Forecasting Passenger Numbers in Saudi Arabian Airlines Flights

Said Ali Hassan¹ And Abdulaziz Tariq Qadi²

¹(Professor, Department of Operations Research and Decision Support, Faculty of Computers and Information, Cairo University, Egypt)
²(HR Consultant, Saudi Arabian Airlines, Jeddah, Kingdom of Saudi Arabia)

Abstract: Forecasting Is A Prediction Of Future After Studying The Past; It Is A Planning Tool That Helps Management For Budgeting, Planning, And Estimating Future Expectations. Exploring The Pattern Of The Data Is The Way To Select Appropriate Forecasting Techniques To The Data. Time Series Data For Passenger In Saudi Arabian Airlines Were Collected In Hijri And Gregorian Calendars, The Data Were Monthly For Ten Years In Gregorian, However, Hijri Data Were Only For About Three Years, The Data Was Transformed From Gregorian To Hijri To Complete The Missing Gap. The Data For Both Calendars Are Of Trend Pattern And Moderate Seasonality. Eight Suitable Time Series Forecasting Models Available In Minitab And Excel Software Are Checked To Apply For Both Hijri And Gregorian Data. These Time Series Forecasting Models Are: Naïve For Trend And Seasonality, Double Exponential Smoothing, Seasonal Exponential Smoothing, Decomposition, Trend Models, ARIMA, Regression For Seasonal Data And Combining Forecasts. Since The Trend Equation With Smaller Measures Of Errors Is Curvilinear, Many Alternatives Of The Regression For Seasonal Data Method Are Proposed With Nonlinear Expressions Of The Time Independent Variable. The Nonlinear Expressions Are Different Power Terms And Also S-Curve Equation Of The Time Independent Variable. Each Forecasting Method Is Checked For Adequacy And For The Measures Of Forecasting Errors And The Best Method Is Used To Perform The Long-Term Forecasting For The Upcoming 5 Years. Keywords: Forecasting, Time Series, Gregorian And Hijri Calendars, Saudi Airlines Passengers.

I. Introduction

Air Passenger Forecasting Provides A Key Input Into Decisions Of Daily Operation Management And Infrastructure Planning Of Airports And Air Navigation Services, And For Aircraft Ordering And Design. Scarpe[1], Planning To The Future Is One Of The Most Important Keys To Success, Forecasting Is The Way. Air Transportation In Saudi Arabia Has Undergone Considerable Expansions And Developments In The Past Years. There Are 25 International And Domestic Airports. The Number Of Passengers (Arriving & Departing) Handled By All Airports Increased At A Considerable Annual Rate.

The Saudi Arabian Airlines (Saudia) Took The Operational Responsibility To Run The Air Transportation Business In The Kingdom, And Has Made A Vital Contribution To The Development Of The Kingdom. Saudials Constantly Endeavoring To Innovate, And To Plan Ahead For Improved Service To Customers. One Of The Improvements To The Current Business Is Forecasting And Analyzing The Air Travel Market.

This Study Try To Attempt The Forecasting Models In Order To Forecast Passenger Numbers In Saudi Arabian Airlines Flights. The Collected Data Were Monthly For Ten Years From January 2007G To July 2017G In Gregorian, However, Hijri Data Were Only From Muharram 1435H To Shawal 1438H, From Muharram 1428H To Dhu AL-Hijjah 1434H, Monthly Numbers Of Passengers Were Calculated.

II. Literature Review

In Airline Industry, There Are Many Research Areas And Applications With Multiple Perspectives And Different Goals. Most Familiar Areas Are: Forecasting Total Passengers In A Specific Airport, Right-Size’ An Airport, Passenger Trips Requested, Passenger Departures, Allocating Traffic Between Airports, Expected Passengers Between Two Cities, Air Passengers Between Two Countries, Passenger Demand At A State Level And The Number Of Domestic And International Airline Passengers. In Addition, Different Forecasting Methods Have Been Applied; The Forecasting Methods Can Be Divided Into Time Series Methods, Causal Methods And Judgmental Methods. All Methods Are Suitable For Different Purposes. Our Area Of Interest Is To Forecast The Number Of Passengers At A Month And Year Levels For A Specific Airline, Saudi Arabian Airlines (Saudia) In A Long-Term Basis.


Another Area Of Research Is About Passenger Trips Requested And Ridership. Passenger Trip Requests Include All Trips Completed, No-Shows, Cancellations And Trips Denied. Ridership Refers To Passenger Trips Completed. It Involves Detailed And Scientific Examination, Both At The System And Regional Levels, Of Trends And Movements In Trip Demand And Its Constitutive Elements Such As Cancellations, No-Shows, Missed Trips, And Trips Completed [8].


Over the coming 20 years, another study is to forecasting air transport passenger demand at a state level. For example, a paper [21] proposed an econometric dynamic model (EDM) to estimate the air transport passenger demand based on economic indicators. The case study is about the Mexican air transport industry identified for each predicted year.

