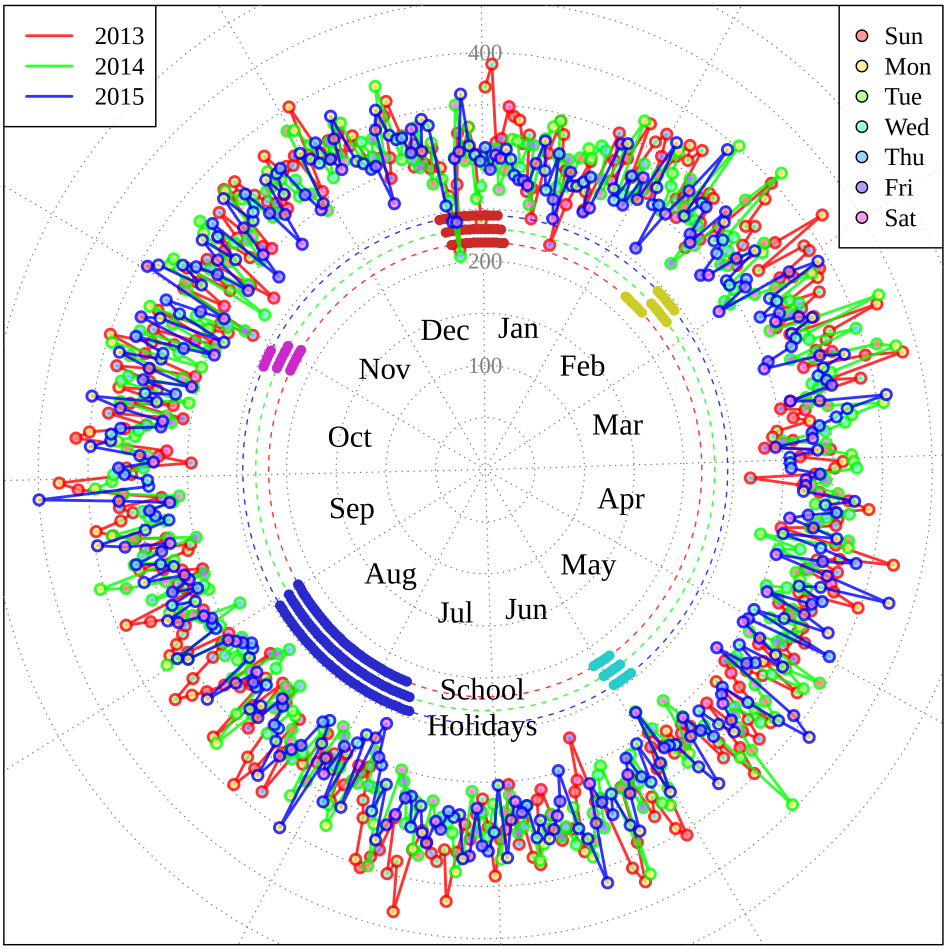

Figure 1: Daily ED attendance of three selected years (2013, 2014, 2015) in polar coordinates.

Figure 1 illustrates the data in polar coordinates such that a full circle corresponds to a year and the distance to the origin matches the number of patient arrivals. Although data is noisy, some systematic structures are visible. The day-of-week effect is highlighted by different fill colours. It is clear that Mondays (yellow fill) have a peak, followed by Sundays (pink fill). We have also highlighted long-date holidays, such as school holidays, for three years. We can observe the decreasing effect of long-date holidays. This will be confirmed later when we discuss the effect of events in 6. We split the data into training and test sets. We consider four years of in-sample data and forecast the next $H = 42$ which corresponds to six weeks and is relevant for operational planning in hospitals in the UK. The rest of data is used as the test set to evaluate forecast accuracy. The six years and three months of data were used in a rolling window forecasting study with re-estimation.

5.2. Benchmark models

We consider four benchmark models to compare the performance of the proposed model. They include a naive benchmark model, a more sophisticated, autoregressive, model, AR(p), an exponential smoothing state space model, and a model with similar design to that proposed to highlight the impact of special events. These models are described in this section.

1. Naive model:

   The naive model is a very simple model that corresponds to historic simulation. Thus, we have the assumption that there is no structure to ED arrivals and the observed past continues into the future. Formally, this model is given by $Y_t \sim F$ where F is the unknown distribution that we estimate by the empirical distribution function (ecdf) of the given sample.

2. AR(p):

   The second benchmark model is an AR(p) with p selected on the grid $1, \dots, p_{max} = 365$. Formally this is

   $$Y_t = \phi_0 + \sum_{k=1}^{p} \phi_k Y_{t-k} + \epsilon_t \tag{2}$$

   where ϕ_k denote the autoregressive parameters, ϕ_0 the intercept and $_t$ the error term. We select the optimal p by minimising the AIC (Akaike information criterion) as in Haben et al. (2019). To simulate from the process, we apply residual based bootstrap as in the proposed method above. Note that every ARIMA process (also SARIMA) can be rewritten as an AR(∞). As every AR process has an exponentially fast decaying memory, it can be approximated well by a high order AR process. Thus, many of the aforementioned ARIMA-type models can be regarded as nested into the considered model.

3. Exponential smoothing state space model:

   The exponential smoothing model is incorporated using corresponding implementation of the forecast package in R. We use the ets() function in the forecast package to generate daily forecasts (Hyndman and Khandakar, 2008). Note that the structure identification is done automatically, but the input data is assigned with a weakly and annual seasonality.

4. Proposed model without date and holiday effects:

   Here, we simply apply the same model and estimation method as for the proposed model (1), but remove the term for the date and holiday effects.

5.3. Probabilistic forecasting setup

In this study, we provide probabilistic forecasts in addition to point forecasts. Probabilistic forecasting quantifies uncertainties around generated forecasts that will allow planners in ED to better manage risks in operational planning. The six years and three months of data were used in a rolling window forecasting study with re-estimation. We considered four years of in-sample data and forecast the next $H = 42$ which corresponds to six weeks and is relevant for operational planning. Formally, we have $Y_{1+n}, \dots, Y_{T+n}$ (with $T = 4 \times 365 + 1 = 1461$) as in-sample data for the n's rolling window and want to forecast $Y_{T+n+1}, \dots, Y_{T+n+H}$. Thus, we are interested in the full H-dimensional probability distribution of the random vector $(Y_{T+n+1}, \dots, Y_{T+n+H})$. In total we have $N = 780$ rolling windows, hence $n = 0, ..., N-1$ covers an out-of-sample time range of more than two years.

