Statistica Sinica **36** (2026), 1-20
doi:https://doi.org/10.5705/ss.202023.0379

A WARPED SELF-NORMALIZED TWO-SAMPLE TEST FOR TIME SERIES WITH STAGGERED OBSERVATION PERIODS

Weiliang Wang and Ting Zhang*

Boston University and University of Georgia

Abstract: We consider the problem of two-sample testing for time series with staggered observation periods, where the two time series can have different starting and ending observation times and can be of different lengths. In addition, we allow the two time series to depend on each other in a general way, which makes the staggered observation periods nontrivial to deal with as it now requires accommodating the joint dependence in the presence of overlapping and nonoverlapping segments when designing a valid inference protocol. This also makes existing self-normalization methods inapplicable to the current problem. To address this, we propose a warped self-normalized two-sample test, which uses warped self-normalized subsamples to provide uncertainty quantification of the global two-sample statistic. The method can be readily applied to compare quantities beyond the mean such as the variance or quantiles, and the associated asymptotic theory has been established. Numerical experiments including a simulation study and a real data analysis are also provided to further illustrate the proposed method.

Key words and phrases: Self-normalization, staggered time series, subsampling, two-sample test.

1. Introduction

Applications from various scientific problems often require the comparison of data from two populations, such as clinical trial data from treatment and control groups, temperature record data from different countries or regions, and electricity usage data from different industries, among many others. The problem is often phrased as a two-sample test in statistical analysis, which has been widely studied in the literature; see, for example, Hotelling (1931), Cressie and Whitford (1986), Hall and Martin (1988), Bai and Saranadasa (1996), Keselman et al. (2004), Chen and Qin (2010), Chen, Dou and Qiao (2013), Cai, Liu and Xia (2014), Gregory et al. (2015), Xu et al. (2016), Städler and Mukherjee (2017), Chen, Li and Zhong (2019), Zhang et al. (2020), and references therein. The aforementioned works mostly concerned the situation when the data can be viewed as independent samples from an underlying distribution. For time series data, Dette and Weißbach (2009) considered testing the difference between

*Corresponding author. E-mail: tingzhang@uga.edu