Journal of Multivariate Analysis 150 (2016) 91–104

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

A local factor nonparametric test for trend synchronism in multiple time series

Vyacheslav Lyubchich [a], Yulia R. Gel [b,c,*]

[a] *Chesapeake Biological Laboratory, University of Maryland Center for Environmental Science, USA*
[b] *Department of Statistics and Actuarial Science, University of Waterloo, Canada*
[c] *Department of Mathematical Sciences, University of Texas at Dallas, USA*

ARTICLE INFO

Article history:
Received 23 July 2015
Available online 30 May 2016

AMS subject classification:
62M10
91B84
62M07

Keywords:
Multiple mean curves comparison
Nonmonotonic trend detection
Time series
Trend synchronism
Resampling and bootstrap

ABSTRACT

The problem of identifying joint trend dynamics in multiple time series, i.e., testing whether two or more observed processes follow the same common trend, is essential in a wide spectrum of applications, from economics and finance to climate and environmental studies. However, most of the available tests for comparing multiple mean functions either deal with independent errors or are applicable only to a case of two time series, which constitutes a substantial limitation in many modern, typically high-dimensional, studies. In this paper we propose a new nonparametric test for synchronism of trends exhibited by multiple linear time series where the number of time series N can be large but fixed. The core idea of our new approach is based on employing the local regression test statistic, which allows to detect possibly non-monotonic nonlinear trends. The finite sample performance of the new synchronism test statistic is enhanced by a nonparametric hybrid bootstrap approach. The proposed methodology is illustrated by simulations and a case study on insurance claims due to extreme weather.

1. Introduction

The problem of detecting joint trend dynamics in time series is essential in a variety of applications, ranging from analysis of macroeconomic indicators [18,46] to assessing patterns in ice phenology measurements from multiple locations [16,28] to evaluating yields of financial instruments at various maturity levels [36] and cell phone download activity at different area codes [14]. The extensive research on comparing trend patterns follows two main directions, namely, analysis of joint stochastic trends, which is closely linked to the cointegration notion by Engle and Granger [17], and testing for joint mean functions. In this paper, we are interested in the second direction, that is, to assess whether N observed time series follow the same hypothesized parametric trend. Our interest in this problem was motivated by a climate adaptation project designed to reduce the adverse impact of extreme weather events on the insurance industry [41,42]. In particular, we wanted not only to align the data on weather-related insurance claims with the observed meteorological conditions at particular cities, but also to assess the joint long-term dynamics of insurance claims across these cities, over a large territory.

There exist many tests for comparing mean functions, primarily in the nonparametric and semiparametric regression literature, and most of the developed methodology focuses on independent errors. Among such approaches are, for example, the large class of the L^2-distance based methods [15,27,29,34], the wavelet-based methods [23], the local factor type

---

[*] Corresponding author at: Department of Statistics and Actuarial Science, University of Waterloo, Canada.
E-mail addresses: ygl@uwaterloo.ca, ygl@utdallas.edu (Y.R. Gel).

http://dx.doi.org/10.1016/j.jmva.2016.05.004

methods [35,50] and the regression-depth procedures [49]. Substantially less is known about testing for joint deterministic trends in a time series framework. Fan and Lin [19] propose an adaptive Neyman test for comparing two groups of curves embedded into a stationary Gaussian linear error process. Li [30] employs the idea of partial sum process to test the equality of two nonparametric regression functions with long memory moving average errors. Vilar-Fernández et al. [45] discuss bootstrap extensions of various methods for two curves comparison whose asymptotic properties are developed under an independence assumption and provide an empirical comparative study of such resampling-based extensions under weak linear dependence. Park et al. [36] develop a scale-space visualization tool (SiZer) for testing the equality of means of two weakly stationary time series and extend this idea to comparison of more than two time series using residual analysis [35]. Vilar-Fernández and González-Manteiga [43] propose to use L^2-distance, or the integrated squared error (ISE) based statistic to test the equality of multiple nonparametric curves embedded into noise following a moving average structure. More recently, Vilar-Fernández and Vilar-Fernández [44] investigate finite sample performance of the Vilar-Fernández–González-Manteiga statistic enhanced by a bootstrap. Finally, Degras et al. [14] and Zhang [52] extend the ISE based approach of Fernández and González-Manteiga [43] to a case of multiple time series with weakly dependent (non)stationary errors. (For a comprehensive literature review of available methodology for comparing mean functions embedded into independent errors in a time series framework, see [14,35].) However, most of these methods either focus on aligning only two curves, or involve selection of multiple nuisance parameters, such as bandwidth, level of smoothness, and window size for a long-run variance function. As mentioned by Park et al. [35], the choice of such multiple nuisance parameters is challenging for a comparison of curves (even under independent and identically distributed setup) and often leads to inadequate performance, especially in samples of moderate size.

