Journal of Multivariate Analysis 151 (2016) 37–53

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

A randomness test for functional panels

Piotr Kokoszka [a], Matthew Reimherr [b,*], Nikolas Wölfing [c]

[a] Colorado State University, United States

[b] Pennsylvania State University, United States

[c] Centre for European Economic Research, ZEW Mannheim, Germany

ARTICLE INFO

Article history:
Received 24 November 2015
Available online 20 July 2016

AMS subject classifications:
62G10
62G20
62H25

ABSTRACT

Functional panels are collections of functional time series, and arise often in the study of high frequency multivariate data. We develop a portmanteau style test to determine if the cross-sections of such a panel are independent and identically distributed. Our framework allows the number of functional projections and/or the number of time series to grow with the sample size. A large sample justification is based on a new central limit theorem for random vectors of increasing dimension. With a proper normalization, the limit is standard normal, potentially making this result easily applicable in other FDA context in which projections on a subspace of increasing dimension are used. The test is shown to have correct size and excellent power using simulated panels whose random structure mimics the realistic dependence encountered in real panel data. It is expected to find application in climatology, finance, ecology, economics, and geophysics. We apply it to Southern Pacific sea surface temperature data, precipitation patterns in the South-West United States, and temperature curves in Germany.

1. Introduction

We define a *functional panel* as a stochastic process of the form

$$\mathbf{X}_n(t) = (X_{1,n}(t), \dots, X_{I,n}(t))^\top, \quad 1 \le n \le N, \tag{1}$$

where each $X_{i,n}$ is a function of time t. The dimension I can increase with the series length N, with examples discussed below. For the applications that motivate the present research, it is enough to think of the $X_{i,n}$ as curves defined on the same time interval, but in principle, functions on more general domains, e.g., volumes or surfaces, can be considered. The discrete time index n refers to a unit like a day, week or year. The index t is the continuous time argument of the function $X_{i,n}$. The index i refers to the ith time series in the panel. This paper develops a test of the null hypothesis

$\mathcal{H}_0$: the random elements $\mathbf{X}_n$, $1 \le n \le N$, are independent and identically distributed.

Our test is designed to detect serial dependence, and we assume stationarity across n even under the alternative.

To illustrate the functional panel concept, Fig. 1 shows four curves, $I = 4$, for a fixed n. The index n refers to years, and the four curves describe the sea surface temperature in four regions used to measure the El Niño climatic phenomenon. Fig. 2 shows another example, now with i fixed. The data point $X_{i,n}(t)$ is the log-precipitation at location i on day t of year n. The construction of this series is explained in detail in Section 3. Data structures of this type are very common in climate studies;

* Corresponding author.
E-mail addresses: Piotr.Kokoszka@colostate.edu (P. Kokoszka), mreimherr@psu.edu (M. Reimherr), Woelfing@zew.de (N. Wölfing).

http://dx.doi.org/10.1016/j.jmva.2016.07.002

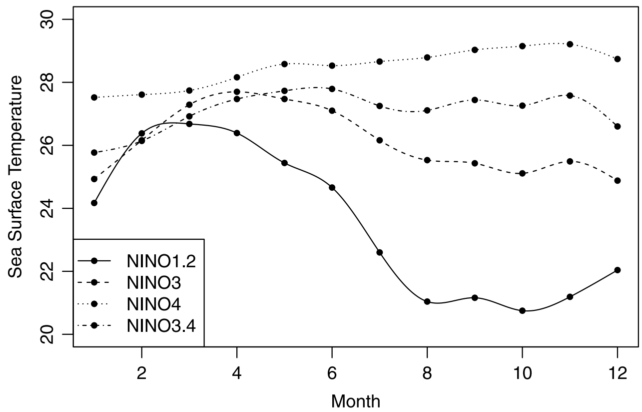

Fig. 1. Sea surface temperature curves of El Niño regions in 2012.

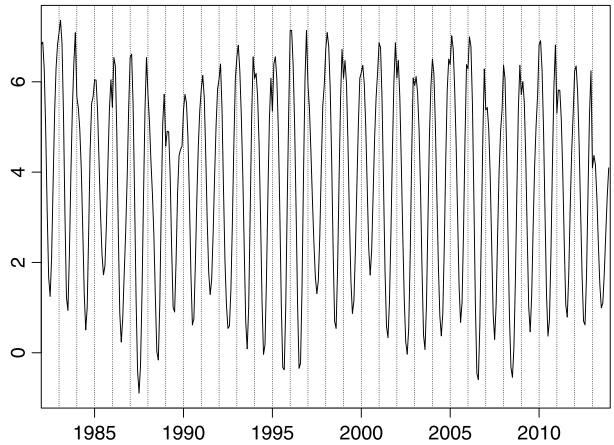

Fig. 2. Smoothed log-precipitation, Santa Cruz, California, 1982–2013.

$X_{i,n}(t)$ can be total precipitation or maximum temperature on day t, $1 \leq t \leq 365$, of year n at location i in some region. In such climate applications, I is comparable to N because records often start at the end of the 19th or towards the middle of the 20th century, thus, they are about 60–120 years long ($N \approx$ 60–120), and there are several dozen measurement stations in a region ($I \approx$ 40–120). (The United States Historical Climatology Network – USHCN – contains weather data collected at 1218 stations across the 48 contiguous states, starting from ca. 1900.)

Climate data do not exhaust possible applications. Intraday financial data typically come in panels. For example, $X_{i,n}(t)$ can be the exchange rate (against the US dollar) of currency i, $1 \leq i \leq I$, at minute t of the nth trading day. Panels of exchange rates contain information on the intraday strength of the US dollar. Corporate bond yield curves are large panels because a bond portfolio includes hundreds of companies, $I \sim 10^3$; in economic studies, government bond yields curves from small panels because only a few countries are considered to assess risk in a region, see, e.g., Härdle and Majer [18]. At the intersection of climate and financial panels, Härdle and Osipienko [19] use a functional panel framework in which i refers to a spatial location, and the interest lies in pricing a financial derivative product whose value depends on the weather at location i. Modeling electricity data involves functional panels indexed by regions or power companies with the daily index n, see Liebl [32] for an overview. Daily pollution (particulate, oxide or ozone) curves at several locations within a city form a functional panel of moderate size.

In these examples, the dependence between the $X_{i,n}$, $1 \leq i \leq I$, for fixed n, is strong, and, generally, the temporal dependence, indexed by n, cannot be neglected. In specific applications, this dependence is modeled by deterministic trends or periodic functions (climate data) or by common factors (financial data). To validate a model, it is usual to verify that residual curves computed in some manner form a random sample. (See, e.g., Kowal et al. [29] for a model applied to functional panels of government yield curves and neurological measurements with an explicit residual i.i.d. assumption.) It is thus important to develop a test of randomness, i.e., to test the null hypothesis $\mathcal{H}_0$ stated above. Such a test could be viewed as analogous to tests of randomness which are crucial in time series analysis, see, e.g., Section 1.6 of Brockwell and Davis [5]. They can be applied to original or transformed data, or to model residuals. The purpose of this paper is to develop a suitable test for functional panels. Before discussing our approach, we provide some historical background. Our methodology builds on the well-established paradigm of testing for randomness in time series which can be traced back to the work of Box