Forecasting Maximum Daily Temperature in the LR Area for 1960

Dillon Helser, Ella Layton, Benjamin Tutka

Introduction

Let's assume that the date is December 31, 1959. We have been collecting data on the LR area daily maximum and minimum temperatures since 1940, the past 20 years. Our goal is to use this data to forecast the maximum daily temperature for the year 1960.

Methods

Before forecasting the data, we turned it into a time series of the maximum daily temperatures. We treated each year as a season, and separated data from 1959 as a validation set. We performed model selection based on the data that best forecasted 1959. Using the best model that had been refitted with the 1959 data, we then forecasted the maximum daily temperature for 1960.

Holt-Winter's Smoothing

The first method we used to forecast the next year's temperature was an additive Holt Winter's method. Every additive time series is the sum of a level component, a trend component, and a seasonal component. The Holt-Winter's method estimates these components for each day of the next year. The formula for the Holt-Winter's method is shown below where the notation $F_{t+k|t}$ is the forecast for k days away made on day t and m is the observations in a season. In our case, $m = 365$.

$$F_{t+k|t} = l_t + kb_t + S_{t+k-m}(\lceil \frac{k-1}{m} \rceil + 1)$$

Each component is estimated by the updating equations based on the parameters α, β, and γ which are all between 0 and 1 calculated by R to minimize the RMSE errors of one day forecasts in the training set. These represent how influential the previous days' estimates are on the estimate for the current day. The updating equations for each component are shown below.

$$l_t = \alpha(y_t - S_{t-m}) + (1-\alpha)(l_{t-1} + 1)$$

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}$$
$$S_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)S_{t-m}$$

To complete this model, R also estimates the level and trend at $k = 0$ along with the seasonal component for the previous m observations.

SARIMA Modeling

Unlike Holt-Winter's modeling that estimates the components of each observation in the series, SARIMA modeling treats each observation as a linear combination of the past few observations, their errors, and differences. The series must be a stationary series for a SARIMA model to be used, meaning that is has no trend or seasonality. There is a built in differencing process that can convert the series to a stationary one assuming that no other operations are needed for this process. A SARIMA model is notated by the number of terms it contains. The components of a SARIMA$(p, d, q) \times (P, D, Q)_s$ model are shown below.

$$B = \text{ the backshift operator in which } B^x y_t = y_{t-x}$$
$$\Phi(B^S) = 1 - \Phi_1 B^m - \cdots - \Phi_P B^{Pm} \text{ the seasonal autoregressive polynomial}$$
$$\varphi(B) = 1 - \varphi_1 B - \cdots - \varphi_p B^p \text{ the autogressive polynomial}$$
$$\Delta_S = 1 - B^m \text{ seasonal differences}$$
$$\Delta = 1 - B \text{ differences}$$
$$y_t = \text{ the observation at time } t$$
$$\Theta(B^S) = 1 + \Theta_1 B^m + \cdots + \Theta_Q B^{Qm} \text{ the seasonal moving averages polynomial}$$
$$\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q \text{ the moving averages polynomial}$$
$$\epsilon_t = \text{ the error at time } t \text{ with } \epsilon_t \sim^{iid} N(0, \sigma^2)$$

The full SARIMA$(p, d, q) \times (P, D, Q)_s$ model is shown below.

$$\Phi(B^S)\varphi(B)\Delta_S^D \Delta^d y_t = \varphi + \Theta(B^S)\theta(B)\epsilon_t$$

When performing model selection on SARIMA models, the parameters are chosen to minimize the AIC of the model.

Results

Holt Winter's

After performing model selection in R, we found that the best Holt Winter's model had the coefficients of $\alpha = 0.6297, \beta = 0,$ and $\gamma = 0.4822$. The updating equations of the model are shown below.

$$l_t = 0.6297(y_t - S_{t-m}) + (1 - 0.6297)(l_{t-1} + 1)$$

$$b_t = b_{t-1}$$

$$S_t = 0.4822(y_t - l_{t-1} - b_{t-1}) + (1 - 0.4822)S_{t-m}$$

The error measures on the validation set, the year 1959, are shown below.

MAD	MASE	RMSE	MAPE
15.484	1.604	18.013	21.842

The point forecasts for 1960,in blue, are plotted next to the rest of the data, in green.

