Testing the Significance of the Linear Regression Coefficients: Exploring Some Estimators for the Autocorrelation Function

M. Rosário Ramos[1,3] and Marco Costa[2,3]

[1] Universidade Aberta, Portugal; ramos@cii.fc.ul.pt

[2] Escola Superior de Tecnologia e Gestão de Águeda, Universidade de Aveiro, Portugal; marco@ua.pt

[3]CMAF - Centro de Matemática e Aplicações Fundamentais da Universidade Lisboa

Abstract. This work addresses the problem of testing the significance of the slope of a linear trend with and without an eventual seasonal effect. It is assumed that the error term follows an AR(1), and that the autoregressive parameter is unknown. The autoregressive parameter is obtained through some competing estimators, namely, a parametric version, a modified Kendall's correlation coefficient, and another non-parametric counter part developed earlier in the context of the state space models. The accuracy of the estimation of this parameter is also analyzed. The performance of the tests considering the three estimators simultaneously is compared through a Monte Carlo simulation study under different assumptions. The study is extended in order to compare the slopes of two or more periods in the same time series.

Keywords: Trend tests, Ordinary least squares, Autocorrelation, Distribution-free estimators.
PACS: 05.45.Tp, 02.50.Ey.

INTRODUCTION

The significance of an increasing or decreasing change in environmental variables within time is usually based on the test for the linear trend model. The significance test for the slope obtained by Ordinary Least Squares method is very common; some nonparametric tests for trend have been proposed and require few assumptions about the data structure. The Mann-Kendall test for trend is based on Kendall's correlation coefficient and it was originally proposed for a time series of independent observations. This test has shown to have good statistical properties for independent observations and it is a distribution-free test.

The problem with standard statistical tests arises when one is in the presence of the following situations: serial correlation, short or moderate time series, missing values, a highly skewed parent distribution.

This work attempts to focus on trend tests under a model in which the time series Y_t is well described by a linear function of time, can have a seasonal effect, and the error term is an autoregressive AR(1) process.

In previous work [4] and in papers from other authors, it was observed that the poor estimation of the autocorrelation is one important cause of the weak performance of both types of tests, parametric and nonparametric.

Some new and old estimators are explored and compared. The effect of this estimation on the two trend tests is analyzed: the parametric test for the least squares estimator of the slope, and a non parametric test, the Mann-Kendall test, based in the correlation coefficient Kendall's Tau - calculated between the observations and time.

TREND TESTS

Consider a model consisting of a linear trend in time, a seasonal effect plus a series of error terms where these follow a first order autoregressive stationary process, which can be expressed by

$$y_t = a + bt + \sum_{i=1}^{g} x_{ti} s_i + \varepsilon_t ,$$

where g is the number of seasons, s_i are the seasonal coefficients and

$$x_{ti} = \begin{cases} 1 & \text{if} \quad t = g(j-1)+i \\ 0 & \text{other case} \end{cases}, \text{ for some } j = 1,\dots,n .$$

For a non-seasonal model all the x_{ti} will be zero. In the case of a series with two periods of time with different slopes, the simple linear model is

$$y_t = b_0^{(1)} x_{1,t} + b_1^{(1)} t x_{2,t} + b_0^{(2)} x_{3,t} + b_1^{(2)} t x_{4,t} + \varepsilon_t ,$$

where $^{(1)}$ and $^{(2)}$ represent the period.

When the series ε_t follow an AR(1) process, the covariance matrix of the Ordinary Least Squares estimators will be given by the usual matrix $\sigma^2 (X'X)^{-1}$ with each element divided by $(1-\phi)^2$. Since both σ^2 and ϕ are usually unknown they have to be estimated from data.

The Mann- Kendall test was initially proposed by [1], [3] and it is based on the test for Kendall's τ correlation coefficient. It is used for detecting monotonic changes over time. For a time series y_1, y_2, …, y_n, the Mann-Kendall test statistic is:

$$S_n = \sum_{i<j} \operatorname{sgn}\left(y_j - y_i\right) \text{ where } \operatorname{sgn}\left(x\right) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases} .$$

Under the null hypothesis of no trend and independence of the series terms, the variance of the Mann-Kendall statistic is given by

$$Var\left[S_n\right] = \frac{n\left(n-1\right)\left(2n+5\right)}{18}$$

and S_n converges to normality.

