

Great time series analysis example using the "Ages at Death of the Kings of England" Dataset

Posted on Jul 23, Posted by [Tom Reilly](#) Category [Forecasting](#)

This is a great example of how ignoring outliers can make your analysis go very wrong. We will show you the wrong way and then the right way. A quote comes to mind that said "A good forecaster is not smarter than everyone else, he merely has his ignorance better organized".

A fun dataset to explore is the "age of the death of kings of England". The data comes from the 1977 book from McNeill called "Interactive Data Analysis" as is an example used by some to perform time series analysis. We intend on showing you the right way and the wrong way (we have seen examples of this!). Here is the data so you can try this out yourself: 60,43,67,50,56,42,50,65,68,43,65,34,47,34,49,41,13,35,53,56,16,43,69,59,48,59,86,55,68,51,33,49,67,77,81,67,71,81,68,70,77,56

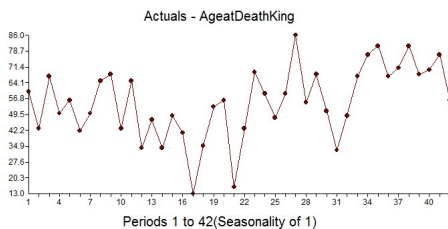
It begins at William the Conqueror from the year 1028 to present (excluding the current Queen Elizabeth II) and shows the ages at death for 42 kings. It is an interesting example in that there is an underlying variable where life expectancy gets larger over time due to better health, eating, medicine, cryogenic chambers???, etc and that is ignored in the "wrong way" example. We have seen the wrong way example as they are not looking for deterministic approaches to modeling and forecasting. Box-Jenkins ignored deterministic aspects of modeling when they formulated the ARIMA modeling process in 1976. The world has changed since then with research done by [Tsay](#), Chatfield/Prothero (Box-Jenkins seasonal forecasting: Problems in a case study (with discussion)" J. Roy Statist soc., A, 136, 295-352), I. Chang, Fox that showed how important it is to consider deterministic options to achieve a better model and forecast.

As for this dataset, there could be an argument that there would be no autocorrelation in the age between each king, but an argument could be made that heredity/genetics could have an autocorrelative impact or that if there were periods of stability or instability of the government would also matter. There could be an argument that there is an upper limit to how long we can live so there should be a cap on the maximum life span.

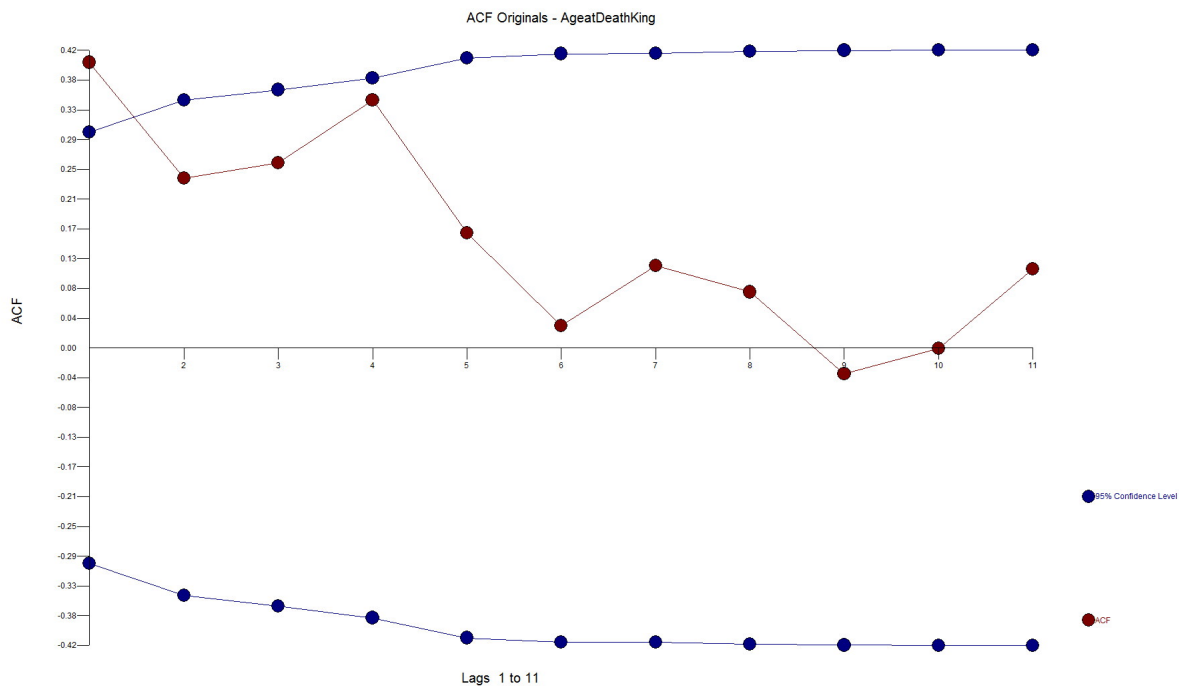
If you look at the dataset knew nothing about statistics, you might say that the first dozen observations look stable and see that there is a trend up with some occasional real low values. If you ignored the outliers you might say there has been a change to a new higher mean, but that

