STA 137

Homework 6

Prof. Burman

Jared Yu (914640019), Danli Zhang (915011728)
3-5-2019

1. An $AR(2)$ model has $n = 95$, with estimated parameters: $\hat{\mu} = 27.0253, \hat{\phi}_1 = 0.5253, \hat{\phi}_2 = -0.8551$, and $\hat{\sigma}^2 = 4.636$ with X_{94} and X_{95} equal to 34.043 and 28.499.
   a) Below are forecasts for $X_{n+h}, h = 1, \cdots, 5$.

$$\hat{X}_{96} - \hat{\mu} = \hat{\phi}_1(X_{95} - \hat{\mu}) + \hat{\phi}_2(X_{94} - \hat{\mu})$$

$$\hat{X}_{96} = \hat{\phi}_0 + \hat{\phi}_1 X_{95} + \hat{\phi}_2 X_{94}, \text{where } \hat{\phi}_0 = \hat{\mu}\left(1 - \hat{\phi}_1 - \hat{\phi}_2\right) = 35.93824$$

$$\hat{X}_{96} = 35.93824 + 0.5253 \times 28.499 + (-0.8551) \times 34.043$$

$$\hat{X}_{96} = 21.7986$$

$$\hat{X}_{97} = \hat{\phi}_0 + \hat{\phi}_1 \hat{X}_{96} + \hat{\phi}_2 X_{95}$$

$$\hat{X}_{97} = 35.93824 + 0.5253 \times 21.7986 + (-0.8551) \times 28.499$$

$$\hat{X}_{97} = 23.01955$$

$$\hat{X}_{98} = \hat{\phi}_0 + \hat{\phi}_1 \hat{X}_{97} + \hat{\phi}_2 \hat{X}_{96}$$

$$\hat{X}_{98} = 35.93824 + 0.5253 \times 23.01955 + (-0.8551) \times 21.7986$$

$$\hat{X}_{98} = 29.39043$$

$$\hat{X}_{99} = \hat{\phi}_0 + \hat{\phi}_1 \hat{X}_{98} + \hat{\phi}_2 \hat{X}_{97}$$

$$\hat{X}_{99} = 35.93824 + 0.5253 \times 29.39043 + (-0.8551) \times 23.01955$$

$$\hat{X}_{99} = 31.69302$$

$$\hat{X}_{100} = \hat{\phi}_0 + \hat{\phi}_1 \hat{X}_{99} + \hat{\phi}_2 \hat{X}_{98}$$

$$\hat{X}_{100} = 35.93824 + 0.5253 \times 31.69302 + (-0.8551) \times 29.39043$$

$$\hat{X}_{100} = 27.45483$$

   b) The model we are given appears as follows.

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + \varepsilon_t$$

First, we will check for the stationarity of the series, this can be done by checking the conditions of the two coefficients ϕ_1 and ϕ_2. This method works because we are currently working with an $AR(2)$ model.

$$-1 < \hat{\phi}_1 < 1$$

$$-1 < 0.5253 < 1$$

$$-1 < \frac{\hat{\phi}_1}{1 - \hat{\phi}_2} < 1$$

$$-1 < \frac{0.5253}{1 - (-0.8551)} < 1$$

$$-1 < 0.2831653 < 1$$

So, the conditions are met, and therefore the series is stationary. Since the series follows the conditions for stationarity, it is also causal and can be written as an $MA(\infty)$ sequence.

Since it has been shown that our series is causal, it can be written as follows.

$$X_t - \mu = \psi(B)\varepsilon_t, \text{ where } \psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \cdots \tag{1}$$

Also, we know that the series $AR(2)$ can be written as follows.

$$\phi(B)(X_t - \mu) = \varepsilon_t, \text{ where } \phi(B) = 1 - \phi_1 B - \phi_2 B^2 \tag{2}$$

With equations (1) and (2), we can then put them together as follows.