Another study [22] employed the back-propagation neural network and genetic algorithm to forecast the air passenger demand in Egypt (international and domestic). Another research is about forecasting the number of domestic and international airline passengers in Saudi Arabia [23]. The method used is the neural network technique to forecast the number of passengers. Results indicated that the oil gross domestic product, population size, and per capita income were found to be the most contributing variables that affect the number of passengers in the Saudi Arabian airline sectors. Another study [24] tests the ability of several time-series models for predicting the monthly number of enplaned/deplaned air passengers in Canada for three market segments: domestic, transborder, and international flights.

### III. Data Collection

Monthly data for passenger numbers in Saudi airlines are collected for the last ten years; the data in the system are only available in Gregorian dates due to the nature of original inputs and the system setup. As an exception case, passenger numbers are provided in Hijri dates only from Muharram 1435H to Shawal 1438H, whereas for Gregorian dates, the data are available from January 2007G to July 2017G based on the request [25].

The Gregorian data are converted to Hijri to complete the missing gap using a special average daily number of passengers. Figure 1 shows the plots for the number of passengers in Gregorian and Hijri calendars. Both figures indicate an upward trend and a possible seasonality. From Fig. 1, it is clear that both time series have an upward trend. To explore the seasonality patterns for the time series data in Gregorian and Hijri calendars, the following methods are used [26]: scatter diagrams, autocorrelation functions, x² goodness-of-fit test, seasonality indexes using ratio to moving average, plot of changes in the seasonal pattern, and autoregression to assess strength of seasonality. New methods are proposed to assess the seasonality of the data and moreover to identify which months of the year have better seasonality. These methods are: month’s orders, measures of dispersion for seasonality indexes (range, quartile deviation, average deviation, and standard deviation). The results obtained from all the applied methods clarify that both data have a weak seasonality and the Gregorian data is slightly better in seasonality than Hijri data.

### IV. Basic Forecasting Techniques

Forecasting techniques are classified into three main divisions: time series, causal and judgmental. The time series forecasting models attempt to predict the future by using the historical data, causal (regression) forecasting models incorporate the variables or factors that might influence the quality being forecasting into the forecasting model. Judgmental forecasts or judgmental adjustments to forecasts are typically based on domain knowledge. Domain knowledge is any information relevant to the forecasting task other than time series data and in principle, more information should lead to better forecasts. Fig. 2 shows the forecasting techniques in Minitab and Excel software [27, 28] which will be used to solve the present problem. Table 1 shows the main characteristics of the suitable forecasting techniques for trend and seasonal time series data [29].
4.1. Naïve Forecasting Method For Trend And Seasonality

The Naïve Forecast Assumes That The Best Predictors Of Future Values Are The Most Recent Data Available. It Is Easy To Apply And Good Way To Start Thinking About Forecasting. This Method Is Used For A Very Short-Term Forecast Of An Event For Which There Is No Much Historical Data. It Can Be Used To Predict Only One Period Ahead For Non-Seasonal Data And One Year Ahead For Seasonal Data. The Method Has A Simplest Formula For Stationary Data And Other Formulas For Trend And Seasonal Data [30]. The Simplest And The Trend Formulas are Used To Predict Only One Period Ahead. Therefore, We Will Use The Trend And Seasonality Formula In Order To Be Able To Forecast One Year Ahead (12 Forecasting Periods).

The Naïve Technique Is Adjusted To Combine Seasonal And Trend Estimates Using The Following Naïve Formula for Monthly Data:

\[
\hat{Y}_{t+1} = \bar{Y}_{t+1} + \frac{(Y_t - \bar{Y}_t) + ... + (Y_{t-11} - \bar{Y}_{t-11})}{12} = Y_{t+1} + \frac{Y_t - Y_{t-12}}{12}
\]

The Formula Is Applied For Both The Gregorian And Hijri Data, Checking The Autocorrelation Functions Concludes That Residuals Are Not Random For Both The Gregorian And Hijri Data.

![Figure 2: Forecasting Techniques InMinitab And Excel Software.](image)

