Alexandra Elg

Time Series student project

Spring 2013

Introduction

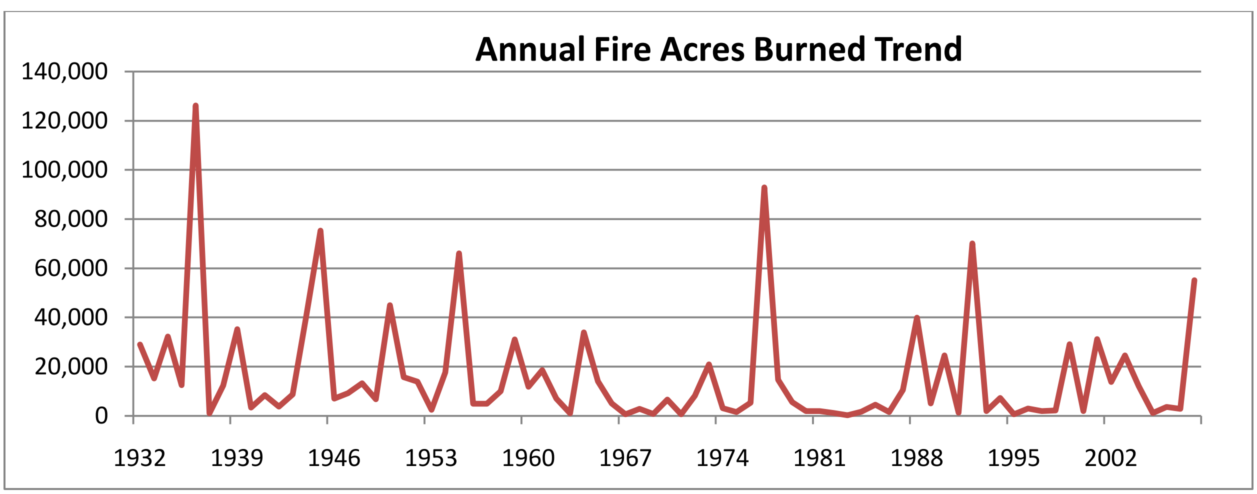
Every year living in California we hear about wildfires throughout our state. Particularly during dry winter seasons like we had this year, wildfires are all but inevitable. For my time series analysis project, I decided to look at historical data specific to California Timberland fires. My project will analyze this annual data, determine whether or not it is stationary, transform it if necessary to make it stationary, assess seasonality, choose an appropriate model that fits the data, analyze its residuals, and finally predict future values.

Data

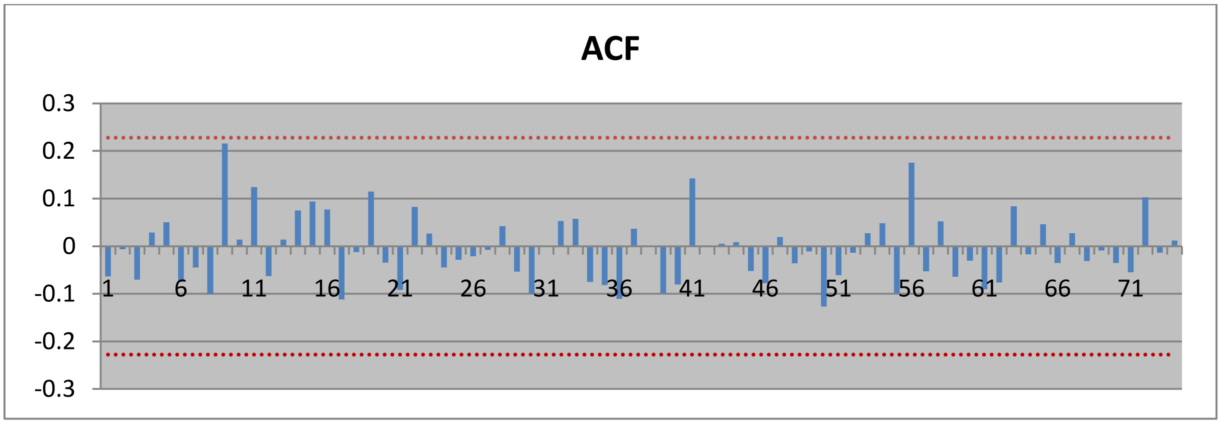
Number of acres burned in Timberland fires in California, 1932-2008
http://cdfdata.fire.ca.gov/incidents/incidents_statsevents

Model Specification

The time series is graphed below. The data is annual, so you don't see seasonality. There are maybe one or two outliers, with particularly high years in 1936 and 1977. It doesn't look like there is any general upward or downward trend.

