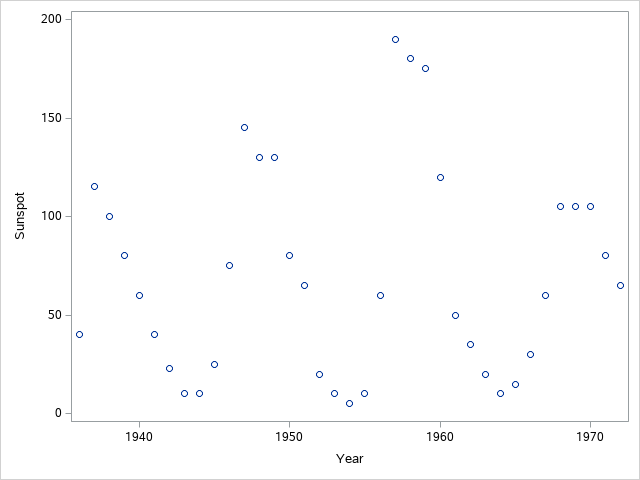
For live session we are going to discuss the general idea of dealing with correlated residuals.  Attached is an excel file that includes Melanoma and Sunspot data over time.  A quick search on google for sunspots may be helpful for a reference. The melanoma variable is the rate of melanoma occurrences. I will discuss melanoma in class but it is optional for the prelive assignment

Sunspot Data

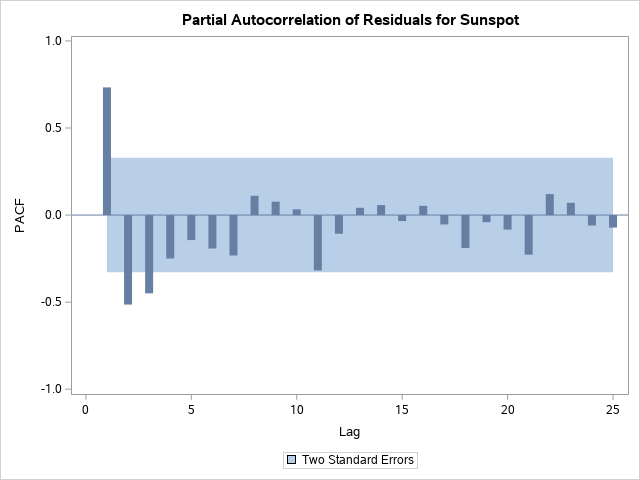
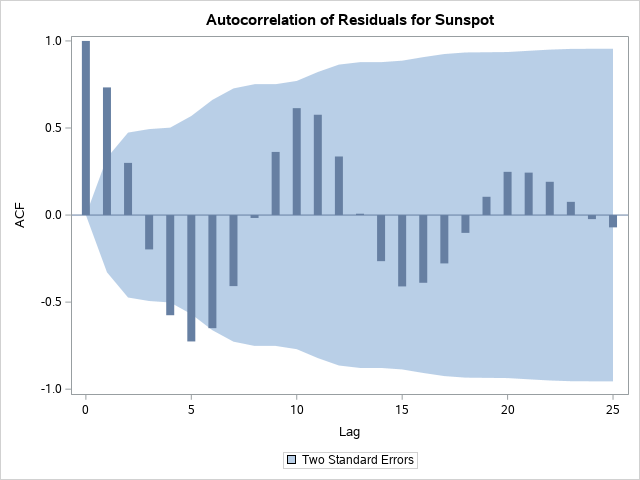
1. The sunspot data has a cyclical behavior. What we are going to do here is explore how an Autoregressive model can actually capture the cyclical behavior without any predictors present.
   * Plot Sunspot versus Years. Visually do you think this time series is stationary. Again do your best based on what you got out of the videos. We will take a deeper dive in class.

