Special Issue Article

Quality and Reliability Engineering International

(wileyonlinelibrary.com) DOI: 10.1002/qre.2095

Published online 6 October 2016 in Wiley Online Library

A Hybrid Approach for Forecasting Patient Visits in Emergency Department

Qinneng Xu,[a] Kwok-Leung Tsui,[a] Wei Jiang[b*†] and Hainan Guo[a]

An accurate forecast of patient visits in emergency departments (EDs) is one of the key challenges for health care policy makers to better allocate medical resources and service providers. In this paper, a hybrid autoregressive integrated moving average–linear regression (ARIMA–LR) approach, which combines ARIMA and LR in a sequential manner, is developed because of its ability to capture seasonal trend and effects of predictors. The forecasting performance of the hybrid approach is compared with several widely used models, generalized linear model (GLM), ARIMA, ARIMA with explanatory variables (ARIMAX), and ARIMA–artificial neural network (ANN) hybrid model, using two real-world data sets collected from hospitals in DaLian, LiaoNing Province, China. The hybrid ARIMA–LR model is shown to outperform existing models in terms of forecasting accuracy. Moreover, involving a smoothing process is found helpful in reducing the interference by holiday outliers. The proposed approach can be a competitive alternative to forecast short-term daily ED volume. Copyright © 2016 John Wiley & Sons, Ltd.

Keywords: ARIMA; emergency department; forecasting; linear regression

1. Introduction

Patients in many countries often suffer from overcrowding when visiting emergency department (ED), which causes increasing length of stay, below standard treatments on patients, intensive stress among doctors and nurses, and increasing health care costs.[1] All of these can directly reduce the healthcare quality. It is urgently needed to make appropriate ED resource planning strategy such as expanding ED, rescheduling staffs' shift plan, allocating physical facilities, etc. to satisfy the growing demand of patients. As patient volumes differ from day to day, as well as from hospital to hospital, one key challenge to improve the healthcare quality is the accurate prediction of patient flows in ED.[2] Accurate predictions contribute to allocating human and physical resources to improve the overall quality of medical services, which benefits both healthcare providers and patients.

The objective of ED forecasting problem is to use statistical forecasting methods to predict weekly, daily, or hourly ED patient volume. Several works have been proposed and compared. For short-term forecasting, Schweigler et al.[1] proposed autoregressive (AR) models with seasonal/sinusoidal adjustment, which are shown to have better performance than the simple average of historical data for forecasting ED occupancy up to 12 h in advance. McCarthy et al.[3] used Poisson regression to predict hourly arrivals in a medical center in U.S. Wargon et al.[4] used generalized linear model (GLM) based on calendar variables to forecast daily visits in four EDs in Paris. Marcilio et al.[5] forecast the daily visits to an ED in Sao Paulo, Brazil, for 7-day and 30-day forecasting horizons. Generally, three models were used: GLM, generalized estimating equations (GEE), and seasonal AR integrated moving average (SARIMA). Empirical results show that the forecasting accuracy of GEE and GLM is similar or better than that of SARIMA. Jones et al.[6] compared liner regression, SARIMA, time series regression, exponential smoothing, and artificial neural network (ANN) for predicting daily ED patient volumes at three diverse hospitals. Forecasts were made for 1-, 7-, 14-, 21-, and 30-day forecasting horizons. The forecasting accuracy of these models is similar, and no model performs consistently better than others. Sun et al.[7] and Kadri et al.[8] used ARIMA model to predict daily patients for three categories of diseases in Singapore and France, respectively. Several statistical methods, including moving average, exponential smoothing, ARIMA, and SARIMA, were also compared to forecast ED arrivals for horizon ranging from 1- to 10-day in advance in a hospital in Melbourne by Abraham et al.[9] The authors concluded that none of the models could make useful predictions for any horizon, and reliable forecasts could only be produced for less than one week in advance. They pointed out that new methods should be studied to enhance the performance for long-term forecasting. Boyle et al.[10] used multiple regression, ARIMA, and exponential smoothing models based on a sliding window of 4 weeks to make forecasts for two EDs in Queensland, and compared the findings and limitations in the related work.

[a]*City University of Hong Kong, Hong Kong, People's Republic of China*
[b]*Shanghai Jiao Tong University, Shanghai, People's Republic of China*
**Correspondence to: Wei Jiang, Shanghai Jiao Tong University, Shanghai, People's Republic of China.*
[†]*E-mail: jiangwei08@gmail.com*

Most ED forecasting applications focus on hospitals in developed countries. Much less work has been applied to EDs in China, in which the performance of the methods and the choice of predictor variables can be significantly different. The existing methods were probably inappropriate because of the differences between health systems and culture. Therefore, it is important and of great interest to investigate the existing models as well as to identify appropriate methods for ED patient forecasting in EDs in China. In this paper, we combine ARIMA and linear regression (LR) models and develop a hybrid ARIMA–LR approach, which applies regression and time series methods in a sequential manner. The hybrid ARIMA–LR model possesses the ability to capture the seasonal trend as well as the effects of predictor variables such as temperature, day-of-week, month-of-year, public holidays, school closure, etc. A smoothing method is also employed in ARIMA–LR model to reduce the inaccuracy caused by certain outliers. Four existing models, GLM, ARIMA, ARIMA with explanatory variables (ARIMAX), and ARIMA–ANN hybrid model, are also employed in the application of forecasting ED patient visits in two EDs in China. In order to investigate the significance of external variables, a variable selection procedure was introduced while applying all the models except ARIMA.

The remainder of the paper is organized as follows. The existing models used in this work are described in Section 2. Section 3 presents our proposed hybrid model. In Section 4, the performance of the proposed hybrid model is compared with existing models using two real-life data sets collected from two hospitals in DaLian, LiaoNing province, China. Discussion and conclusions are given in Section 5 and Section 6.