<table>
<thead>
<tr>
<th>Method</th>
<th>Pattern Of Data</th>
<th>Time Horizon</th>
<th>Type Of Model</th>
<th>Minimal Data Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve For Trend And Seasonality</td>
<td>T.S</td>
<td>S</td>
<td>TS</td>
<td>1 X S</td>
</tr>
<tr>
<td>Double Exponential Smoothing (Holt’s)</td>
<td>T</td>
<td>S</td>
<td>TS</td>
<td>4 X S</td>
</tr>
<tr>
<td>Seasonal Exponential Smoothing (Winter’s)</td>
<td>T.S</td>
<td>S</td>
<td>TS</td>
<td>2 X S</td>
</tr>
<tr>
<td>Classical Decomposition</td>
<td>T.S</td>
<td>S</td>
<td>TS</td>
<td>5 X S</td>
</tr>
<tr>
<td>Exponential Trend Models</td>
<td>T</td>
<td>1, L</td>
<td>TS</td>
<td>10</td>
</tr>
<tr>
<td>S-Curve Fitting</td>
<td>T</td>
<td>1, L</td>
<td>TS</td>
<td>10</td>
</tr>
<tr>
<td>GompertzModels</td>
<td>T</td>
<td>1, L</td>
<td>TS</td>
<td>10</td>
</tr>
<tr>
<td>Growth Curves</td>
<td>T</td>
<td>1, L</td>
<td>TS</td>
<td>10</td>
</tr>
<tr>
<td>ARIMA (Box-Jenkins)</td>
<td>ST, T, C, S</td>
<td>S</td>
<td>TS</td>
<td>24 X S</td>
</tr>
<tr>
<td>Regression With Time Series Data</td>
<td>T.S</td>
<td>1, L</td>
<td>TS</td>
<td>6 X S</td>
</tr>
</tbody>
</table>

Pattern Of Data: ST, Stationary; T, Trended; S, Seasonal; C, Cyclical.
**Time Horizon:** S. Short Term (Less Than Three Months); I, Intermediate; L, Long Term

**Type Of Model:** TS, Time Series; C, Causal.

**Seasonal:** S, Length Of Seasonality.

**Variable:** V, Number Variables.

### 4.2 Exponential Smoothing Methods

Exponential Smoothing Methods Are Based On Averaging (Smoothing) Past Values Of A Series In A Decreasing Exponential Manner, With More Weight Being Given To The More Recent Observations. Double Exponential Smoothing (Also Called Holt’s Method) Smooths The Data When A Trend Is Present. It Uses Two Constants (0<α and β< 1) For The Level And The Slope (Trend) And Is Better At Handling Trends. The Constants α, β and γ Can Be Selected Subjectively, A Grid Of Values Could Be Developed, Then Selecting The Ones Producing The Lowest Forecasting Errors [32]. A Common Approach In Minitab Software Is A Nonlinear Optimization Algorithm To Find Optimal Constants.

When The Data Show Trend And Seasonality, A Third Equation Is Introduced In Order To Take Care Of Seasonality. The Model Will Have Three Constants (0<α, β and γ< 1) To Deal With The Level, Trend And Seasonality. The Constants α, β and γ Can Be Selected Subjectively Or By Minimizing An Error Such As MSE.

There Are Two Variations To This Method That Differ In The Nature Of The Seasonal Component. The Additive Method Is Preferred When The Seasonal Variations Are Roughly Constant Through The Series, While The Multiplicative Method Is Preferred When The Seasonal Variations Are Changing Proportional To The Level Of The Series [33]. The Additive And Multiplicative Methods Are Applied For Both Gregorian And Hijri Data Using Minitab Software. The Additive Model Gives Smaller Measures Of Errors For All Values Of Parameters A, B And Γ. The Smallest Measures Of Errors Are Obtained For The Model With Parameters α, β and γ= 0.2, 0.1 and 0.1 Respectively For The Gregorian Data And For Parameters Of 0.4, 0.1 And 0.1 For The Hijri Data. The Autocorrelation Functions Of Residuals Show That Residuals Are Not Random For All Tested Values Of Parameters α, β and γ.

### 4.3 Classical Decomposition

Time Series Decomposition Involves Separating A Time Series Into Several Distinct Components. There Are Three Components That Are Typically Of Interest: T, A Deterministic, Nonseasonal Trend Component. This Component Is Sometimes Restricted To Being A Linear Trend, Though Higher-Degree Polynomials Are Also Used. S, A Deterministic Seasonal Component With Known Periodicity. This Component Captures Level Shifts That Repeat Systematically Within The Same Period (E.G., Month Or Quarter) Between Successive Years. I, A Stochastic Irregular Component. There Are Two Forms Of Classical Decomposition: An Additive Decomposition And A Multiplicative Decomposition. The Additive Model Treats The Time Series Values As A Sum Of The Components; It Works Best When The Seasonal Variation Is Relatively Constant Over Time. The Multiplicative Model Treats The Time Series Values As The Product Of The Components; It Works Best When The Seasonal Variation Increases Over Time [34].

Both The Multiplicative And Additive Models Are Applied For Both Gregorian And Hijri Data Using Minitab Software. The Additive Model Shows Smaller Measures Of Errors Than The Multiplicative Model. The Autocorrelation Functions For The Additive And Multiplicative Models For Both Gregorian And Hijri Data Show That Residuals Are Not Random.