To tackle this sophisticated forecasting problem, we applied ensemble forecasting methods. Hence, we report a large ensemble of $M = 5000$ simulated paths. These paths are simulated from the estimated model using residual based bootstrap recursively. In detail, we can rewrite our model (1) by $Y_t = f(\boldsymbol{\beta}, t, Y_{t-1}, Y_{t-2}, \dots) + \varepsilon_t$. Therefore, with the in-sample data $Y_{1+n}, \dots, Y_{T+n}$, we can perform a 1-step ahead ensemble forecast by $Y^{[m]}_{T+n+1} = f(\widehat{\boldsymbol{\beta}}, t, Y_{T-n}, Y_{T-n-1}, \dots) + \varepsilon^{[m]}_t$ where $m = 1, \dots, M$ and $\varepsilon^{[m]}_t$ is drawn uniformly from the residuals $\widehat{\varepsilon}_{1+n}, \dots, \widehat{\varepsilon}_{T+n}$. For m's simulation of the 2-step ahead forecast we replace the unknown Y_{T-n+1} by the m's simulated value, thus $Y^{[m]}_{T+n+2} = f(\widehat{\boldsymbol{\beta}}, t+1, Y^{[m]}_{T+n+1}, Y_{T-n}, \dots) + \varepsilon^{[m]}_{t+1}$ where again $\varepsilon^{[m]}_{t+1}$ is drawn uniformly from the residuals $\widehat{\varepsilon}_{1+n}, \dots, \widehat{\varepsilon}_{T+n}$. This procedure continues until the desired forecasting horizon $H = 42$ is reached. The large ensemble of size M is used as an approximation of the forecasting distribution of interest. Every characteristic (e.g. mean, median, variance) can be computed from based on sample statistics. We also use this property for the evaluation.

Figure 2 illustrates an example of probabilistic forecast generated by the proposed model. Here, instead of presenting the $M = 5000$ simulated trajectories, the resulting 99 percentile forecasts are presented.

5.4. Evaluation scheme and metrics

We evaluate the forecast accuracy of our approach and its benchmarks using out-of-sample data containing two years and three months of daily attendance. Models are applied in a rolling window forecasting study with re-estimation. Forecasting horizon is chosen to be $H = 42$-ranging from one day ahead to 42 days ahead- which corresponds to six weeks that are relevant for operational planning in the EDs of NHS hospitals in

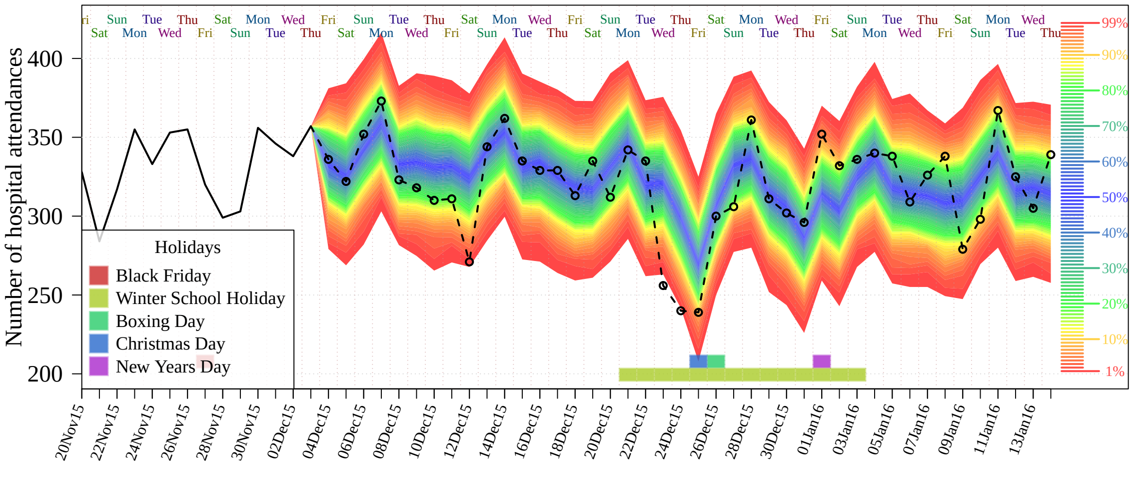

Figure 2: Example, probabilistic forecast

the UK. Daily forecast horizon may vary across EDs in the NHS between seven days (one week) and 42 days (six weeks), which are covered with our H=42 days forecasts. We have considered four different evaluation measures: two point forecasting metrics and two probabilistic metrics. The point forecasting metrics are the MAE (median absolute error) and RMSE (root mean square error) for point forecasts (Hyndman and Koehler, 2006). The first is strictly appropriate for the median, the latter for the mean. Hence, the MAE is minimised only for the optimal median forecast and the RMSE only for the optimal mean forecast. Thus, they can identify the best median and mean forecasting model. To obtain the median and mean from a given ensemble forecast, we evaluate the corresponding sample median and sample mean of the ensemble. As we perform an H-step ahead forecast, we evaluate the MAE and RMSE for each forecasting step separately. Note that we do not apply the MAPE as evaluation criterion, even though it is popular in ED forecasting, due to its disadvantageous statistical properties (Franses, 2016).

In addition to the standard point forecasting measures, we also apply two probabilistic forecasting measures: the pinball score (also known as quantile loss) such as the energy score. The pinball score is a strictly appropriate evaluation criterion for quantiles. We evaluate it on a dense probability grid for all percentiles $(1\%, \dots, 99\%)$ as popular in forecasting (Hong et al., 2016). As such a dense grid closely approximates the underlying H marginal distributions of $(Y_1, \dots, Y_H)$, it can be regarded as a probabilistic evaluation measure. If the distance of the considered probability grid (here 1%) is taken to be zero (as a limit) then the average pinball score across the grid converges to the CRPS (continuous ranked probability score), which is strictly appropriate for univariate distributions. Similarly, as regards the MAE and RMSE, we can evaluate the pinball score for each forecasting horizon, but also across the 99 percentiles.

Finally, we consider the energy score (Gneiting and Raftery, 2007). In contrast to the aggregated pinball score across the percentiles or the CRPS, this is a strictly appropriate evaluation criterion for the full H-dimensional multivariate distribution. It is, therefore, suitable for detecting misspecification in the marginal distribution structure of the forecast, such as dependency structure. Due to the high complexity of the energy score, it has not been very popular in forecasting so far, but in recent years an increasing number of applications have appeared, especially in weather and energy forecasting.

6. Results and discussion

In this section, we present the results of our study. In the subsection 6.1, we analyse the impact of variables considered in our model on the ED attendance; more specifically, we look at the effect of special events. In the subsection 6.2, we then present the forecast accuracy of our model and its benchmarks, evaluated using out-of-sample data.

6.1. Impact of components and interpretation

Prior to the evaluation of all effects, we consider the residuals of the fitted model to determine whether the proposed model captured all genuine patterns presented in the time series. In Figure 3, we observe the sample autocorrelation plot (acf) of the residuals and the squared residuals, such as the histogram of the residuals.

Both autocorrelation plots show no series correlation. Thus, the residuals seem to be uncorrelated and homoscedastic. We can see from the histogram that the residuals have a unimodal distribution without heavy tails. The density is close to a normal distribution with correspondingly matching variance. Nevertheless, note that for the estimation procedure, the normality assumption is not required for an OLS estimation method. However, if the residuals follow a normal distribution, it would underline the suitability of the estimation procedure as the normally distributed likelihood approach coincides with the OLS.

It is important to note that residuals are well behaved. Thus, interpretations of the model output are valid, even in-sample. Moreover, as the autocorrelation functions do not show series correlation, we can conclude that all autoregressive and deterministic information is fundamentally covered. No substantial improvements can be expected unless external regressors are considered.