2. Existing methods

2.1. GLM

The basic idea and model framework of GLM can be found in the statistical literature.[11,12] In this study, the ED visit counts might follow the Poisson distribution, we test both LR and log-LR in Experiment I in the next section. As in literature,[3–6] we choose the following frequently used indicators:

- daily highest and lowest temperature T_t^{high}, T_t^{low} (we used one-day before temperatures in the GLM model);
- day of the week $D_{i,t}$;
- month of the year $M_{j,t}$;
- school holiday indicator I_t^{school};
- public holiday indicator $I_{k,t}^{holiday}$;
- the indicator of the day after a public holiday $I_t^{after_{hol}}$.

Among them, day-of-week and month-of-year are calendar indicators. For example, if day t is on Monday and in January, then we assign $D_{1,t}=1$ and $M_{1,t}=1$, and all remaining $D_{i,t}$ and $M_{j,t}$ equal to 0. School closure, public holiday, and day-after-holiday are other categorical variables in the regression model. On summer and winter holidays, $I_t^{school}=1$, and $I_t^{school}=0$ on all remaining days. Similarly, we consider the following public holidays: New Year's Day; Spring Festival; Tomb-sweeping Day; Labor Day; Dragon Boat Festival; National Day; and Mid-autumn Festival. In these holidays, $I_{k,t}^{holiday}$ represents the k th day in one certain holiday, and we assign $I_t^{after\_hol}=1$ if day t is the day after these public holidays.

In Experiment I, we consider y_t as the

$$y_t = a + \sum_{i=1}^{7}\left(b_i \cdot D_{i,t}\right) + \sum_{j=1}^{12} c_j \cdot M_{j,t} + dI_t^{school} + \sum_{k=1}^{8} e_k \cdot I_{k,t}^{holiday} + f \cdot I_t^{after_{hol}} + g \cdot T_{t-1}^{low} + h \cdot T_{t-1}^{high} + e$$

where a is the constant intercept, $b_1 \dots b_7$ are coefficients for day-of-week indicators, $c_1 \dots c_{12}$ are coefficients for month-of-year indicators, d is the coefficient for school closure indicator, e_k is the coefficient for k th day of a certain holiday, and f is the coefficient for the day after these holidays g and h are coefficients for temperature indicators. Similarly, we also consider the following GLM model when $\ln(y_t)$ is the response variable

$$\ln(y_t) = a + \sum_{i=1}^{7}\left(b_i \cdot D_{i,t}\right) + \sum_{j=1}^{12} c_j \cdot M_{j,t} + d \cdot I_t^{school} + \sum_{k=1}^{8} e_k \cdot I_{k,t}^{holiday} + f \cdot I_t^{after_{hol}} + g \cdot T_{t-1}^{low} + h \cdot T_{t-1}^{high} + e.$$

However, it might be possible that some of those widely used variables do not have a significant effect on the response variable in every case. Therefore, a variable selection procedure is introduced in Experiment I to choose the most significant variables. The external variables are divided into three groups according to their natural characteristics: Group 1: temperature variables (T_t^{high} and T_t^{low}); Group 2: holiday variables (I_t^{school}, $I_{k,t}^{holiday}$, and $I_t^{after_{hol}}$); and Group 3: calendar variables ($D_{i,t}$ and $M_{j,t}$).

In the variable selection process, we use different combinations of these three groups of variables as external variables in each model and choose the significant variables in terms of forecasting accuracy.

2.2. ARIMA and ARIMAX

Consider ARIMA (p,d,q) models where p is the number of AR terms, d is the number of nonseasonal differences needed for stationarity, and q is the number of lagged forecast errors in the prediction equation. To apply ARIMA, we need to identify d first. Then,

p and q can be obtained by using the autocorrelation function (ACF) and partial autocorrelation function (PACF).[13–15] ARIMAX model is an extension of ARIMA in attempt to forecast the ED visits using the explanatory variables, which are also used in GLM. The same variable selection procedure as used in GLM is also employed in ARIMAX.

3. A hybrid approach

Hybrid models have been developed and shown superior forecasting performance in many applications: Zhang[16] used it to forecast the sunspot series; Cadenas et al.[17] and Liu et al.[18] used it to forecast the wind speed; Areekul et al.[19] used it to forecast the price in deregulated market. In particular, ARIMA–ANN hybrid model is widely used. However, it is shown ineffective as there are many parameters in ANN needed to be tuned. Choices of parameters could have remarkable influences on forecasting result. Therefore, we here consider a hybrid ARIMA–LR model to integrate the above ARIMA and LR (or GLM) models to forecast ED patient visits. ARIMA–LR is easier to fit and interpreted in real life applications.

As the 371-day sliding window is used, in each fitting process, we need to forecast $\hat{y}_{t+1}$, ED visits on day $t+1$, based on previous ED visits ($y_t, y_{t-1}, \dots\dots, y_{t-370}$), and predictor variables ($X_{t+1}, X_t, X_{t-1}, \dots\dots, X_{t-370}$) which are the same as those used in LR method. The procedure of applying hybrid ARIMA–LR (or ARIMA–ANN) model in each fitting process can be generalized as follows:

Step 1 use ARIMA to fit ($y_t, y_{t-1}, \dots\dots, y_{t-370}$), and calculate residuals (training error), ($e_t, e_{t-1}, \dots\dots, e_{t-370}$), where $e_i = y_i - \hat{L}_i$, $i = t, t-1, \dots\dots, t-370$, $\hat{L}_i$ is the forecasting value on day i. Based on the fitted ARIMA model, $\hat{L}_{t+1}$ can be forecasted.

Step 2 use LR (or ANN) to fit the new training set and estimate the residual on day $t+1$, $\hat{e}_{t+1}$. In this step, ($e_t, e_{t-1}, \dots\dots, e_{t-n}$) are outputs, and if LR is employed, e_t is calculated as

$$e_t = a + \sum_{i=1}^{7}\left(b_i \cdot D_{i,t}\right) + \sum_{j=1}^{12} c_j \cdot M_{j,t} + d \cdot I_t^{school} + \sum_{k=1}^{8} e_k \cdot I_{k,t}^{holiday} + f \cdot I_t^{after_{hol}} + g \cdot T_{t-1}^{low} + h \cdot T_{t-1}^{high} + \varepsilon.$$

Similarly, ANN is similar to LR, with the same output and input variables. The forecasted residual on day $t+1$, $\hat{e}_{t+1}$, can be calculated by the fitted model.