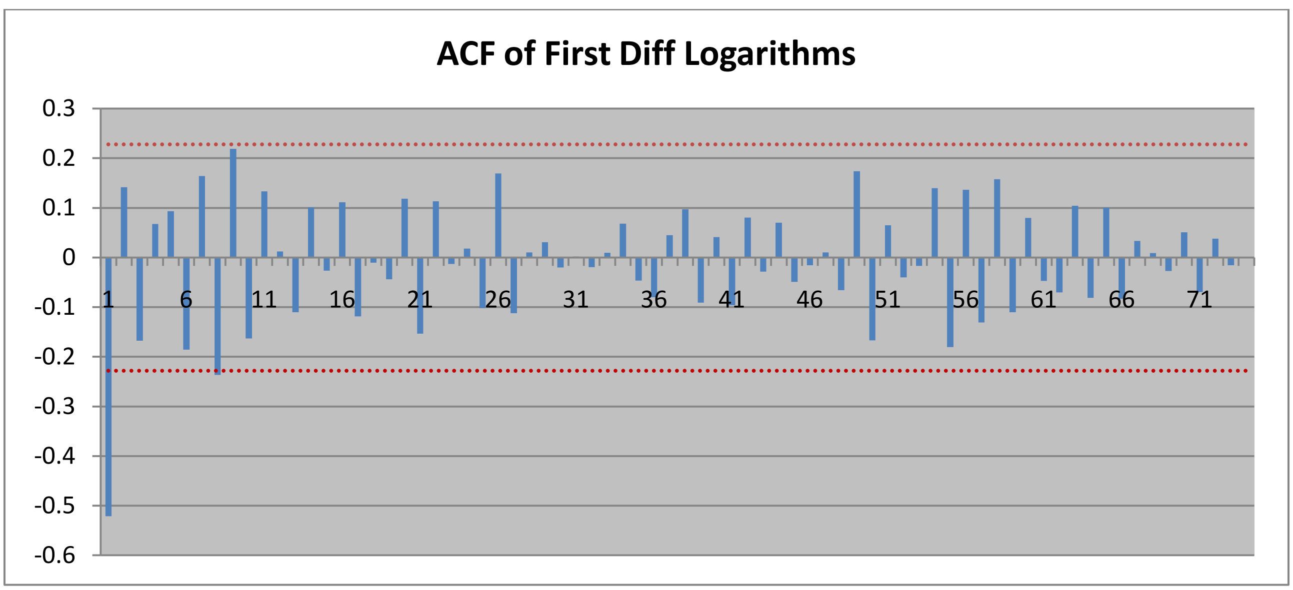


The autocorrelation function for this original series is graphed below



$$\pm 2/\sqrt{n} \quad with\ n = 77$$

Since this looks like a stationary series, we could use the original data to fit a model to. However, I will look at a transformed set as well. Taking logarithms and then first differences, the transformed series' ACF below I like better than the original series, because of the negative autocorrelation at lag 1.



I will fit a model to both the original series, and the first differences of the logarithms. I will refer to the original series at Y_t and the transformed series as $W_t = \ln(Y_t) - \ln(Y_{t-1})$

Model Fitting

To fit both series to the $AR(1)$ model, I used the regression tool in excel to estimate the parameters. For the original series Y_t

$AR(1): Y_t = \varphi(Y_{t-1}) + e_t + \theta_0$

$$Y_t = -0.66(Y_{t-1}) + e_t + 17{,}296$$

This is not a good model, as the parameter is close to zero.

Fitting the transformed W_t series to the $AR(1)$ model:

$AR(1): W_t = \varphi W_{t-1} + e_t + \theta_0$

Regression Statistics	
Multiple R	0.530973
R Square	0.281933
Adjusted R Square	0.272096
Standard Error	1.566112
Observations	75

