



Effect of Persistence on the Significance of Kendall Tau

Dr. K. H. Hamed

Cairo University, Faculty of Engineering, Irrigation and Hydraulics Department, Giza, Egypt
(hamedkhaled@hotmail.com; khamed@squ.edu.om)

Although persistence in natural data is generally admitted, its effect on the significance of various statistical tests has not been extensively studied, and it is sometimes overlooked or simply ignored in practice. In particular, modified tests that are robust in the presence of persistence are still lacking. In many situations, need may arise to test the significance of correlation between two observed natural time series. Although the estimation of the classical product-moment correlation coefficient is a straightforward task, classical significance testing depends on two major assumptions. The first assumption is that the data are Gaussian, which is violated by many natural time series. In this case, a distribution-free measure of correlation, such as Kendall tau should be used. The second, and often overlooked assumption, is that the observations in each time series are not autocorrelated, which is also violated by most natural time series. Similar to the case of trend testing, which has received some attention recently, the existence of persistence increases the chance of significant correlation when the two series are actually uncorrelated. In this paper, the effect of both short- and long-term persistence (STP and LTP) on the distribution of Kendall tau as a distribution-free measure of correlation between two time series is investigated, and an exact expression for its variance under persistence is derived. The implications of these results for natural data are illustrated through the study of spurious correlation between a 133-year Nile flow time series from A.D. 1871 to A.D. 2003 and independent segments of a reconstruction of the North Hemisphere temperature time series from A.D. 1000 to A.D. 1980, both of which exhibit LTP. It is shown that significant spurious correlation between completely unrelated segments of the two time series is on average three times as common as in random series of the same length at the 10% significance level, which is consistent with the theoretical results. It is also shown that accounting for LTP by using the correct variance of the test statistic effectively reduces the probability of false identification to near its expected nominal rate of 10%. Similar results were obtained at other significance levels.