### 4.4 Trend Forecasting Models

Time Series May Show Gradual Shifts Or Movements To Relatively Higher Or Lower Values Over A Longer Period. If A Time Series Plot Exhibits This Type Of Behavior, Then A Trend Pattern Exists. Trend Analysis Can Be Explored Using The Autocorrelation Function As Follows [35]:

- A Significant Relationship Exists Between Successive Time Series Values.
- The Autocorrelation Coefficients Are Large For The First Several Time Lags, And Then Gradually Drop Toward Zero As The Number Of Periods Increases.
- The Autocorrelation For Time Lag 1: Is Close To 1, For Time Lag 2: Is Large But Smaller Than For Time Lag 1.

Data That Increase By A Constant Amount At Each Successive Time Period Shows A Linear Trend. Data That Increase By Increasing Amounts At Each Successive Time Period Show A Curvilinear Trend. Exponential Trend Can Be Fitted When A Time Series Starts Slowly And Then Appears To Be Increasing At An Increasing Rate Such That The Percentage Difference From Observation To Observation Is Constant. It Is Achieved By Allowing The Level And The Slope To Be Multiplied Rather Than Added [36]. The S-Curve Model Fits The Pearl-Reed Logistic Trend Model. This Accounts For The Case Where The Series Follows An S-Shaped Curve. S-Shaped Curves Are Numerous And Different Not Only By Essence, But Mostly By Names.
[37] The Gompertz Model is one of the most frequently used sigmoid models fitted to growth data and other data. It is well known and widely used in many aspects of biology. It has been frequently used to describe the growth of animals and plants, as well as the number or volume of bacteria and cancer cells [38].

Both figures of the autocorrelation functions for the Gregorian and Hijri data are closely similar and show that there exists a significant relationship between successive time series values; there is a trend pattern for both the data in Gregorian and Hijri calendars.

Minitab software has these types of trend models: linear, quadratic, exponential, and S-shaped curves. The comparison between the basic measures of forecasting errors for the trend methods for Hijri and Gregorian data is shown in Table 2.

Table 2: Measures of Errors for the Trend Methods.

<table>
<thead>
<tr>
<th>Forecasting Trend Methods</th>
<th>Measurement of Forecasting Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hijri</td>
</tr>
<tr>
<td></td>
<td>MAPE</td>
</tr>
<tr>
<td>Linear</td>
<td>8.15899</td>
</tr>
<tr>
<td>Quadratic</td>
<td>7.68115</td>
</tr>
<tr>
<td>Exponential</td>
<td>7.71733</td>
</tr>
<tr>
<td>S-Curve</td>
<td>7.66434</td>
</tr>
</tbody>
</table>

From the measures of forecasting errors, the quadratic trend model has two smallest measures out of three for the Hijri data, while the S-Curve has two smallest measures out of three for the Gregorian data.

4.5 Autoregressive Integrated Moving Average (ARIMA)

The Box-Jenkins autoregressive integrated moving average (ARIMA) method is one of the more complicated methods used to forecast data. It is popular because it can forecast for all data patterns, and because it is often one of the most accurate forecasting tools available. It does not assume any particular pattern in the historical data. The model fits well if the residuals are small and randomly distributed. If the specified model is not satisfactory, the process is repeated using a new model, and the final model is used for forecasting.

ARIMA command appears as ARIMA(p, d, q) where p is the order of the autoregressive part (AR), d is amount of differencing (integration, I) and q is the order of the moving average (MA). The choice of the parameters will depend on the patterns in the autocorrelation functions (ACF) and the partial autocorrelation functions (PACF).

An ARIMA model including seasonal parameters is written as ARIMA(p, d, q)(P, D, Q). It is sometimes called SARIMA, the S stands for the word seasonal. The second set of parameters refers to the seasonal autoregression (P), seasonal differencing (D), and seasonal moving average (Q) [39].

Since the data of the no. of passengers for both Gregorian and Hijri calendars are non-stationary (trend), thus, differencing process is required and the data seems to be stationary after one lag differencing. Figs. 3 and 4 show graphs for the autocorrelation function (ACF) and partial autocorrelation function (PACF) for Gregorian and Hijri data after one lag differencing.

Figure 3: ACF and PACF for Gregorian data after one lag differencing.
Both the Autocorrelation and Partial Autocorrelation Function for the two types of data don’t give clear ARIMA models. For Gregorian data, the Autocorrelation function has spikes at lags: 1, 6, 12 and 24, and the Partial Autocorrelation function has spikes at lags: 1, 6, 7, 9, 10, 11 and 12. So a possible suitable ARIMA model is (1,1,1)(1,0,2). For Hijri data, the Autocorrelation function has spikes at lags: 1, 6 and 12, and the Partial Autocorrelation function has spikes at lags: 1, 6, 7 and 10. So a possible suitable ARIMA model is (1,1,1)(0,0,1). Other ARIMA models are tested also, and results are compared with respect to MS values and Autocorrelation function of residuals. Minitab software is used to carry out the calculations, but because the number of passengers are large, the results cannot be converged. As a solution, the number of passengers is divided by 1000 and then corresponding results concerning errors and forecasting will be multiplied by 1000.