Step 3 the ED visits on day $t+1$ is the sum of the two forecasting values, that is, $\hat{y}_{t+1} = \hat{L}_{t+1} + \hat{e}_{t+1}$.

However, when applying these models, abnormal errors could be introduced by expected outliers (especially in some official holidays) and unexpected outliers in the raw data. These errors would lead to inaccurate prediction in either method and large accumulative errors in the final forecasting value. To alleviate this shortcoming, we consider smoothing the raw holiday data before applying these models.

Figure 1 depicts the hybrid model with smoothing. On day t, we first use 7-day moving average to smooth the time series ($y_t, y_{t-1}, \dots\dots, y_{t-370}$) only for those public holidays: if day i, $i = t-370, t-369, \dots\dots, t$, is a public holiday, $y_i' = \frac{\sum_{j=i-7}^{i-1} y_j}{7}$, which means we use the 7-day moving average to replace the actual ED visits on day i; if day i, $i = t-370, t-369, \dots\dots, t$, is not a holiday, we do not make any changes, and $y_i' = y_i$. This process aims at eliminating the disturbance of holiday outliers before applying time series models. ARIMA is then used to fit the smoothed time series data ($y_t', y_{t-1}', \dots\dots, y_{t-370}'$). Residuals, ($e_t, e_{t-1}, \dots\dots, e_{t-370}$), are calculated as follows: $e_i = y_i - \hat{L}_i$, $i = t, t-1, \dots\dots, t-370$. The reason for using y_i rather than y_i' is that the eliminated holiday effects should be compensated. Same procedure is employed before using GLM and ARIMA.

4. A case study

4.1. Data collection

The study ED data are collected from two representative hospitals in DaLian, LiaoNing Province, China. ED A is one of the biggest hospitals in southern LiaoNing Province, which has an annual volume of around 200 000 ED visits. ED B is a relatively small hospital with an annual volume of around 15 000 ED visits. Both EDs receive ambulances carrying patients from the surrounding districts, as well as walk-in patients. In China, the health system is different from that in other countries: patients can visit either ED or specialists in

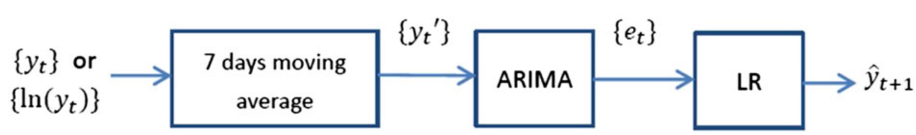

Figure 1. Flow chart of hybrid models with 7-day moving average smoothing.

working days (normally Monday to Friday, 8:00 to 12:00, 14:00 to 17:30) without an appointment, and patients can only visit ED at other times.

We extract daily ED visits in both hospitals from January 1, 2012 to December 31, 2013. We also collect the daily highest and lowest temperature in Dalian from China Metrological Administration for the same time period. In this study, New Year's Day, Spring Festival, Tomb-sweeping Day, Labor Day, Dragon Boat Festival, National Day, and Mid-autumn Festival are considered as public holidays. School closure includes summer holidays and winter holidays. The summer school holiday is from July 21 to August 31 in 2012 and from July 22 to August 31 in 2013. The winter school holiday was from January 15 to February 29 in 2012 and from January 19 to February 28 in 2013.

On average, there are about 700 daily attendances in ED A and about 40 attendances in ED B during the period from the year 2012 to 2013. The daily volume plot (Figure 2) shows the annual seasonal fluctuations in ED A: there were more patients in July and August; there were slightly more patients in December. In contrast, the right plot shows unobvious seasonal trends in ED B.

The boxplot of the distribution of ED arrivals by day-of-week, month-of-year, holiday, and school closure (Figure 3) shows similar patterns of daily patient volumes in both EDs: the daily count of ED admissions in each day is similar, and there are slightly more patients on Monday and Sunday; the patient volumes are much higher in July and August; more patients visit ED during school closure. But for those holidays, there are differences between ED A and ED B: there are more patients visiting ED A during public holidays, and in several days during these periods, there are much more patient arrivals, while this phenomenon is not obvious in ED B.

4.2. Study design

Two experiments are conducted to forecast daily ED visits 1 day in advance. In Experiment I, we compare forecast performance of GLM, ARIMA, ARIMAX, hybrid ARIMA–ANN, and ARIMA–LR models. In order to verify the Poisson assumption in literature, we consider two approaches in Experiment I: in the first approach, the count of ED patient visits is used as the response variable; in the second approach, log-transformation of the ED visits is used. A variable selection procedure is also introduced; we regroup the external variables according to their natural characteristics and produce forecast values by using all the possible combinations of different groups of variables. Through Experiment I, we aim to determine the significant variables in each ED and compare five models and test the effect of the Poisson assumption. In Experiment II, to alleviate holiday outliers, we further investigate the effect of using 7-day moving average smoothing with the best models in Experiment I.

The data set is divided into training set and test set. The training set covers the period from January 1, 2012 to June 31, 2013 (547 days), and the test set covers the period from July 1, 2013 to December 31, 2013 (184 days). Normally, GLM uses the training set to initialize the regression coefficients, and the test set is used to estimate the forecasting value. Only one computing process is needed for forecasting the last half year's patient volume. As Burkom et al.[20] pointed out, this fitting method is based on one major assumption, that is, the relationship between the chosen indicators and the number of ED admissions did not vary over time from training data and could be reasonably used for out-of-sample forecasting. However, this assumption might not hold in our case, because the data behaviors are always changing. Therefore, we use an adaptive approach to overcome this weakness.