Figure 4 shows the model fit (conditional mean) with the corresponding unconditional mean estimate. We see clearly the annual pattern in the data; especially, a strong winter

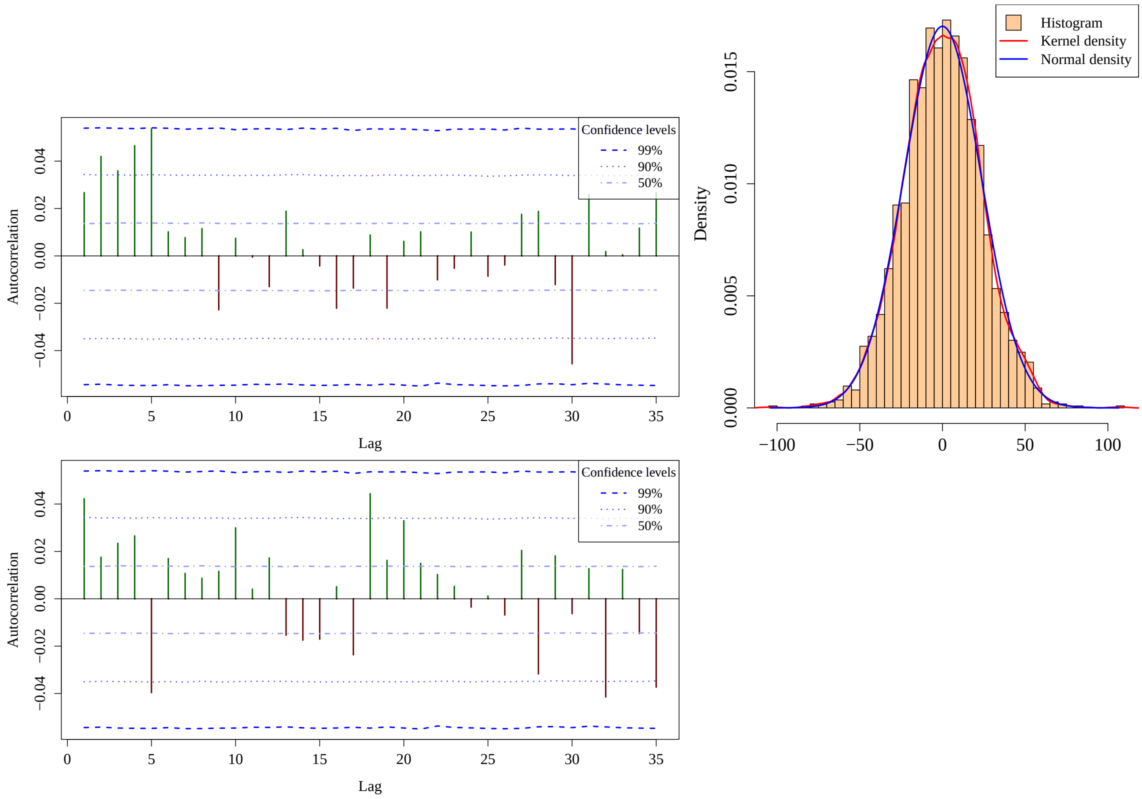

Figure 3: Sample autocorrelation function of $\widehat{\varepsilon}_t$ and $\widehat{\varepsilon}_t^2$, with histogram of $\widehat{\varepsilon}_t$

holiday in December/January dip and a smaller summer holiday period. Looking carefully at the graph, weekly variations are also observable. It therefore appears that weekly effects are also captured.

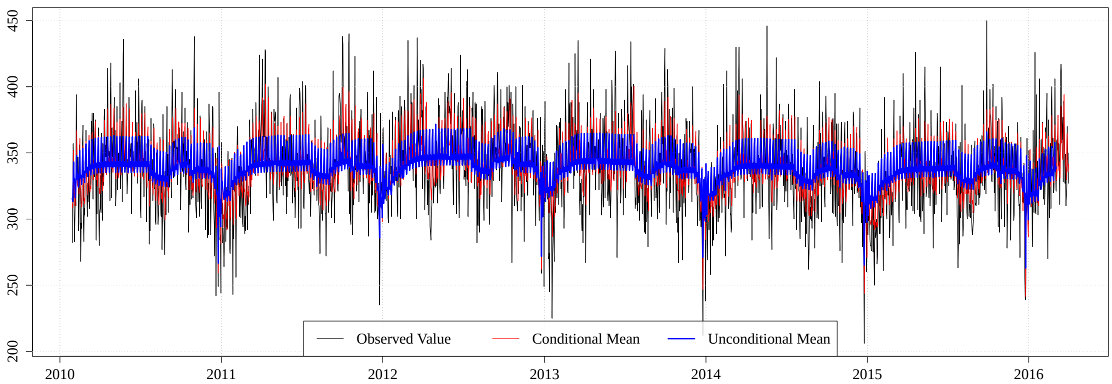

Figure 4: Observed values of Y_t with estimates of the mean $\mathbb{E}(Y_t)$ and conditional mean $\mathbb{E}(Y_t|Y_{t-1}, Y_{t-2}, \ldots)$.

In Figure 5, the fitted long-term trend is visualised. We observe that some long-trend components are estimated. However, in absolute terms these effects are very small, and barely visible in Figure 4.

Figure 6 presents the autoregressive effects, including the weekly periodically autore-

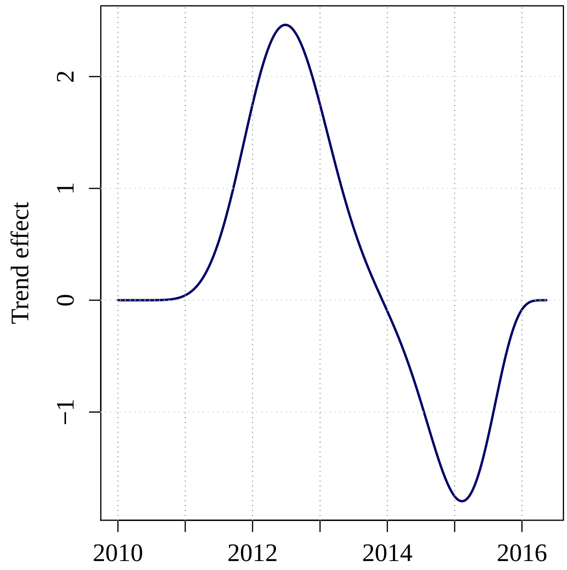

Figure 5: Long-term trend effect.

gressive effects. It is apparent that the major effects are driven by the non-periodic parameters. Here, the most recent value (lag=1) has the strongest impact. In the memory, we also observe the weekly structure, so lag multiples of 7 (14, 21, 28) show some impacts. Next to the standard autoregressive effects, we also observe some periodic effects. The short term memory, specifically, is adjusted for Mondays and Sundays. These effects are the only explanation for actual weekly seasonal structure in Figure 4. It can be seen, therefore, that even those small autoregressive parameters can have a substantial impact on the model performance. Here, the authors wish to add that, if the weekly periodically autoregressive parameters are removed from the model, then next to a small reduction in the model performance, the weekday dummies show positive effects.

Finally, we discuss the effect of special events. We begin by looking at the impact of fixed-date events.