$$1 = \psi(B)\phi(B)$$

$$1 = (\psi_0 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \psi_4 B^4 + \cdots)(1 - \phi_1 B - \phi_2 B^2)$$

$$\begin{aligned} 1B^0 + (0)B^1 + (0)B^2 + (0)B^3 + (0)B^4 + \cdots \\ &= \psi_0 + (\psi_1 - \phi_1\psi_0)B + (\psi_2 - \phi_1\psi_1 - \phi_2\psi_0)B^2 + (\psi_3 - \phi_1\psi_2 - \phi_2\psi_1)B^3 \\ &+ (\psi_4 - \phi_1\psi_3 - \phi_2\psi_2)B^4 + \cdots \end{aligned}$$

Then the following terms can be derived from the above equation.

$$1 = \psi_0$$

$$0 = \psi_1 - \phi_1\psi_0$$

$$0 = \psi_2 - \phi_1\psi_1 - \phi_2\psi_0$$

$$0 = \psi_3 - \phi_1\psi_2 - \phi_2\psi_1$$

$$0 = \psi_4 - \phi_1\psi_3 - \phi_2\psi_2$$

Now to find solutions for each of the $\psi_i's, i = 0, \cdots, 4$.

$$\psi_0 = 1$$

$$\psi_1 = \phi_1\psi_0 = \phi_1 \times 1 = \phi_1$$

$$\psi_2 = \phi_1\psi_1 + \phi_2\psi_0 = \phi_1(\phi_1) + \phi_2(1) = \phi_1^2 + \phi_2$$

$$\psi_3 = \phi_1\psi_2 + \phi_2\psi_1 = \phi_1(\phi_1^2 + \phi_2) + \phi_2(\phi_1) = \phi_1^3 + 2\phi_1\phi_2$$

$$\begin{aligned} \psi_4 = \phi_1\psi_3 + \phi_2\psi_2 = \phi_1(\phi_1^3 + 2\phi_1\phi_2) + \phi_2(\phi_1^2 + \phi_2) &= \phi_1^4 + 2\phi_1^2\phi_2 + \phi_1^2\phi_2 + \phi_2^2 \\ &= \phi_1^4 + 3\phi_1^2\phi_2 + \phi_2^2 \end{aligned}$$

Plugging in our estimates $\hat{\phi}_1 = 0.5253, \hat{\phi}_2 = -0.8551$ we find the following results for $\psi_i, i = 0, \cdots, 4$

$$\hat{\psi}_0 = 1$$

$$\hat{\psi}_1 = 0.5253$$

$$\hat{\psi}_2 = 0.5253^2 + (-0.8551) = -0.5791599$$

$$\hat{\psi}_3 = 0.5253^3 + 2 \times 0.5253 \times (-0.8551) = -0.7534167$$

$$\hat{\psi}_4 = 0.5253^4 + 3 \times 0.5253^2 \times (-0.8551) + (-0.8551)^2 = 0.09946983$$

Now applying the formula for the variance of the forecast error:

$$\hat{\sigma}^2(h) = Var\left(X_{n+h} - \hat{X}_{n+h}\right) = E\left(X_{n+h} - \hat{X}_{n+h}\right)^2$$

$$\begin{aligned}\hat{\sigma}^2(1) &= E\left(X_{n+1} - \hat{X}_{n+1}\right)^2 = E\left((\varepsilon_{n+1} + \psi_1\varepsilon_n + \psi_2\varepsilon_{n-1} + \cdots) - (\psi_1\varepsilon_n + \psi_2\varepsilon_{n-1} + \cdots)\right)^2 \\ &= E\varepsilon_{n+1}^2 = Var(\varepsilon_{n+1}) + (E(\varepsilon_{n+1}))^2 = \sigma^2 + 0 = \sigma^2 = 4.636\end{aligned}$$

$$\begin{aligned}\hat{\sigma}^2(2) &= E\left(X_{n+2} - \hat{X}_{n+2}\right)^2 = E(\varepsilon_{n+2} + \psi_1\varepsilon_{n+2})^2 = E(\varepsilon_{n+2}^2) + E((\psi_1\varepsilon_{n+2})^2) = \sigma^2 + \psi_1^2\sigma^2 \\ &= (1 + \psi_1^2)\sigma^2 = 5.915258\end{aligned}$$