is when you ignore outliers and fall prey to [Simpson's paradox](#) or simply put "local vs global" inferences.

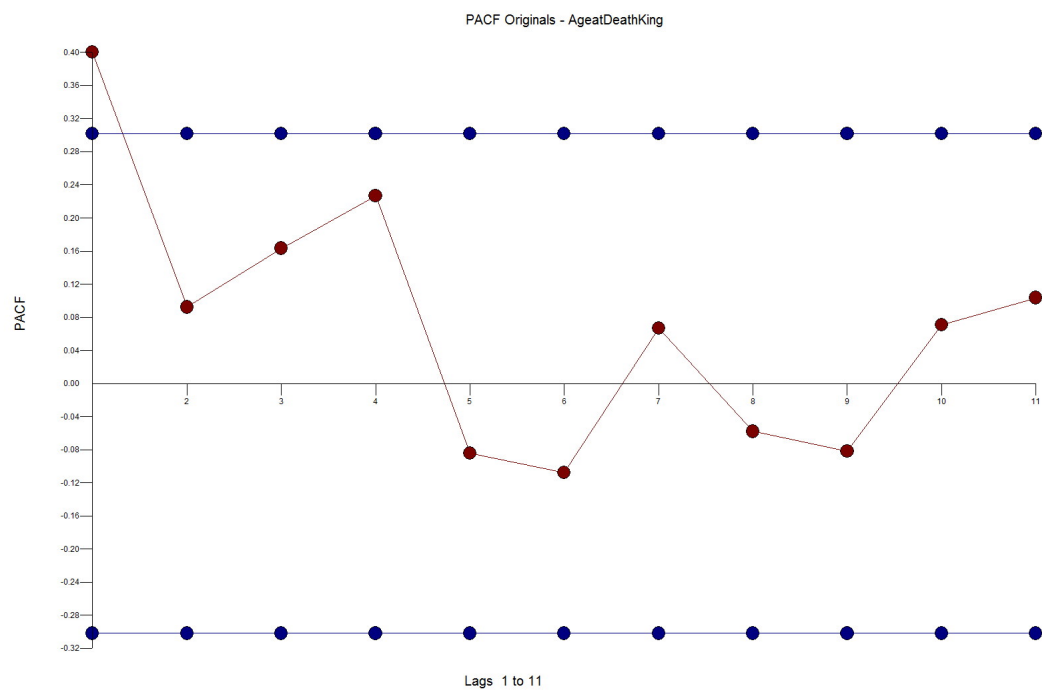
If you have some knowledge about time series analysis and were using your "rule book" on how to model, you might look at the ACF and PACF and say the series has no need for differencing and an AR1 model would suit it just fine. We have seen examples on the web where these experts use their brain and see the need for differencing and an AR1 as they like the forecast.



You might (incorrectly), look at the Autocorrelation function and Partial Autocorrelation and see a spike at Lag 1 and conclude that there is autocorrelation at lag 1 and then should then include an AR1 component to the model. Not shown here, but if you calculate the ACF on the first 10 observations the sign is negative and if you do the same on the last 32 observations they are positive supporting the "two trend" theory.