The core idea of our approach is based on generalizing the nonparametric local factor (LF) test statistic, which allows to assess whether N weakly dependent time series exhibit a joint (non)monotonic (non)linear trend that belongs to a pre-specified parametric family of functions [31,47,48,51]. The test procedure employs an artificial balanced one-way analysis of variance where each distinct time point is viewed as a category and an associated cell includes all observations within a surrounding local window. Following Lyubchich et al. [31], we propose to choose an optimal window for the local factor using the data-driven heuristic m-out-of-n bootstrap selection algorithm of Bickel et al. [5]. We show that the new trend synchronism statistic is asymptotically normally distributed. However, as convergence to the asymptotic distribution might be slow, we enhance finite sample properties of the test statistic by a hybrid bootstrap procedure. Hence, our test is fully data-driven, free from any distributional assumptions on the observed data, and easy to compute, which makes it particularly attractive in applications. Our numerical study shows that the new synchronism test delivers accurate type I error estimates and competitive power performance. Moreover, our test yields noticeably more accurate estimates of the size of the test, compared with the test of Degras et al. [14], especially when there is either a small number of observed time series or when the autocorrelation structure of an individual time series is allowed to include negative terms, which is a frequent situation for economic and environmental studies.

2. Synchronism test statistic

Let us observe N time series processes

$$Y_{it} = \mu_i(t/T) + \epsilon_{it} \quad (i = 1, \ldots, N;\ t = 1, \ldots, T), \tag{1}$$

where $\mu_i(u)$ $(0 < u \leq 1)$ are unknown smooth trend functions, and the noise ϵ_{it} satisfies the following assumptions:

Assumption A. The noise is a finite order autoregressive process

$$\epsilon_{it} = \mathrm{AR}(p_i) = \sum_{k=1}^{p_i} \phi_{ik}\epsilon_{i,t-k} + e_{it} \quad (i = 1, \ldots, N;\ t = 1, \ldots, T),$$

where conditions on e_{it} are specified in Assumption B, and the polynomial $\phi_i(\lambda) = 1 - \phi_{i1}\lambda^1 - \cdots - \phi_{ip_i}\lambda^{p_i}$ has all its roots outside the closed unit disk: $\phi_i(\lambda) \neq 0$, for all $|\lambda| \leq 1$.

Assumption A′. The noise is an infinite order autoregressive process

$$\epsilon_{it} = \mathrm{AR}(\infty) = \sum_{k=1}^{\infty} \phi_{ik}\epsilon_{i,t-k} + e_{it} \quad (i = 1, \ldots, N;\ t = 1, \ldots, T), \tag{2}$$

where conditions on e_{it} are specified in Assumption B, and (2) does not degenerate to a finite dimensional autoregressive representation of order p, that is, $\nexists K > 0$ such that $\forall k > K$, $\phi_{ik} = 0$; $\phi_i(\lambda) = 1 - \phi_{i1}\lambda - \cdots$ is bounded away from zero within the closed unit disk: $\phi_i(\lambda) \neq 0$, for all $|\lambda| \leq 1$; and $\sum k^{1/2}|\phi_{ik}| < \infty$ (see Berk [3] for discussion on properties of such AR(∞) processes).

Assumption B. In Assumptions A and A′, e_{it} $(i = 1, \ldots, N)$ are independent and identically distributed random variables with $\mathrm{E}(e_{it}) = 0$, $\mathrm{E}(e_{it}^2) = \sigma_i^2$, $\mathrm{E}(e_{it}^4) < \infty$, and $\{e_{it}\}_{t=1}^T$ and $\{e_{jt}\}_{t=1}^T$ are independent if $i \neq j$.