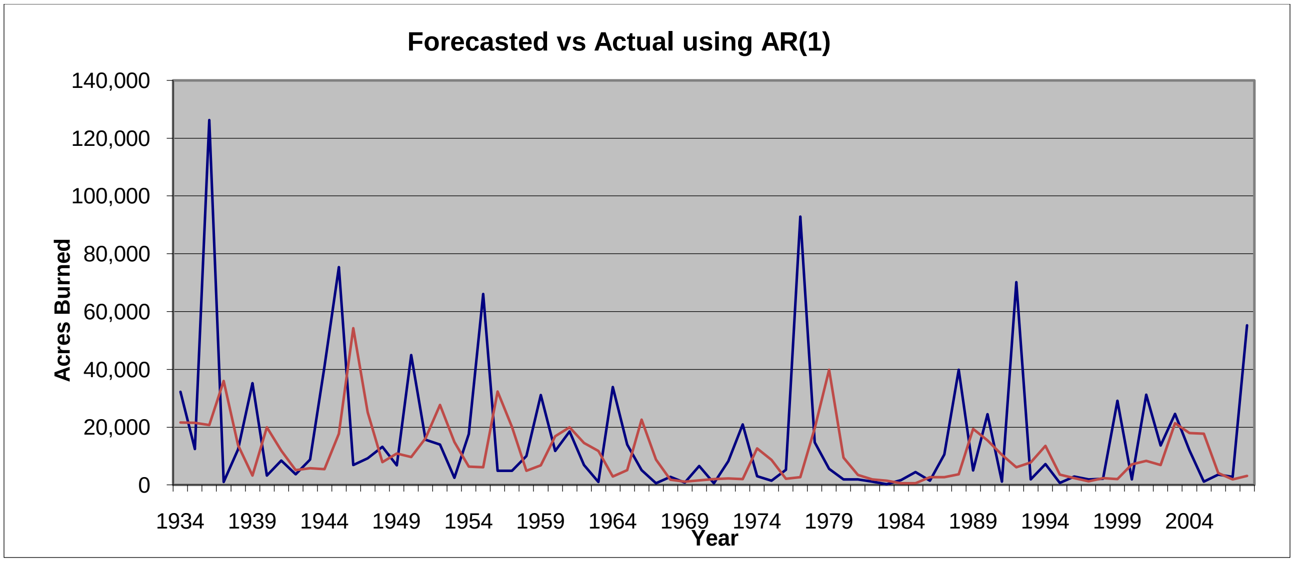
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	70.29895829	70.29895829	28.6618	9.55E-07
Residual	73	179.0475315	2.452705912		
Total	74	249.3464898			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.000254	0.180866585	0.001403934	0.998884	-0.36021	0.360721
X Variable 1	-0.54036	0.100932184	-5.353671332	9.55E-07	-0.74152	-0.3392

$$W_t = -0.54W_{t-1} + e_t + 0.00025$$

A graph of the forecasted values against the actual values is below.



The forecasted values are in red. This model tends to be somewhat reactive, and doesn't handle the peaks of the actual series well.

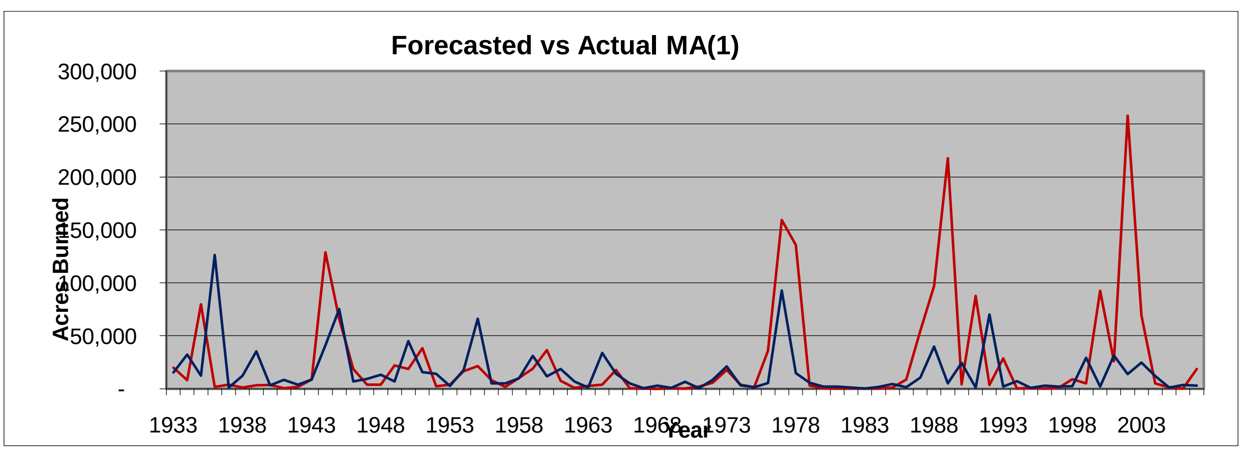
To consider another model, I am fitting the transformed W_t series to the $MA(1)$ model

