Forecasting Models

There are many ways to forecast the number of visitors, including random walk, simple exponential smoothing, trend exponential smoothing, Holt-winters smoothing, etc. In the following analysis, Holt-winters exponential smoothing method is used since "it extended trended corrected method to seasonal data and allows for both additive and multiplicative seasonality". Multiplicative Holt-winters smoothing is conducted for the number of visitors since it has a seasonality proportional to the level of the time series, while additive Holt-winters smoothing is conducted for the log-transformed number of visitors since it has a constant seasonality.

Additive Holt-winters Smoothing (AHW) produces exponentially smoothed values for the level component (ℓ_t), the trend component (b_t), and the seasonal component (S_t), with smoothing parameters α, β and δ. "S_t is the weighted average between the current seasonal index and the seasonal index of the same season last year and S_t will add up to approximately zero within each year, and ℓ_t is seasonally adjusted by subtracting S_t ".
The forecast equation is given as follow:

$$\widehat{y_{t+1}} = \ell_t + b_t + S_{t+1-L}$$

$$\ell_t = \alpha(y_t - S_{t-L}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1}$$

$$S_t = \delta(y_t - \ell_t) + (1-\delta)S_{t-L}$$

for a seasonal frequency L, which is typically 12 for monthly data, initialize ℓ_0, b_0 and S_{i-L} for $i = 1, \dots \dots, L$, where $0 \le \alpha, \beta, \delta \le 1$.
To implement this method, we first formulate an observation equation, after which the α, β and δ are estimated by least square.

$$\widehat{y_{t+1}} = \ell_t + b_t + S_{t+1-L} + \varepsilon_{t+1} \quad (\varepsilon_{t+1} \text{ is i.i.d with variance } \sigma^2.)$$

And the forecast equations will then be

$$\widehat{y_{t+1}} = \ell_t + hb_t + S_{t-L+(h \bmod L)} \text{ (mod is the Modula operator)}$$

Multiplicative Holt-Winters Smoothing (MHW) is similar to AHW, while S_t is expressed in percentages and it will sum up to approximately L within each year. The forecast equation is given as follow:

$$\widehat{y_{t+1}} = (\ell_t + b_t) \times S_{t+1-L}$$

$$\ell_t = \alpha(y_t / S_{t-L}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1}$$

$$S_t = \delta(y_t / \ell_t) + (1-\delta)S_{t-L}$$

for a seasonal frequency, initialize ℓ_0, b_0 and S_{i-L} for $i = 1, \dots \dots, L$, where $0 \le \alpha, \beta, \delta \le 1$.
The implementation of MHW is also similar to AHW. The difference is the observation equation and the forecast equation as follows:

$$\widehat{y_{t+1}} = (\ell_t + b_t) \times S_{t+1-L} + \varepsilon_{t+1} \; (\varepsilon_{t+1} \text{ is i.i.d with variance } \sigma^2)$$

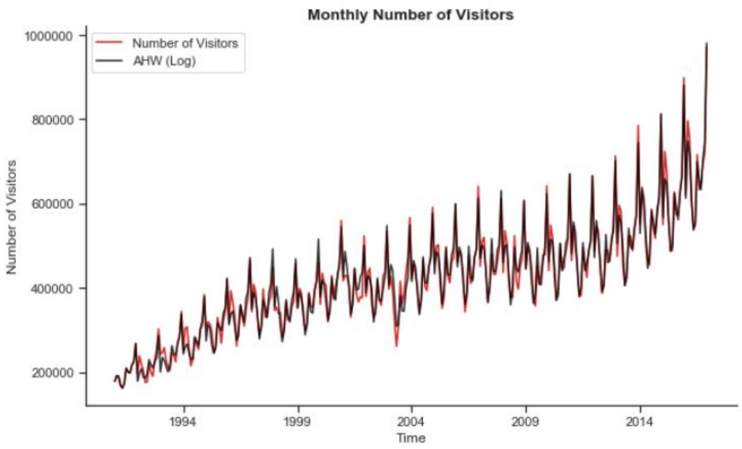
$$\widehat{y_{t+1}} = (\ell_t + hb_t) \times S_{t-L+(h \bmod L)}$$