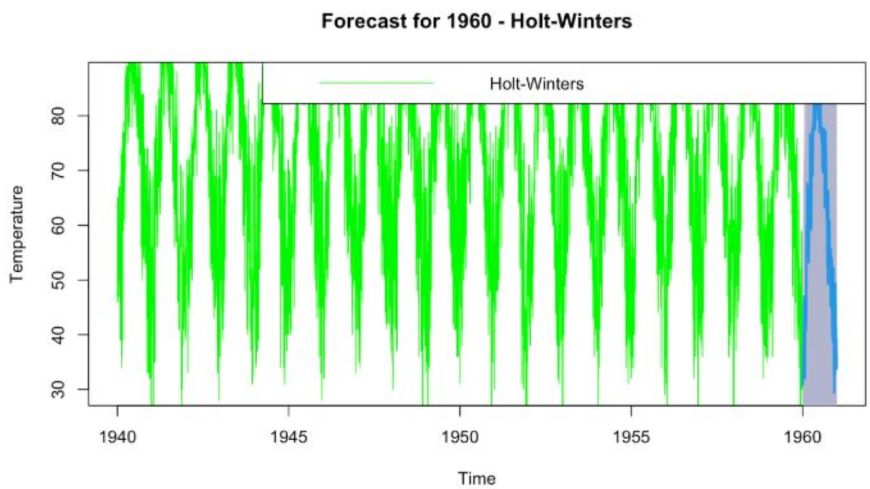

Figure 1: Forecasting 1960 maximum temperatures with the Holt-Winter's Model

SARIMA

The SARIMA model that best fits the data is a SARIMA$(1,0,4) \times (0,1,0)_s$ model. To make the series stationary, the model took seasonal differences once. The best model is one such that the forecast of the next observation is based on the previous observation and the errors of the four previous observations. The formula for the model is shown below.

$$(1 - \phi_1 B)(1 - B^{365})y_t = \phi + (1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3 + \theta_4 B^4)\epsilon_t$$

The error measures on the validation set, the year 1959, are shown below.

MAD	MASE	RMSE	MAPE
9.085	0.941	12.13	14.055

The point forecasts for 1960,in blue, are plotted next to the rest of the data, in red.

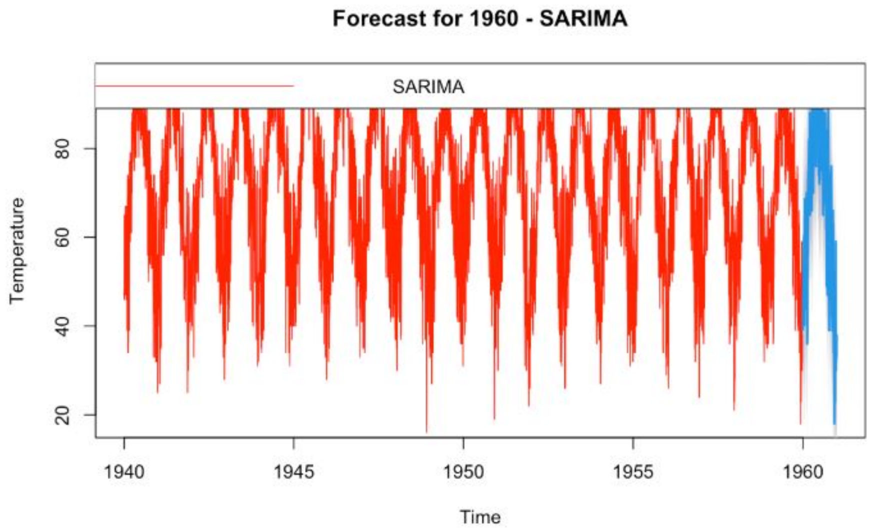

Figure 2: Forecasting 1960 maximum temperatures with the SARIMA Model

Conclusion

To forecast the maximum daily temperature in the LR area in the future, we will want to use the best model. Since every error measure for the SARIMA model is less than the error measures in the Holt Winter's model, the SARIMA model best forecasted the validation set. In the plots, we can see that the SARIMA model better captured the variability in temperature, so we can claim that it will offer more accurate forecasts. To create both models, we needed a constant, integer frequency. To achieve this, we removed the date February 29th from the data set; this did remove some of the data, but what was removed was a very small portion of the dataset. Roll forward validation could be used in future work to create a new SARIMA model with updated parameters. Even with the limitation, the SARIMA model was able to create an accurate forecast for 1960 and should be used in any future work.