$MA(1)$: $W_t = e_t - \theta e_t + \mu$

Using the least squares estimate, by using Solver in Excel, I get the resulting process:

$$W_t = e_t - 0.8712e_t + 0.0085$$

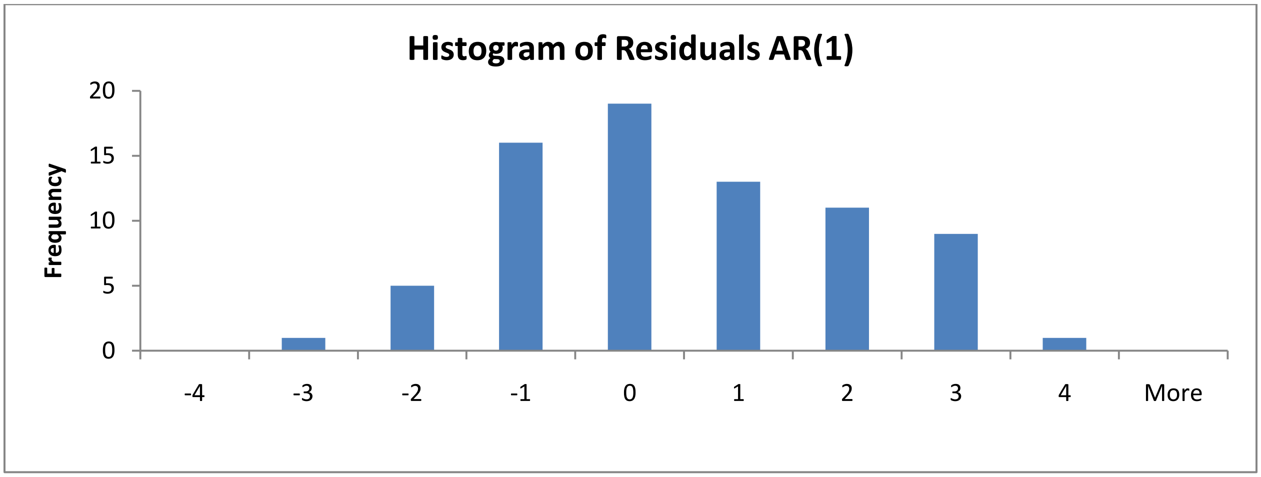
A graph of the forecasted values against the actual values is below. Again forecasted values are in red.

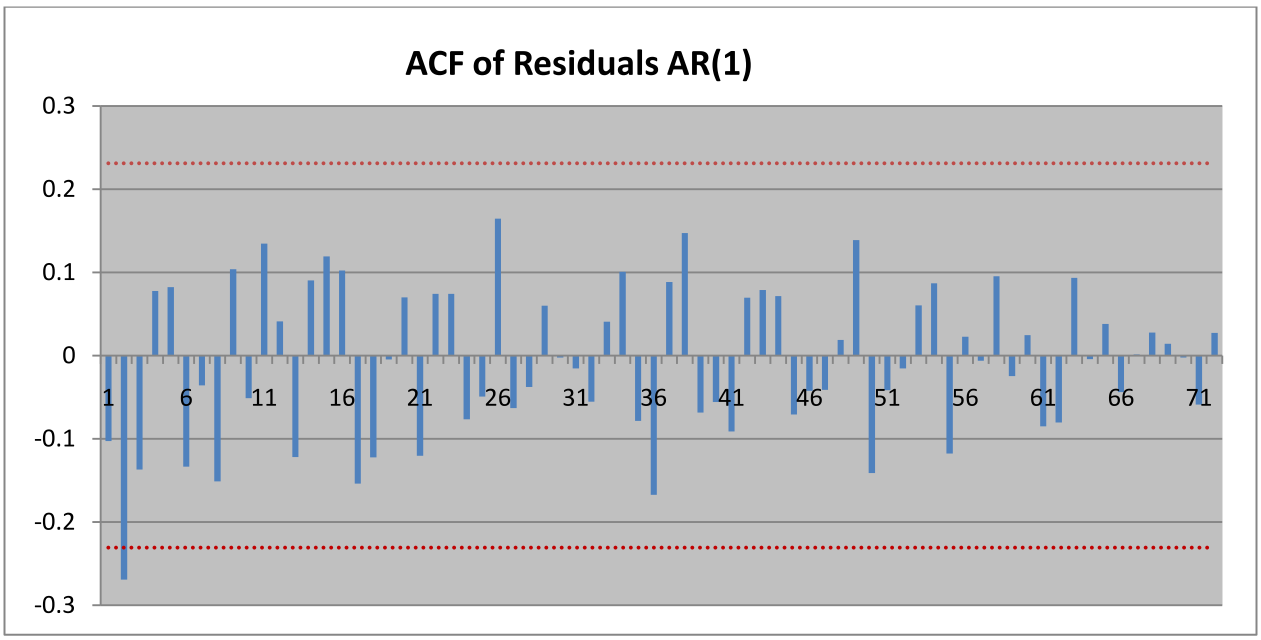


This process tends to overestimate the peaks, particularly in the later years.