$$\begin{aligned}\hat{\sigma}^2(3) &= E\left(X_{n+3} - \hat{X}_{n+3}\right)^2 = E\left((\varepsilon_{n+3} + \psi_1\varepsilon_{n+2} + \psi_2\varepsilon_{n+1} + \cdots) - (\psi_3\varepsilon_n + \psi_4\varepsilon_{n-1} + \cdots)\right)^2 \\ &= E(\varepsilon_{n+3} + \psi_1\varepsilon_{n+2} + \psi_2\varepsilon_{n+1})^2 = Var(\varepsilon_{n+3} + \psi_1\varepsilon_{n+2} + \psi_2\varepsilon_{n+1}) \\ &= \sigma^2 + \psi_1^2\sigma^2 + \psi_2^2\sigma^2 = (1 + \psi_1^2 + \psi_2^2)\sigma^2 = 7.470294\end{aligned}$$

$$\begin{aligned}\hat{\sigma}^2(4) &= E\left(X_{n+4} - \hat{X}_{n+4}\right)^2 = E\left((\varepsilon_{n+4} + \psi_1\varepsilon_{n+3} + \psi_2\varepsilon_{n+2} + \cdots) - (\psi_4\varepsilon_n + \psi_5\varepsilon_{n-1} + \cdots)\right)^2 \\ &= E(\varepsilon_{n+4} + \psi_1\varepsilon_{n+3} + \psi_2\varepsilon_{n+2} + \psi_3\varepsilon_{n+1})^2 \\ &= Var(\varepsilon_{n+4} + \psi_1\varepsilon_{n+3} + \psi_2\varepsilon_{n+2} + \psi_3\varepsilon_{n+1}) = \sigma^2 + \psi_1^2\sigma^2 + \psi_2^2\sigma^2 + \psi_3^2\sigma^2 \\ &= (1 + \psi_1^2 + \psi_2^2 + \psi_3^2)\sigma^2 = 10.10186\end{aligned}$$

$$\begin{aligned}\hat{\sigma}^2(5) &= E\left(X_{n+5} - \hat{X}_{n+5}\right)^2 = E\left((\varepsilon_{n+5} + \psi_1\varepsilon_{n+4} + \psi_2\varepsilon_{n+3} + \cdots) - (\psi_5\varepsilon_n + \psi_4\varepsilon_{n-1} + \cdots)\right)^2 \\ &= E(\varepsilon_{n+5} + \psi_1\varepsilon_{n+4} + \psi_2\varepsilon_{n+3} + \psi_3\varepsilon_{n+2} + \psi_4\varepsilon_{n+1})^2 \\ &= Var(\varepsilon_{n+5} + \psi_1\varepsilon_{n+4} + \psi_2\varepsilon_{n+3} + \psi_3\varepsilon_{n+2} + \psi_4\varepsilon_{n+1}) \\ &= \sigma^2 + \psi_1^2\sigma^2 + \psi_2^2\sigma^2 + \psi_3^2\sigma^2 + \psi_4^2\sigma^2 = (1 + \psi_1^2 + \psi_2^2 + \psi_3^2 + \psi_4^2)\sigma^2 = 10.14773\end{aligned}$$

c) To calculate a 95% confidence interval, we will use the following formula.

