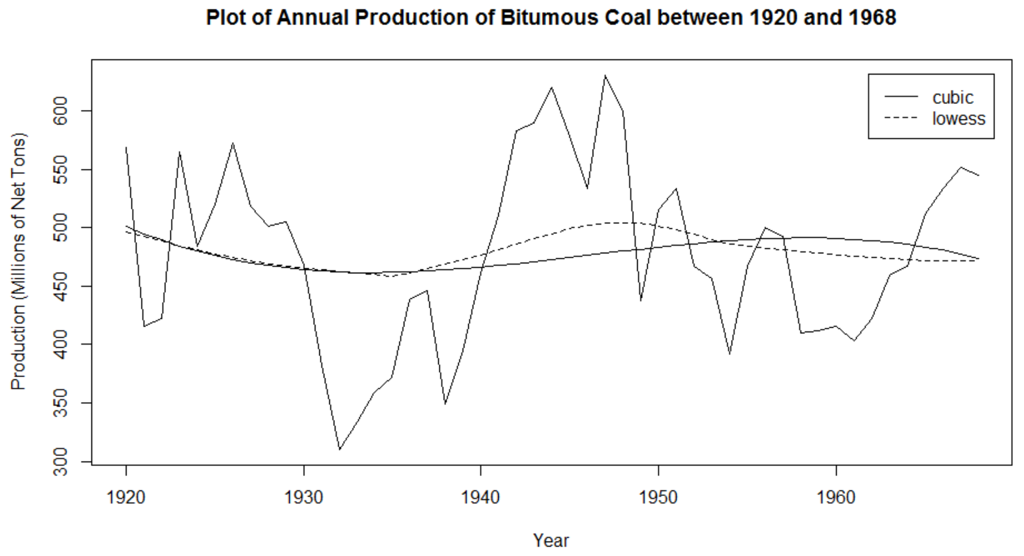
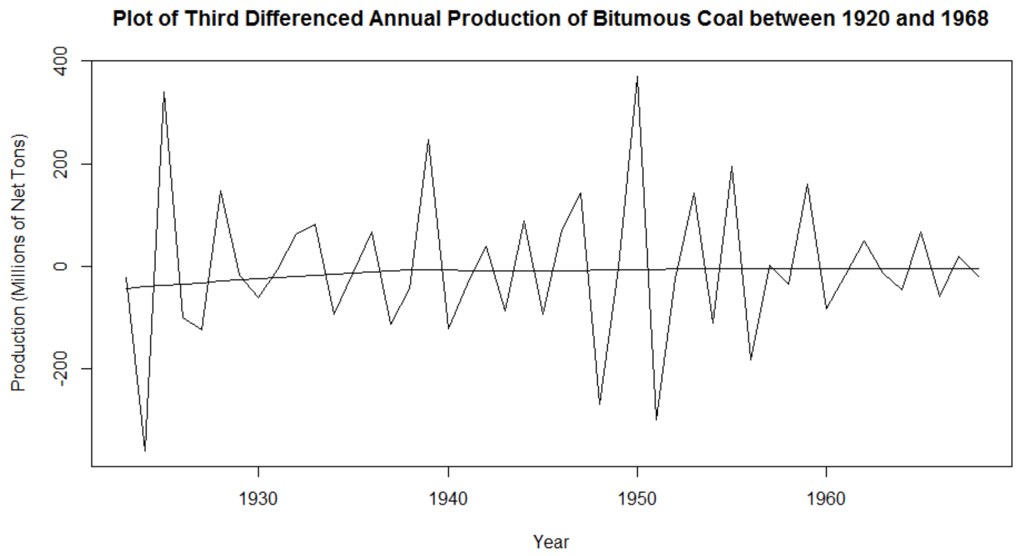
The time series object bicoal.tons stores the annual production in millions of net tons of bitumous coal between 1920 and 1968.