Figure 7 illustrates the effects of the date based dummies over a year with high- lighted fixed-date events. We observe that there is a reduction in baseline intensity during the colder periods of the year (end of October to end of February) when fewer patients visit the Emergency Department. Around the Christmas/New Year period (mid December to mid January), there appears to be an additional reduction in the effect. Moreover, we see that not all fixed-date events have an effect on hospital attendance. Nevertheless, there are some interesting patterns.

We observe strong impacts of Christmas, which is not surprising. We have a clear We observe strong impacts from Christmas, which is not surprising. We have a clear reduction on Christmas Eve (24 Dec) and Christmas Day (25 Dec) but in contrast a strong positive

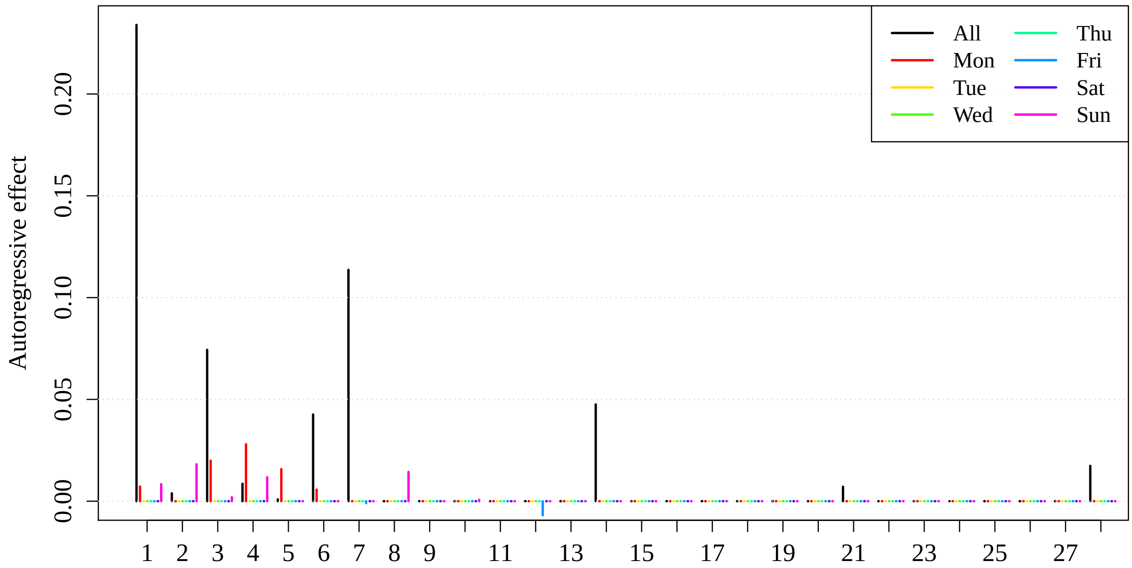

Figure 6: Estimated autoregressive parameters, including periodically autoregressive effects

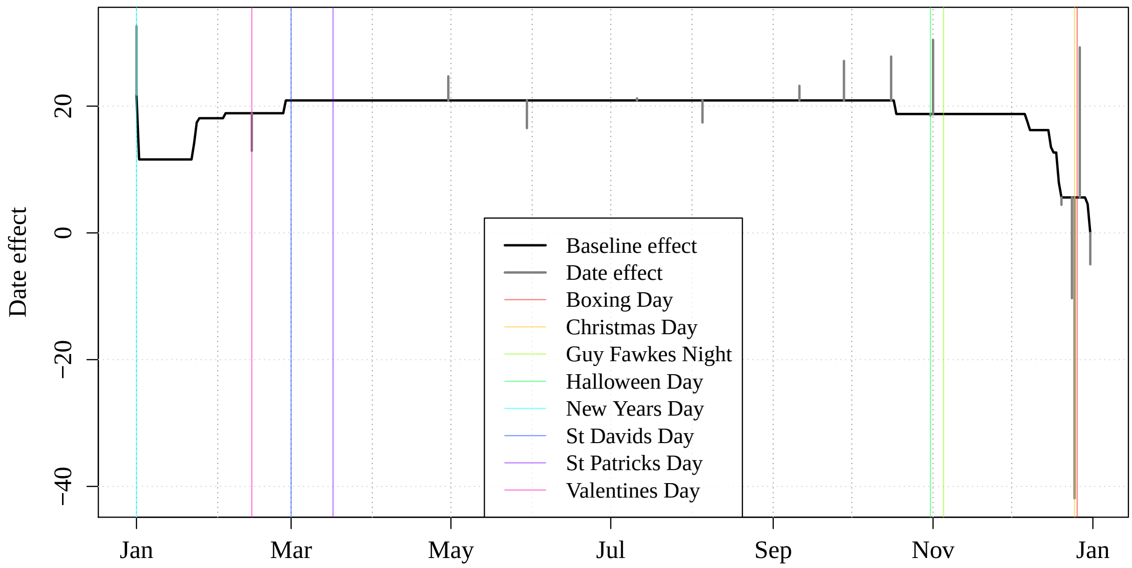

Figure 7: Estimated date based effects with highlighted fix-date events.

peak on 27 Dec., the day after boxing day (26 Dec). This peak, following the strong reduction around Christmas, might be attributed to ailing people attending hospital after the Christmas feast. Furthermore, we see a reducing effect on New Years Eve (31 Dec) but a strong increase in the attendance on New Years Day (1st Dec). A similar effect can be observed at Halloween (31 Oct). There is a small reduction in attendance on Halloween followed by a large increase the next day. Possibly, some people celebrated too much overnight, resulting in a hospital visit the following day. However, there are other examples of events effects. For instance, at Valentines (14 Feb), there is a significant reduction in hospital visits. It is possible that directly love-related events lead to less violence in everyday life and thus fewer hospital visits. Moreover, some other fixed-date effects are clearly visible. We observe that the following fixed-date events have an increasing effect on ED attendance: 13 March, 30 April, 11 July, 11 September, 28 September and 16 October; while the following dates are shown to have a decreasing effect: 30 May, 5 August, 20 August, 25 October, 20 December.

Figure 8 illustrates the effect of flexible-date and long-date events. Most of the events in these categories have no significant impact. However, if there is an effect, it is a reduction effect. We see more impact from holiday periods, namely Autumn Half-Term Holiday, and Summer Holiday. In Figure 4, it is apparent that the Summer Holiday dummy is the main reason for the visible summer reduction effect.