The adaptive GLM fits the observations in a sliding baseline interval. In this study, we use a 371-day window and cover a whole year in order to capture the yearly trend and seasonal pattern. In other words, the regression coefficients are recomputed for each forecast by using only the training data from the 53 weeks before the forecasting day. Thus, we have to refit the model for 184 times in order to forecast the ED admissions from July 1, 2013 to December 31, 2013 (184 days). To make the comparison fair and reasonable, the ARIMA, ARIMAX, ARIMA–ANN, and ARIMA–LR models all use the same adaptive method with a 371-day window in this study.

Two commonly used metrics are employed to measure the forecast accuracy: the root mean square error (RMSE) and the mean absolute percentage error (MAPE). RMSE measures the average error magnitudes and is sensitive to large errors. It is useful for our case that large errors are particularly undesirable. MAPE, on the other hand, gives the percentage error of the series. For a series of forecasting values $(\hat{y}_1, \hat{y}_2, \dots \dots, \hat{y}_n)$ and the corresponding observed values $(y_1, y_2, \dots \dots, y_n)$

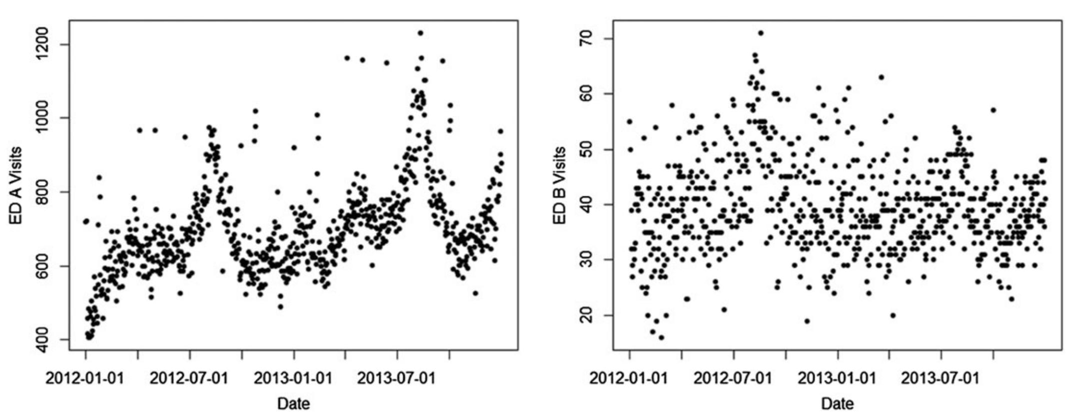

Figure 2. Daily attendances in ED A (left) and ED B (right).

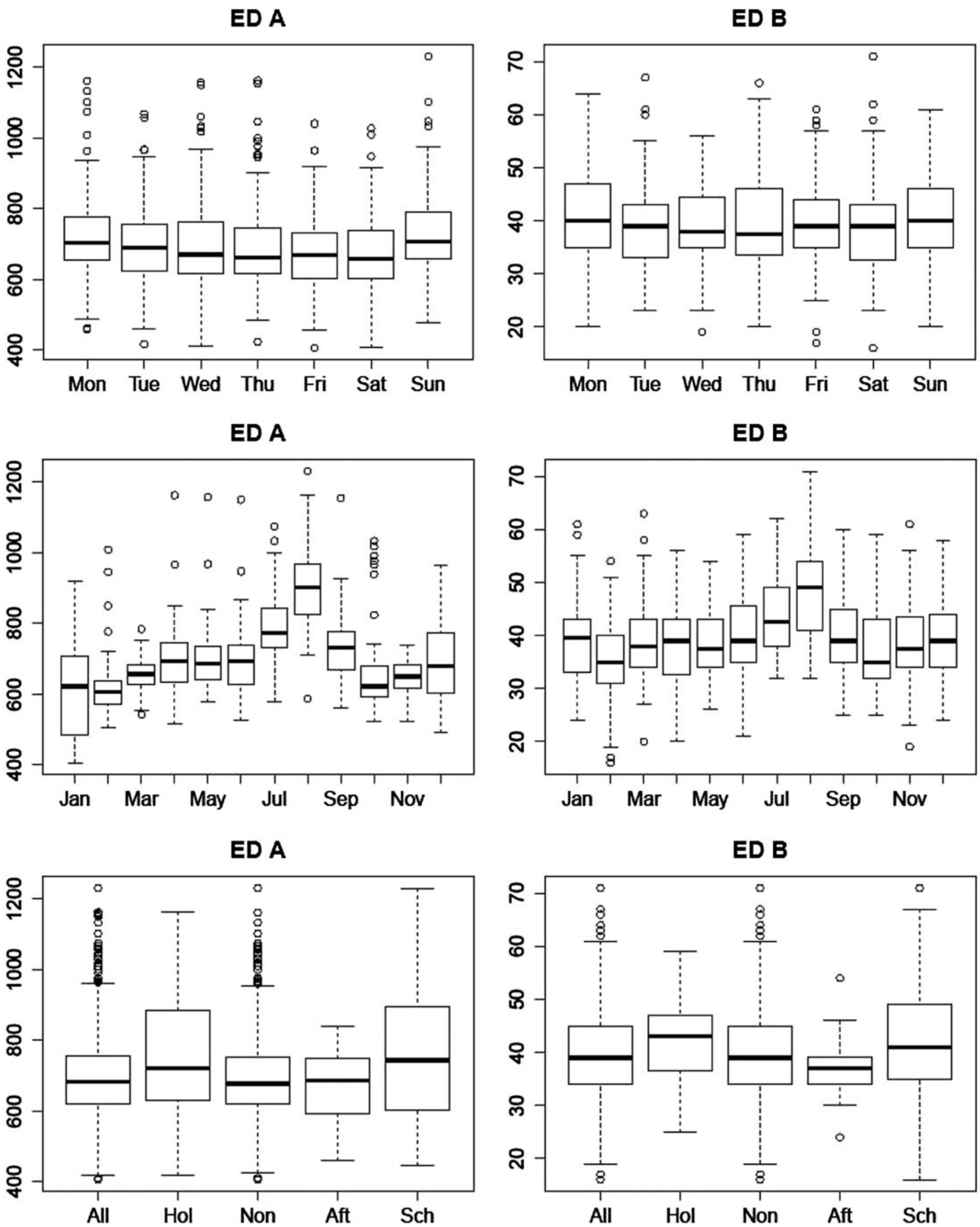

Figure 3. Boxplot of patient volumes in ED A and ED B (Hol: holidays; Non: non-holidays; Aft: the days after holidays; Sch: school closure).