From the results, it is concluded that the best ARIMA model with the smallest MS value and with random residuals is (1,1,3)(1,0,3) for Gregorian data and (1,1,1)(1,0,3) for Hijri data. Figure 5 presents the Autocorrelation functions for residuals of the best Gregorian and Hijri models.

**4.6 Using Regression To Forecast Seasonal Data**

A regression model can be used in order to forecast the seasonal data. The model is closely aligned with an additive decomposition is considered. In this model, the seasonality is handled by using dummy variables in the regression function. A seasonal model for monthly data with a time trend is:

\[
Y_t = \beta_0 + \beta_1 t + \beta_2 S_2 + \beta_3 S_3 + \ldots + \beta_{12} S_{12} + \epsilon_t
\]

where:
- \(Y_t\) = The variable to be forecast,
- \(T\) = The time index,
- \(S_2\) = A dummy variable that is 1 for the second month of the year; 0 otherwise,
- \(S_3\) = A dummy variable that is 1 for the third month of the year; 0 otherwise,
- \(S_{12}\) = A dummy variable that is 1 for the twelfth month of the year; 0 otherwise,
- \(\epsilon_t\) = The errors, assumed to be independent and normally distributed with mean zero and a constant variance,
- \(\beta_0, \beta_1, \beta_2, \beta_3, \ldots, \beta_{12}\) = The coefficients to be estimated.
The Twelve Months Are Described With Only Eleven Dummy Variables; The First Month Is Handled By The Intercept Term $B_0$ In Equation (A1). To See This, For First Month Data, $S_2 = S_3 = \ldots = S_{12} = 0$, And The Expected Level Is:

$$E(Y_t) = B_0 + B_1t$$

For Second-Month Data, $S_2 = 1$ And $S_3 = S_4 = \ldots = S_{12} = 0$, Where The Expected Level Is:

$$E(Y_t) = B_0 + B_1t + B_2$$

Similar Patterns Emerge For The Third, Forth, \ldots, And Twelfth Months. The Forecast For The First Month Of An Upcoming Year Can Be Computed By Putting $S_2$, $S_3$, \ldots, $S_{12} = 0$ As Follows:

$$Y_t = B_0 + B_1t$$

Where $t$ Is The Serial Number Of The Period To Be Forecasted.

An Examination Of Residuals Is A Crucial Component, It Is Important To Compute The Residual Autocorrelations To Check The Independence Assumption. Lack Of Independence Can Drastically Distort The Conclusions Drawn From $T$ Tests. Moreover, The Independence Assumption Is Particularly Risky For Time Series Data [29].

The Minitab Software Is Used For The Gregorian Data, The Results Show That The F Statistic And Its P-Value (0.000) Clearly Indicate That The Regression Is Significant, But The T Statistic For X2, X4, X6, X9, X10 And X12 are Small With Relatively Large P-Value (> 0.05). Therefore, The Regression Is Significant, But Each Of These Predictor Variables Are Not Significant. Then, The Regression Model Cannot Be Used In This Manner; There Are Three Methods To Deal With This Situation: Regression Method, All-possible Regressions (Best Subsets Regression) And Stepwise Regression [29]. The Same Procedure Is Repeated For The Hijri Data. Table 8 Shows The Dummy Variables For Hijri Data For The Three Methods.

**Regression Diagnostics And Residual Analysis**

An Examination Of The Residuals Is A Crucial Component For The Determination Of Model Adequacy. The Two Assumptions For The Regression Model Are:

- The Errors Are Normally Distributed.
- The Errors Should Be Independent.

All The Regression Equations For Both Gregorian And Hijri Data Are Checked To Perform Diagnostics Of The Regression And Analyze The Residuals. The Histogram Of The Residuals Provides A Shape Check On The Normality Assumptions; Fig. 6 Shows An Example Of The Normality Test Of Residuals For The Regression Method For The Gregorian And Hijri Data.

![Image](image.png)

**Figure 6:** Normality Test Of Residuals For Regression Method For Gregorian And Hijri Data.

Anderson-Darling Normality Test Is Used To Determine If Data Follow A Normal Distribution. If The P-Value Is Lower Than A Pre-Determined Level Of Significance (e.g. < 0.05), The Data Do Not Follow A Normal Distribution [27]. From Figure 6 and According To The P-Values, Then The Hijri Data Residuals Are Normally Distributed But That For The Gregorian Data Are Not. For Randomness, The Sample Autocorrelations Of The Residuals Should All Be Small For All Time Lags $k = 1, 2, \ldots, k = n/4$, Where $n$ Is The Number Of Residuals.

