Task B Report - Time Series Forecasting

SID: 490411065

1 Exploratory Data Analysis

The *Visitors* data set contains the monthly number of visitors to a country from January 1991 up till December 2016. We would like to produce 24-month forecasts for the number of visitors using different time series models.

Below is a quick summary of the data. Note that it only contains one variable (the historical number of visitors by month):

	Number of Visitors
count	312
mean	419407.3718
std	132443.0593
min	161400
median	412950
max	971800

Figure 1.1

We have 312 months worth of data, and the number of visitors ranges from 161,400 up to 971,800. The mean is a bit higher than the median, which could indicate positive skew in the number of visitors in recent years.

In any case, the best way to get a feel for the data is to plot it. We apply a decomposition algorithm from the `statsmodels` library to decompose the data into different time series components:

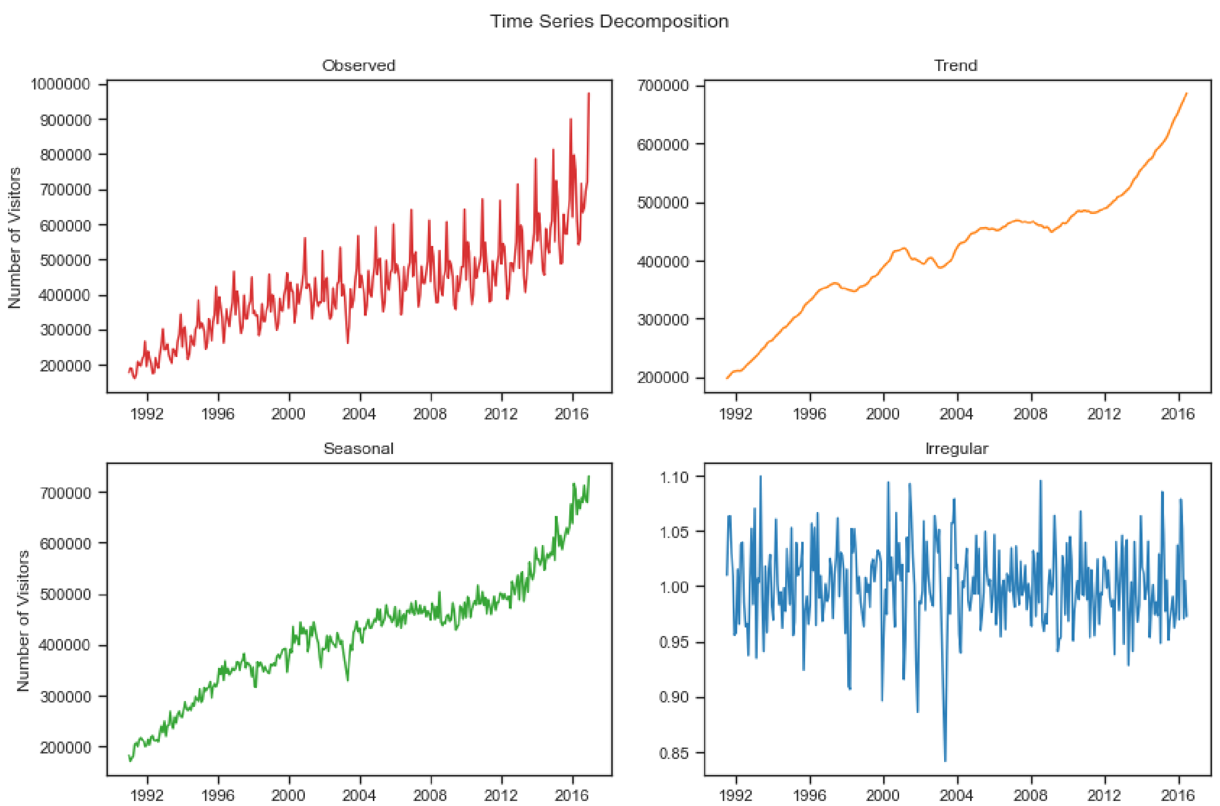


Figure 1.2

The 'observed' plot is just a plot of our time series data. It clearly shows a periodic, upward trend in the number of visitors. This is reflected in the 'trend' and 'seasonal' plots, which try to separate out these components. Additionally, the size of the seasonal effect appears to increase over time, because the spikes become more jagged in the 'observed' plot. This suggests that a multiplicative model may model the data well. Finally, the noise—represented by the 'irregular' plot—is quite random and contributes at most ~15% to the deviation of the observed values from what we might have expected them to be.

