Greg Sollenberger
Time Series
Spring 2012
VEE Student Project

Introduction

For my student project, I chose the time series "Age of Death of the World's Oldest Person, Mid 20th Century Forward." This data was taken from the Wikipedia entry titled "Oldest People" at (http://en.wikipedia.org/wiki/Oldest_people). The table my data was taken from is titled "Chronological list of the verified oldest living person since 1955." The data is not in a format that can be easily entered into a spreadsheet. I had to enter each value by hand. Note that this is not a time series with constant increments. A slightly different project might be "Age of oldest person at end of year." Since the world's oldest person can hold that title for several years (or less than a year), I thought that such a series would contain some obvious and uninteresting correlations (or in the case of the titleholder dying before year end, would actually miss some data points). I thought "Age at death" would be a much more interesting series to analyze.

Analysis

Below is a graph of the data, which I plotted to look for obvious clues as to what kind of time series to model.

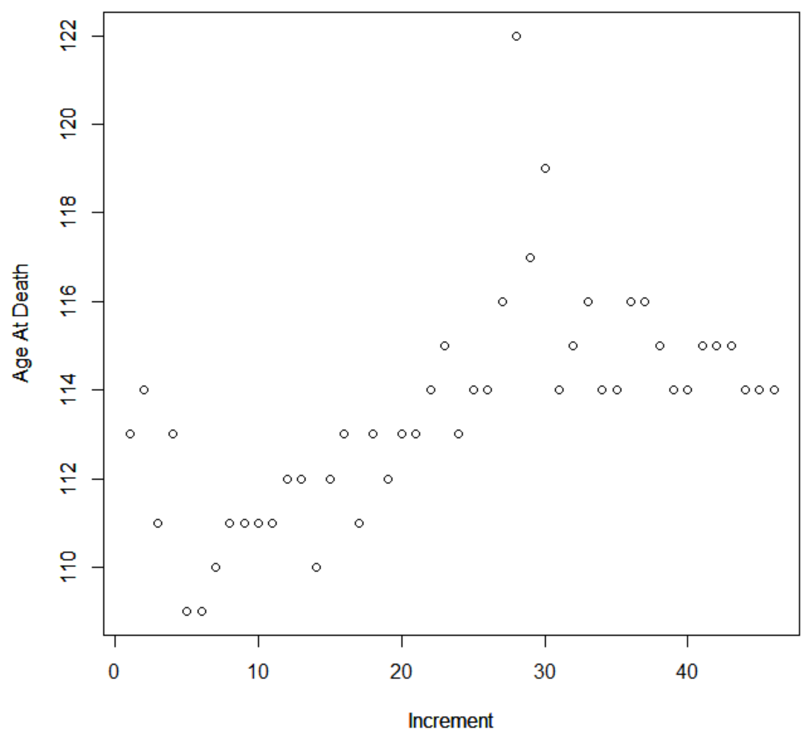

Figure 1. Age at Death

Greg Sollenberger
Time Series
Spring 2012
VEE Student Project

Since I am using the rounded age at death, note many point have the exact same vertical value. Recall that “increment” means the next person in the series, not the next year. There is a visible upward trend, although with so few points one can’t discern if it’s an exponential increase, linear increase, or some other scaling. I will fit some ARIMA(p,0,q) models for the sake of comparison, but this series appears non-stationary. A 1^{st} or 2^{nd} difference will probably provide a better fit. Also see the correlogram for evidence that this is probably an ARIMA(p,x,q) model of non-zero x.