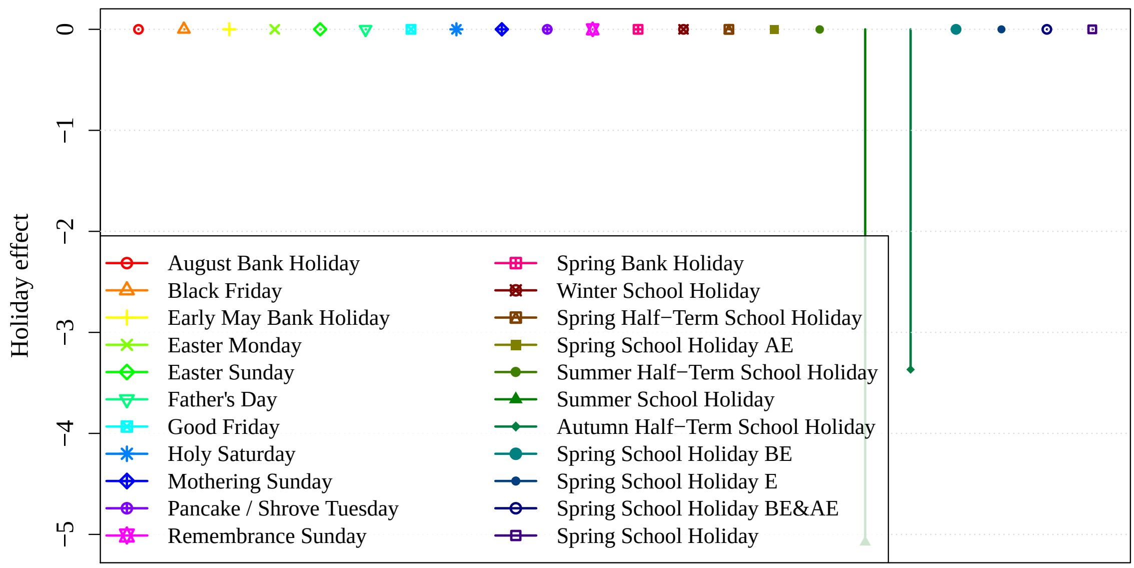

Figure 8: Estimated flexible-date and long-date effects.

It appears that the effect of School holidays and flexible-date holidays varies from year to year as the holidays shift. This effect overlaps with the general effect that weekdays shift from year to year. This means that if 1 January is on a Monday in a certain year, it will be on Tuesday or Wednesday the next year, if the recent year is a leap year. The overall behaviour of the model is illustrated in Figure 9 based on the structure captured in the model where noise has been filtered out. Here, the estimated, expected value of the arrivals is visualised in polar coordinates with annual cycles. The distance from the centre matches the corresponding number of arrivals. The weekday shift, mentioned above, is clearly visible. Additionally, the strong Christmas effect is apparent, where expected arrivals drop clearly below 300, whereas for the remainder of the year arrivals are clearly above 300. We can also observe that the effect of long-date events is captured.

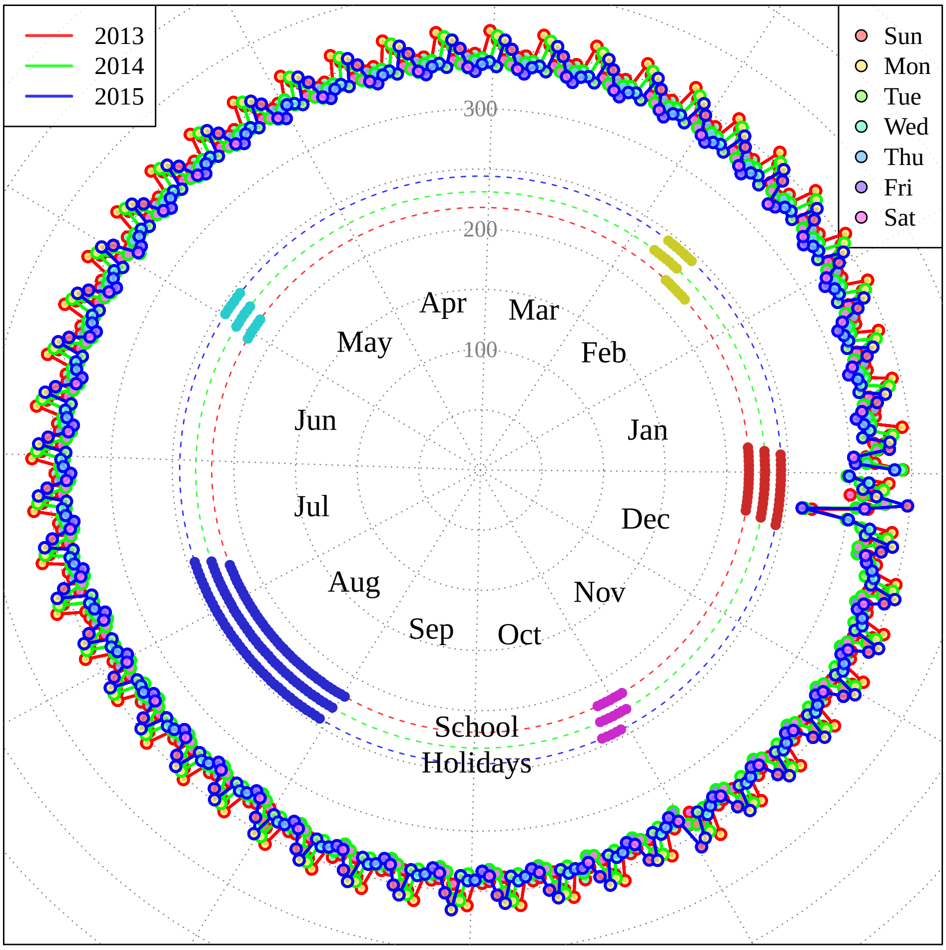

Figure 9: Estimated expected value of the daily hospital arrivals across three selected years in polar coordinates.

6.2. Forecast accuracy evaluation

In this section, we present the forecast accuracy of our model and its benchmarks for both point and probabilistic forecasts using Out-of-sample data. Next to graphs discussed in this section, detailed tables of forecasting measures are presented in the Appendix. First, we focus on the point forecasting results. Figure 10 shows the MAE and RMSE values across the forecasting horizon for all five models considered. We observe that, overall, forecasting behaviour regarding the MAE and RMSE are relatively similar. This is intuitive because the structure of errors looks relatively symmetric (see Fig. 4) whwhich yields similar median and mean forecasts. Overall, we observe that the proposed model consistently outperforms all benchmarks for all forecasting horizons. For longer forecasting horizons, the proposed model retains high forecasting accuracy compared to other methods. Furthermore, we see that the impact of special events effects in the proposed model is especially important for longer forecasting horizons, whereas the proposed model without special events effects consistently reduces its forecasting performance. Clearly,