The Number Of Residuals For Gregorian Data Are 127, So $k = n/4 = 127/4 = 31.75$ (Round It Up To Be 32), And For Hijri Data Are 130, So $k = n/4 = 130/4 = 32.5$ ( Rounded It Up To Be 33). The Autocorrelation Functions Of Residuals For Gregorian And Hijri Data Show That Residuals Are Not Random. Table 3 Represents The Results For Normality And Randomness Tests For Different Regression Methods.
Forecasting Passenger Numbers In Saudi Arabian Airlines Flights

Table 3: Normality And Independence Tests Of Residuals For Different Regression Methods.

<table>
<thead>
<tr>
<th>Data</th>
<th>Method</th>
<th>Residuals</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Normality</td>
<td>Independence</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Test</td>
<td>Test</td>
<td></td>
</tr>
<tr>
<td>Gregorian</td>
<td>Regression Method</td>
<td>×</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Best Subset</td>
<td>√</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stepwise Regression</td>
<td>√</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>Hijri</td>
<td>Regression Method</td>
<td>√</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Best Subset</td>
<td>√</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stepwise Regression</td>
<td>√</td>
<td>×</td>
<td></td>
</tr>
</tbody>
</table>

4.7 Using Regression With Nonlinear Expressions Of Time Variable

Since the trend equation with smaller measures of errors is curvilinear, many alternatives of the regression for seasonal data method are proposed with nonlinear expressions of the time independent variable. The nonlinear expressions are different power terms and also S-Curve equation of the time independent variable. The nonlinear expressions T, T², T³, …T⁶ and the S-Curve equation are considered. While the seasonality is handled by using dummy variables in the regression function as in the previous model. The proposed seasonal model for monthly data with a quadratic time trend will be as follows:

\[ Y_t = \beta_0 + \beta_1 t + \beta_1 t^2 + \beta_2 S_2 + \beta_3 S_3 + \ldots + \beta_{12} S_{12} + \epsilon_t \]

The proposed seasonal model for monthly data with a time trend with power 3 will be as follows:

\[ Y_t = \beta_0 + \beta_1 t + \beta_1 t^2 + \beta_1 t^3 + \beta_2 S_2 + \beta_3 S_3 + \ldots + \beta_{12} S_{12} + \epsilon_t \]

Other time trend with t, t², t³ are also checked.

Also, a regression model with S-Curve expression of time variable is checked. The equations for the S-curves for both the Gregorian and Hijri data are obtained from the trend equations as follows:

Gregorian: \[ Y(t) = \frac{10^7}{0.952042 + 8.09059 \times (0.991933 t)} \]

Hijri: \[ Y(t) = \frac{10^7}{-2.16360 + 11.3278 \times (0.995106 t)} \]

Table 4: Forecasting Performance Of The Best Regression Equations

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>0.66532</td>
<td>98498</td>
<td>1.6E+10</td>
<td>×</td>
<td>0.66022</td>
<td>100190</td>
<td>1.71E10</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T, T² &amp; T³</td>
<td>0.04103</td>
<td>94688</td>
<td>1.51E+10</td>
<td>×</td>
<td>0.05869</td>
<td>979785</td>
<td>1.62E10</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T, T², T³ &amp; T⁴</td>
<td>0.04252</td>
<td>74954</td>
<td>9.68E09</td>
<td>×</td>
<td>0.847219</td>
<td>819455</td>
<td>1.19E10</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T, T², T³ &amp; T⁴</td>
<td>The Same Results As T, T² And T³</td>
<td>×</td>
<td></td>
<td></td>
<td>The Same Results As T, T² And T³</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T, T², T³ &amp; T⁴</td>
<td>0.043248</td>
<td>74939</td>
<td>9.65E09</td>
<td>×</td>
<td>0.047742</td>
<td>82472</td>
<td>1.24E10</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T, T², T³ &amp; T⁴</td>
<td>0.043316</td>
<td>75028</td>
<td>9.67E09</td>
<td>×</td>
<td>0.050529</td>
<td>82616</td>
<td>1.23E10</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-Curve</td>
<td>0.058884</td>
<td>100159</td>
<td>1.64E10</td>
<td>×</td>
<td>1.64E10</td>
<td>0.058408</td>
<td>99588</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.8 Combining Forecasts

A developing branch of forecasting study involves the combination of two or more forecasting methods to produce the final forecasts. Considerable literature has accumulated over the years regarding the combination of forecasts. The primary conclusion of this line of research is that forecast accuracy can be substantially improved through the combination of multiple individual forecasts. Furthermore, simple combination methods often work reasonably well relative to more complex combinations. The gains in accuracy from combining increased with the numbers of forecasts used, especially when these forecasts were based on different methods and different data, and in situations involving high levels of uncertainty [40].
In practice, the method most commonly used is the simple averaging of the forecasts [41]. If \( \hat{y}_{11}, \hat{y}_{12}, \ldots, \hat{y}_{1m} \) are one-step-ahead forecasts produced by \( M \) different methods, then the one-step-ahead combined forecast, \( \hat{y}_{1C} \), obtained by simple averaging is:

\[
\hat{y}_{1C} = \frac{1}{m} (\hat{y}_{11} + \hat{y}_{12} + \ldots + \hat{y}_{1m})
\]

The common practice, however, is to obtain a weighted average of forecasts, with the weights adding up to unity. In this case, each one-step-ahead forecast, \( \hat{y}_{1i} \), receives weight \( 0 < w_i < 1 \) where \( \sum_{i=1}^{m} w_i = 1 \).

