Trend tests: a tendency to resampling

Conference Paper · June 2014

2 authors:

Maria Do Rosário Ramos
Universidade Aberta

Clara Cordeiro
Universidade do Algarve

Trend tests: a tendency to resampling

M.Rosário Ramos[1] and Clara Cordeiro[2]

[1] Universidade Aberta, Portugal and Centro de Matemática e Aplicações Fundamentais(CMAF), Universidade de Lisboa, 1749-016 Lisboa, Portugal
[2] Faculdade de Ciência e Tecnologia da Universidade Algarve and Centro de Estatística e Aplicações (CEAUL), Universidade de Lisboa, Portugal

Abstract. Trend analysis is an important problem in time series. Many studies have been developed to investigate this issue, with special attention to its application to environmental and hydrological time series. The presence of autocorrelation and missing observations affects the significance and power of trend tests, parametric or non-parametric. This study assesses the performance of two trend tests, t-test and the Mann-Kendall, through an appropriate resampling technique. A new procedure based on subsampling is proposed in order to assure good statistical properties of these tests. A comparison was established between this new approach and others already developed, such as bootstrap-based tests. In order to evaluate the performance of the new method, a simulation study is conducted considering a set of underlying slopes, different values of autocorrelation and different fractions of randomly missing data. The order of autocorrelation structure is estimated by the best fitting model obtained through the Akaike information criterion. Inspection of the data to detect missing observations is required, before applying the trend tests. In case of missing observations, their estimation and replace is performed by an imputation method available in R software.

Keywords: Trend tests · time series · sieve bootstrap · subsampling · serial correlation · missing values

1 Introduction

Detecting the presence of a trend is an important issue in time series analysis. Research related to this subject has been carried out and particular attention was given to their application to environmental and hydrological time series. The standard methods used for detecting trends are the parametric t test test and the nonparametric Mann-Kendall test. Besides its good statistical properties, in case of independent data and a small amount of missing observations, a suitable modification to accommodate the effect of autocorrelation is needed. The original tests results in false rejections of the null hypothesis of no trend, i.e. increases the size of the test (α) and affects also the power of the tests. Missing observations can lead to similar consequences. Papers of van Belle and Hugues [1], Hirsh et al.[2], Lettenmaier [3], El Shaarawi [4, 5], Harcum et al.[6], Sun and Pantula [7], Woodward and Gray [8], or Noguchi et al.[9], address this issue. The bootstrap is a very popular methodology and widely used in many research fields. In [9], the sieve bootstrap was chosen. Now, besides this resampling scheme, also a new one is explored, the subsampling. A simulation study is developed to deal with autocorrelation and missing observations, in which the performance of using the two resampling techniques and trend tests is compared and analysed. In our case, one considers that the underlying model is a time series with a linear trend in time plus an error term that is a weak stationary process. The simulation study considers autocorrelation coefficients estimated from data and different fractions of randomly missing data from the original time series.

All the computational work was performed with software [10].

2 Testing for trend in time series

One assumes that is interested in testing the existence of a general monotonic trend in the time series and, in such a case, it is common to make the assumption of a model with a linear trend plus an error term. When dealing with time series of hydrologic or environmental data, a significant autocorrelation structure exists in the series and can be modeled by a weak stationary autoregressive moving average process, $ARMA(p,q)$. By convenience, the following model with autoregressive errors is considered,

$$y_t = a + bt + \epsilon_t \tag{1}$$

where $\epsilon_t = \sum_{j=1}^{p} \alpha_j \epsilon_{t-j} + a_t$ and $\{a_t\}$ is a white noise with zero mean and constant variance, $E(a_t^2) = \sigma_t^2$.