**This data does seem to be stationary/periodic/cyclical/serially correlated.**

**Could be modeled by some sort of sinusoidial function rather easily**



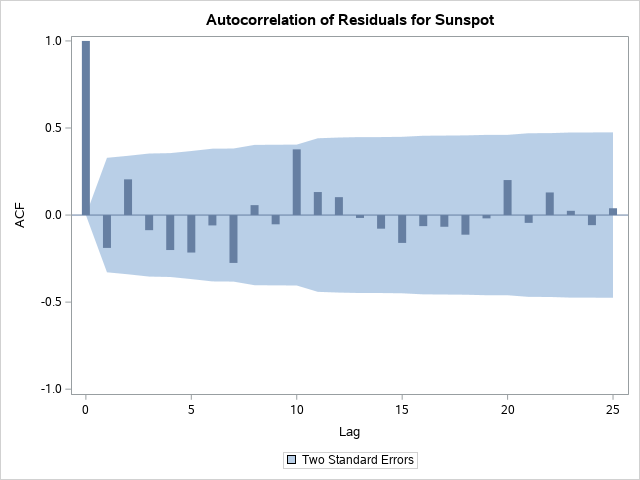
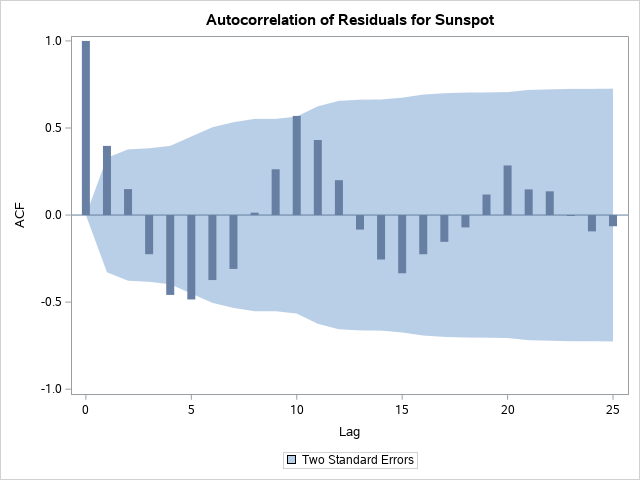
* + Using the code below as an example, fit a simple regression model to Sunspot with just an intercept (model sunspot= / nlag= in SAS) Comment on the ACF and PCF plots. Note that with no nlag option, it is just fitting a regression model with an intercept and nothing more.

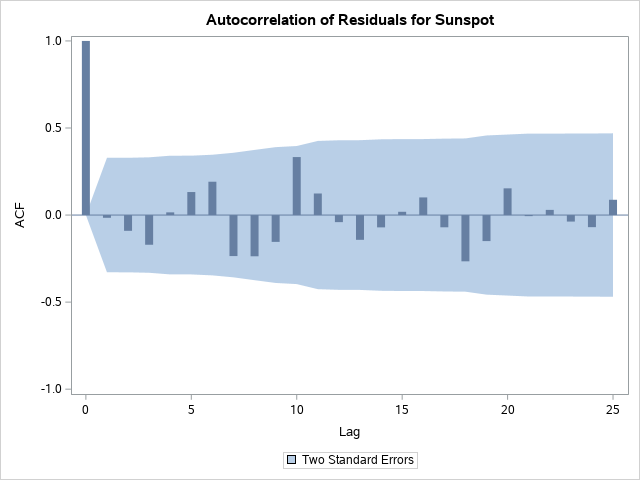
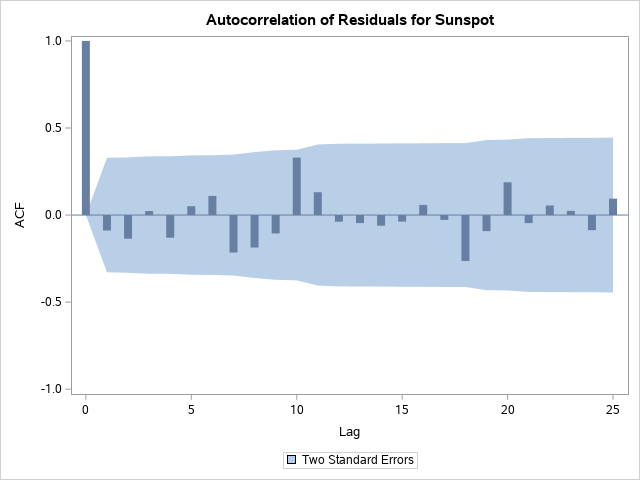


**The ACF plot shows large changes in serial correlation – varying from positive to negative – also has a very wide prediction interval as the number of lags increases.**