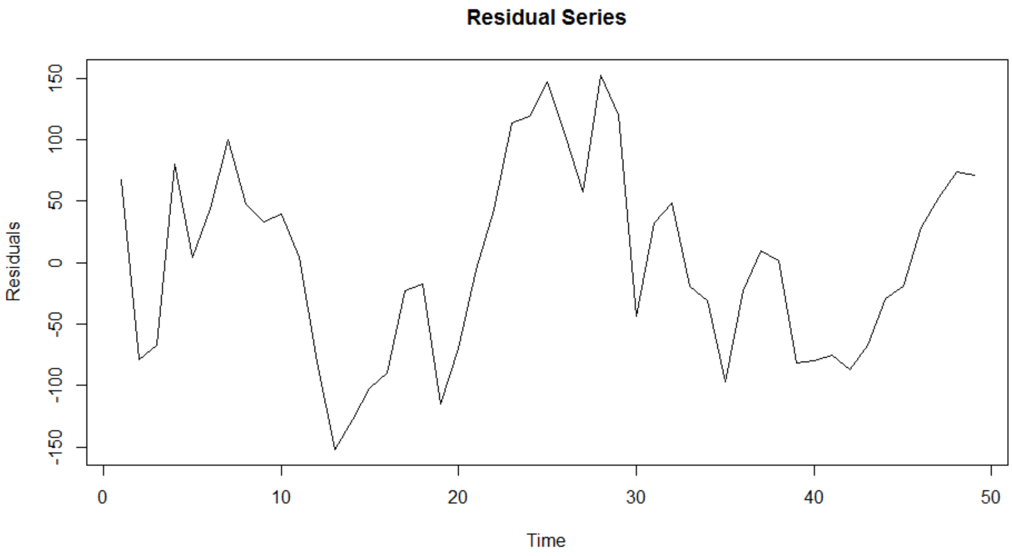
The figure above shows the time series plot for the bicoal.tons dataset. There are no obvious signs of heteroscedasticity. Furthermore, the above plot looks very similar to the plot of log(bicoal.tons) and also sqrt(bicoal.tons) and hence, there is no need for any transformation.

A significant dip can be observed in the production of bitumous coal in the early 1930s (most probably due to the Great Depression which took place during that time) and then production rose again quite steadily starting from the late 1930s. It stayed quite high during the 1940s and then fell again slightly in the 1950s. This leads to the consideration of using a cubic trend to model the series. The dotted curve in the above plot is a LOWESS curve fitted and the solid curve is a cubic trend fitted to the series. It can also be observed that the cubic trend matches the LOWESS curve quite well at the beginning of the series but deviates a little from the LOWESS curve in the middle and towards the end. Despite slight inaccuracies in modelling the trend of the second half of the series, fitting a cubic trend is the optimal choice because fitting a quadratic makes the model even more inaccurate and fitting a quartic only improves the result very slightly but results in a bigger model used instead.

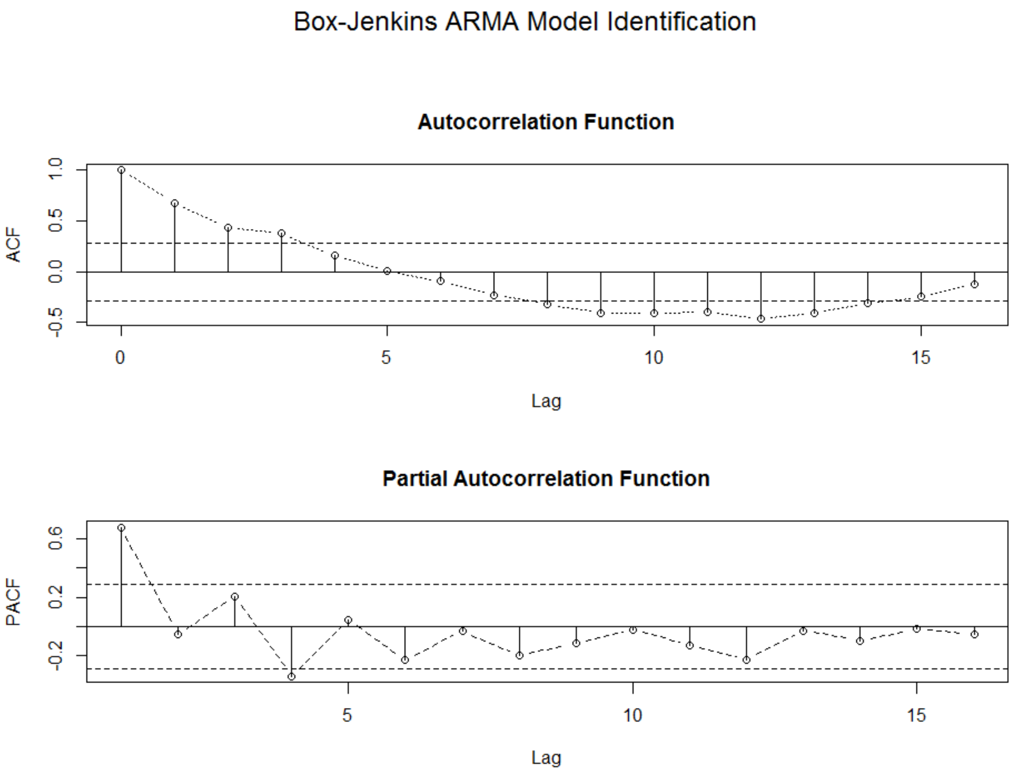


The figure above shows a LOWESS curve fitted after taking first differences three times and this curve looks almost like a constant straight line. This strengthens the choice to fit a cubic polynomial as the trend component.