$$\hat{X}_{n+h} \pm 1.96 \times \sqrt{\hat{\sigma}^2(h)}$$

So below are the confidence intervals for $h = 1, \cdots, 5$

$$\hat{X}_{n+1} \pm 1.96 \times \sqrt{\hat{\sigma}^2(1)}$$

$$21.7986 \pm 1.96 \times \sqrt{4.636} \to (17.57845, 26.01875)$$

$$\hat{X}_{n+2} \pm 1.96 \times \sqrt{\hat{\sigma}^2(2)}$$

$$23.01955 \pm 1.96 \times \sqrt{5.915258} \to (24.03339, 34.74747)$$

$$\hat{X}_{n+3} \pm 1.96 \times \sqrt{\hat{\sigma}^2(3)}$$

$$29.39043 \pm 1.96 \times \sqrt{7.470294} \to (25.46347, 37.92257)$$

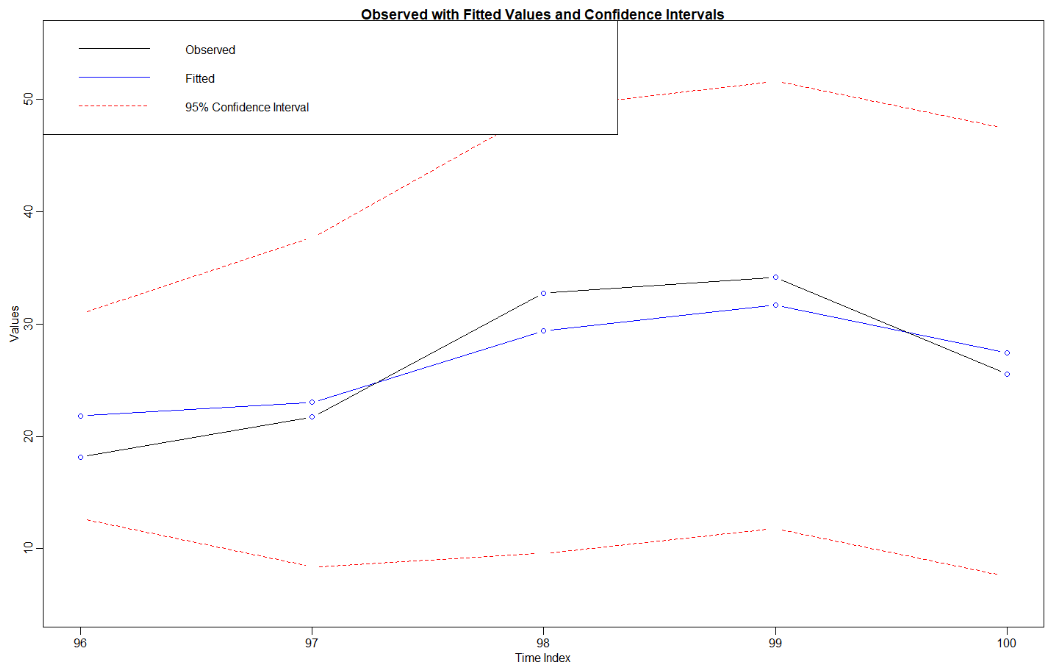
$$\hat{X}_{n+4} \pm 1.96 \times \sqrt{\hat{\sigma}^2(4)}$$

$$31.69302 \pm 1.96 \times \sqrt{10.10186} \to (18.25258, 27.78653)$$

$$\hat{X}_{n+5} \pm 1.96 \times \sqrt{\hat{\sigma}^2(5)}$$

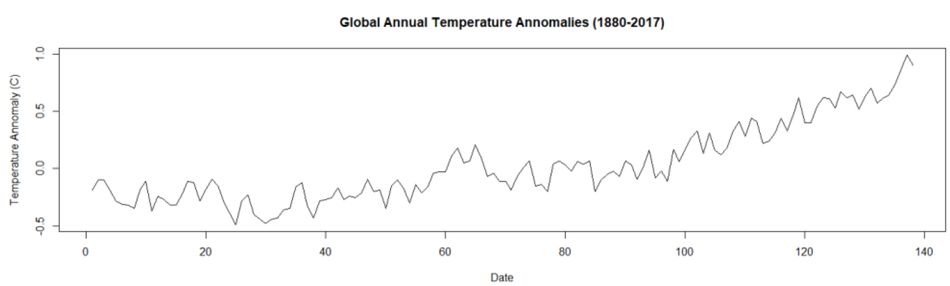
$$27.45483 \pm 1.96 \times \sqrt{10.14773} \rightarrow (21.21115, 33.6985)$$

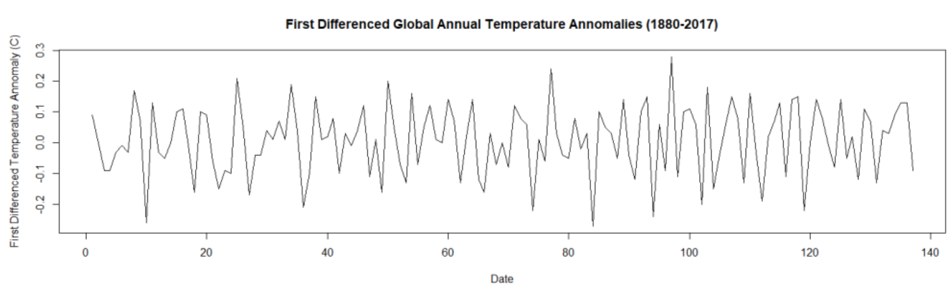
d) Below is a plot of the observed values, fitted values, and prediction bands.