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right|$$

4.3. Forecasting accuracy

Tables I and II show the results of variables selection in ED A and ED B, respectively. In ED A (Table I), all of the three groups of variables are important when using GLM, ARIMAX, and ARIMA–LR, no matter the response variables is $\{y_t\}$ or $\{\ln(y_t)\}$. Therefore, temperature, holidays, and calendar indicators all have important effects on patient visits in ED A. While for ARIMA–ANN model, it is difficult to determine the significant variables. The reason is that we used only one set of parameters in each fitting process as there are few

Table I. The result of variable selection in ED A—RMSE (MAPE) (*Group 1*: *temperature variables* (T_t^{high} *and* T_t^{low}); *Group 2*: *holiday variables* (I_t^{school}, $I_{k,t}^{holiday}$, *and* $I_t^{after_{hol}}$); *Group 3*: *calendar variables* ($D_{i,t}$ *and* $M_{j,t}$))

Output	Applied method	Group 1	Group 2	Group 3	Group 1, 2	Group 1, 3	Group 2, 3	Group 1, 2, 3
y_t	GLM	128.8 (12.2%)	131.2 (11.6%)	110.9 (10.3%)	112.4 (10.7%)	109.2 (10.0%)	100.7 (9.6%)	98.2 (9.2%)
	ARIMAX	115.6 (10.6%)	103.6 (9.5%)	113.8 (10.3%)	103.3 (9.6%)	114.3 (10.3%)	98.8 (7.0%)	98.2 (9.3%)
	ARIMA–LR	77.9 (7.2%)	71.1 (6.9%)	76.5 (7.0%)	70.8 (6.9%)	77.2 (7.1%)	69.0 (6.7%)	67.1 (6.5%)
	ARIMA–ANN	79.1 (7.6%)	73.0 (7.4%)	82.5 (7.9%)	72.1 (7.1%)	82.1 (7.6%)	72.3 (7.3%)	72.0 (7.2%)
$\ln(y_t)$	GLM	130.2 (11.9%)	136.5 (11.9%)	112.3 (10.3%)	112.4 (10.5%)	111.0 (10.0%)	102.9 (9.6%)	97.8 (9.2%)
	ARIMAX	116.7 (10.6%)	104.7 (9.6%)	115.0 (10.3%)	103.0 (9.6%)	115.2 (10.3%)	101.2 (9.5%)	99.7 (9.3%)
	ARIMA–LR	76.2 (7.0%)	69.2 (6.7%)	74.0 (6.7%)	69.0 (6.7%)	75.0 (6.7%)	64.6 (6.2%)	64.8 (6.2%)
	ARIMA–ANN	79.6 (7.4%)	68.9 (6.6%)	80.2 (7.5%)	74.9 (7.0%)	99.3 (9.5%)	79.3 (7.3%)	94.2 (8.0%)

Table II. The result of variable selection in ED B—RMSE (MAPE) (*Group 1*: *temperature variables* (T_t^{high} *and* T_t^{low}); *Group 2*: *holiday variables* (I_t^{school}, $I_{k,t}^{holiday}$, *and* $I_t^{after_{hol}}$); *Group 3*: *calendar variables* ($D_{i,t}$ *and* $M_{j,t}$))

Output	Applied method	Group 1	Group 2	Group 3	Group 1, 2	Group 1, 3	Group 2, 3	Group 1, 2, 3
y_t	GLM	6.26 (14.5%)	5.73 (13.1%)	6.82 (15.3%)	6.19 (14.2%)	6.85 (15.3%)	6.68 (15.0%)	6.73 (15.1%)
	ARIMAX	6.01 (14.0%)	5.62 (12.9%)	6.64 (14.8%)	5.96 (13.8%)	6.69 (14.9%)	6.56 (14.7%)	6.62 (14.9%)
	ARIMA–LR	5.67 (12.8%)	5.49 (12.3%)	5.68 (12.6%)	5.55 (12.6%)	5.67 (12.7%)	5.58 (12.6%)	5.58 (12.6%)
	ARIMA–ANN	5.94 (13.3%)	5.61 (12.5%)	6.42 (14.0%)	6.42 (14.3%)	7.20 (15.1%)	7.04 (15.0%)	7.00 (14.8%)
$\ln(y_t)$	GLM	6.08 (14.0%)	5.66 (12.7%)	6.54 (14.5%)	6.00 (13.7%)	6.57 (14.6%)	6.41 (14.3%)	6.77 (15.2%)
	ARIMAX	5.86 (13.4%)	5.55 (12.5%)	6.38 (14.2%)	5.81 (13.3%)	6.42 (14.3%)	6.31 (14.0%)	6.37 (14.2%)
	ARIMA–LR	5.57 (12.3%)	5.41 (11.9%)	5.54 (12.3%)	5.47 (12.2%)	5.53 (12.3%)	5.46 (12.2%)	5.45 (12.2%)
	ARIMA–ANN	5.80 (13.2%)	5.54 (12.2%)	6.25 (13.6%)	8.40 (15.1%)	8.39 (16.2%)	6.66 (14.1%)	9.11 (12.7%)

efficient auto tuning algorithms appropriate in this case, and it is difficult to find the best combination of parameters for each fit process in an efficient and reasonable way. It is also difficult to find a general set of parameters for every fitting process which can produce better forecasting results than ARIMA–LR. Therefore, ARIMA–ANN model might not be an appropriate model in this problem.