2 Modelling (MHW)

We saw in our EDA that the time series data exhibited an upwards trend and also yearly seasonal effects that increased with time. It therefore makes sense to model the data using the multiplicative Holt-Winters (MHW) method.

2.1 Methodology

The MHW method using a seasonal length of 12 (months) produces forecasts using the equations below:

$$\begin{aligned} \text{forecasts} \quad \hat{y}_{t+1} &= (l_t + b_t) \times S_{t+1-12} \\ \text{level} \quad l_t &= \alpha(y_t / S_{t-12}) + (1-\alpha)(l_{t-1} + b_{t-1}) \\ \text{trend} \quad b_t &= \beta(l_t - l_{t-1}) + (1-\beta) b_{t-1} \\ \text{seasonal} \quad S_t &= \delta(y_t / l_t) + (1-\delta) S_{t-12} \end{aligned}$$

The parameters α, β, δ are estimated by the 'L-BFGS-B' optimisation algorithm, which is implemented using the `minimize()` function from the `scipy.optimize` package. The output after estimating the parameters is shown below:

```
Multiplicative Holt-winters exponential smoothing

Smoothing parameters:
alpha (level)    0.310 (0.078)
beta  (trend)    0.012 (0.009)
delta (seasonal) 0.362 (0.050)

In-sample fit:
MSE               434674463.731
RMSE              20848.848
Log-likelihood   -3545.566
AIC               7099.131
BIC               7114.103
```

Figure 2.1.1

These parameters result in the smoothed series below. It captures the trend and seasonality effects well and thus follows the real data quite closely:

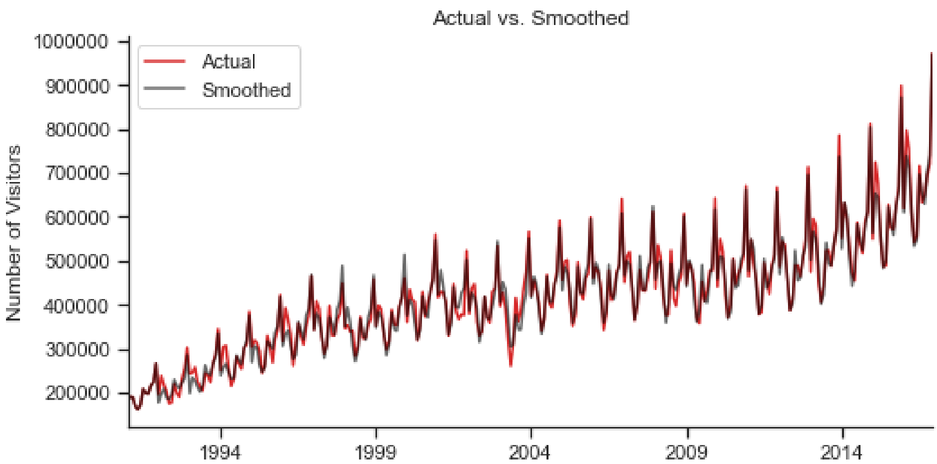


Figure 2.1.2

2.2 Model Diagnostics

Our model assumes that the errors $\varepsilon_t \sim N(0, \sigma^2)$ are independent. Below is a plot of our model's residuals and an autocorrelation plot of these residuals:

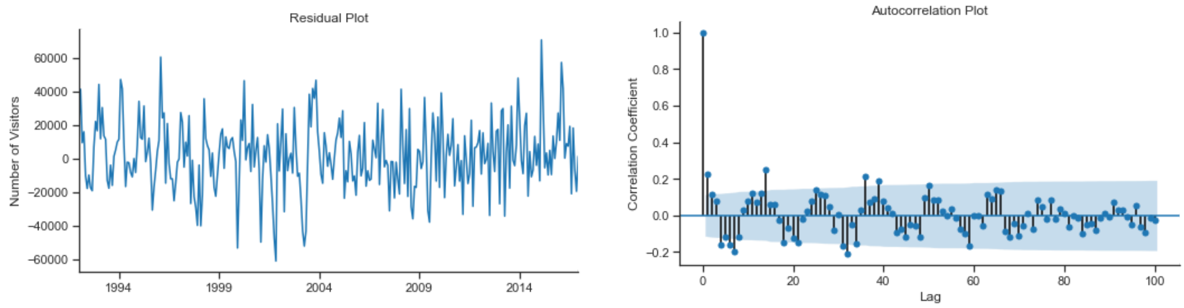


Figure 2.2.1

The residual plot suggests that the residuals exhibit equal variance (σ^2), so we conclude that this assumption is satisfied. With the autocorrelation plot, there are some values of lag that show correlations of roughly $\pm$ 0.2. But apart from these, there doesn't seem to be much correlation between the residuals and the lagged versions of themselves. This is important because it suggests that the errors are random and thus *independent* of each other.