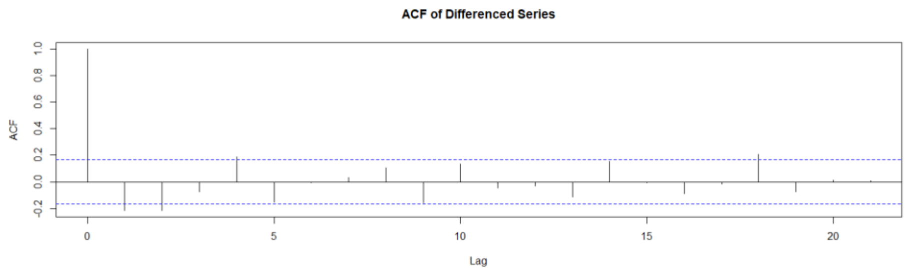
The fitted values follow quite closely to the observed values, staying both near and going above or below the observed values. The prediction band also does a good job of containing both the predicted values and the observed values. This may indicate that the model is properly specified, and it shows how the prediction bands can contain the observed and fitted values if there is a simple trend that is not making sudden swings in any direction.

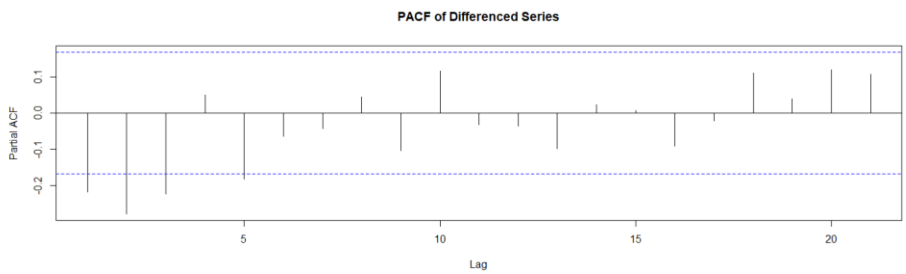
2. a) Below is a plot of the series, along with a first differenced plot of the series.





Initially the temperature data showed a trend upwards, going from negative to positive over time. Taking the first difference of the series shows a much different plot. The differenced data shows data which resembles a rough, where there is a horizontal band and some random fluctuations throughout it. Below is an ACF and PACF of the differenced data.





The ACF model shows that there is some possible significance at lags 1 and 2, but they look rather slight. There are also spikes which seem to cross the line at lags 4 and 18. However, these lags look even less significant than the first two. The PACF model shows that there are some possibly significant lags at lags 1, 2, 3, and 5. If the series were an $ARMA(p,q)$ model, it would be expected that tailing off would occur in both plots, however this does not seem evident in either. It is interesting that there is no clear tailing off on either or both plots, making it difficult to be sure what sort of model is best. In this case, it may be better to fit several models and judge them based upon a certain criterion such as AICc.