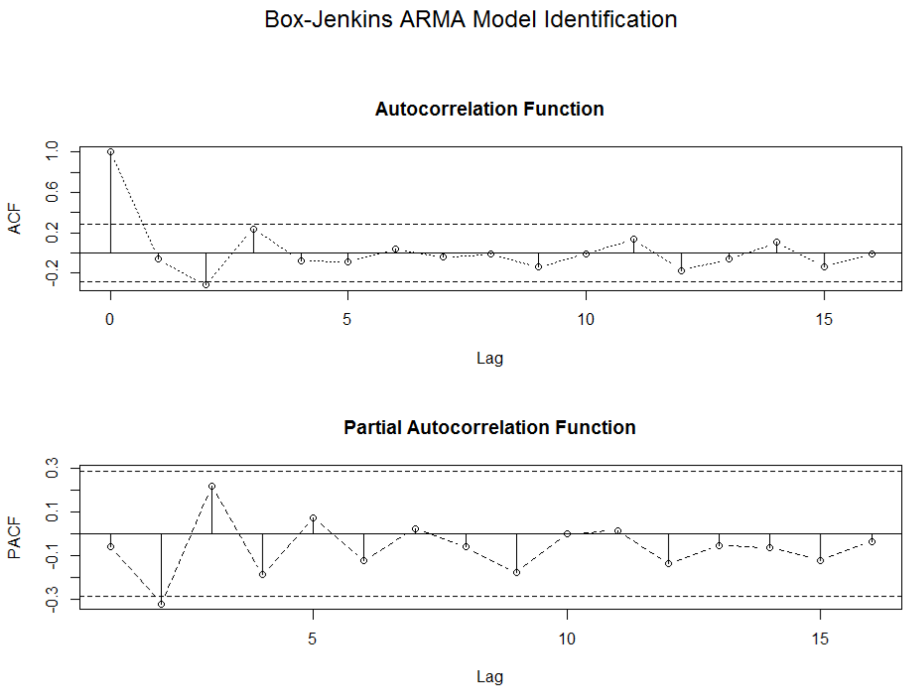
There is no seasonal component for this time series because there is no regular repeating pattern within the series.



The next thing to look at is the irregular component of the series. The figure above shows the residual series and it can be deduced that this is a dependent set of residuals because the series tends to move in the direction it is headed towards instead of just wriggling up and down randomly.

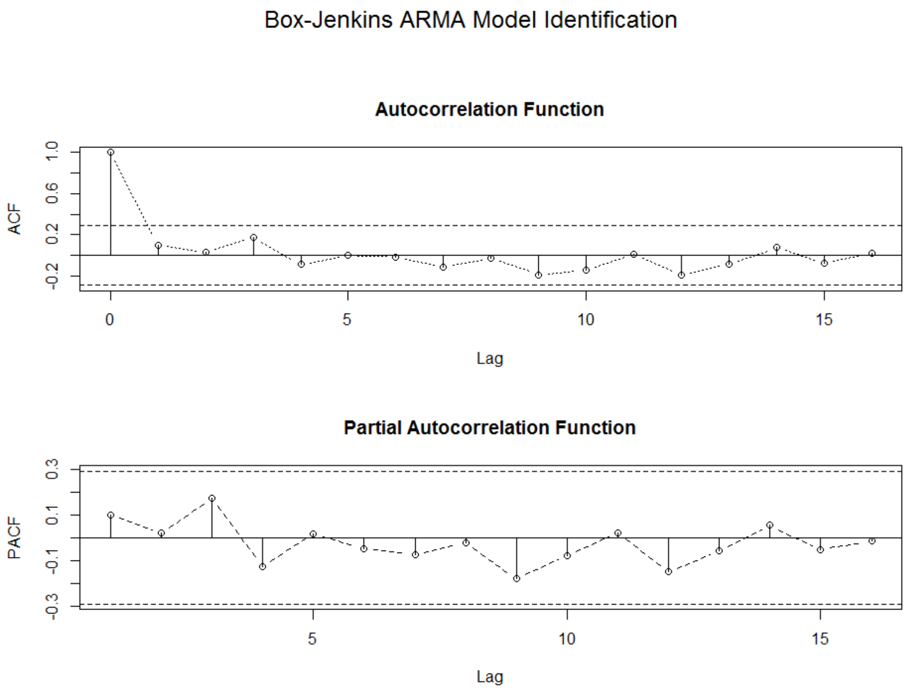


To test for the stationarity of the series, the Box-Jenkins ARMA Model Identification is used. From the Autocorrelation Function (ACF) plot above, the series does not seem stationary as it does not damp down within the dotted lines and stay there after a while (spikes from lag = 8 to lag = 14 are not within the dotted lines). Hence, first differences is taken.