Therefore, with \( M \) one-step-ahead forecasts produced by \( M \) different methods, the combined forecast \( \hat{y}_{1C} \) is:

\[
\hat{y}_{1C} = w_1 \hat{y}_{11} + w_2 \hat{y}_{12} + \ldots + w_m \hat{y}_{1m}
\]

Past errors of each of the original forecasts are used to determine the weights to attach to these two original forecasts in forming the combined forecasts, one would wish to give greater weight to the set of forecasts which seemed to contain the lower (mean-square) errors, i.e., set weights proportional to the inverse of the models’ MSE-values [42]:

\[
w_i = \frac{1}{\sum_{i=1}^{m} MSE_i^{-1}}
\]

This method is applied to the problem by combining the two most promising forecasting methods. Those methods are: ARIMA\((1,1,1)(1,0,3)\) for Hijri data (which is the only method having random residuals) with the regression method with nonlinear expressions of time variable \( T, T^2, T^3, T^4 \) & \( T^5 \) (which has the smallest measures of errors for Hijri data, but errors are not random). The combination is done by considering different weights as shown in Table 5. From the table, measures of errors in the best combining forecasts with equal weights are smaller than both combined methods, but residuals of that method are not random; moreover, the residuals for all the combined models are correlated.

<table>
<thead>
<tr>
<th>Forecasting Methods</th>
<th>Hijri Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA</td>
<td>Regression</td>
</tr>
<tr>
<td>0.1</td>
<td>0.9</td>
</tr>
<tr>
<td>0.2</td>
<td>0.8</td>
</tr>
<tr>
<td>0.3</td>
<td>0.7</td>
</tr>
<tr>
<td>0.32</td>
<td>0.68</td>
</tr>
<tr>
<td>0.34</td>
<td>0.66</td>
</tr>
<tr>
<td>0.36</td>
<td>0.64</td>
</tr>
<tr>
<td>0.38</td>
<td>0.62</td>
</tr>
<tr>
<td>0.4</td>
<td>0.6</td>
</tr>
<tr>
<td>0.42</td>
<td>0.58</td>
</tr>
<tr>
<td>0.44</td>
<td>0.56</td>
</tr>
<tr>
<td>0.46</td>
<td>0.54</td>
</tr>
<tr>
<td>0.47</td>
<td>0.53</td>
</tr>
<tr>
<td>0.48</td>
<td>0.52</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>0.6</td>
<td>0.4</td>
</tr>
<tr>
<td>0.7</td>
<td>0.3</td>
</tr>
<tr>
<td>0.8</td>
<td>0.2</td>
</tr>
<tr>
<td>0.9</td>
<td>0.1</td>
</tr>
<tr>
<td>0.46757</td>
<td>0.52243</td>
</tr>
</tbody>
</table>

Table 5: Measures of Errors for the Combining Forecasts Method.
V. Results

A Comparison Between Measures Of Forecasting Errors For The Used Forecasting Methods Is Presented In Table 6. From The Measurement Of Forecasting Errors, It Can Be Seen That Methods With Minimum Measures Of Errors And Uncorrelated Residuals Are ARIMA(1,1,3)(1,0,3) For Gregorian Data And ARIMA(1,1,1)(1,0,3) For Hijri Data. These Two Methods Will Be Used For Long-Term Forecasting Of The Number Of Passengers In Saudi Arabian Airlines. It Can Be Seen Also That All Measures Of Error In The Gregorian Data Are Smaller Than That Measures In The Hijri Data.

6. LONG TERM FORECASTING OF PASSENGER NUMBERS

Forecasting Using Gregorian Data Is More Suitable Than Forecasting With Hijri Data For These Reasons:

A) The Gregorian Data Are Actual Observations, While The Hijri Data Are Calculated Using The Daily Average Number Of Passengers For The Missing Part Of The Data.

Table 6: Performance Of The Forecasting Methods.