Assuming a linear regression model with AR(1) errors, the performance of the trend tests will be compared through a simulation study under different assumptions.

ESTIMATION OF THE AUTOREGRESSIVE PARAMETER

Aiming to find estimators of the autoregressive parameter, ϕ having good statistical properties and adequate to the insert in the trend tests we study some alternatives for the parametric and nonparametric tests, as follows

i) the first choice is the very well known consistent estimator, the first order autocorrelation of the residual series

$$\hat{\phi} = \frac{\sum_{t=1}^{n-1} e_{t+1} e_t}{\sum_{t=1}^{n-1} e_t^2}$$

To perform the trend test, the variance σ^2 is obtained, as usual, by removing the estimated trend, and using the residual series.

ii) a second estimator of the autoregressive parameter is a distribution-free estimator based on the covariance structure of the residual series suggested in [2],

$$\hat{\phi} = \frac{\sum_{k=1}^{\ell} \hat{\gamma}_e(k+1)\hat{\gamma}_e(k)}{\sum_{k=1}^{\ell} \hat{\gamma}_e^2(k)}.$$

The number ℓ is the best choice for the number of equations necessary to estimate ϕ and must be based on the sample size and $\hat{\gamma}_e(k)$ is the sample auto-covariance function of the residual series. In original work [2], it is suggested to take $\ell = 45,80,60,50$ according to sample size $n = 50,100,200$ and 500.

When performing the trend test, given that some properties assumed are asymptotic, it may be necessary to access their validity in each situation.

iii) [6] proposed to use the Mann-Kendall statistic to test for trend applying a correction on the variance for the independent case based in the concept of Effective Sample Size (ESS). So, the adjusted variance of the Mann-Kendall statistic S_n is given by

$$Var^*\left[S_n\right] = Var\left[S_n\right] \times \frac{n}{n^*} \quad \text{where} \quad \frac{n}{n^*} = 1 + 2\frac{\phi^{n+1} - n\phi^2 + (n-1)\phi}{n(\phi-1)^2}$$

and n^* is the ESS for the AR(1) case.

Again, since ϕ is often unknown we may obtain an estimate through the following algorithm suggested in [5] and which follows from the same procedure used in i).

Estimate the slope using the nonparametric Theil-Sen estimator

$$\tilde{\beta} = median_{i<j}\left\{\frac{y_j - y_i}{t_j - t_i}\right\},$$

and then remove the estimated trend from the original series.

Estimate ϕ as the sample autocorrelation of the resulting residual series, e_1, e_2,…, e_t, that is,

$$\hat{\phi} = \frac{\sum_{t=1}^{n-1}(e_{t+1} - \bar{e})(e_t - \bar{e})}{\sum_{t=1}^{n-1}\left(e_t - \bar{e}\right)^2}.$$

SIMULATION STUDY

A Monte Carlo simulation study was carried out in order to assess the relative performance of both trend tests, parametric and non parametric, and the quality of the estimation of the autoregressive parameter. The study includes the distribution of the test statistics, the evaluation of the empirical power and significance level of the tests, the effect of estimating the autocorrelation on the trend tests, and the statistical properties of the different estimators of autocorrelation regarding a range values for ϕ and for de shape of the parent distribution, symmetric and asymmetric.

REFERENCES

1. H.B. Mann, Nonparametric tests against trend, *Econometrica* **13**, 245-249 (1945).
2. M. Costa and T. Alpuim, Parameter estimation of state space models for univariate observations *J. Stat. Plan. Infer.* **140**, 1889-1902 (2010).
3. M.G. Kendall, Rank Correlation Methods; 1955; Griffin, London

4. M. R. Ramos, *Testes de tendência com aplicação à avaliação da qualidade da água. Estudo comparativo entre métodos paramétricos e não paramétricos*, *PhD Thesis*, Universidade de Lisboa, 2006.
5. S. Yue and C. Y. Wang, The Mann-Kendall test modified by Effective Sample Size to detect trend in serially correlated hydrological series, *Water Resour. Manag.* **18**, 201-218 (2004).