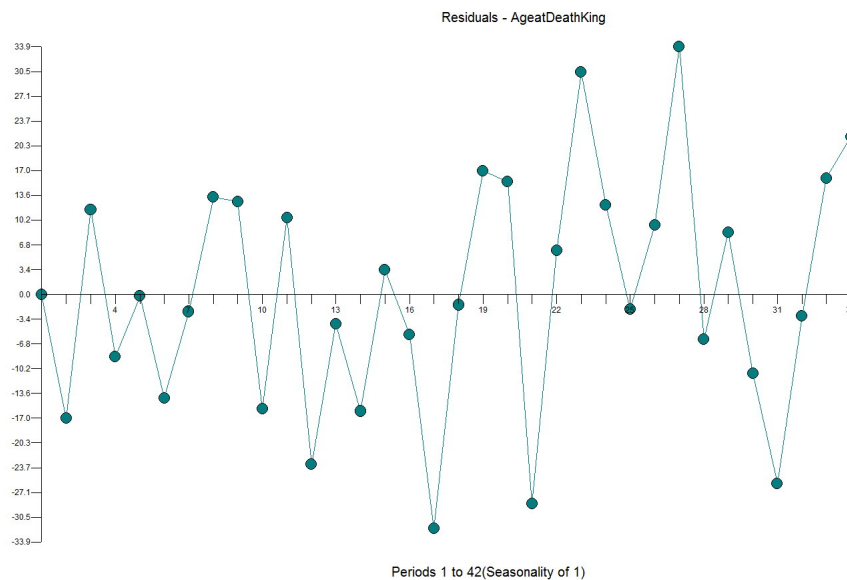
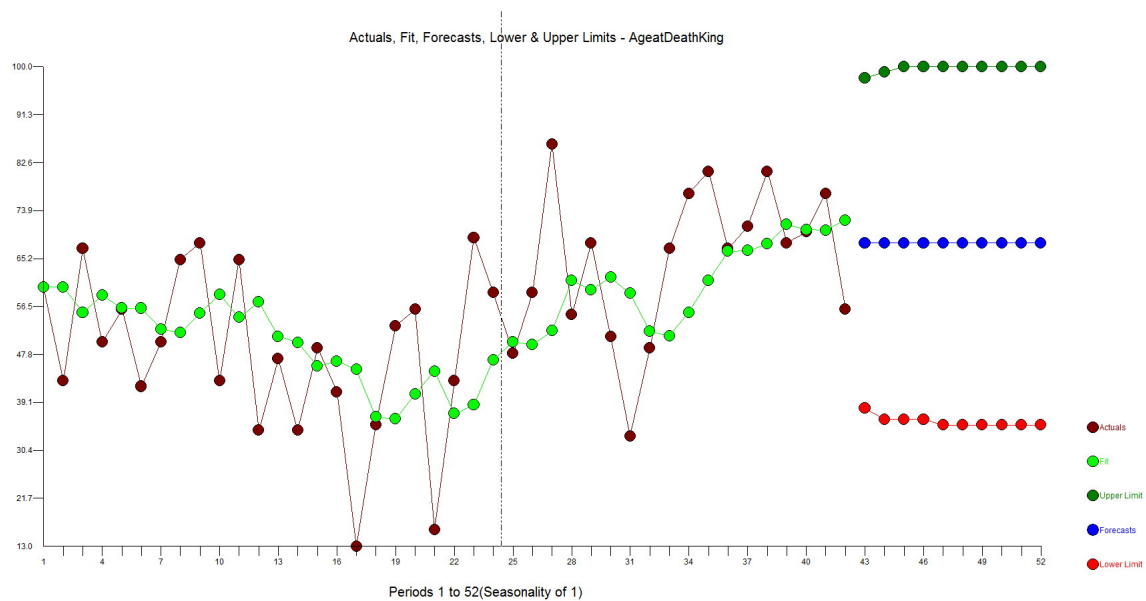


The PACF looks as follows:



Here is the forecast when using differencing and an AR1 model.