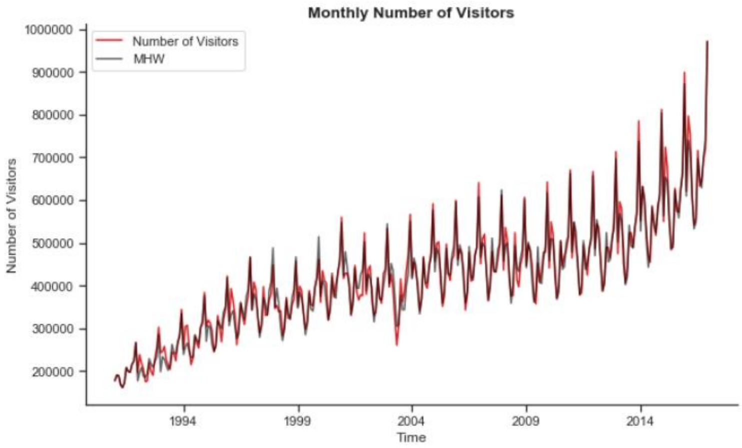
After implementing the methods in python, we get the following result. It is important to notice that the MSE and RMSE of Log AHW and MHW are not comparable, since the variable in Log AHW is the log-transformed number of visitors, while in MHW, it is the number of visitors.

	AHW	MHW

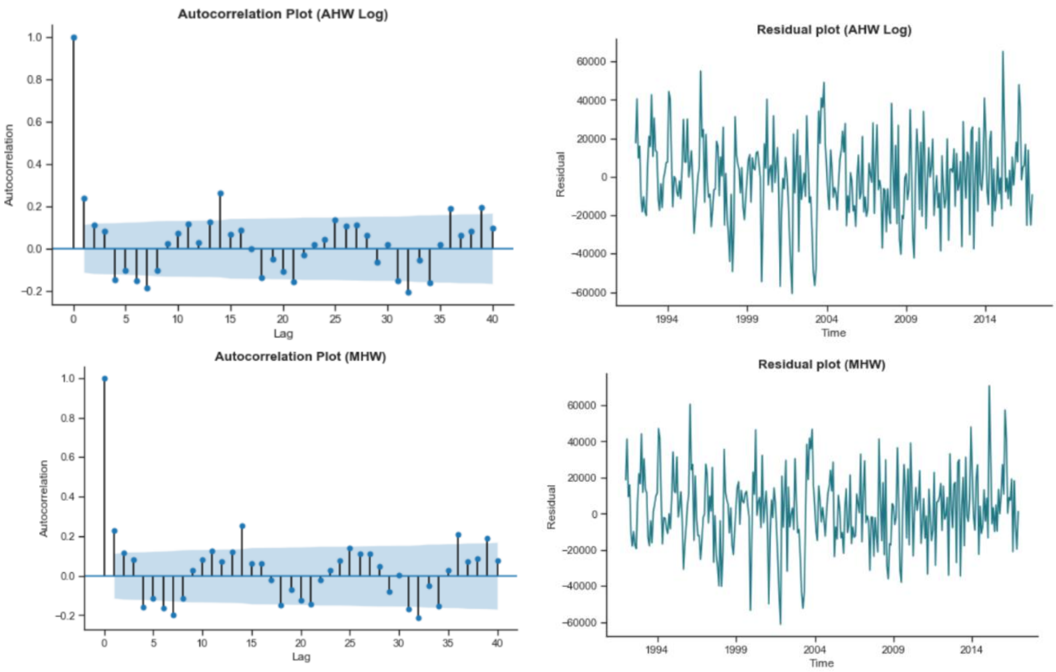
α	0.300	0.310
ℓ_t	0.052	0.078
β	0.009	0.012
b_t	0.007	0.008
δ	0.426	0.362
S_t	0.074	0.050
MSE	0.003	434674481.526
RMSE	0.054	20848.848

The following figures show that the smoothed series based on both methods tracks the original series very closely.





The residual and autocorrelation plots indicate that residuals have roughly constant variance over time, and both methods adequately capture the series, leading to small and insignificant residual correlations.