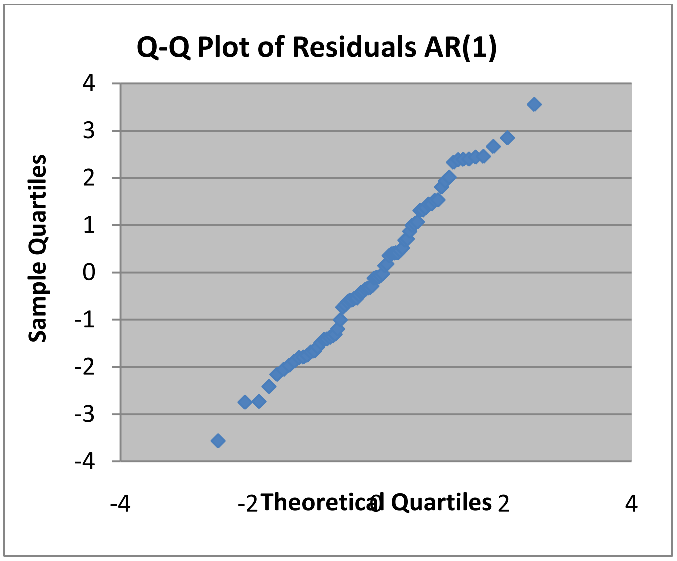
Residual Analysis

For the $AR(1)$ model: $W_t = -0.54W_{t-1} + e_t + 0.00025$, below is a histogram for the residuals. It looks like they are distributed somewhat normally, with a bit of a left skew. The ACF is within the CI, except for one value.