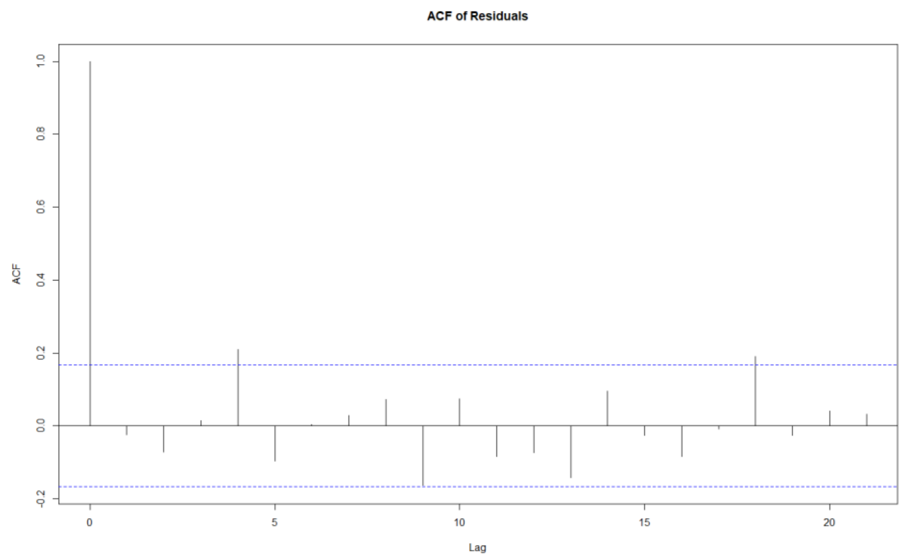
b) Below is a table of the AICC values for all the possible models

$ARMA(p,1,q)$	$AICC$
$ARIMA(1,1,1)$	-3.484091
$ARIMA(1,1,2)$	-3.478999
$ARIMA(1,1,3)$	-3.508047
$ARIMA(2,1,1)$	-3.482160
$ARIMA(2,1,2)$	-3.517923
$ARIMA(2,1,3)$	-3.479208
$ARIMA(3,1,1)$	-3.466324
$ARIMA(3,1,2)$	-3.486478
$ARIMA(3,1,3)$	-3.478076

The smallest value and hence optimal model chosen by AICC is $ARIMA(2,1,2)$. Below are the estimated parameters and the standard errors.

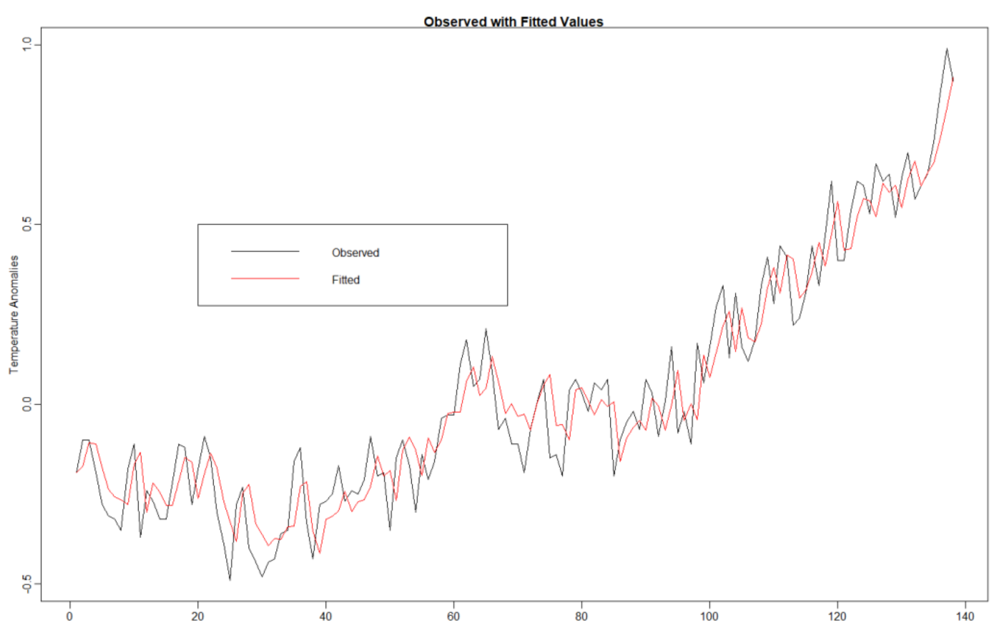
	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$
	0.3056	-0.2109	-0.6317	0.0626
$Standard\ error$	0.4069	0.1933	0.4106	0.2912

Below is an ACF of the residuals from the model $ARIMA(2,1,2)$.