Below is the histogram and the normal Q-Q plot of the residuals:

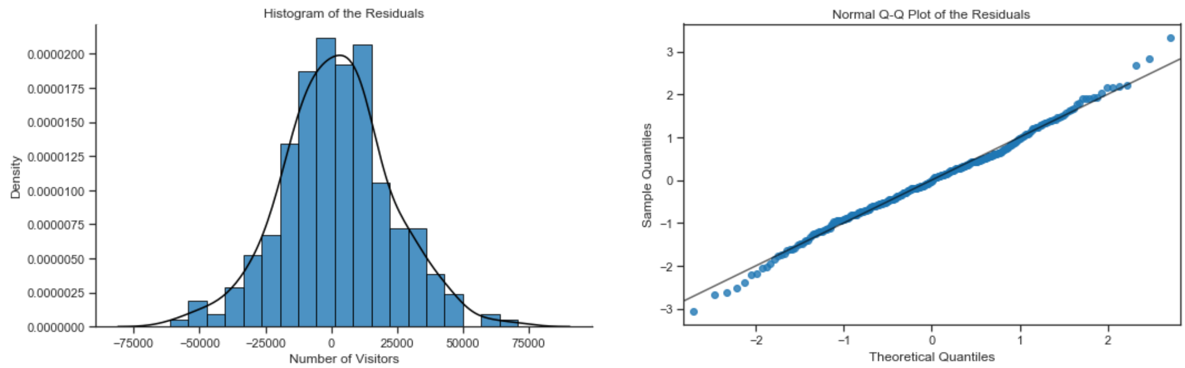


Figure 2.2.2

The histogram resembles a normal distribution that is centred at 0. The Q-Q plot confirms this as the residual quantiles line up with theoretical normal quantiles quite well.

Our plots together summarise that the errors independently follow a $N(0, \sigma^2)$ distribution. Hence, our model satisfies the assumptions that it makes about the errors.

3 Modelling (MHW with damping)

The 'observed' plot in our EDA doesn't teeter off at the end and therefore suggests that there may be no need for damping. However, we would like to see if it helps anyway, because no other models (simple exponential smoothing, additive Holt-Winters, etc.) make sense with our data.

3.1 Methodology

The forecast equations for a damped MHW model are almost the same as the non-damped model, except that we multiply the trend terms (b_t) by a damping parameter ϕ:

$$
\begin{aligned}
\text{forecasts} \quad \hat{y}_{t+1} &= (l_t + b_t) \times S_{t+1-12} \\
\text{level} \quad l_t &= \alpha(y_t / S_{t-12}) + (1-\alpha)(l_{t-1} + \phi b_{t-1}) \\
\text{trend} \quad b_t &= \beta(l_t - l_{t-1}) + (1-\beta)\phi b_{t-1} \\
\text{seasonal} \quad S_t &= \delta(y_t / l_t) + (1-\delta) S_{t-12}
\end{aligned}
$$

These parameters are also estimated by the 'L-BFGS-B' optimisation algorithm using the `minimize()` function from the `scipy.optimize` package. We obtain the estimated parameters below:

```
Multiplicative Holt-winters exponential smoothing (damped trend)

Smoothing parameters:
alpha (level)    0.482 (0.077)
beta  (trend)    0.000 (0.029)
delta (seasonal) 0.412 (0.086)
phi (damping)    0.792 (0.054)

In-sample fit:
MSE               445818494.913
RMSE              21114.414
Log-likelihood   -3549.515
AIC               7109.029
BIC               7127.744
```

Figure 3.1.1

What's interesting here is that the β parameter is now 0. From the equations above, this means that the trend does not depend on the level anymore; only on the previous values of the trend. We see that this model also tracks the real data quite closely:

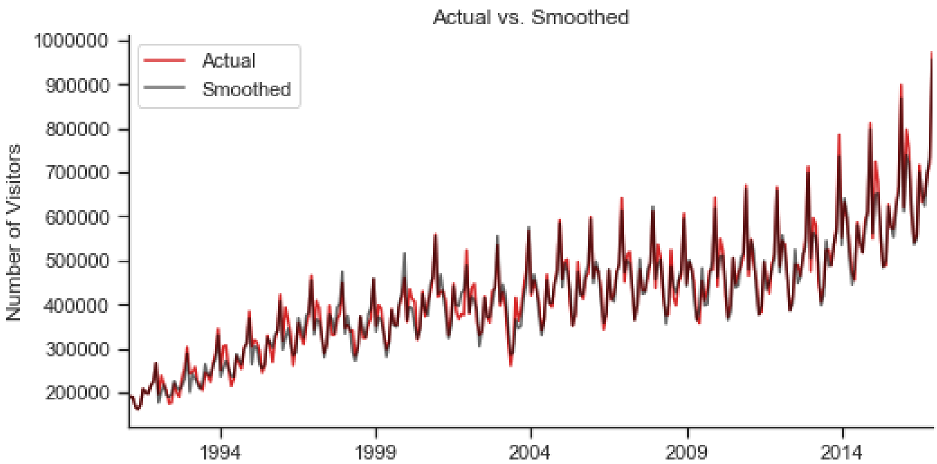
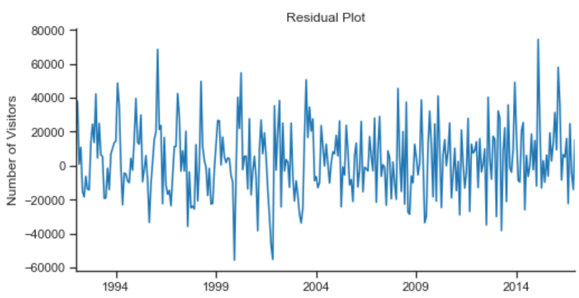


Figure 3.1.2

3.2 Model Diagnostics

Again, our model assumes that the errors follow a $N(0, \sigma^2)$ distribution and are independent of each other. Below is a residual plot and an autocorrelation plot of the MHW model with damping:



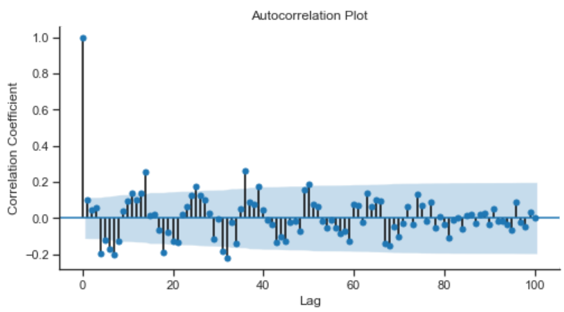


Figure 3.2.1

The residuals appear to satisfy the equal variance assumption as we don't see any fanning going on in the residual plot. With the autocorrelation plot, there is again little correlation between the residuals and the lagged versions of themselves. We conclude that the errors are independent of each other.

Finally, we look at the histogram and the normal Q-Q plot of the residuals:

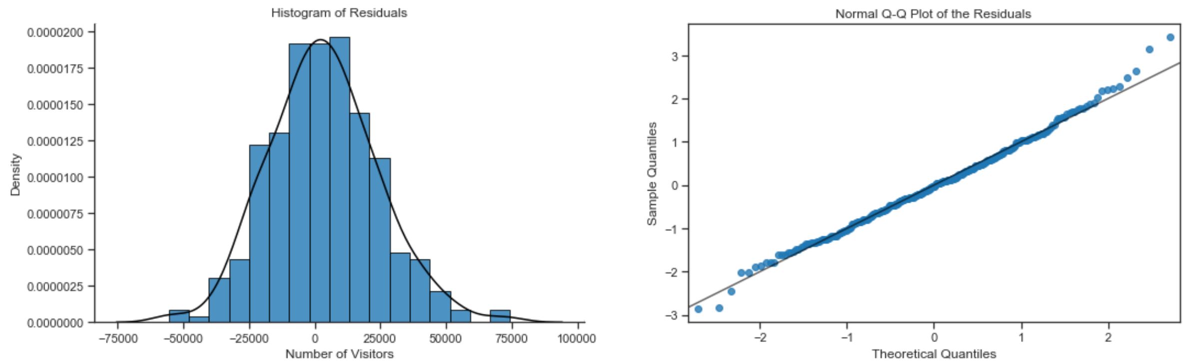


Figure 3.2.2

The histogram resembles a normal distribution that is centred at 0. The Q-Q plot confirms this as the residual quantiles line up with theoretical normal quantiles quite well. There is some departure from the line at the ends, but this is not too severe.