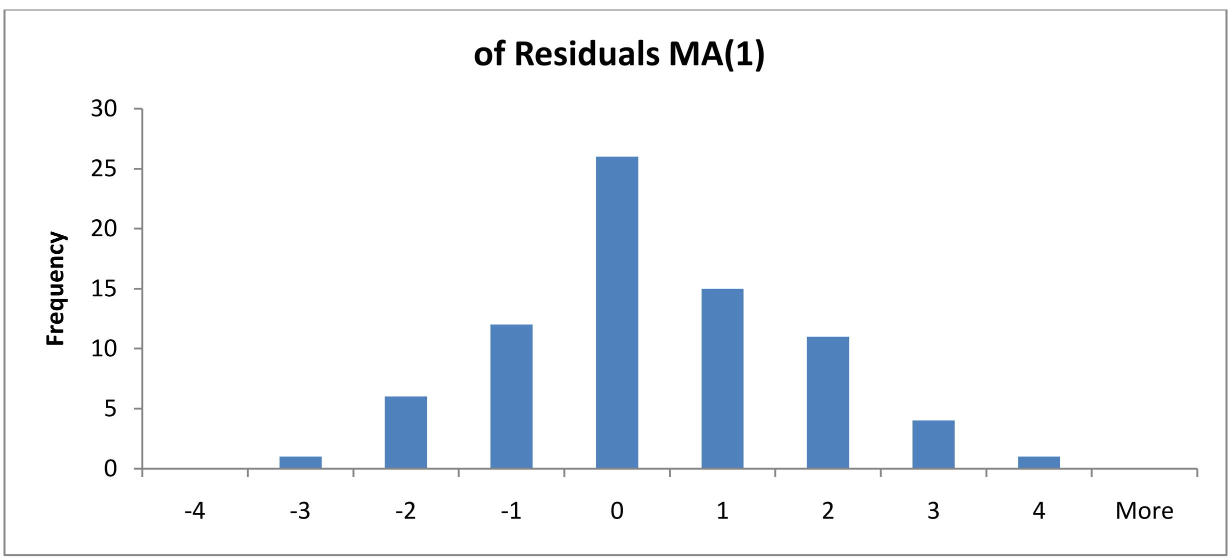


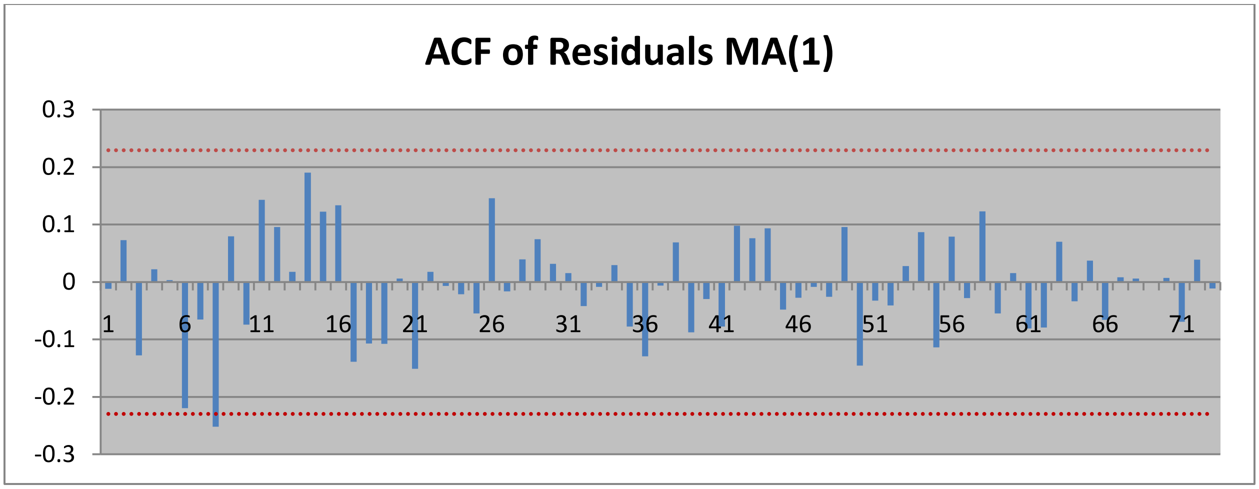
$$\pm 2/\sqrt{n} \text{ with } n = 75$$



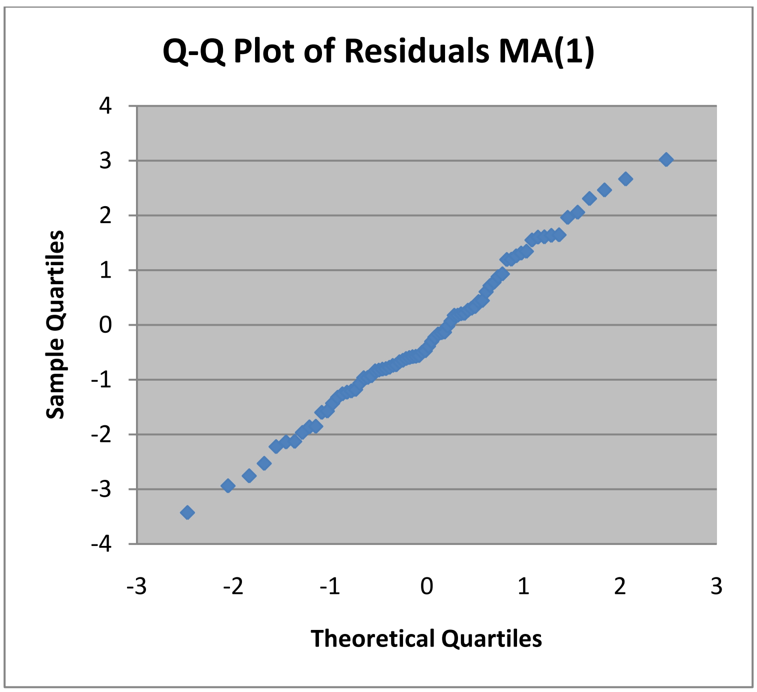
The Q-Q plot looks like the residuals are distributed normally.

For the $MA(1)$ model: $W_t = e_t - 0.8712e_t + 0.0085$, below is a histogram for the residuals. It looks like they are distributed somewhat normally. The ACF is within the CI, except for one value.





$$\pm 2/\sqrt{n} \quad with\ n = 76$$



The Q-Q plot looks like the residuals are distributed normally.

Conclusion

This analysis shows that the Fire data can be modeled using an ARIMA process. The first differences of logarithms would be a better series for modeling. Using the transformed series, the $AR(1)$ and $MA(1)$ models both have reasonably normal residuals, making them good candidates for forecasting future values.