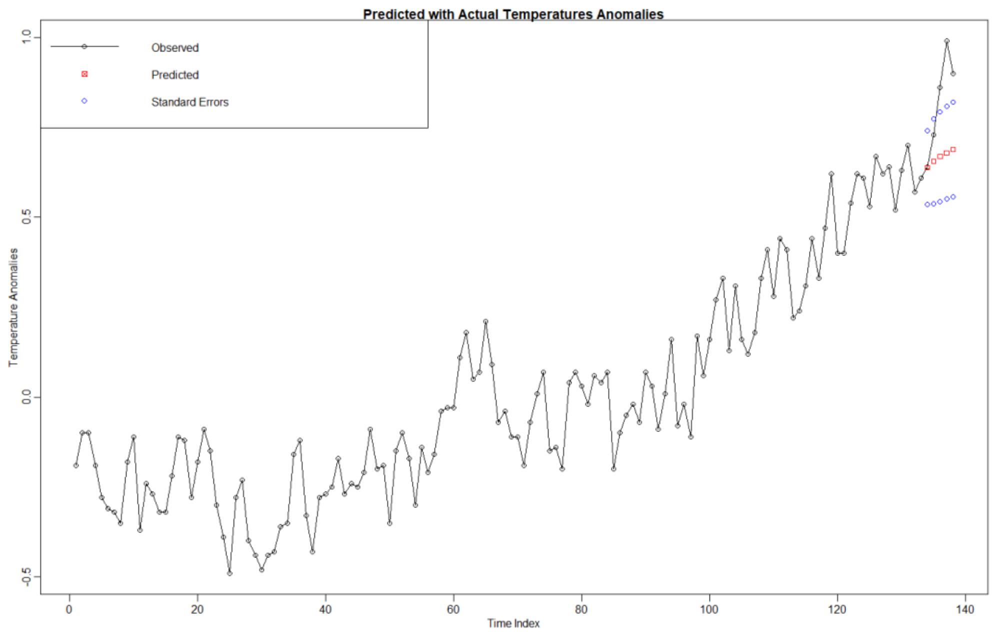
The ACF of the residuals show that there doesn't seem to be significant tails, except possibly at lag 4 and 18. The former seems somewhat possible, and the latter looks even less likely. In other words, it is possible that their significance comes from randomness. However, it is not certain that either are the case. It is possible that the $ARIMA(2,1,2)$ has done as close to a good job as possible for modeling the data with the given options, but there may be some part of the model that is not fully understood or incorrectly specified.

c) Below is a plot of the observed values with the fitted values.



The fit seems quite close, and so it could be said that the model of $ARIMA(2,1,2)$ is accurate. The $ARIMA(2,1,2)$ model is supposed to be able to account for the trend in the data. When looking at the fitted line as it moves along with the observed line, it seems that the fitted line does a good job of staying behind the observed data as it twists up and down. In other words, the fitted line can move with the data without being swayed too much in one direction or the other, counter towards where the trend is going.

d) Below is a plot of the observed values along with predictions in red for the last 5 points.



There seems to be a rather large difference between the observed values of the last 5 points and the predicted values. Looking at the data, it is apparent that there is some narrowing in the observations. In other words, it could be seen beforehand that the observations are condensing into a narrow region before the observations make a breakout of this pattern. This breakout shows a large sudden acceleration in the points, making them quite different from the predicted values. It could be believed that the prediction is not good, but at the same time it is possible that the prediction would have done well prior to the sudden acceleration upwards. To analyze this sudden change, the standard errors have been plotted as well in blue. It is apparent that by the second prediction point, the observed values have jumped up to highly close to the boundary of the standard error. This sort of behavior may indicate that an $ARIMA(2,1,2)$ model may only be adequate at modeling certain timeframes of the observations and may start to break down at other parts such as when there are sudden large swings in the data. Towards the end, the observations make a large sudden move upwards, it would be difficult to predict such an event based on past observations.