Our plots together summarise that the errors independently follow a $N(0, \sigma^2)$ distribution. Just like the MHW model without damping, this model also satisfies the assumptions that it makes about the errors.

4 Results

4.1 Model Validation

Our validation procedure is as follows:

1. Estimate the parameters of the model using the data from January 1991 – December 2009
2. Make a prediction for the number of visitors in January 2010
3. Re-estimate the parameters of the model using the data from January 1991 – January 2010
4. Make a prediction for the number of visitors in February 2010
5. Repeat until we have predictions for January 2010 – December 2016
6. Calculate RMSEs and SEs using our predictions and the actual values

We do this for both of our models. The results are collected in the table below:

	RMSE	**SE**
MHW	23485.9423	2702.3208
MHW with damping	24322.3280	2724.1950

Figure 4.1.1

We see that the RMSE for the model *without* damping is actually lower, which means that it performed better. Its empirical standard error is also lower, which means that its predictions are slightly more consistent. This was expected because our data didn't showcase a damping trend. However in the future, it might make more sense to use the model with damping because visitors to a country are finite.

4.2 Forecasting

We now use both of our time series models to forecast the number of visitors for the next 24 months (January 2017 – December 2018):

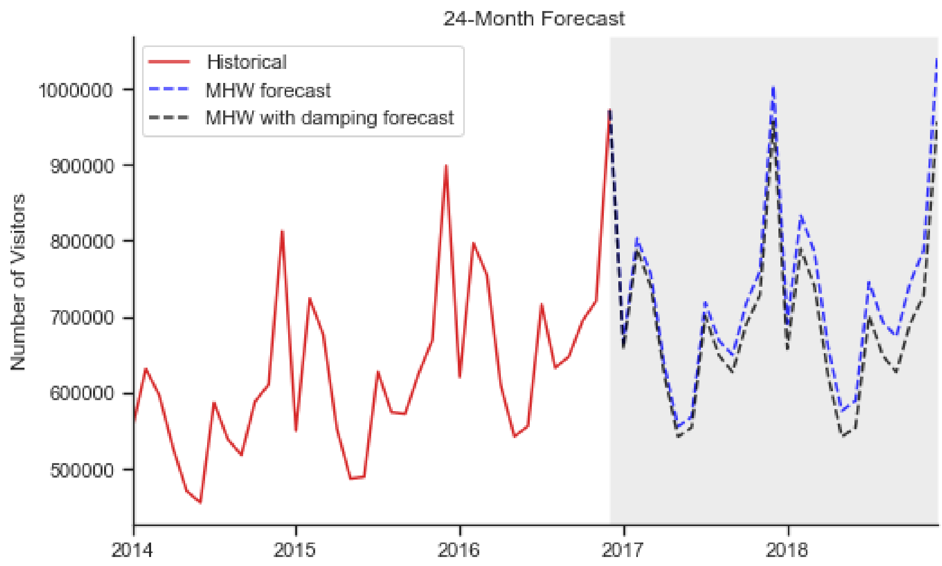


Figure 4.2.1

With both models, we see that the seasonality aspect of the original data is captured quite well in the forecasts. However, the MHW model with damping does not capture the trend as well as the standard MHW model. This reflects the trend parameter $\beta = 0$ that we saw from before.

As a result, we conclude that the MHW model *without* damping is the better model. It appears to capture the trend well in the forecast plot, and it also has better validation results.

The actual forecasted values are available to view in the appendix (Figure A6)

5 Conclusion

We used two different time series models to forecast the number of visitors to a country—the multiplicative Holt-Winters (MHW) model and the MHW model with a damping. From our validation and forecasting results, we concluded that the standard MHW model was the better model.

One limitation was that the data kept trending upwards even though we know that visitors are finite. If we were to redo our modelling a few years into the future, we may find that the damped MHW model works better. We also saw some slight departure in our residuals from the normal Q-Q line, which suggests that maybe we could bootstrap the error distribution instead of assuming normality.

Finally, it would be interesting to try more advanced time-series models such as ARIMA, which could potentially give more accurate forecasts. Not only that, we could factor in other time series data into our model building (multivariate time series), for example the stock market index. Our intuition tells us that visitors would be less likely to leave their own country when the global economy is in recession and the stock market is down. We could build innovative models by finding novel ways to include domain expertise into our analysis.