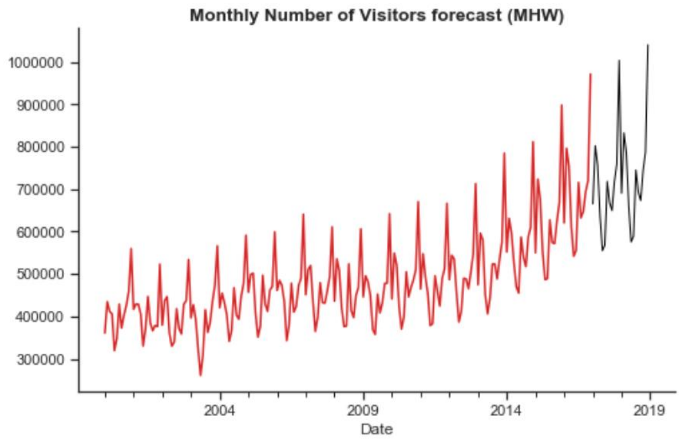
For AHW, the skewness and kurtosis of the residual are -0.017 and 0.409. For MHW, the skewness and kurtosis of the residual are 0.059 and 0.059. Thus, we can conclude that both methods have a distribution of residuals that is close to normal.

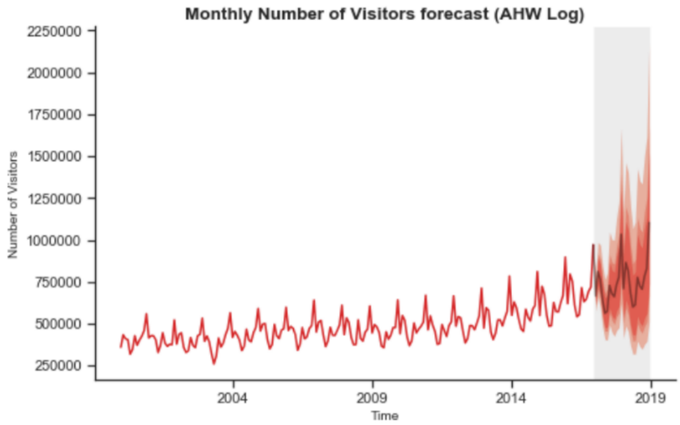
After model diagnostic, the validation process was done using expanding windows cross validation. Although k-fold cross-validation is widely used in machine learning, it ignores the temporal components inherent in the time series problem. In expanding window cross-validation, the size of the training set is expanded from an initial size to a maximum size depending on the data's frequency.

Two other models, including seasonal random walk and AHW for the number of visitors, are used for performance evaluation. Among them, AHW for log-transformed number of visitors has the smallest RMSE and SE, while MHW has the second-lowest RMSE and SE.

	RMSE	**SE**
Seasonal Random Walk	49114.158	3049.602
AHW	31970.874	3741.803
MHW	25815.898	3372.705
AHW (log)	23705.442	3032.109

Finally, a two-year forecast is generated. The predictions of AHW for log-transformed number of visitors are retransformed to the original number under the assumption of normality. The following figures and table illustrate the result of the point forecast and interval forecast for two models. Fan chart cannot be produced for MHW because the algorithm requires h to be larger than w. The result of the point forecast can be found in Appendix J.





Conclusion and Limitations and Further Suggestions

The forecast results indicate that both models are appropriate for this time series dataset, however, Log AHW is better than MHW since it has a lower RMSE.

There are still some limitations in the methodology used in this report.

1. Although the residual distributions for both models are approximately normal, we should use alternative assumptions for computing prediction intervals. And there are a minority of the in-sample forecast errors exceeds the significance bounds, Ljung-Box test can be conducted to test if there is evidence of non-zero autocorrelation in the in-sample forecast errors.
2. The Holt-winters exponential smoothing forecasting models do not consider the effects of external factors.
3. The starting values in these two methods are set as 0.1,0.1, 0.05 for α, β, δ. Although there are many ways of providing starting values, there is little advice in academic research as to which to choose. And the choice of starting values will change the optimum smoothing parameters.

While limited, these two models provide a starting point for future exploration of more complex models such as AR, ARIMA, etc. Furthermore, more time-series variables could be included in the forecast to find the interactions between different time-dependent variables.