Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

A Gini-based unit root test

Amit Shelef*

Department of Industrial Engineering and Management, SCE—Shamoon College of Engineering, Basel Street, Beer-Sheva 84100, Israel
Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, P.O.B. 653, Beer-Sheva 8410501, Israel

HIGHLIGHTS

- A Gini-based unit root test is developed.
- The test relies on the semi-parametric Gini regression.
- Includes an in-depth numerical comparison of the Gini-based test and existing tests.
- Simulations indicate the superiority of the Gini-based test in some design settings.

ARTICLE INFO

Article history:
Received 23 January 2014
Received in revised form 12 August 2014
Accepted 16 August 2014
Available online xxxx

Keywords:
Time series analysis
Unit root tests
Gini regression
Bootstrap

ABSTRACT

A Gini-based statistical test for a unit root is suggested. This test is based on the well-known Dickey–Fuller test, where the ordinary least squares (OLS) regression is replaced by the semi-parametric Gini regression in modeling the AR process. A residual-based bootstrap is used to find critical values. The Gini methodology is a rank-based methodology that takes into account both the variate values and the ranks. Therefore, it provides robust estimators that are rank-based, while avoiding loss of information. Furthermore, the Gini methodology relies on first-order moment assumptions, which validates its use for a wide range of distributions. Simulation results validate the Gini-based test and indicate its superiority in some design settings in comparison to other available procedures. The Gini-based test opens the door for further developments such as a Gini-based cointegration test.

1. Introduction

In most of the literature dealing with time series analysis, underlying dependencies of the time series are modeled based on variance and covariance as measures of variability and association, respectively. This research develops a unit root test that is based on the Gini Mean Difference (hereafter GMD) as an alternative index of variability. The GMD index shares many properties of the variance, but the former can be more appropriate for distributions that depart from normality or symmetry. This measure is less sensitive to extreme observations than the variance because it takes into account both the values of the random variable and its ranks. In addition, the GMD is defined solely under first-order moment assumptions. To clarify the notations, we distinguish between population parameters and estimators by using upper-case letters in the population version and lower-case letters in the sample version.

1.1. Autoregressive unit root tests

We refer to a first-order univariate autoregression, denoted by AR(1), which satisfies

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \varepsilon_t, \tag{1}$$

* Correspondence to: Department of Industrial Engineering and Management, SCE—Shamoon College of Engineering, Basel Street, Beer-Sheva 84100, Israel. Tel.: +972 54 777 6321.
E-mail addresses: amit.shelef@gmail.com, amitshe@post.bgu.ac.il.

http://dx.doi.org/10.1016/j.csda.2014.08.012
0167-9473/

where ϕ_0 is the constant of the model, ϕ_1 is the parameter of the model and ε_t is an independent and identically distributed (i.i.d.) innovation process. If $\phi_1 = 1$, then Y_t is nonstationary. Testing for stationarity by detecting a unit root is an important task in the analysis and modeling of time series. Dickey and Fuller (1979) developed a procedure to test for the presence of a unit root (hereafter referred to as the DF test). The main objective of the DF test is to determine whether $H_0 : \phi_1 = 1$ or, alternatively, $H_1 : \phi_1 < 1$.

The distribution of the appropriate t-statistic, based on applying the OLS estimator for the nonstationarity parameter, is nonstandard and cannot be analytically evaluated. Dickey and Fuller (1979) used the Monte Carlo simulation method to tabulate the percentiles of the DF t-statistic distribution based on $\varepsilon_t \sim i.i.d\ N(0, \sigma_\varepsilon^2)$ innovations. The DF t-statistic is

$$DF_{t-stat} = \frac{\hat{\phi}_1^{OLS} - 1}{\hat{S}D(\hat{\phi}_1^{OLS})}, \tag{2}$$

where $\hat{\phi}_1^{OLS} = \frac{\mathrm{cov}(Y_t, Y_{t-1})}{\mathrm{cov}(Y_{t-1}, Y_{t-1})}$ is the OLS estimator for ϕ_1, $\hat{S}D(\hat{\phi}_1^{OLS}) = \left(\hat{\sigma}_\varepsilon^2 \left(\sum_{t=1}^n Y_{t-1}^2\right)^{-1}\right)^{0.5}$ and $\hat{\sigma}_\varepsilon^2$ is the least squares estimator.