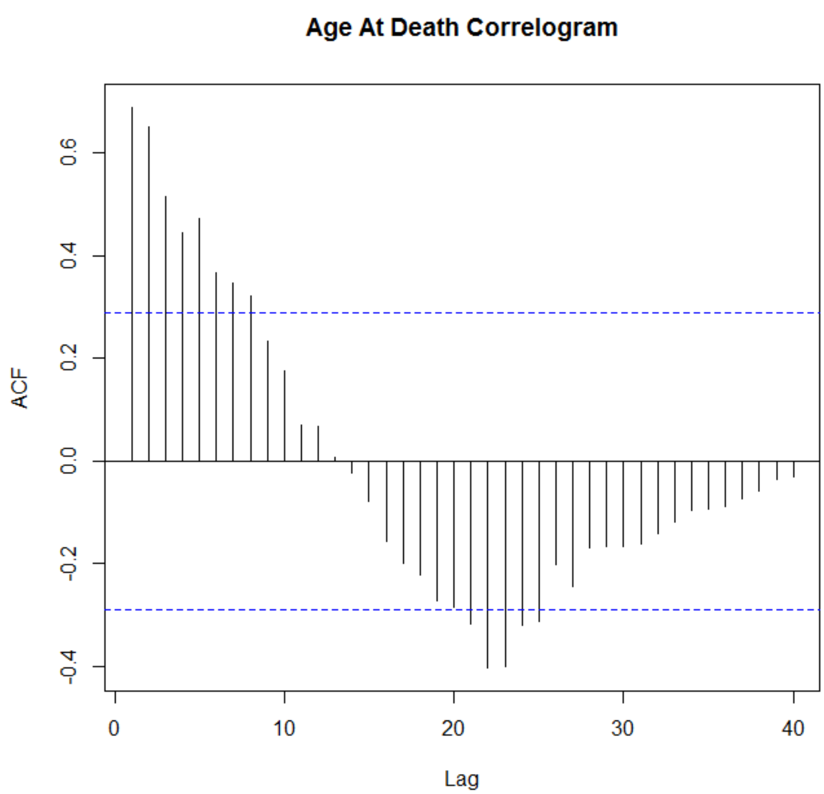

Figure 2. Autocorrelogram of data, no differencing.

The autocorrelations don't fall off rapidly enough. We should try a 1st difference and see how it looks.

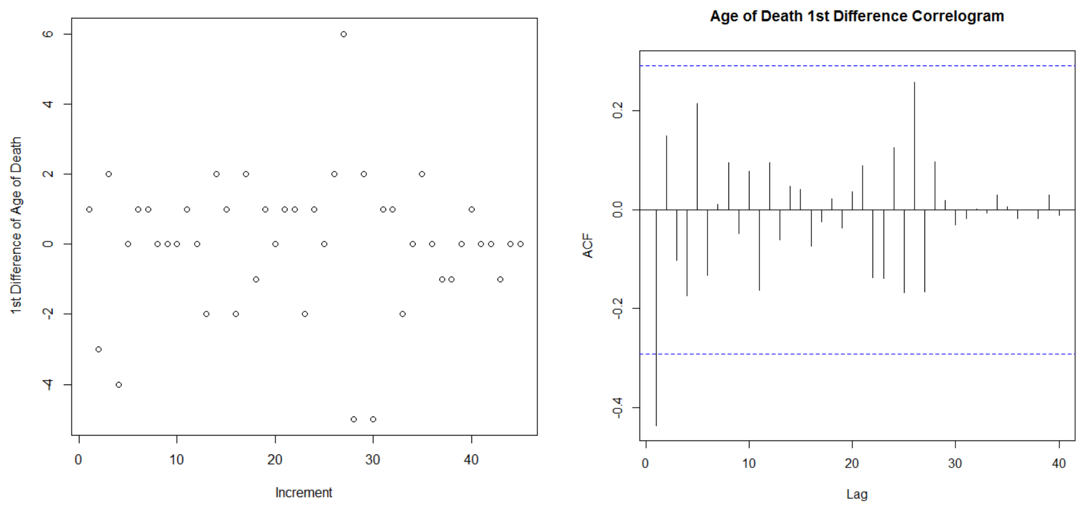

Figure 3. Age of Death 1st difference. Data and Correlogram.

