Journal of Multivariate Analysis 145 (2016) 208–223

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Set-valued and interval-valued stationary time series

Xun Wang [a,b], Zhongzhan Zhang [a], Shoumei Li [a,*]

[a] Department of Statistics, College of Applied Sciences, Beijing University of Technology, 100124 Beijing, China
[b] Laboratory M2P2 UMR7320, Centrale Marseille, CNRS, Aix-Marseille University, 13451 Marseille, France

ARTICLE INFO

Article history:
Received 19 November 2014
Available online 30 December 2015

AMS subject classifications:
62M10
65G40

Keywords:
Set-valued stationary process
Interval-valued time series
D_p metric
Uncertainty quantification

ABSTRACT

Stationarity is a key tool in classical time series. In order to analyze the set-valued time series, it must be extended to the set-valued case. In this paper, stationary set-valued time series is defined via D_p metric of set-valued random variables. Then, estimation methods of expectation and auto-covariance function of stationary set-valued time series are proposed. Unbiasedness and consistency of the expectation estimator and asymptotic unbiasedness of the auto-covariance function estimator are justified. After that, a special case of the set-valued time series, known as interval-valued time series, is considered. Two forecast methods of the stationary interval-valued time series are explicitly presented. Furthermore, the interval-valued time series is contextualized in the Box–Jenkins framework: an interval-valued autoregression model, along with its parameter estimation method, is introduced. Finally, experiments on both simulated and real data are presented to justify the efficiency of the parameters estimation method and the availability of the proposed model.

1. Introduction

Classical time series models [8,9] have played an important role in a wide range of areas, such as signal processing and data analysis of stock markets [16,26]. However, many problems cannot be explained and handled by the traditional time series. In some particular cases, time series with set-valued or interval-valued observations have to be taken into account.

Set-valued or interval-valued data may represent uncertainty or variability. In the former case, they represent incomplete observations, i.e., we just know that the actual data belong to a range, instead of precise values. For instance, when economists predict the economic increasing rate, they tend to give their answers like "5% to 7%" or "5.8% to 7.2%", which stand for the uncertainty in the future. By contrast, in the variability case, a set or an interval is not interpreted as containing a single true value, but the observation itself is set-valued or interval-valued. For example, when the traditional time series models are used to forecast stock prices, people always take the closing prices as the data. In this case, the information of stock price fluctuations during each day cannot be utilized; the forecast of stock price is also a single-value, such that the information provided to the decision makers is lack of flexibility. Alternatively, we may take the highest and lowest prices of each day as the upper and lower boundaries of an interval, which represents the variability of stock price over this period. Another example of interval-valued data representing the variability is the weather forecast: it always provides the highest and lowest temperatures of the next day, which form an interval including almost all the useful information regarding tomorrow's temperature. This interval reflects the temperature variability of each day.

* Corresponding author.
E-mail address: lisma@bjut.edu.cn (S. Li).

http://dx.doi.org/10.1016/j.jmva.2015.12.010

Interval-valued random variables are a special kind of set-valued random variables, taking values as compact convex subsets of $\mathbb{R}^1$. Since we have had many results in the theory of set-valued random variables [2,12,18–20], in this paper we deal with the problem in the set-valued framework first, then consider the interval-valued time series as a special case of the set-valued time series.

Wang and Wang [31] defined the set-valued strong (or strict) stationary processes: the joint probability distribution of the set-valued stochastic process does not change when shifted in time. However, a weaker form of stationarity is more commonly employed. In this case, the random processes only require that expectation and covariance do not vary with respect to time. The main purpose of this paper is to define and investigate the set-valued weak stationary time series. For convenience, this paper always calls them set-valued stationary time series.

Since the hyperspace of sets (e.g., the space composed of all the closed intervals) is not linear with respect to addition and scalar multiplication, there has been only a few works discussing the variance and covariance of set-valued random variables until recently such that the definition of stationarity cannot be straightforwardly generalized to the set-valued case. Vital [28] investigated the metric for compact convex sets via support functions. Yang and Li [32] studied D_p metric in the space of set-valued random variables. They proposed to define the variance and covariance of set-valued random variable using the D_p metric. After that, Blanco-Fernandez et al. [7] defined the d_K-variance of interval-valued random variables with underlying space $\mathbb{R}^1$, which is a special case of [32].