The result of ED B (Table II) is inconsistent with that of ED A. Holiday variables (Group 2), including holidays, after holidays and school closure, are most significant in all the models. Temperature (Group 1) and calendar (Group 3) variables have no important effects on patient visits in ED B. The partial reason is that the patient flows of ED B do not show clear seasonal patterns and calendar variables are not important then. On the other hand, the temperature is also kind of relating to the calendar date. This phenomenon can be explained by the real situation that Chinese people prefer visiting bigger hospitals when they are sick even though they may only catch a cold and they choose smaller hospitals only when the situation is critically emerging and they need to be treated as soon as possible. Therefore, the patient flows in smaller hospitals are more randomized, and it is difficult to find a relatively consistent seasonal trend from year to year. The seasonal trend in the bigger hospital is more obvious as there are many diseases (such as cold and influenza) usually have a seasonal epidemic and the outbreak of the epidemic is highly related to meteorological or calendar variables.

Table III displays the MAPE and RMSE of each model in Experiment I. In both EDs, the performances of GLM, ARIMA, and ARIMAX to forecast $\{y_t\}$ and $\{\ln(y_t)\}$ are similar, while the forecasting accuracy of ARIMA–LR is better and the accuracy of ARIMA–ANN is worse when the log-transformed ED patient visits $\{\ln(y_t)\}$ are the intrinsic time series of interest. When comparing these five models, ARIMA–LR is the best in both EDs, no matter the response variable is $\{y_t\}$ or $\{\ln(y_t)\}$. This is more obvious in ED A than ED B, as MAPE and RMSE of ARIMA–LR and other models are closer in ED B. When considering overall accuracy, ARIMA–LR has superiority in ED A and not evident advantage in ED B. This can be also explained by the fact that the patient flow of ED B consists of randomness and uncertainty.

Table III also shows the comparison of forecasting accuracy for holidays and non-holidays in ED A and ED B. During the forecasting period, from July 1, 2013 to December 31, 2013, there were 10 holidays and 174 non-holidays. For non-holidays, ARIMA–LR always outperforms other methods and ARIMA is the second best, no matter $\{y_t\}$ or $\{\ln(y_t)\}$ is used as the response variable. For holidays, ARIMA–LR still outperforms others in most cases, but in ED B ARIMA produces the most accurate forecasted values. When considering

Table III. Comparison of one-day-ahead forecasting for ED A and ED B-RMSE (MAPE)							
		ED A			ED B		
Output	Applied method	Overall	Holidays	Non-holidays	Overall	Holidays	Non-holidays
y_t	GLM	98.2 (9.2%)	164.5 (13.1%)	92.9 (9.0%)	5.73 (13.1%)	5.75 (13.3%)	5.73 (13.1%)
	ARIMA	78.1 (7.2%)	190.9 (16.4%)	65.9 (6.7%)	5.64 (12.7%)	6.50 (12.5%)	5.60 (12.7%)
	ARIMAX	98.2 (9.3%)	163.6 (13.5%)	93.1 (9.0%)	5.62 (12.9%)	5.64 (13.0%)	5.62 (12.9%)
	ARIMA–LR	67.1 (6.5%)	136.4 (13.0%)	60.8 (6.2%)	5.49 (12.3%)	5.66 (13.1%)	5.48 (12.3%)
	ARIMA–ANN	72.0 (7.2%)	137.1 (12.9%)	66.3 (6.9%)	5.61 (12.5%)	7.53 (16.6%)	5.48 (12.3%)
$\ln(y_t)$	GLM	97.8 (9.2%)	167.5 (13.5%)	92.2 (8.9%)	5.66 (12.7%)	5.86 (13.0%)	5.54 (12.6%)
	ARIMA	76.3 (7.0%)	190.5 (16.3%)	63.9 (6.5%)	5.53 (12.2%)	6.61 (12.6%)	5.50 (12.2%)
	ARIMAX	99.7 (9.3%)	175.4 (14.4%)	93.5 (9.0%)	5.55 (12.5%)	5.62 (12.9%)	5.53 (12.9%)
	ARIMA–LR	64.8 (6.2%)	131.4 (12.1%)	58.7 (5.8%)	5.41 (11.9%)	5.54 (12.8%)	5.37 (11.7%)
	ARIMA–ANN	94.2 (8.0%)	253.6 (17.8%)	75.4 (7.5%)	5.54 (12.2%)	7.53 (16.6%)	5.48 (12.3%)

ARIMAX, which incorporates ARIMA model with explanatory variables in one step, and ARIMA–ANN, both of them cannot produce satisfactory forecasting results especially ARIMA–ANN. When comparing ED A and ED B, including external variables in an appropriate way would improve forecasting performances more significantly in ED A than that in ED B, which implies that external variables are helpful in ED A while ED B visits are probably driven by serial correlations (inertia) instead of exogenous indicators.

Table IV displays the MAPE and RMSE of each model in Experiment II. We choose the best two models in Experiment I and apply a 7-day moving average smoothing step before using these two models. The result shows that integrating the smoothing step improves the overall forecasting accuracy of ARIMA–LR model in both EDs. Among all models, the smoothing ARIMA–LR model to forecast $\{\ln(y_t)\}$ performed the best in both ED A and ED B. When considering holidays and non-holidays separately, results are different: during non-holidays, the smoothing step improves both ARIMA–LR models in ED A and only slightly improves forecasting accuracy in ED B; during non-holidays, the smoothing even deteriorates the forecasting accuracy in ED A but slightly improves the two models in ED B. Therefore, although the smoothing step is helpful in improving the forecast accuracy in both EDs, the effect on holidays and non-holidays could be different in different EDs.

5. Discussions

In this study, we employ several statistical models to forecast one-day-ahead ED patient visits in two EDs in China. We investigate GLM, ARIMA, ARIMAX, hybrid ARIMA–ANN, and ARIMA–LR model when the output time series data is $\{y_t\}$ or $\{\ln(y_t)\}$. A variable selection procedure is introduced to choose the most significant variables for both EDs. A 7-day moving average smoothing process for public holidays is also integrated in order to eliminate the negative effect of holiday outliers. Adaptive fitting process is applied to capture the recent time series trend and relationships between ED patient visits and predictor variables, which depends on the fact that patterns of ED volume differed from year to year. One motivation of our study is that no existing model is applicative for every ED even though each existing method has been proved to be effective and accurate in certain forecasting cases, and EDs in this study located in China have not been extensively studied. Another motivation is to demonstrate the feasibility and effectiveness of hybrid model with smoothing in forecasting ED visits.