I fit several ARIMA(p,x,q) models using R software. The "arima" function makes it quite easy to do this. One can fit whatever model is needed and then extract model outputs such as the parameters, the Akaike Information Criterion, and the log-likelihood of the fit. I ran several arima(p,0,q) models and several arima(p,1,q) models. I did not attempt any 2nd difference models. From the correlograms and plots above, it appears that a 1st difference will be sufficient, and a 2nd difference may lead to "over-differencing" problems. Based on the AIC, the ARIMA(0,1,1) process appears to fit the data best. (One could also pick this model by looking at the few models with the smallest log-likelihood and picking the one with the fewest parameters on the grounds of parsimony. Essentially, the AIC adjusts does this numerically by penalizing extra parameters.) Note that this is a 1st difference model; the 1st difference of the series is an MA(1) model with a negative theta. Its form is given by the following equation.

$$Y_t - Y_{t-1} = \varepsilon_t - \theta * \varepsilon_{t-1}$$

Note that theta is negative, so a positive residual from the previous period will increase our estimate for the next period, and a negative residual will decrease our estimate for the next period. In other words, if the oldest person in the world died older than we expected, the next "oldest person" should have a higher estimated age of death. This makes some intuitive sense, because if the oldest person is especially old at time of death, it's more likely that there are a lot of very old candidates to take his/her place. If the oldest person dies relatively young, all contenders for the next "world's oldest person" will necessarily be younger still, and it's more likely that the next person in line dies relatively young.

Greg Sollenberger
Time Series
Spring 2012
VEE Student Project

Model	Parameters	AIC	Log-likelihood
ARIMA(1,0,0)	phi_1 = 0.676, intercept = 113.53	186.771	-91.386
ARIMA(0,0,1)	theta_1 = 0.45, intercept = 13.53	200.761	-98.381
ARIMA(2,0,0)	phi_1 = 0.449, phi_2 = 0.324, intercept = 113.56	183.486	-88.743
ARIMA(1,0,1)	phi_1 = 0.875, theta_1 = -0.403, intercept = 113.53	184.295	-89.148
ARIMA(0,0,2)	theta_1 = 0.478, theta_2 = 0.637, intercept = 113.58	188.051	-91.025
ARIMA(0,1,0)	N/A	186.714	-93.357
ARIMA(1,1,0)	phi_1 = -0.430	179.318	-88.659
ARIMA(1,1,1)	phi_1 = 0.075, theta_1 = -0.564	180.962	-88.481
ARIMA(0,1,1)	theta_1 = -0.494	178.983	-88.491
ARIMA(2,1,0)	phi_1 = -0.449, phi_2 = -0.043	181.234	-88.617
ARIMA(0,1,2)	theta_1 = -0.493, theta_2 = -0.01	180.978	-88.489

Table 1. Model outputs for various arima models.