Some preliminary studies on the interval-valued time series have been done. Hsu and Wu [15] investigated the interval-valued time series and introduced three different evaluation criteria for estimation and forecast efficiency of the interval-valued time series. Maia et al. [21] investigated the forecast problem of the interval-valued time series. Wang and Li [29] introduced a new type of interval-valued time series, known as interval autoregression time series model, and proposed a parameter estimation and forecast method based on the evaluation criteria in [15]. Cappelli et al. [11] used the interval-valued time series to solve the problems of urban air pollution and agricultural product prices respectively. However, all of the above works consider the interval-valued time series as two separate real-valued time series (left- and right-endpoints or center and radius of intervals) and forecast the interval-valued time series separately. This approach may generate some meaningless forecast such that the left-endpoint is greater than the right-endpoint, since these two series are unrelated. In order to avoid this drawback, this paper views the interval-valued time series as a whole.

Some other interval-valued and set-valued statistical models have been investigated, as follows. Tanaka and Lee [25] introduced the interval linear regression model, which is not based on the interval-valued random variable framework but via two point-valued models. They estimated the linear regression coefficients using a quadratic optimization method. Blanco-Fernandez et al. [6] and Sinova et al. [23] investigated the linear relationship between two interval-valued random variables, by considering the input variable as two real-valued random variables (center and radius of interval). They proposed a least square estimation of model coefficients under the d_2 metric of intervals. Blanco-Fernandez et al. [5] studied strong consistency and asymptotic distribution of the least square estimate. Gonzalez-Rivera and Lin [14] introduced a constrained condition for the regression models of upper and lower bounds of intervals, which guarantees the nature order of interval in the forecast problem. Beresteanu and Molinari [4] investigated the inference problem for partially observed models via an asymptotic approach; they assumed the observations to be uncertain and proposed an estimation method for the real-valued parameters. Le-Rademacher and Billard [17] defined the likelihood function of the interval-valued random variable by assuming that mean and variance of each interval-valued variable follow Gaussian and exponential distributions, respectively. Therefore, the parameters of the distributions, as well as the overall mean and variance of interval-valued random variable, can be estimated via maximum likelihood estimation (MLE). Wang et al. [30] investigated the linear relationship between an interval-valued output variable and a few real-valued input variables, based on the definitions of variance and covariance of set-valued random variables introduced in [32]. Manski and Tamer [22] investigated the inference problem of regression model, in which one of the observed variables is not precise but in the form of intervals. D'Urso [13] studied the linear regression models for fuzzy/crisp input and fuzzy/crisp output data and developed the parameter estimation methods via least square. Song and Chissom [24], Tseng et al. [27], and Cappelli et al. [10] investigated the fuzzy-valued time series and the corresponding model estimation and forecast problems.

In this paper, we build the framework of stationary set-valued time series by using the moments of set-valued random variables, which are different from the above mentioned references. Furthermore, we investigate some more theoretical results of interval-valued time series and interval-valued autoregression model, as well as numerical and empirical studies. The organization of this paper is as follows. In Section 2, the variance and covariance of set-valued random variables are defined via the D_p metric. In Section 3, the stationarity of set-valued time series is defined. Then, an unbiased and consistent estimate of the expectation and an asymptotically unbiased estimate of the auto-covariance function of stationary set-valued time series are given in Section 4. After that, Section 5 focuses on the interval-valued time series: two forecast methods are proposed. Then, the interval-valued autoregression (IAR) model, along with its parameter estimation method, is introduced. Section 6 presents a simulation experiment and a real case application to demonstrate the advantage of the proposed approach. Finally, the conclusions are given in Section 7.

2. Variance and covariance of set-valued random variables

In this section, we first recall some concepts related to the set-valued random variables, then the variance and covariance are defined.