3 Resampling in time series

Resampling is the process of sampling from the observations in a sample, in order to obtain estimates for population parameters without making assumptions about the form of the population distribution. In the case where a set of observations can be assumed independent and identically distributed (i.i.d.), this can be implemented by constructing a number of resamples of the observed data set, each of which is obtained by random sampling with replacement from the original data set, this is called by bootstrap. The first formulation of the bootstrap was given by Efron [11]. This technique is a computer-intensive method that presents solutions in situations where the traditional methods fail.

3.1 Sieve Bootstrap

Although the bootstrap is a very simple and comprehensive method used for many statistical problems, is not appropriate for dependent data. A bootstrap approach that overcomes this problem is the sieve bootstrap, proposed by Peter Bühlmann [12]. In a few words, the sieve bootstrap considers first an autoregressive process $AR(p)$ that is fitted to a time series and then a resampling mechanism is performed in the residuals.

3.2 Subsampling

Subsampling appears after many researchers, inspired by Efrons approach, having tried to innovate and create new resampling schemes. In this creative phase, the subsampling was proposed by Politis and Romano [13]. Similar to the Efron's bootstrap, it can be applied in the i.i.d. context, but considers that the observations are drawn without replacement from the original data set and a smaller resample size [14].

3.3 Resampling within trend tests

In [9] the sieve bootstrap was used with the objective of combining the most well-known trend tests with sieve bootstrap, under minimal assumptions about the dependence structure and distribution of hydrological data. The proposed procedure is a combination between the trend tests and the sieve bootstrap and subsampling, as described below:

Step 1. Fit an $AR(p)$ to the time series using the AIC criterion;
Step 2. Obtain the residuals;
For B=1,000 replicates
Step 3. Resample the centered residuals;
Step 4. Obtain a new time series by recursion, using the residuals (**Step 3**) and the autoregressive coefficients (**Step 1**);
Step 5. Obtain the test statistics (t-test/Mann-Kendall) for the new series (**Step 4**) .

Note that when one considers the sieve bootstrap, the **Step 3** is performed *with* replacement and, in the subsampling case is done *without* replacement. The order of autocorrelation structure is estimated by the best fitting model obtained through the Akaike Information Criterion (AIC), which takes into account the complexity of the model.

To conclude the procedure, p-values (significance of the test) were estimated as in [9], that is, based on the number of times the statistics values at **Step 5** exceeds the statistical value obtain for the observed data set, divided by B, the number of replicates.

4 Data example

The Fig. 1 shows the time series of the Level of Lake Huron [15], Canada, and corresponds to the annual measurements of the level, in feet, from 1875 to 1972. By inspection, this time series seems to show a weak downward trend.

This is the time series under study and before applying the test for trend detection, a procedure is done in order to evaluate the performance of the proposed tests in series which exhibit these sample characteristics. Inspection of the data to detect missing observations is required, before applying the trend tests. In case of missing observations, their estimation and replacement are performed

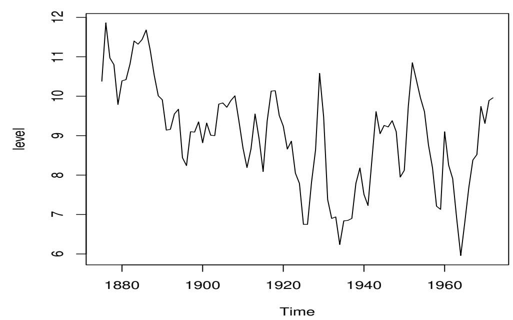

Fig. 1. Time series of Lake Huron.

by an imputation function `na.interp()` [16]. One used the Student t-test statistic and the Mann-Kendall test available in package Kendall [15]. So, we start by analysing the results of the simulation study. The procedure scheme presented in Sect. 3.3 was repeated for $1,000$ replicates of a model with normal distributed errors, with and without trend, in order to estimate the size and power of the tests. The first results are achieved and shown in Table 1 and Table 2.