Extensive attempts to improve this test and to find alternative or superior tests followed. Leybourne (1995) suggested using the maximum of DF t-statistics based on applying the OLS regressions twice, looking both forward and backward at the series. The critical values for the test are obtained in a manner similar to that performed in the original DF test discussed above using Monte-Carlo simulation. Elliott et al. (1996) and Ng and Perron (2001) suggested a class of unit root tests that are based on generalized least squares detrending of the series and then applying the DF test on the detrended data.

An important development in unit root tests includes the use of resampling methods for calculating critical values; see, for example, a survey of such methods in Palm et al. (2008). A first residual-based bootstrap version of the DF test was proposed by Ferretti and Romo (1996). Later, Moreno and Romo (2000) used a bootstrap procedure based on the LAD estimator, and more recently, Moreno and Romo (2012) suggested a family of unit root bootstrap tests for the infinite variance case.

Another important development appears in Muller and Elliott (2003), who revealed that the initial condition (Y_0) has a non-negligible influence on the finite sample performances of unit root tests. In general, it is difficult to rule out, a priori, the existence of small or large values of Y_0. Elliott and Müller (2006) suggest statistics whose power is less sensitive to the size of Y_0. Harvey and Leybourne (2005) and Harvey et al. (2009) recommend a union of test rejection decision rules to improve the performances of the statistical procedure.

Hallin et al. (2011) propose a class of distribution-free rank-based tests for the null hypothesis of a unit root that takes into account several initial values for the series. They use three test statistics that are based on a choice of a reference density function, which need not be the unknown actual density of the innovations. The first test statistic is based on the Gaussian reference density and is defined as $T_{vdW}^{(n)} = \frac{1}{\sqrt{n}} \sum_{t=1}^n \left(\frac{t}{n+1} - \frac{1}{2}\right) \Phi^{-1}\left(\frac{R_t}{n+1}\right)$, where R_t are the ranks of the increments $\Delta Y_t = Y_t - Y_{t-1}$ and Φ denotes the standard normal distribution. This statistic is also known as the normal or van der Waerden score. The second test statistic is based on the double-exponential distribution (Laplace or sign test scores), $T_L^{(n)} = \frac{1}{\sqrt{n}} \sum_{t=1}^n \left(\frac{t}{n+1} - \frac{1}{2}\right) sign\left(\frac{R_t}{n+1} - \frac{1}{2}\right)$. The third test statistic is based on the logistic distribution (Wilcoxon scores), $T_W^{(n)} = \frac{\pi}{\sqrt{3n}} \sum_{t=1}^n \left(\frac{t}{n+1} - \frac{1}{2}\right) \left(\left(1 - \frac{n+1-R_t}{R_t}\right) / \left(1 + \frac{n+1-R_t}{R_t}\right)\right)$. These three test statistics are denoted here as HAW-*vdW*, HAW-*Laplace* and HAW-*Wilcoxon*, respectively. Asymptotic results and simulated quantiles for these statistics for several sample sizes are given in Hallin et al. (2011). The results for finite samples indicate that for a broad range of non-zero initial values and for a variety of heavy-tailed innovation densities, the suggested rank-based tests outperform a variety of unit root tests.

Rank-based tests, which intrinsically involve some loss of information of the real values, are expected to perform well under heavy-tailed distributions. In this paper, we propose a Gini-based unit root test that relies on both the real values and the ranks, while avoiding loss of information.

1.2. The Gini methodology

The GMD is an alternative index of variability that is used in this research instead of the variance. The most prevalent presentation of the GMD index is the expected absolute difference between two independent and identically distributed (i.i.d.) variables X_1 and X_2 (Gini, 1914). Formally, the GMD of X is defined as

$$G_X = E\left|X_1 - X_2\right|. \tag{3}$$

Alternatively, the GMD can be expressed as a special case of a covariance, i.e., four times the covariance of X, a random variable, and $F_X(X)$, its cumulative distribution function (Lerman and Yitzhaki, 1984). Formally,

$$G_X = 4COV\left(X, F_X(X)\right), \tag{4}$$

where $F_X(X)$ is the cumulative distribution function of X.

The GMD index shares many properties of the variance, but the former can be more informative for distributions that depart from normality or symmetry. Both measures are based on weighted averages of the distances between each pair of i.i.d. variables. The fundamental difference is the method used to measure the distance. The GMD distance function is referred to as the "city block" distance, which allows one to move only in the vertical and horizontal directions. The variance distance function is Euclidean, which allows one to move in any desired direction.

Please cite this article in press as: Shelef, A., A Gini-based unit root test. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.08.012