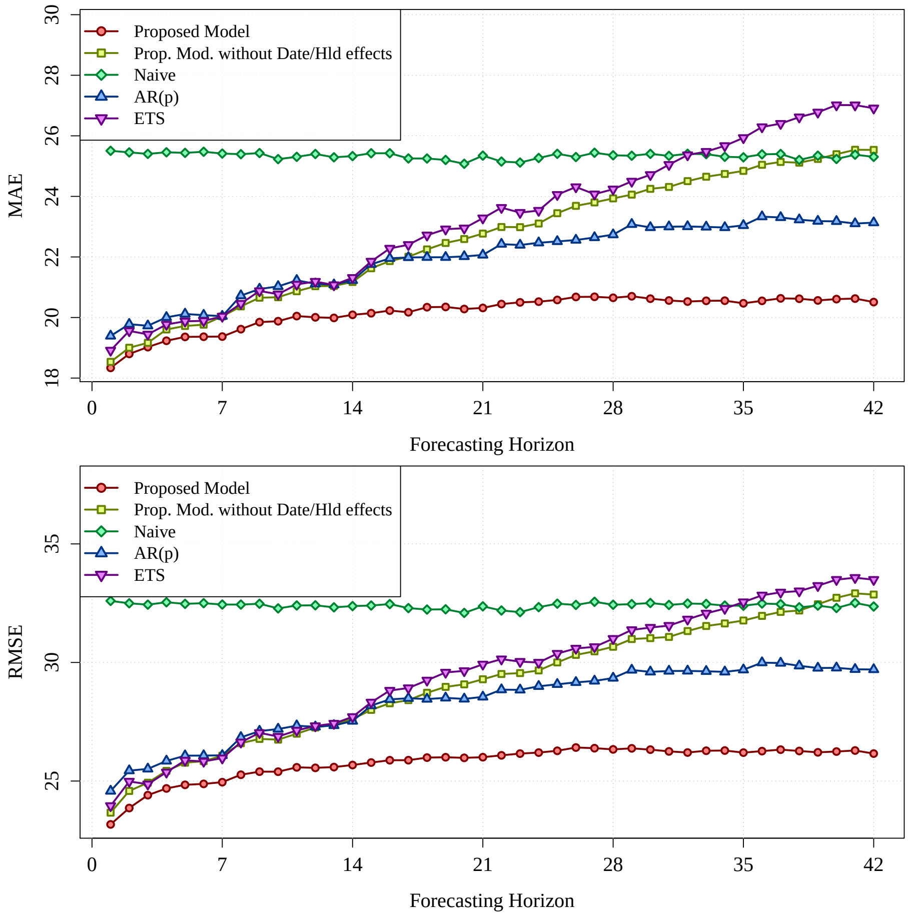

Figure 10: MAE and RMSE of considered models across the forecasting horizon.

the proposed model not only remains accurate across the forecasting horizon, but is also more robust than other methods against increasing horizons.

In evaluating the probabilistic forecast accuracy of our model and its benchmark, we first focus on the marginal distributional characteristics evaluated by the corresponding pinball score. Figure 11(top) shows the corresponding pinball score aggregated across the quantile grid for each forecasting horizon. Figure 11(bottom) (bottom) presents pinball score aggregated across forecasting horizons for each quantile grid.

The results show that the proposed model captures the density structure for all horizons consistently better than the benchmark models. When studying the results regarding the accuracy for single quantiles, we observe that the major area of improvement lies in the centre quantiles. For the extremes (e.g. 1% and 99% quantile), the proposed model presents no clear improvement in forecasting accuracy compared to other models. This might be improved by considering a proper modelling of the (conditional) variance of the error process, see e.g. Ziel (2016).

Finally, we study the results regarding fully multivariate forecast distribution evaluated

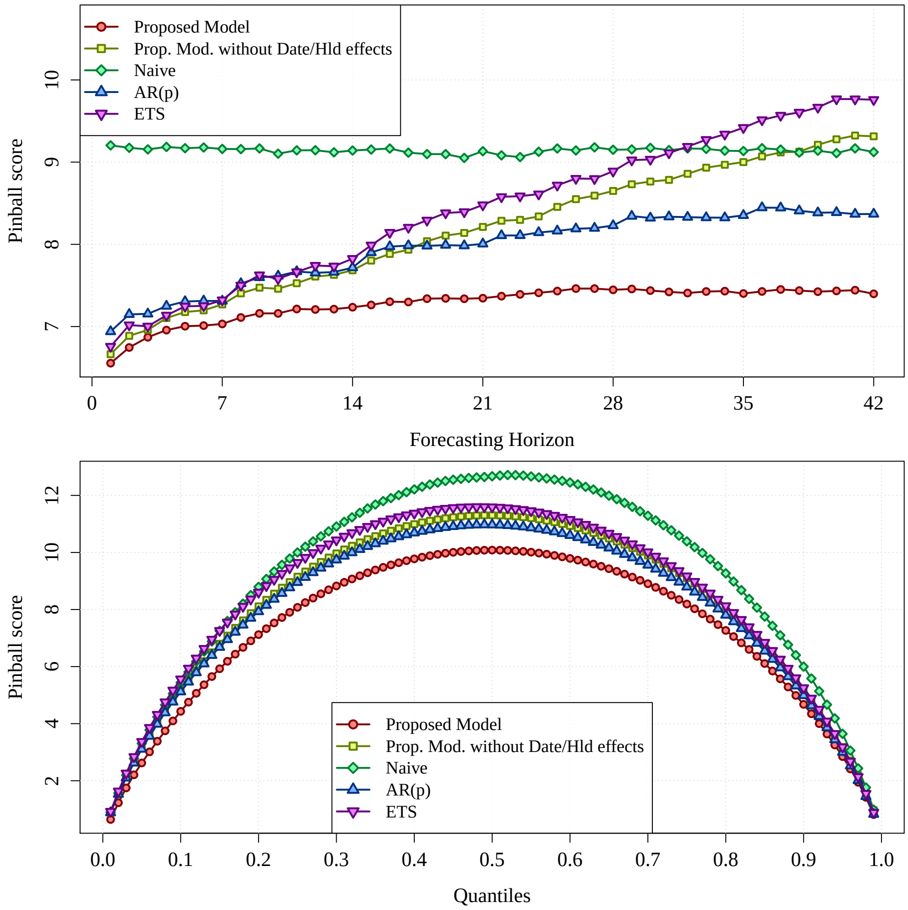

Figure 11: Pinball score across the forecasting horizon (top) and across the quantile levels (bottom) of all considered models.

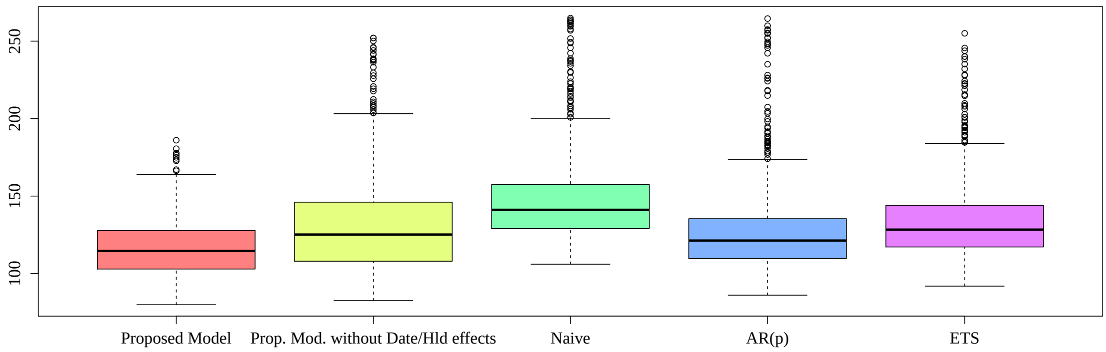

Figure 12: Box plot of energy score for all considered models.

by the energy score given in Figure 12.It is apparent that we have only one energy score for the full horizon. Therefore, the box plot of all individual scores across the forecasting horizon is chosen for illustration. We see that, overall, the proposed model has the best energy score. It appears that the main improvement in accuracy compared to the proposed model without holiday and date effects refers to the reduction in the upper tail. For example, there is no single day with an energy score above 200, whereas there are several such days for the remaining models.