Greg Sollenberger
Time Series
Spring 2012
VEE Student Project

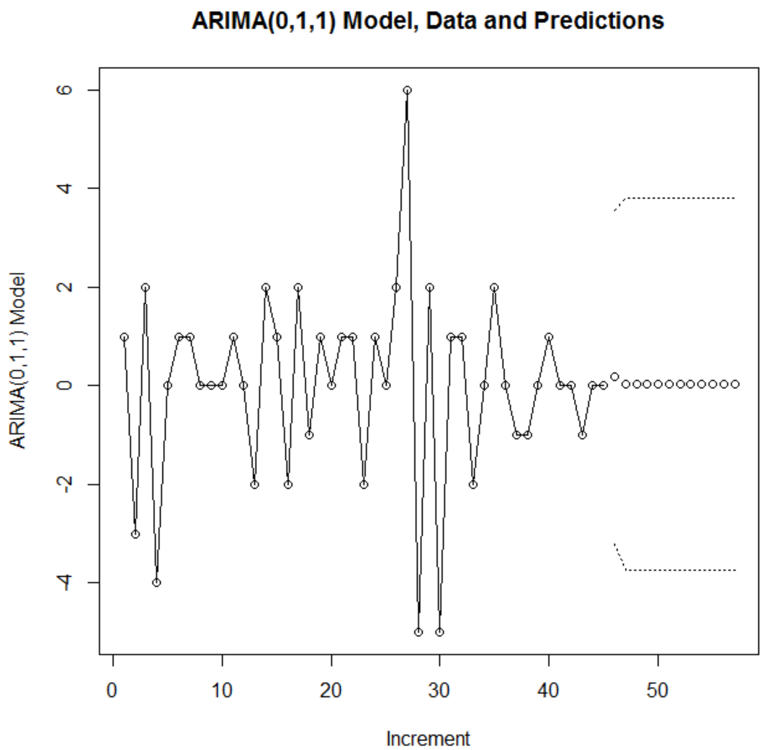

Figure 4. ARIMA(0,1,1) Model, 1st Differences and Predictions

Because this is an MA(1) model, there is a prediction for one period beyond the data, and all subsequent predictions are the same.

Conclusion

An ARIMA(0,1,1) model (aka an IMA(1,1) model, aka an MA(1) model on the first differences) is the appropriate model for fitting the "Age of the World's Oldest Person at Time of Death" data. The theta value for this model is about -0.494. This model implies that a larger than expected first difference predicts a larger first difference in the next period. An interesting refinement would be to run a similar model on the unrounded age of death. The age in years and days of the oldest person is available from the same source I pulled this data from. It would be interesting to know if the conclusion of this study is robust to such refinements. It would also be an interesting exercise to determine if some assumption about the distribution of ages of the world's oldest people, along with some assumptions about their mortality patterns, predicts this model. From such an assumption, one can construct predictions such as, "GIVEN that the age of the world's oldest person is X, the distribution of ages of his successors is the following…" Such an exercise might predict an ARIMA model, possibly the one suggested by this project or possibly some other ARIMA model. That exercise is beyond the scope of this project.

Greg Sollenberger
Time Series
Spring 2012
VEE Student Project

Appendix: R code

This project was heavily dependent on the R language, particularly the "arima" function and some associated function. I will supply some sample code in this appendix. Descriptions of code are preceded by the # character, actual code preceded by the > character.

```
#First, call the TSA library, which has some useful time series functions.

>library(TSA)

#Next, read in the data from a csv datafile.

>Age_Data<-read.csv("//SPIPFILE04/HO-Home/solleng1/My Documents/Time Series
Class/Age_At_Death_Only.csv",header=TRUE)

#Plot the data and a correlogram of the series. The function "acf" plots a correlogram.

>plot(Age_Data['Increment'][,1] ,Age_Data['Age_At_Death'][,1],xlab="Increment",ylab="Age At Death")

>acf(Age_Data['Age_At_Death'], main = "Age At Death Correlogram",lag.max = 40)

#Obtain the 1st difference series, then plot that series and its correlogram.

>Lag_1 <- diff(Age_Data['Age_At_Death'][,1])

>plot(Lag_1, xlab = "Increment",ylab="1st Difference of Age of Death")

>acf(Lag_1,main= "Age of Death 1st Difference Correlogram",lag.max = 40)

#Fit an ARIMA model to the data, then extract the details of the model. This is one sample model.
#Obviously many models were fitted. I will not crowd this paper with all the code necessary to generate
#every single model that was fitted.

>ARIMA_100_Model <- arima(Age_Data['Age_At_Death'][,1],order = c(1,0,0))

>ARIMA_100_Model$coef

>ARIMA_100_Model$loglik

>ARIMA_100_Model$aic

#Finally, plot the data along with a prediction from the appropriate model, with error bars.

>plot(ARIMA_011_Model, xlab = "Increment", ylab = "ARIMA(0,1,1) Model",main = "ARIMA(0,1,1)
Model, Data and Predictions")
```