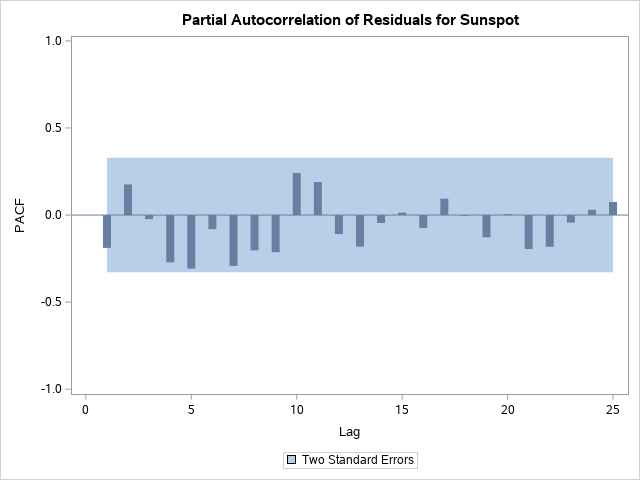
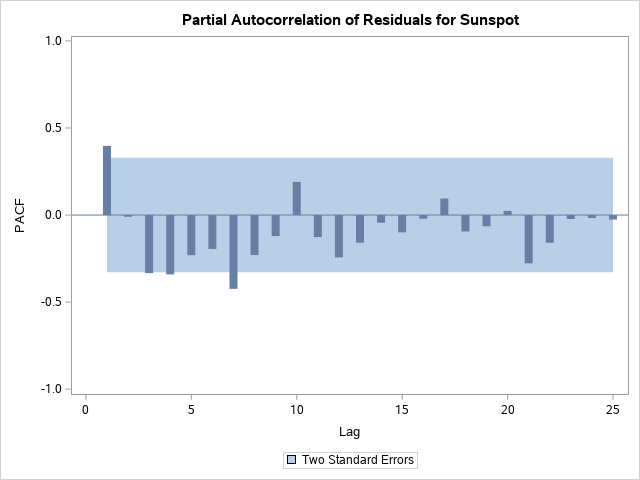
**The PACF plots strips the effects of intermediate lags and generally shows that the main serial effect can be captured by one**

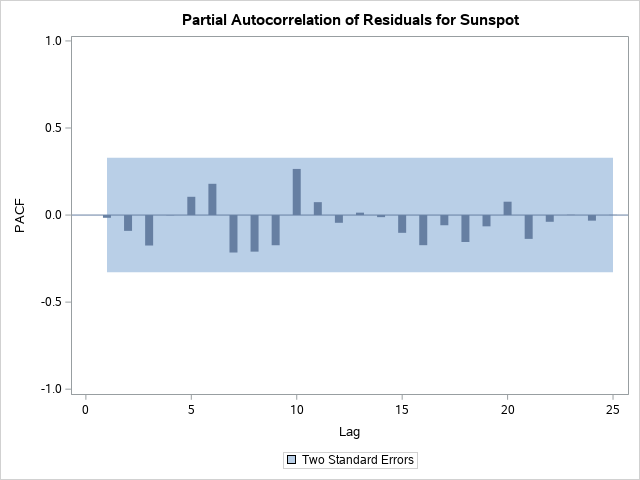
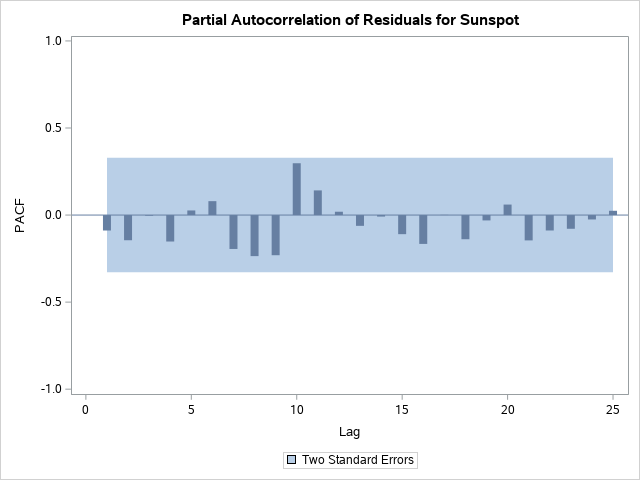
* + Fit an AR(1), AR(2), AR(3), and AR(4) model by specifying the nlag option to 1,2,3, or 4.
    1. Examine and compare the ACF and PACF plots for each model. What do you make of them, say AR(1) model compared to the AR(4)?