Our results show that the proposed model not only outperforms benchmarks in terms of generating the correct value of future ED attendance, but is also the most accurate when providing consistent information about uncertainty around ED attendance. The model provides the entire probability distribution for ED attendance for a given day. Our model offers the ability to capture a range of possibilities regarding ED attendance for any given day that is not contained in the point forecasts. This is very important for decision makers and planners in EDs, as it helps them to assess the risk and make better decisions in operational planning. Further research attempts are required on how to use probabilistic forecasts and interpret output to inform decisions in the Emergency Department.

7. Concluding remarks

Accurate forecasts of ED attendance can help roster planners to manage the risk of over and under capacity resource allocation better. This is crucial for health service management as both cases have a significant impact on all health service stakeholders, such as patients, ED itself, ambulance services and social care services. Over and under capacity resource allocation may increase cost, increase staff sickness, waste resources, put pressure on other health services and consequently deteriorate the efficiency of the entire health service. This paper proposes a forecasting model to estimate daily attendance in EDs. The model is high dimensional and includes many potential predictor variables which are classified into: i) auto-regressive effect; ii) weekday effect; iii) long term trend effect v) special events effects.

Not all potential predictor variables are present in the forecasting model. We use an estimation procedure that decides which parameter is significantly different from zero that is kept in the model and will be used to analyse the impact of selected predictors on ED attendance and to generate the daily forecast of ED attendance. In generating both point and probabilistic forecasts, we compare the effectiveness of our model in terms of forecasting performance against four time series method and a regression model without accounting for the effect of special events. The benchmarks are: 1) Naive; 2) Auto-regressive order p; 3) exponential smoothing state space (ETS) model; and 4) a regression model without special events effects. In summary, we believe that the problem setting we have considered is a very realistic one. We have used daily attendance data from a major ED in the UK for a period of over six years. Four different accuracy measures have been used to evaluate the forecast accuracy performance of all models in generating a daily forecast, 42 days in advance. The main findings of this study can be summarised as follows:

- We propose a forecasting model that outperforms its benchmarks in generating both points and probabilistic forecasts for daily ED attendance using an empirical study analysing real-life ED data. Moreover, the model handles the impact of special events on ED attendance. To the best of our knowledge, probabilistic forecasts have never been used to forecast ED attendance. They provide more information about uncertainty of future ED attendance, which can be used for risk management in allocating resources for staffing. More investigation is required to link probabilistic forecasts to operational planning in the Emergency Department.
- We provide evidence that incorporating special events into the forecasting model improves forecast accuracy. We not only show the increase or decrease effects of special events, but also provide the estimated impact. We strongly recommend decision makers and planners to take account of the impact of special events in generating forecasts for EDs. Managers should also trust the size effect of special events estimated by the model.
  - We show that the following date based events have a decreasing effect on ED attendance: Valentine's Day, Halloween, Christmas Eve, Christmas Day, New Year's Eve, Autumn school holiday, Summer school holiday, 30$^{\text{th}}$ May, 5$^{\text{th}}$ August, 20$^{\text{th}}$ August, 25$^{\text{th}}$ October and 20$^{\text{th}}$ December.
  - We observe that the following events have an increasing effect on ED attendance: New Year's Day; day after Halloween; day after Boxing Day; fresher's week (28$^{\text{th}}$ September); 13$^{\text{th}}$ March; 30$^{\text{th}}$ April; 11$^{\text{th}}$ July; 11$^{\text{th}}$ September; and 16$^{\text{th}}$ October.
- One of our primary objectives during the development phase was to create a model that could be generalized. Our model can easily be adapted to different settings not only in EDs but also in other health services such as ambulance service or hospitals trying to generate accurate forecasts by incorporating special events. Our model could also be generalized to include more special events if necessary.

The proposed model takes i) the historical daily ED attendance and ii) the time series of the type of events as inputs and generates point and probabilistic forecasts as outputs.

We provide the R codes for the proposed model and benchmarks used in this paper. They will also be available in an open source R package.

It is important to note that translating forecasts to utility measures is not a straightforward task. The link between forecast accuracy and staffing in the emergency department requires further investigation to quantify its impact on performance indicators, such as calling for external resources, operational cost, staff sickness, patient waiting time and ambulance queue. Moreover, Emergency Department managers are not familiar with the benefits of probabilistic forecasts. It is crucial to investigate how probabilistic forecasts might be used to inform decisions in the emergency department and how they should be linked to operational planning. We will investigate these two important avenues in our future works. works.

References

Alhusain, F.A.M., Alhassan, N.F.A., Aljohi, W.A.A., Alrumaih, F.I., Al Jerian, N.A., Alharthy, N.A., et al., 2017. Paediatric emergency department during the holidays: Findings from a 10-year analysis of visit rates and trauma patterns. Journal of Health Specialties 5, 142.

Baer, A., Elbert, Y., Burkom, H.S., Holtry, R., Lombardo, J.S., Duchin, J.S., 2011. Usefulness of syndromic data sources for investigating morbidity resulting from a severe weather event. Disaster medicine and public health preparedness 5, 37–45.

Batal, H., Tench, J., McMillan, S., Adams, J., Mehler, P.S., 2001. Predicting patient visits to an urgent care clinic using calendar variables. Academic Emergency Medicine 8, 48–53.

Blunt, I., 2014. Focus on: Ae attendances. Quality Watch .

Booth, A., Harrison, C., Gardener, G., Gray, A., 1992. Waiting times and patient satisfaction in the accident and emergency department. Emergency Medicine Journal 9, 162–168.

Box, G.E., Jenkins, G.M., Reinsel, G.C., Ljung, G.M., 2015. Time series analysis: forecasting and control. John Wiley & Sons.

Boylan, J.E., 2016. Reproducibility. IMA Journal of Management Mathematics 27, 107–108.

Boylan, J.E., Goodwin, P., Mohammadipour, M., Syntetos, A.A., 2015. Reproducibility in forecasting research. International Journal of Forecasting 31, 79–90.

Calegari, R., Fogliatto, F.S., Lucini, F.R., Neyeloff, J., Kuchenbecker, R.S., Schaan, B.D., 2016. Forecasting daily volume and acuity of patients in the emergency department. Computational and Mathematical Methods in Medicine 2016.

Champion, R., Kinsman, L.D., Lee, G.A., Masman, K.A., May, E.A., Mills, T.M., Taylor, M.D., Thomas, P.R., Williams, R.J., 2007. Forecasting emergency department presentations. Australian Health Review 31, 83–90.

Cheang, B., Li, H., Lim, A., Rodrigues, B., 2003. Nurse rostering problems: a bibliographic survey. European Journal of Operational Research 151, 447–460.

DaGar, S., SAHIN, S., YILMAZ, Y., DURAK, U., 2014. Emergency department during long public holidays. Turkish journal of emergency medicine 14, 165–171.

Diehl, A.K., Morris, M.D., Mannis, S.A., 1981. Use of calendar and weather data to predict walk-in attendance. Southern medical journal 74, 709–712.

Durbin, J., Koopman, S.J., 2012. Time series analysis by state space methods. volume 38. Oxford University Press.

Faryar, K.A., 2013. The effects of weekday, season, federal holidays, and severe weather conditions on emergency department volume in montgomery county, ohio .

Franses, P.H., 2016. A note on the mean absolute scaled error. International Journal of Forecasting 32, 20–22.

Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102, 359–378.