3. I. Let X_t be a mean zero series following an $MA(1)$ model.

$$X_t = \varepsilon_t + \theta\varepsilon_{t-1}$$

The forecasted value of X_{t+1} can be written as follows,

$$\hat{X}_{t+1} = \theta\hat{\varepsilon}_t$$

Then let $X_t - \hat{X}_t = \hat{\varepsilon}_t$, which is known as an *innovation*. The forecasted value then can be rewritten as follows.

$$\hat{X}_{t+1} = \theta(X_t - \hat{X}_t)$$

Q.E.D.

II. Continuing from part I, we can say that the $MA(1)$ series is the differenced version of series Y_t which follows an $ARIMA(0,1,1)$ model. So, from part I we have written:

$$\hat{X}_{t+1} = \theta(X_t - \hat{X}_t)$$

This can be expanded using the following formula from differencing.

$$X_t = \nabla Y_t = Y_t - Y_{t-1}$$

The above can be rewritten as follows:

$$Y_{t+1} = Y_t + X_{t+1}$$

Now considering that the observations $Y_t, Y_{t-1}, \cdots$ are from the past, we can write the forecast as follows:

$$\hat{Y}_{t+1} = Y_t + \hat{X}_{t+1}$$

$$\hat{Y}_{t+1} = Y_t + \theta\big(X_t - \hat{X}_t\big)$$

The value X_t can be rewritten as $Y_t - Y_{t-1}$.

$$\hat{Y}_{t+1} = Y_t + \theta\big(Y_t - Y_{t-1} - \hat{X}_t\big)$$

Now, using the formula $Y_{t+1} = Y_t + X_{t+1}$, we can write the forecast as follows:

$$\hat{Y}_{t+1} = Y_t + \hat{X}_{t+1}$$

The reason is that Y_t is from the past, and therefore is already known. Applying the same to $\hat{X}_t$, the value can be rewritten below.

$$\hat{Y}_{t+1} = Y_t + \theta\left(Y_t - Y_{t-1} - \big(\hat{Y}_t - Y_{t-1}\big)\right)$$

$$\hat{Y}_{t+1} = Y_t + \theta(Y_t - \hat{Y}_t)$$

Q.E.D.

II. Continuing from part I, we can say that the $MA(1)$ series is the differenced version of series Y_t which follows an $ARIMA(0,1,1)$ model. So, from part I we have written:

$$\hat{X}_{t+1} = \theta(X_t - \hat{X}_t)$$

This can be expanded using the following formula from differencing.

$$X_t = \nabla Y_t = Y_t - Y_{t-1}$$

The above can be rewritten as follows:

$$Y_{t+1} = Y_t + X_{t+1}$$

Now considering that the observations $Y_t, Y_{t-1}, \cdots$ are from the past, we can write the forecast as follows:

$$\hat{Y}_{t+1} = Y_t + \hat{X}_{t+1}$$

Then we can substitute the equation for $\hat{X}_{t+1}$ from part I into the forecast.

$$\hat{Y}_{t+1} = Y_t + \theta\left(X_t - \hat{X}_t\right)$$

The value X_t can be rewritten as $Y_t - Y_{t-1}$.

$$\hat{Y}_{t+1} = Y_t + \theta\left(Y_t - Y_{t-1} - \hat{X}_t\right)$$

The reason is that Y_t is from the past, and therefore is already known. Applying the same to $\hat{X}_t$, where $\hat{Y}_t = Y_{t-1} + \hat{X}_t$, the value can be rewritten below.

$$\hat{Y}_{t+1} = Y_t + \theta\left(Y_t - Y_{t-1} - \left(\hat{Y}_t - Y_{t-1}\right)\right)$$

$$\hat{Y}_{t+1} = Y_t + \theta(Y_t - \hat{Y}_t)$$

Q.E.D.

Reference:

https://stackoverflow.com/questions/21148498/remove-last-n-rows-in-data-frame-with-the-arbitrary-number-of-rows

http://www.sthda.com/english/wiki/add-legends-to-plots-in-r-software-the-easiest-way

https://piazza.com/class/jqk73rklz4k5ld?cid=69

https://www.stat.berkeley.edu/~bartlett/courses/153-fall2010/lectures/5.pdf

https://stackoverflow.com/questions/5182238/replace-x-axis-with-own-values

https://eml.berkeley.edu/~powell/e241b_f06/TS-StatInv.pdf