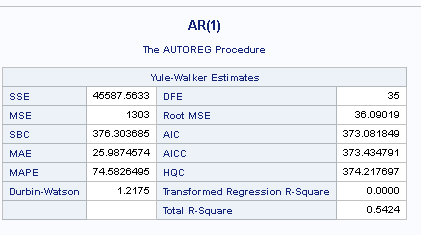


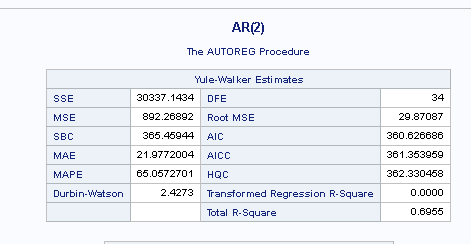
**Of the four ACF plots – AR(3) (bottom left) has the lowest values for autocorrelation for each lag. AR(1) is consistently higher in autocorrelation across all lags than AR(4) and consequently has a much wider prediction band.**

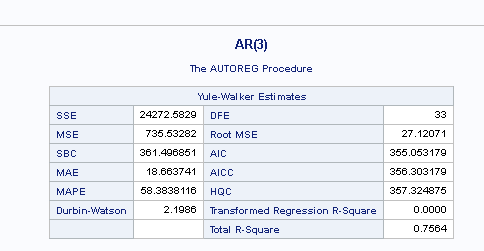


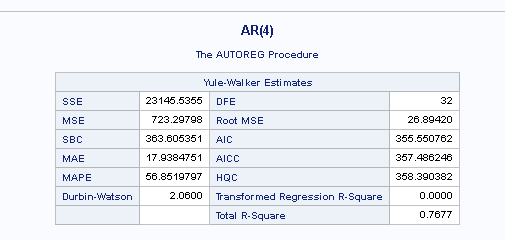


* + 1. Locate the AIC statistic for each of the models and compare them



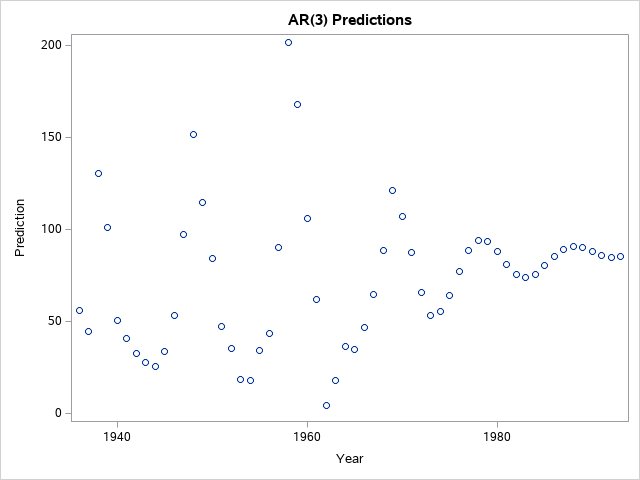






**AR(3) has the lowest AIC**

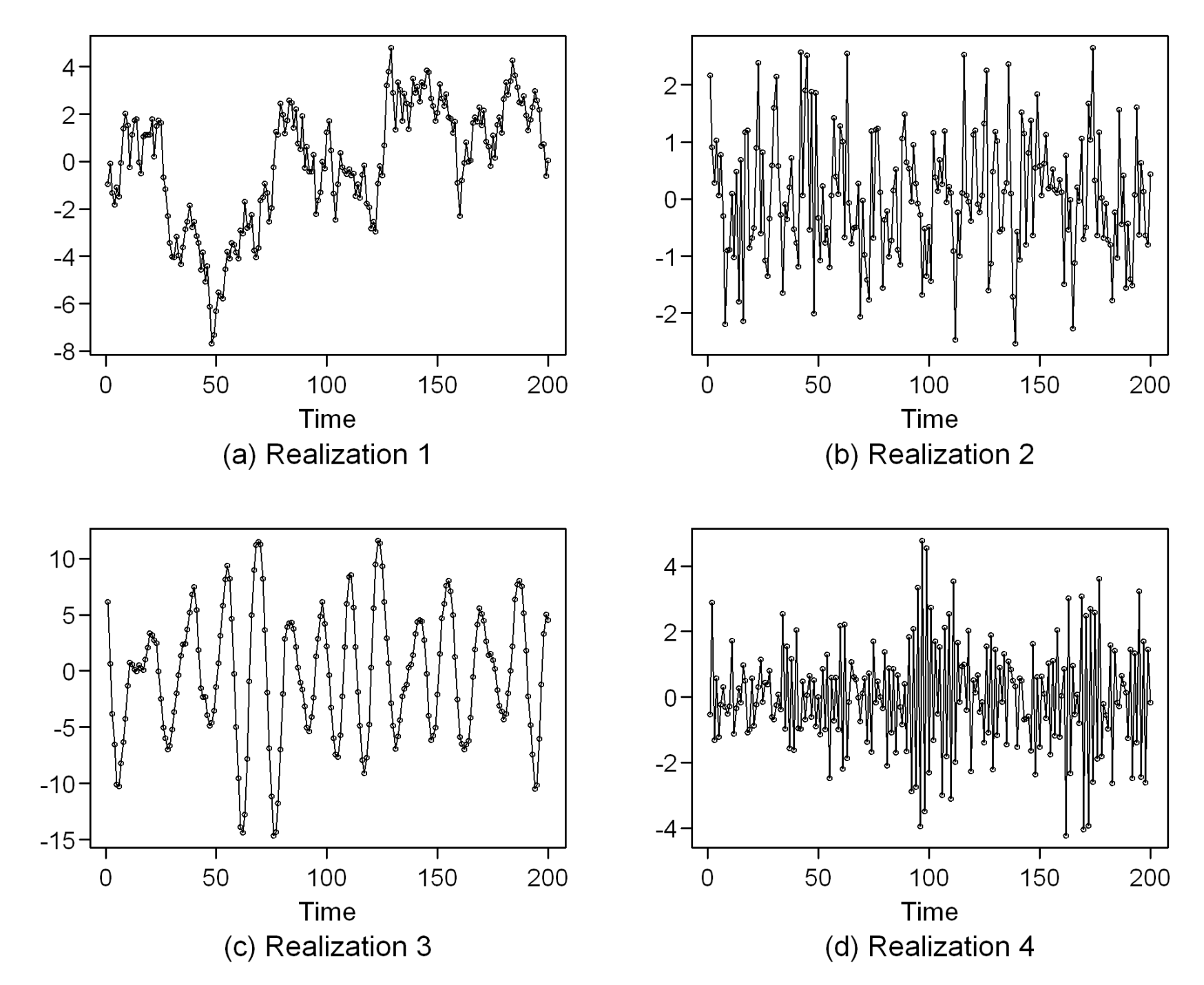
* + Try to forecast the next 10-20 years using the model that has the lowest AIC. Once you have the predictions, try to add them to your plot from part A so we can see what is going on. If you are stumped on how to predict future values of the time series, check out the Output and Predicted statement and options within Proc Autoreg or some of the examples:

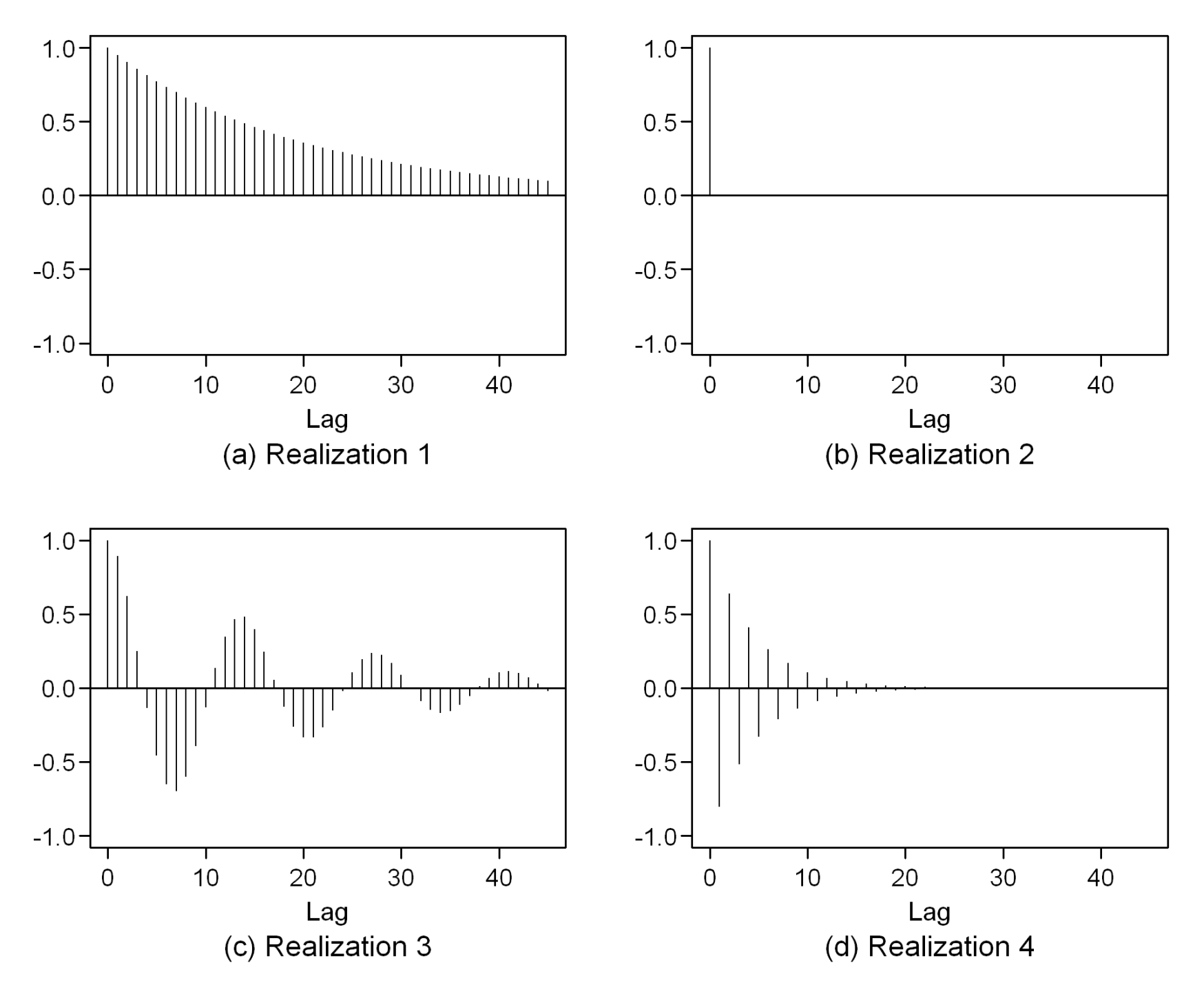
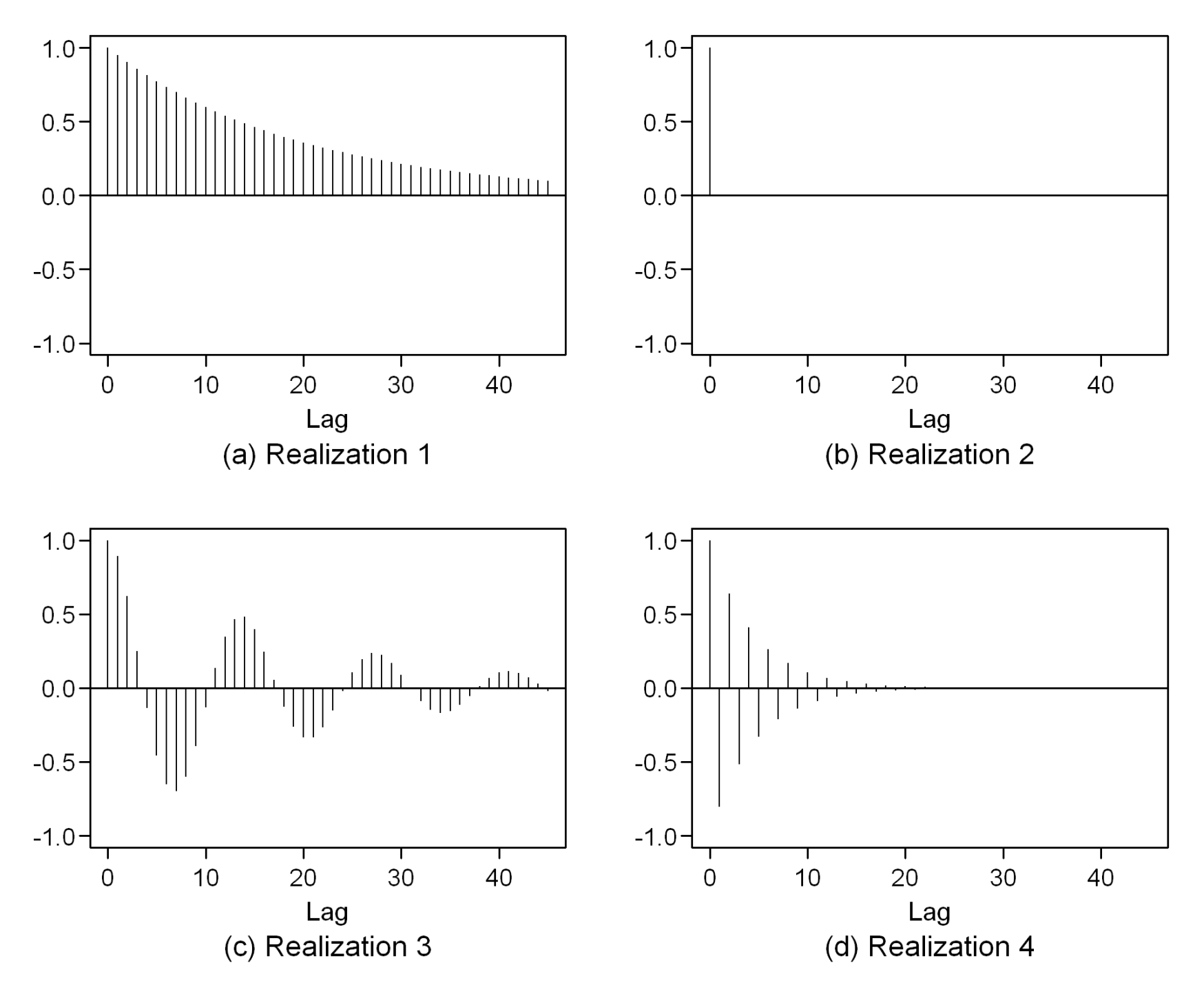
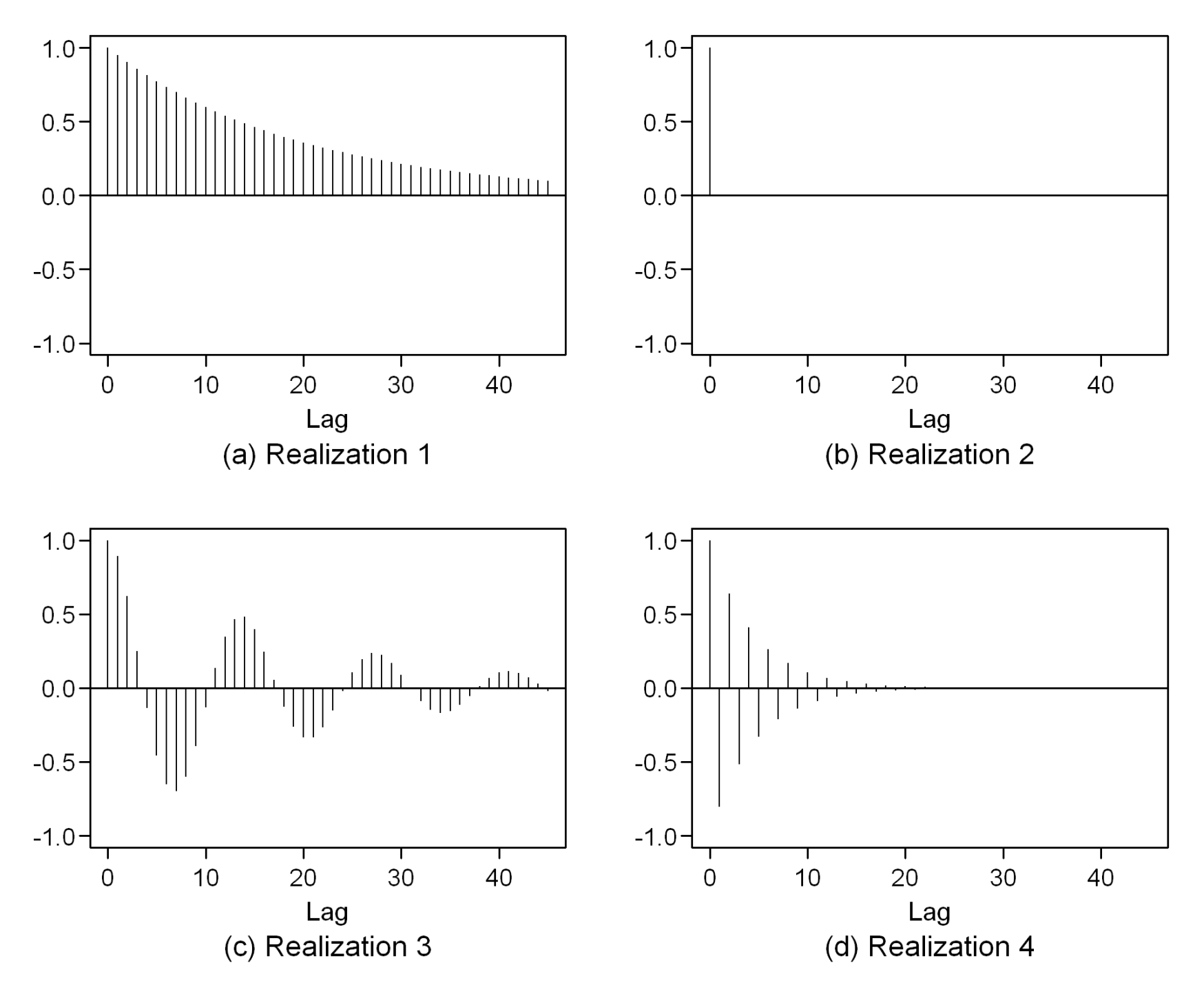
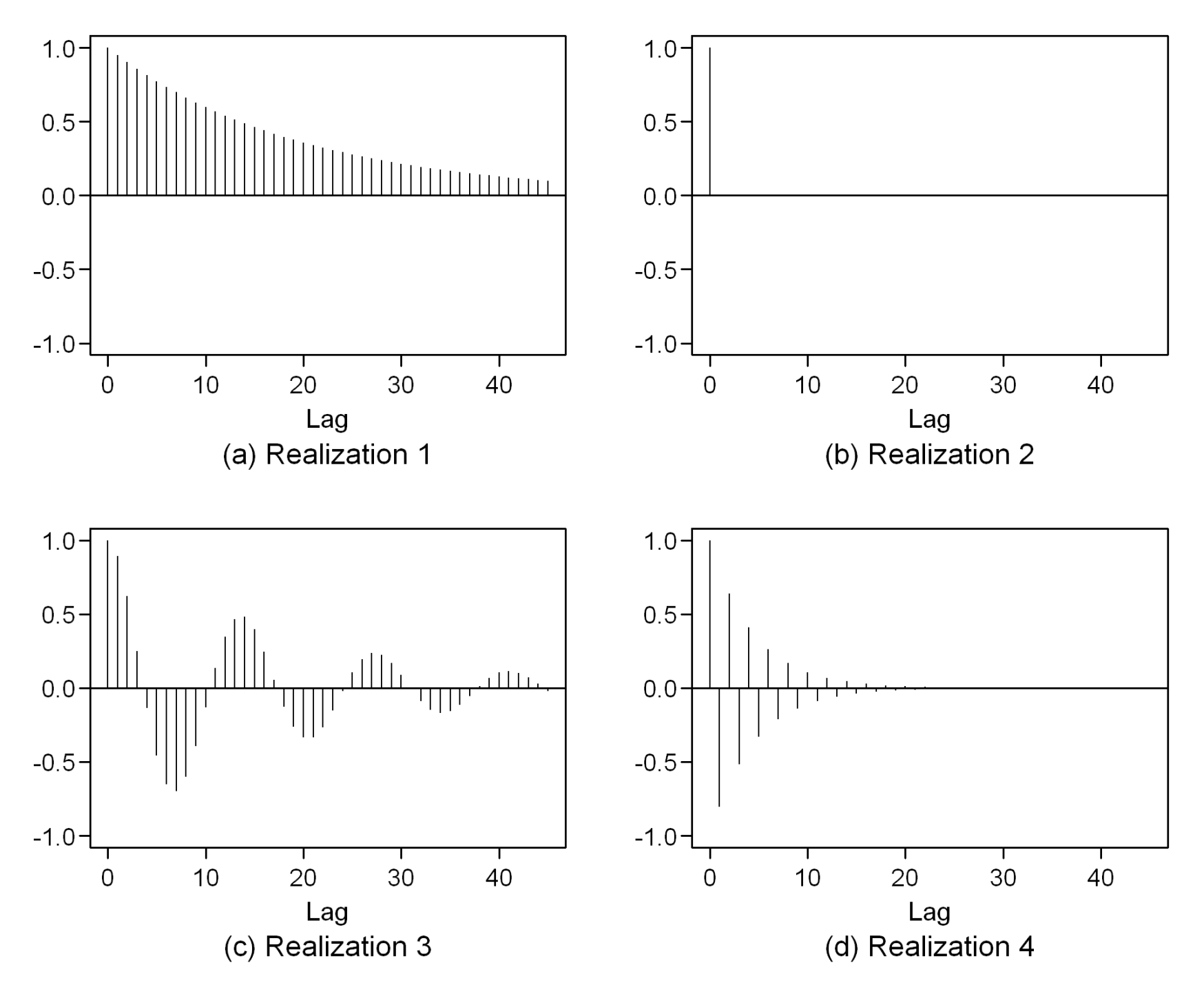


Give this next question a good college try. I recognize this one is a little tough without hearing me speak directly on it.

The autocorrelation plot tool is an extremely useful tool for diagnosing fits for time series models. My experience in the MSDS program is that this graph and the idea of stationary gives the most fits.

* + Go to my power point slide deck and take a look at slides 9-28 in presentation mode so you can see the automations. I will discuss and record these slides during office hours on Saturday so check it out if you want to hear commentary. The point is to illustrate more step-by-step how the acf graph is created.
  + Once you feel a little bit better about the acf plot. Match the following four time series with their corresponding acf. Zoom in





1

4

3

2

**A = 3**

**C = 2**

**D=1**

**B=4**