Haben, S., Giasemidis, G., Ziel, F., Arora, S., 2019. Short term load forecasting and the effect of temperature at the low voltage level. International Journal of Forecasting 35, 1469–1484.

Hastie, T., Tibshirani, R., Wainwright, M., 2015. Statistical learning with sparsity: the lasso and generalizations. CRC press.

Holleman, D.R., Bowling, R.L., Gathy, C., 1996. Predicting daily visits to a waik-in clinic and emergency department using calendar and weather data. Journal of General Internal Medicine 11, 237–239.

Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., Hyndman, R.J., 2016. Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. International Journal of Forecasting 32, 896–913.

Hyndman, R.J., Khandakar, Y., 2008. Automatic time series forecasting: the forecast package for R. Journal of Statistical Software 26, 1–22. URL: http://www.jstatsoft.org/article/view/v027i03.

Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. International journal of forecasting 22, 679–688.

Izady, N., Worthington, D., 2012. Setting staffing requirements for time dependent queueing networks: The case of accident and emergency departments. European Journal of Operational Research 219, 531–540.

Kam, H.J., Sung, J.O., Park, R.W., 2010. Prediction of daily patient numbers for a regional emergency medical center using time series analysis. Healthcare informatics research 16, 158–165.

Kamali, M.F., Tezcan, T., Yildiz, O., 2018. When to use provider triage in emergency departments. Management Science .

Marcilio, I., Hajat, S., Gouveia, N., 2013. Forecasting daily emergency department visits using calendar variables and ambient temperature readings. Academic emergency medicine 20, 769–777.

Reis, B.Y., Mandl, K.D., 2003. Time series modeling for syndromic surveillance. BMC Medical Informatics and Decision Making 3, 2.

Rotstein, Z., Wilf-Miron, R., Lavi, B., Shahar, A., Gabbay, U., Noy, S., 1997. The dynamics of patient visits to a public hospital ed: a statistical model. The American journal of emergency medicine 15, 596–599.

Schweigler, L.M., Desmond, J.S., McCarthy, M.L., Bukowski, K.J., Ionides, E.L., Younger, J.G., 2009. Forecasting models of emergency department crowding. Academic Emergency Medicine 16, 301–308.

Sun, Y., Heng, B.H., Seow, Y.T., Seow, E., 2009. Forecasting daily attendances at an emergency department to aid resource planning. BMC emergency medicine 9, 1.

Tai, C.C., Lee, C.C., Shih, C.L., Chen, S.C., 2007. Effects of ambient temperature on volume, specialty composition and triage levels of emergency department visits. Emergency Medicine Journal 24, 641–644.

Uniejewski, B., Nowotarski, J., Weron, R., 2016. Automated variable selection and shrinkage for day-ahead electricity price forecasting. Energies 9, 621.

Wargon, Guidet B., H.T.D.H.G., 2009. A systematic review of models for forecasting the number of emergency department visits. Emergency Medicine Journal 29, 395–399.

Xu, K., Chan, C.W., 2016. Using future information to reduce waiting times in the emergency department via diversion. Manufacturing & Service Operations Management 18, 314–331.

Zheng, W., Muscatello, D.J., Chan, A.C., 2007. Deck the halls with rows of trolleys... emergency departments are busiest over the christmas holiday period. Med J Aust 187, 630–3.

Ziel, F., 2016. Iteratively reweighted adaptive lasso for conditional heteroscedastic time series with applications to ar–arch type processes. Computational Statistics & Data Analysis 100, 773–793.

Ziel, F., 2019. Quantile regression for the qualifying match of gefcom2017 probabilistic load forecasting. International Journal of Forecasting 35, 1400–1408.

Ziel, F., Liu, B., 2016. Lasso estimation for gefcom2014 probabilistic electric load forecasting. International Journal of Forecasting 32, 1029–1037.

Ziel, F., Weron, R., 2018. Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. Energy Economics 70, 396–420.

Appendix

Models→ h↓	Proposed Model	Proposed Model without Date/Hld effects	Naive	AR(p)	ETS
1	18.34	18.53	25.51	19.39	22.99
2	18.80	19.00	25.45	19.78	23.76
3	19.02	19.16	25.40	19.73	23.34
4	19.23	19.60	25.46	20.00	23.45
5	19.36	19.72	25.44	20.12	23.18
6	19.37	19.76	25.47	20.08	22.53
7	19.37	20.04	25.41	20.04	22.85
14	20.09	21.17	25.33	21.23	24.04
21	20.32	22.77	25.35	22.07	25.56
28	20.65	23.93	25.36	22.74	26.67
35	20.47	24.84	25.29	23.04	27.91
42	20.51	25.54	25.31	23.13	28.66

Table 2: MAE results for selected forecasting horizons h.

Models→ h↓	Proposed Model	Proposed Model without Date/Hld effects	Naive	AR(p)	ETS
1	23.17	23.66	32.43	24.64	29.50
2	23.86	24.57	32.34	25.47	30.58
3	24.41	24.95	32.26	25.55	30.13
4	24.71	25.42	32.37	25.89	30.23
5	24.84	25.75	32.35	26.10	30.10
6	24.90	25.84	32.36	26.10	29.40
7	24.97	26.07	32.31	26.11	29.74
14	25.71	27.62	32.21	27.56	31.35
21	26.01	29.29	32.23	28.61	32.99
28	26.32	30.62	32.28	29.39	34.08
35	26.21	31.74	32.24	29.74	35.41
42	26.16	32.81	32.21	29.74	36.10

Table 3: RMSE results for selected forecasting horizons h.

Models→ h↓	Proposed Model	Proposed Model without Date/Hld effects	Naive	AR(p)	ETS
1	6.55	6.66	9.20	6.94	8.31
2	6.75	6.89	9.17	7.15	8.59
3	6.87	6.96	9.16	7.15	8.46
4	6.96	7.10	9.19	7.25	8.49
5	7.00	7.18	9.17	7.30	8.43
6	7.01	7.20	9.18	7.31	8.23
7	7.03	7.27	9.16	7.31	8.36
14	7.24	7.68	9.14	7.72	8.78
21	7.35	8.21	9.13	8.01	9.27
28	7.45	8.65	9.15	8.23	9.66
35	7.40	9.00	9.13	8.35	10.09
42	7.40	9.31	9.12	8.37	10.32

Table 4: Pinball score results for selected forecasting horizons h.

Models→ q↓	Proposed Model	Proposed Model without Date/Hld effects	Naive	AR(p)	ETS
0.01	0.64	0.89	0.89	0.88	0.96
0.05	2.62	3.15	3.31	3.11	3.55
0.1	4.43	5.18	5.44	5.12	5.95
0.2	7.12	8.11	8.79	7.93	9.29
0.3	8.83	9.97	10.91	9.74	11.42
0.4	9.78	10.99	12.21	10.69	12.68
0.5	10.08	11.30	12.67	10.97	13.09
0.6	9.79	10.93	12.45	10.60	12.72
0.7	8.90	9.89	11.29	9.56	11.53
0.8	7.27	8.08	9.27	7.81	9.44
0.9	4.68	5.19	6.00	4.99	6.19
0.95	2.85	3.18	3.64	3.00	3.81
0.99	0.81	0.97	0.99	0.83	1.06

Table 5: Pinball score results for selected probability levels q.