The results show that forecasting accuracy of ED A is better than that of ED B, because more obvious trend-like effects (such as month-of-year effect, holiday effect, etc.) appear in ED A. As noise in series data of ED B is larger, time series models could not easily capture the short-term trend and relationships between predictor variables and ED arrivals could not be accurately estimated. Therefore, the count of ED patient visits with low volume or large variance is more difficult to forecast. Temperature and calendar variables might not have effects on patient visits in small EDs.

Table IV. Comparison of involving 7-day moving average smoothing—RMSE (MAPE)							
		ED A			ED B		
Output	Applied method	Overall	Holidays	Non-holidays	Overall	Holidays	Non-holidays
y_t	ARIMA–LR (smoothing)	65.4 (6.1%)	147.7 (12.0%)	57.2 (5.8%)	5.38 (12.0%)	5.58 (12.9%)	5.35 (11.6%)
$\ln(y_t)$	ARIMA–LR (smoothing)	64.4 (6.0%)	145.9 (11.9%)	56.3 (5.7%)	5.36 (11.7%)	5.59 (12.9%)	5.33 (11.4%)

Table V. Comparison of 7-day-ahead forecasting—RMSE (MAPE)		
Applied method	ED A	ED B
GLM	70.5 (6.8%)	5.12 (11.3%)
ARIMA	109 (10.6%)	6.21 (13.7%)
ARIMA–LR	96.1 (9.6%)	5.83 (13.0%)
ARIMA–LR (smoothing)	104 (9.6%)	6.37 (13.6%)

The log transformation of patient visits before using ARIMA, ARIMAX and ARIMA–LR can improve the accuracy, while log transformation does not effectively improve the result of GLM and ARIMA–ANN. The partial reason may be that the log transformation converted nonstationary raw data with respect to the variance into relatively stationary time series (variation stabilization), and ARIMA could perform better when applied to stationary time series data. ARIMA–ANN performs worse than ARIMA–LR because of the difficulties in choosing the best parameters in all the 184 fits. This reason could also partially explain why ARIMAX does not perform well.

Selection of predictor variables still remains challenging and the variables chosen in this study may not be applicable or representative in other studies. On the other hand, short-term changes of patient volumes are usually not large, and time series models could capture the short-term trends better.

We suggest that certain noises in the raw data, such as holiday outliers, might cause the inaccuracy. To eliminate these noises, we add in a smoothing procedure to smooth patient visits in holidays. After smoothing, ARIMA could better estimate the short-term trend without disturbances caused by holiday outliers, but the removed component, $\{y_t - y_t'\}$, could not be compensated. But in ARIMA–LR model, the removed part would be compensated when applying LR model to forecast residuals. The result (Table II) shows that hybrid ARIMA–LR model with smoothing produces the overall best result in both EDs, and we could conclude that smoothing process did help resolve the problem because of holiday outliers. This smoothing method can be also extended to other outliers: if a certain factor is found to be related with outliers (such as school closure), we can smooth these days before applying ARIMA and compensate it when forecasting residuals.

When considering holidays and non-holidays separately, the result of ED A is consistent with the result of ED B during non-holidays but inconsistent during holidays. Embedding smoothing improved original models as holiday outliers are replaced by average ED count of previous 7 days. As there are 174 non-holidays and only 10 holidays in the test period, the model which has the best performance during non-holidays also shows the best overall performance. The limited number of holidays in the test set could lead to large variances of forecasting values and make the results different from ED A and ED B. It is uncertain if the smoothing step could improve forecasting accuracy during holidays. Forecasting accuracy of holidays is much worse than that of non-holidays, which is because of the fact that only 20 to 30 days' data are used to fit regression models if the forecasted day is a holiday. The data in the training set is not enough to calculate the coefficients as 31 predictor variables existed, and overfitting might exist. Thus, our hybrid model with smoothing could be a reliable method to forecast Chinese ED patient volume in non-holidays, but for holidays, more datasets should be collected and investigated to draw further conclusions.

To understand the performances of our methods in long-term forecasting, Table V displays the MAPE and RMSE of GLM, ARIMA, ARIMA–LR, and ARIMA–LR with smoothing for seven-day-ahead forecasting. Different from the result of one-day-ahead forecasting problem, adaptive GLM shows the lowest MAPE and RMSE when forecasting patient volumes seven days in advance. ARIMA–LR also improved the forecasting accuracy when compared to ARIMA, but employing smoothing procedures could not help improve ARIMA–LR model. Therefore, any method incorporating ARIMA cannot produce reliable long-term predictions. This is because of the characteristics of time series methods. Time series methods are suited to short-term forecasts and relatively stable situations. When substantial fluctuations are common and underlying conditions are subject to extreme change, then time series methods may give relatively poor results. In both EDs, the change during 7 days' period can be quite large and patient flows are not stable. Thus, when forecasting seven-day-ahead patient visits, there are more uncertainty and interference factors, and the simple method like adaptive GLM which does not incorporate time series method may be more suitable.

6. Conclusions

Our study investigates several models to forecast daily ED admissions in China. The hybrid ARIMA–LR model shows better results than existing models and involving smoothing helps reduce the interference by holiday outliers. The proposed approach can be a feasible alternative to forecast short-term daily ED volume in China or even other countries and helps policymakers to generate better strategies to improve healthcare quality in ED.

However, there are also limitations in this study. The dataset contains only two years' data, and the pattern of patient volume in the year 2012 is different from that in the year 2013. The data may be not enough for regression models to capture the long-term relationships between patient visits and predictor variables. The knowledge of selecting predictor variables in China is limited, and there are probably other important factors we have not taken into consideration. On the other hand, predictor variables in our models may also cause overfitting. Besides, there are only two EDs tested in this study, more EDs in China should be included in this study to demonstrate the generalization of our methods. These are all the potential research directions in the future.