With a QF status, a QF graduate and leader can



MODEL COMPONENT	LAG	COEFF	STANDARD	P	T
Intercept		0.000000	0.000000	0.000000	0.000000
Constant		0.000000	0.000000	0.000000	0.000000
AR(1)	1	0.000000	0.000000	0.000000	0.000000
AR(2)	2	0.000000	0.000000	0.000000	0.000000
AR(3)	3	0.000000	0.000000	0.000000	0.000000
AR(4)	4	0.000000	0.000000	0.000000	0.000000
AR(5)	5	0.000000	0.000000	0.000000	0.000000
AR(6)	6	0.000000	0.000000	0.000000	0.000000
AR(7)	7	0.000000	0.000000	0.000000	0.000000
AR(8)	8	0.000000	0.000000	0.000000	0.000000
AR(9)	9	0.000000	0.000000	0.000000	0.000000
AR(10)	10	0.000000	0.000000	0.000000	0.000000
AR(11)	11	0.000000	0.000000	0.000000	0.000000
AR(12)	12	0.000000	0.000000	0.000000	0.000000
AR(13)	13	0.000000	0.000000	0.000000	0.000000
AR(14)	14	0.000000	0.000000	0.000000	0.000000
AR(15)	15	0.000000	0.000000	0.000000	0.000000
AR(16)	16	0.000000	0.000000	0.000000	0.000000
AR(17)	17	0.000000	0.000000	0.000000	0.000000
AR(18)	18	0.000000	0.000000	0.000000	0.000000
AR(19)	19	0.000000	0.000000	0.000000	0.000000
AR(20)	20	0.000000	0.000000	0.000000	0.000000
AR(21)	21	0.000000	0.000000	0.000000	0.000000
AR(22)	22	0.000000	0.000000	0.000000	0.000000
AR(23)	23	0.000000	0.000000	0.000000	0.000000
AR(24)	24	0.000000	0.000000	0.000000	0.000000
AR(25)	25	0.000000	0.000000	0.000000	0.000000
AR(26)	26	0.000000	0.000000	0.000000	0.000000
AR(27)	27	0.000000	0.000000	0.000000	0.000000
AR(28)	28	0.000000	0.000000	0.000000	0.000000
AR(29)	29	0.000000	0.000000	0.000000	0.000000
AR(30)	30	0.000000	0.000000	0.000000	0.000000
AR(31)	31	0.000000	0.000000	0.000000	0.000000
AR(32)	32	0.000000	0.000000	0.000000	0.000000
AR(33)	33	0.000000	0.000000	0.000000	0.000000
AR(34)	34	0.000000	0.000000	0.000000	0.000000
AR(35)	35	0.000000	0.000000	0.000000	0.000000
AR(36)	36	0.000000	0.000000	0.000000	0.000000
AR(37)	37	0.000000	0.000000	0.000000	0.000000
AR(38)	38	0.000000	0.000000	0.000000	0.000000
AR(39)	39	0.000000	0.000000	0.000000	0.000000
AR(40)	40	0.000000	0.000000	0.000000	0.000000
AR(41)	41	0.000000	0.000000	0.000000	0.000000
AR(42)	42	0.000000	0.000000	0.000000	0.000000
AR(43)	43	0.000000	0.000000	0.000000	0.000000
AR(44)	44	0.000000	0.000000	0.000000	0.000000
AR(45)	45	0.000000	0.000000	0.000000	0.000000
AR(46)	46	0.000000	0.000000	0.000000	0.000000
AR(47)	47	0.000000	0.000000	0.000000	0.000000
AR(48)	48	0.000000	0.000000	0.000000	0.000000
AR(49					

#	MODEL COMPONENT	LAG (BOP)	COEFF	STANDARD ERROR	F VALUE	T VALUE
1	CONSTANT		33.0	8.25	.0003	4.01
2	Autoregressive-Factor # 1	1	.401	.143	.0078	2.80

DIAGNOSTIC CHECK #3: THE TIAO TEST FOR CONSTANCY OF THE MEAN OF THE RESIDUALS

(ASSUMING THAT THE ORIGINAL SERIES FOLLOWS AN ARMAX PROCESS)

