Statistica Sinica **37** (2027), 1-19
doi:https://doi.org/10.5705/ss.202024.0106

ON RUNS TESTS FOR DIRECTIONAL DATA AND THEIR LOCAL AND ASYMPTOTIC OPTIMALITY PROPERTIES

Maxime Boucher[1], Christian Francq[2,3], Yuichi Goto[4] and Thomas Verdebout*[1]

[1] *Université libre de Bruxelles,* [2] *Université de Lille,*
[3] *CREST and* [4] *Kyushu University*

Abstract: In the present paper, we tackle the problem of detecting serial correlation in directional data. We introduce a concept of runs properly adapted to the directional context. We then show that tests based on the latter runs enjoy some local and asymptotic optimality properties against local alternatives with serial dependence. We evaluate the finite-sample performances of our tests using Monte Carlo simulations and show their usefulness on a real data illustration that involves the analysis of sunspot locations for various solar cycles.

Key words and phrases: Directional data, hypothesis testing, randomness, runs tests.

1. Introduction

The problem of testing randomness of a series of observations is one of the most important problems in time series analysis. To tackle the problem, runs tests are nowadays very classical tools in statistics. In particular, the runs tests for randomness of Wald and Wolfowitz (1940) is very popular. More recently, runs have been used, for instance, in Henze and Penrose (1999) and Biswas, Mukhopadhyay and Ghosh (2014) to compare samples; in McWilliams (1990), Corzo and Babativa (2013), and Dyckerhoff, Ley and Paindaveine (2015) to test for symmetry; in Dufour, Hallin and Mizera (1998) to test for nonhomogeneous white noise; and in Paindaveine (2009), Cho and White (2011), and Hentati-Kaffel and de Peretti (2015) to test for randomness.

Classical runs tests typically reject the null hypothesis of randomness when the number of *runs* in the sequence is too large or too small. In a sequence of observations, a *run* is defined as a consecutive series of observations with the same sign. From a univariate sample $X_1, \ldots, X_n$, the number of runs can be computed from the quantity $\sum_{t=2}^{n} U_t(\theta)U_{t-1}(\theta)$, where $U_t(\theta) := \text{sign}(X_t - \theta)$ is the sign of a centered version of X_t; θ plays the role of a location parameter here. More precisely, if $X_1, \ldots, X_n$ are mutually independent random variables with median

*Corresponding author. E-mail: tverdebout@gmail.com