Acknowledgements

We would like to thank the Associate Editor and two anonymous referees for their helpful comments. We thank Tie Niu, the director of the information center of the First Affiliated Hospital of Dalian Medical University, and Gang Qu, the vice president of Dalian University Affiliated Xinhua Hospital, for providing and annotating the data. This research is supported by the Research Grants Council Collaborative Research Fund (Ref. CityU8/CRF/12G) and Theme-Based Research Scheme (Ref. T32-102/14-N). Wei Jiang's research is partially supported by the National Natural Science Foundation of China (Grants 71531010, 71172131, 71325003) and Program of Shanghai Subject Chief Scientist (Grant 15XD1502000).

References

1. Schweigler LM *et al.* Forecasting models of emergency department crowding. *Academic Emergency Medicine* 2009; **16**(4):301–308.
2. Davidson SJ, Koenig KL, Cone DC. Daily patient flow is not surge: "management is prediction". *Academic Emergency Medicine* 2006; **13** (11):1095–1096.
3. Mccarthy ML *et al.* The challenge of predicting demand for emergency department services. *Academic Emergency Medicine* 2008; **15**(4):337–346.
4. Wargon M, Casalino E, Guidet B. From model to forecasting: a multicenter study in emergency departments. *Academic Emergency Medicine* 2010; **17** (9):970–978.
5. Marcilio I, Hajat S, Gouveia N. Forecasting daily emergency department visits using calendar variables and ambient temperature readings. *Academic Emergency Medicine* 2013; **20**(8):769–777.
6. Jones SS *et al.* Forecasting daily patient volumes in the emergency department. *Academic Emergency Medicine* 2008; **15**(2):159–170.
7. Sun Y *et al.* Forecasting daily attendances at an emergency department to aid resource planning. *BMC Emergency Medicine* 2009; **9**(1):1.
8. Kadri F *et al.* Time series modelling and forecasting of emergency department overcrowding. *Journal of Medical Systems* 2014; **38**(9):1–20.
9. Abraham G, Byrnes GB, Bain CA. Short-term forecasting of emergency inpatient flow. *IEEE Transactions on Information Technology in Biomedicine* 2009; **13**(3):380–388.
10. Boyle J *et al.* Predicting emergency department admissions. *Emergency Medicine Journal* 2012; **29**(5):358–365.
11. Nelder JA, Wedderburn RWM. Generalized linear models. *J Roy Stat Soc Ser A* 1972; **135**(3):370–384.
12. Mccullagh P, Nelder JA, Mccullagh P. Generalized Linear Models. Chapman and Hall: London, 1989.
13. George Box. Time Series Analysis: Forecasting & Control, 3/e. Pearson Education India, 1994.
14. Hamilton JD. Time Series Analysis. Princeton university press: Princeton, 1994.
15. Chatfield C. Time-Series Forecasting. CRC Press: Boca Raton, Florida, 2000.
16. Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 2003; **50**:159–175.
17. Cadenas E, Rivera W. Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA–ANN model. *Renewable Energy* 2010; **35** (12):2732–2738.
18. Liu H, H-q T, Y-f L. Comparison of two new ARIMA–ANN and ARIMA–Kalman hybrid methods for wind speed prediction. *Applied Energy* 2012; **98**:415–424.
19. Areekul P *et al.* Notice of violation of IEEE publication principles a hybrid ARIMA and neural network model for short-term price forecasting in deregulated market. *IEEE Transactions on Power Systems* 2010; **25**(1):524–530.
20. Burkom HS, Murphy SP, Shmueli G. Automated time series forecasting for biosurveillance. *Statistics in Medicine* 2007; **26**(22):4202–4218.

Authors' biographies

Qinneng Xu received his Bachelor's degree in Automation from Zhejiang University, China, in 2010. He is currently pursuing the Ph.D. degree in the Department of Systems Engineering and Engineering Management at the City University of Hong Kong. His research interests include data mining and statistics, especially their application to real problems.

Professor Kwok-Leung Tsui is Head and Chair Professor of Industrial Engineering in the Department of Systems Engineering and Engineering Management at the City University of Hong Kong and the founder and Director of Center for Systems Informatics Engineering. Prior to joining City University of Hong Kong, professor Tsui was a professor at the School of Industrial and Systems Engineering at the Georgia Institute of Technology. Professor Tsui was a recipient of the National Science Foundation Young Investigator Award. He is a fellow of the American Statistical Association, American Society for Quality, and International Society of Engineering Asset Management; a U.S. representative to the ISO Technical Committee on Statistical Methods. Professor Tsui was a Chair of the INFORMS Section on Quality, Statistics, and Reliability and the Founding Chair of the INFORMS Section on Data Mining. Professor Tsui's current research interests include data mining, surveillance in healthcare and public health, prognostics and systems health management, calibration and validation of computer models, process control and monitoring, and robust design and Taguchi methods.

Wei Jiang is the head and distinguished professor of Operations Management at Antai College of Economics and Management, Shanghai Jiao Tong University. Prior to joining Shanghai Jiao Tong University, he worked in AT&T Labs, Stevens Institute of Technology, and Hong Kong University of Science and Technology for more than 20 years. His research interests include big data analytics and innovation, Industry 4.0 and quality management, logistics and operations management, and information finance and risk management. He was a recipient of NSF CAREER award in 2006 and China National Funds for Distinguished Young Scientists award from NSFC in 2013. He was recognized as an expert of the "1000 Talents Program" in 2012 and "Program of Shanghai Subject Chief Scientist" in 2015 by Shanghai Municipal government. He is currently serving as Associate Editor for *Naval Research Logistics, IIE Transactions, Decision Sciences Journal*, etc.

Hainan Guo received the B.S. degree in Automation from the Northeastern University, Shenyang, China, in 2011 and the M.S. degree in Systems Engineering from the Northeastern University, Shenyang, China, in 2013. She is currently pursuing the Ph.D. degree in the Department of Systems Engineering and Engineering Management, City University of Hong Kong, Hong Kong. She is a student member of INFORMS.