Table 1. Empirical significance level α of the tests.

		t-test			Mann-Kendall		
n		original	sieve	subsampling	original	sieve	subsampling
100	AR(1)	0.451	0.069	0.061	0.438	0.069	0.054
	AR(2)	0.369	0.052	0.048	0.350	0.053	0.042
	ARMA(1,1)	0.492	0.066	0.066	0.463	0.053	0.056
200	AR(1)	0.476	0.056	0.048	0.455	0.054	0.044
	AR(2)	0.391	0.048	0.053	0.376	0.053	0.053
	ARMA(1,1)	0.469	0.065	0.058	0.443	0.056	0.054

According to the results seen in previous studies, we found that both tests, Mann-Kendall and t test, are strongly affected on the empirical significance level (type I error) when applied to the original expressions of the test statistic (for independence). The findings for sieve bootstrap and subsampling are similar. In both, the empirical α is closer to the usual nominal level α=5%. Regarding the power of the tests in Table 2, it appears that the sieve bootstrap has a slightly higher power when compared to the subsampling approach. Better results are shown in $AR(1)$. In the $AR(2)$ and $ARMA(1,1)$ cases, the power is low. These

Table 2. Empirical power of the tests, under the empirical significance level α.

		t-test			Mann-Kendall		
n		original	sieve	subsampling	original	sieve	subsampling
100	AR(1)	1	0.909	0.841	1	0.904	0.824
	AR(2)	0.981	0.715	0.692	0.981	0.702	0.665
	ARMA(1,1)	0.954	0.553	0.504	0.946	0.544	0.482
200	AR(1)	1	1	0.999	1	1	0.999
	AR(2)	1	1	1	1	1	1
	ARMA(1,1)	1	0.999	0.994	1	0.999	0.992

findings are in agreement with the studies conducted by Noguchi et al [9] and in the case of normal innovations.

In case of missing data the results are shown in Table 3 and Table 4.

Table 3. Empirical significance level α of the tests: incomplete time series.

			t-test			Mann-Kendall		
n	% missing		original	sieve	subsampling	original	sieve	subsampling
100	10%	AR(1)	0.469	0.067	0.049	0.467	0.065	0.048
		AR(2)	0.362	0.045	0.043	0.349	0.050	0.042
		ARMA(1,1)	0.507	0.066	0.050	0.471	0.073	0.054
	25%	AR(1)	0.475	0.047	0.037	0.458	0.049	0.035
		AR(2)	0.366	0.048	0.046	0.349	0.048	0.046
		ARMA(1,1)	0.495	0.057	0.048	0.482	0.055	0.044
200	10%	AR(1)	0.463	0.059	0.051	0.443	0.061	0.049
		AR(2)	0.374	0.039	0.041	0.366	0.043	0.042
		ARMA(1,1)	0.480	0.056	0.056	0.460	0.058	0.052
	25%	AR(1)	0.482	0.059	0.050	0.485	0.063	0.049
		AR(2)	0.382	0.054	0.050	0.385	0.052	0.046
		ARMA(1,1)	0.499	0.061	0.057	0.492	0.066	0.060

One concludes that, in general, the empirical α increases when compared to the test on the full sample, although in small amounts. In the study of power for $AR(2)$ and $ARMA(1,1)$: in $n = 100$ the power was lower in both resampling techniques, as for the complete time series, and $ARMA(1,1)$ was the worst; in $n = 200$ the power was good, even in the 25% fraction of missing data.

Table 4. Empirical power of the tests, under the empirical α: incomplete time series.

			t-test			Mann-Kendall		
n	% missing		original	sieve	subsampling	original	sieve	subsampling
100	10%	AR(1)	1	0.883	0.775	0.999	0.871	0.755
		AR(2)	0.978	0.737	0.710	0.971	0.736	0.701
		ARMA(1,1)	0.964	0.563	0.490	0.959	0.547	0.464
	25%	AR(1)	1	0.852	0.712	0.999	0.845	0.685
		AR(2)	0.983	0.732	0.690	0.980	0.716	0.668
		ARMA(1,1)	0.959	0.564	0.476	0.949	0.557	0.457
200	10%	AR(1)	1	1	0.999	1	1	0.997
		AR(2)	1	1	1	1	1	1
		ARMA(1,1)	1	1	0.993	1	0.999	0.990
	25%	AR(1)	1	1	0.996	1	1	0.994
		AR(2)	1	1	0.996	1	1	0.997
		ARMA(1,1)	1	0.999	0.992	1	0.999	0.989