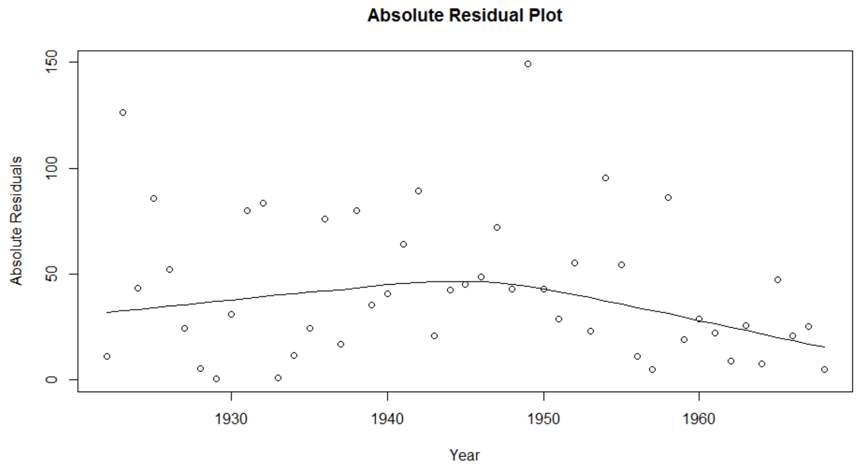
After taking first differences

The figure above shows the Box-Jenkins ARMA Model Identification after first differences was taken. From the ACF plot, we can now deduce that the series is stationary as we can observe it damping down within the dotted lines and staying there after lag = 2. Based on the Partial Autocorrelation Function (PACF) plot, an AR(2) should be fitted as the spike at lag = 2 is the last spike which is not within the dotted lines.

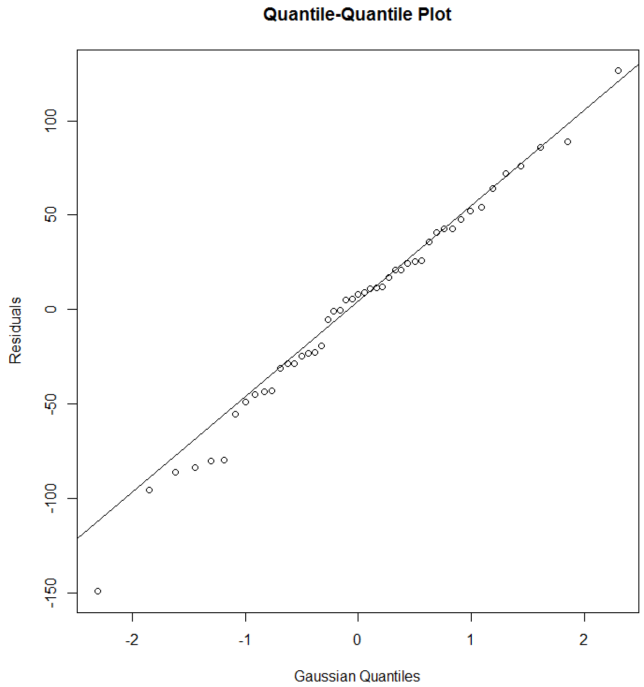


AR(2) Residuals

After fitting the model, we can see from the above figure that the residuals from an AR(2) have no more features, i.e. the only spike not within the dotted lines for the ACF plot is the spike at lag = 0 and all of the spikes are within the dotted lines for the PACF plot.



Although the LOWESS curve through the absolute residual plot above is not exactly a straight line, the slight drop in the end can be ignored because overall, the absolute residuals look mostly like random noise and hence, we can assume that the residuals have constant variance.



The quantile-quantile plot above looks roughly like a straight line and hence we can say that the residuals are normally distributed.

Model:

Structure (prediction):

Zt = 472.7710 + 1.7750(t - 1944) + 0.0257(t - 1944)^2 - 0.0041(t - 1944)^3 + Xt, where Xt is irregular

Dependence (description):

Diff(X(t)) = -0.09934837Diff(X(t - 1)) - 0.26854079Diff(X(t - 2))

The bicoal.tons series had gone down and up and down again over the period we had data, i.e. from 1920 to 1968 (trend). There were no regular repeating patterns within the series (no seasonal). There is serial dependence in the errors and this model describes it (irregular).

Cubic trend is overfit, it would be worth examining a constant trend and choosing the smaller model the diagnostics look OK.