<table>
<thead>
<tr>
<th>Forecasting Methods</th>
<th>Measures Of Forecasting Errors</th>
<th>Gregorian</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAPE</td>
<td>MAD</td>
</tr>
<tr>
<td>Naive For Trend And</td>
<td>0.07364</td>
<td>131710</td>
</tr>
<tr>
<td>Seasonality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Double Exp. Smoothing</td>
<td>6.26151</td>
<td>1.0671E5</td>
</tr>
<tr>
<td>Seasonal Exp. Smoothing</td>
<td>4.36101</td>
<td>7.9016E4</td>
</tr>
<tr>
<td>Decomposition</td>
<td>6.09046</td>
<td>9.8804</td>
</tr>
<tr>
<td>Trend Quadratic</td>
<td>7.68115</td>
<td>125341</td>
</tr>
<tr>
<td>S.Curve</td>
<td>7.64334</td>
<td>128221</td>
</tr>
<tr>
<td>ARIMA (1,1,3)(1,0,3)</td>
<td>0.04352</td>
<td>73248</td>
</tr>
<tr>
<td>Regress. seasonal Data</td>
<td>0.04205</td>
<td>74954</td>
</tr>
<tr>
<td>Combining Forecasts</td>
<td>0.04000</td>
<td>59885</td>
</tr>
</tbody>
</table>

B) The Residuals Of The Gregorian Data For The Used Forecasting Method Are Clearly Random, But Those For The Hijri Data Are Not.

C) All The Three Measures Of Errors For The Forecasting Method In Gregorian Data Are Slightly Smaller Than Those Of The Hijri Data.

D) Saudi Arabian Airlines Already Uses The Gregorian Calendar For All The Reservation, Which Is Standard Calendar Around The World.

Minitab Software Is Used To Long-Term Forecast The Expected Number Of Passengers In Saudi Arabian Airlines For The Upcoming 5-Years (60 Months) For The Gregorian Calendar. Figures 7 And 8 Represent The Original Observations And Fitted Values And The 5 Years Forecasts Respectively.

Figure 7: Graph For The Original Observations And Fitted Values For The Gregorian Data.
VI. Conclusions

The Main Purpose Of This Paper Is To Long-Term Forecast The Number Of Passenger In Saudi Arabian Airline Flights. The Collected Data Were Monthly For Ten Years From January 2007G To July 2017G. In Gregorian, However, Hijri Data Were Only From Muharram 1435H To Shawal 1438H. The Gregorian Data Are Converted To Hijri To Complete The Missing Gap Using A Special Average Daily Number Of Passengers.

The Data For Both Calendars Are Of Trend Pattern And Moderate Seasonality. Nine Candidate Time Series Forecasting Models Available In Minitab And Excel Software Are Selected To Apply For Both The Gregorian And Hijri Data Of The Number Of Passengers In Saudi Airlines. These Time Series Forecasting Models Are: Naïve For Trend And Seasonality, Double Exponential Smoothing, Seasonal Exponential Smoothing, Decomposition, Trend Models, ARIMA, Regression For Seasonal Data And Combining Forecasts.

The Main Conclusions From This Research Cab Be Summarized As Follows:
1. The Autocorrelation Functions Indicate That Residuals Are Not Random For All The Used Forecasting Models Except For One Model Of ARIMA Method For Each Of The Gregorian And Hijri Data.
4. The Forecasting Method Having The Smallest Measures Errors And Random Residuals Is ARIMA(1,1,3)(1,0,3) For The Gregorian Data And ARIMA(1,1,1)(1,0,3) For The Hijri Data.
5. The Regression Method With Nonlinear Expressions Of Time Variable $t$, $t^2$, $t^3$, $t^4$ & $t^5$ For Hijri Data Has The Smallest Measures Of Errors, But Errors Are Not Random. This Method Is Combined With ARIMA(1,1,1)(1,0,3) By Considering Different Weights. Although The Measures Of Errors For The Resulted Combining Forecasts Are Smaller Than Measures Of Each Of The Combined Methods, But The Errors Are Not Random.
6. All The Three Measures Of Errors For The Best Forecasting Method In Gregorian Data Are Slightly Smaller Than Those Of The Hijri Data.
7. A Regression Model Can Be Used In Order To Forecast The Seasonal Data. In This Model, The Trend Is Handled By Adding An Independent Variable $T$ Representing The Time And The Seasonality Is Handled By Using Dummy Variables In The Regression Function.
9. Long-Term Forecasting For 5 Years (60 Months) Is Done Using ARIMA(1,1,3)(1,0,3) For The Gregorian Data And ARIMA(1,1,1)(1,0,3) For The Hijri Data. The Graph For The Fitted Values Shows A Good Match Between The Predicted And The Real Data, Which Reassure The Results Of The Future Forecasts.

References

Forecasting Passenger Numbers In Saudi Arabian Airlines Flights


[22] M. M. Mohie El-Din, M. S. FaragAnd A. A. Abouzeid, Airline Passenger Forecasting In EGYPT (Domestic And International), International Journal Of Computer Applications (0975 - 8887), Volume 165 - No.6, May 2017.

[23] Abdullah Omer Bafail, Applying Data Mining Techniques To Forecast Number Of Airline Passenger In Saudi Arabia (Domestic And International Travels), Journal Of Air Transportation, Volume 9, Issue Number 1, Pp.100-115, Aviation Institute, 2004. ISSN: 1544-6980