For the original time series data of Lake Huron, both sieve bootstrap based test and subsampling based test reject the null hypothesis of no trend.

5 Closing Comments

This study relies on the comparison of the performance of the trend tests with resampling techniques for two cases: complete time series and incomplete time series. As general conclusions, we observe that the results of both trend tests after the imputation are in accordance with the one obtained for the original times series. Both tests detect the presence of a trend although the power is lower for some dependence structures and size of the time series. The reconstructed time series maintains the structure and order of the stochastic term estimated from the original series $AR(1)$ and $AR(2)$, respectively. These are the first steps on extending and evaluating the use of other resampling scheme in trend analysis context. Other models for the distribution of the error term must be studied as well as some weak points related with the power of the tests.

Acknowledgments. Research of M. Rosário Ramos and Clara Cordeiro were partially funded by Fundação para a Ciência e a Tecnologia (FCT) through the projects FCT-PEst/OE/MAT/UI0209/2013 and PEst-OE/MAT/UI0006/2014, respectively.

References

1. van Belle G., Hughes J. P.: Nonparametric tests for trend in water quality. Water Resour. Res. 20, 127–136 (1984)
2. Hirsch, R. M., Slack J. R., Smith R. A.: Techniques of trend analysis for monthly water quality data. Water Resour. Res. 18, 107–121 (1982)
3. Lettenmaier, D. P.: Multivariate nonparametric tests for trend in water quality. Water Resour. Bull. 24, 505–512 (1988)
4. El Shaarawi A.H., Damsleth E.: Parametric and nonparametric tests for dependent data. J. Am. Water Resour. Assoc. 24, 513–519 (1988)
5. El Shaarawi A.H., Piegorsh W. W.: Encyclopedia of Environmetrics, vol I. John Wiley and Sons, England (2002)
6. Harcum J. B., Loftis J. C., Ward R. C.: Selecting trend tests for water quality series with serial correlation and missing values. Water Resour. Bull. 28, 469–478 (1992)
7. Sun H., Pantula S. G.: Testing for trends in correlated data. Stat. Probabil. Lett. 41, 87–95 (1999)
8. Woodward W. A., Gray H. L.: Selecting a model for detecting the presence of a trend. J. Climate 8, 1929–1937 (1995)
9. Noguchi, K., Gel, Y.R., Duguay, C.R.: Bootstrap-based tests for trends in hydrological time series, with application to ice phenoloy data. J. Hydrol. 410, 150–161 (2011)
10. R Develpment core team: A Language and Environment for Statistical Computing. Software available at `http://www.r-project.org`
11. Efron, B.: Bootstrap Methods: Another Look at the Jackknife. Ann. Stat. 7, 1–26 (1979)
12. Bühlmann, P.: Sieve bootstrap for time series. Bernoulli **3**, 123–148 (1997)
13. Politis, D.N., Romano, J.P.: Large sample confidence regions based on subsamples under minimal assumptions. Ann. Stat. 22 (4), 1105–1124 (1994)
14. Politis, D.N., Romano, J.P., Wolf, M.: Subsampling. Springer-Verlag, New York (1999)
15. McLeod, A.I.: kendall: Kendall rank correlation and Mann-Kendall trend test. R package version 2.2, `http://cran.r-project.org/package=kendall`
16. Hyndman, R.: **forecast**: Forecasting Functions for Time Series. R package version 5.3, `http://cran.r-project.org/package=forecasting`