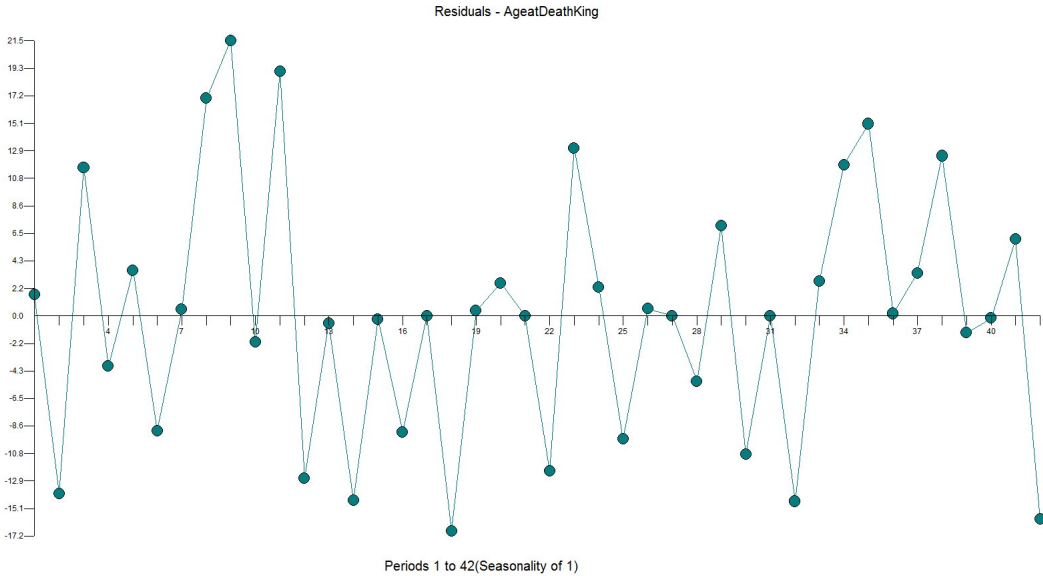
MODEL COMPONENT	LAG	COEFF	STANDARD	P	T
ADDITIVE					
PULSE					
NA	31				
	31				
	-31,307				
					.0487

#	MODEL COMPONENT	LAP (SOP)	CORR?	STANDARD ERROR	VALUE	VALUE
1	CONSTANT		59.8	6.01	.0000	9.95
2	INPUT SERIES X1 I-T00001		TIME		2	
3	20omega (Input) -Factor # 1	0	-1.48	.755	.0582	-1.96
4	INPUT SERIES X2 I-T00011		TIME		12	
5	30omega (Input) -Factor # 2	0	2.32	.856	.0105	2.70
6	INPUT SERIES X3 I-P00021		PULSE		22	
7	40omega (Input) -Factor # 3	0	-38.3	9.87	.0004	-3.88
8	INPUT SERIES X4 I-P00031		PULSE		18	
9	50omega (Input) -Factor # 4	0	-37.9	9.94	.0005	-3.81
10	INPUT SERIES X5 I-P00041		PULSE		32	
11	60omega (Input) -Factor # 5	0	-29.6	9.89	.0050	-3.00
12	INPUT SERIES X6 I-P00027		PULSE		28	
13	70omega (Input) -Factor # 6	0	26.7	9.85	.0103	2.71

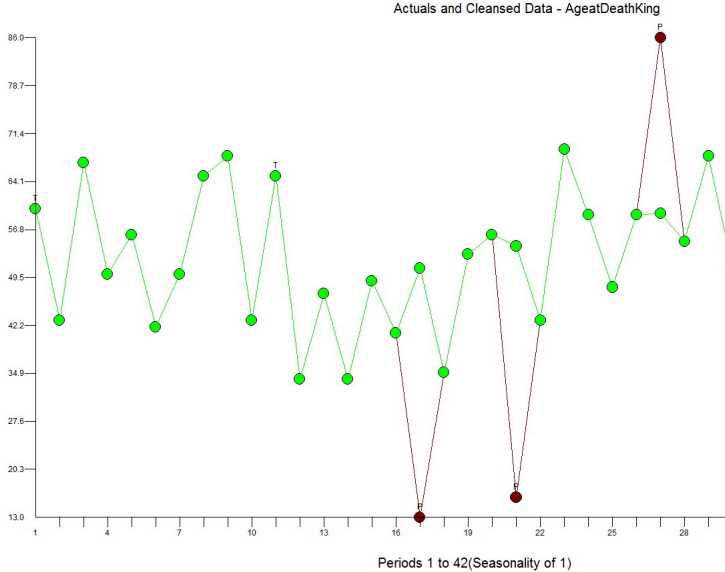
If you consider deterministic variables like outliers, level shifts, time trends

Elisabeth I, 37 years old. Peter II, 4 years old. James I, 49 years old. Charles I, 49 years old. Charles II, 49 years old. James II, 49 years old. George I, 49 years old. George II, 49 years old. George III, 49 years old. George IV, 49 years old. William IV, 49 years old. Victoria, 49 years old. Edward VII, 49 years old. George V, 49 years old. Edward VIII, 49 years old. George VI, 49 years old. Elizabeth II, 49 years old. This

the data, and instead of this, you get a flat line. This is the flat fitting line that is the result of



early is the only Autobox clearing history is coming. It's when you correct for outliers that you can



Tags: Tagged in: outliers for time series analysis outliers sas outliers sysstat