Set-valued and interval-valued stationary time series

Xun Wang\textsuperscript{a,b}, Zhongzhan Zhang\textsuperscript{a}, Shoumei Li\textsuperscript{a,}\textsuperscript{*}

\textsuperscript{a} Department of Statistics, College of Applied Sciences, Beijing University of Technology, 100124 Beijing, China
\textsuperscript{b} Laboratory M2P2 UMR 7320, Centrale Marseille, CNRS, Aix-Marseille University, 13451 Marseille, France

\textbf{ARTICLE INFO}

\textbf{Article history:}
Received 19 November 2014
Available online 30 December 2015

\textbf{AMS subject classifications:}
62M10
65G40

\textbf{Keywords:}
Set-valued stationary process
Interval-valued time series
$D_p$ metric
Uncertainty quantification

\textbf{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.

© 2015 Elsevier Inc. All rights reserved.

\section{Introduction}

Classical time series models\textsuperscript{[8,9]} have played an important role in a wide range of areas, such as signal processing and data analysis of stock markets\textsuperscript{[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.

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

http://dx.doi.org/10.1016/j.jmva.2015.12.010
0047-259X/© 2015 Elsevier Inc. All rights reserved.
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.
2.1. $d_p$ metric of sets

Assume that $(\Omega, A, Pr)$ is a complete probability space, $(\mathcal{X}, \| \cdot \|_{\mathcal{X}})$ is a Banach space, $\mathbf{K}(\mathcal{X})$ is the family of all nonempty closed subsets of $\mathcal{X}$, and $\mathbf{K}_{cc}(\mathcal{X})$ is the family of all nonempty compact convex subsets of $\mathcal{X}$. For any $A, B \in \mathbf{K}(\mathcal{X})$, $\lambda \in \mathbb{R}$, the addition and scalar multiplication are defined by $A + B = \{a + b : a \in A, b \in B\}$ and $\lambda A = \{\lambda a : a \in A\}$, respectively. For any $A, B \in \mathbf{K}_{cc}(\mathcal{X})$, $A + B \in \mathbf{K}_{cc}(\mathcal{X})$. Note that $\mathbf{K}_{cc}(\mathcal{X})$ is not a linear space, since $A + (-1)A \neq \emptyset$.

For each $A \in \mathbf{K}_{cc}(\mathcal{X})$, its support function is defined by

$$s(x^* , A) = \sup \{x^*(a) : a \in A\}, \quad x^* \in \mathcal{X}^* ,$$

(1) in which $\mathcal{X}^*$ is the dual space of $\mathcal{X}$, i.e. the set of all bounded linear functionals on $\mathcal{X}$. The support functions have the following properties:

$$s(x^*, A + B) = s(x^*, A) + s(x^*, B),$$

(2)

$$s(x^*, \lambda A) = \lambda s(x^*, A), \quad \lambda \geq 0.$$  

(3)

For $1 \leq p < \infty$, the metric $d_p$ of sets $A, B \in \mathbf{K}_{cc}(\mathcal{X})$ is defined by

$$d_p(A, B) = \left( \int_{\mathcal{X}^*} |s(x^*, A) - s(x^*, B)|^p \, d\mu \right)^{1/p},$$

(4) in which $S^*$ is the unit sphere of $\mathcal{X}^*$, i.e. $S^* = \{x^* \in \mathcal{X}^* : \|x^*\|_{\mathcal{X}^*} = 1\}$, $\mu$ is a measure on $(\mathcal{X}^*, \mathcal{B}(\mathcal{X}^*))$. For any $p \geq 1$, $(\mathbf{K}_{cc}(\mathcal{X})^d, d_p)$ is a complete and separable metric space [32].

**Example 1.** Let $\mathcal{X} = \mathbb{R}^1$, then $\mathbf{K}_{cc}(\mathbb{R}^1) = \{[a, b] : -\infty < a \leq b < \infty\}$ represents all the closed intervals, and $\mathcal{X}^* = \mathcal{X}$. Let $\mu$ be the Lebesgue measure, and denote $A_1, A_2 \in \mathbf{K}_{cc}(\mathbb{R}^1)$ as $A_1 = [a_1, b_1], A_2 = [a_2, b_2]$, then the $d_p$ metric of the intervals is

$$d_p(A_1, A_2) = \left( |a_2 - a_1|^p + |b_2 - b_1|^p \right)^{1/p}.$$  

(5)

2.2. $D_p$ metric space of set-valued random variables

A set-valued mapping $F : \Omega \to \mathbf{K}(\mathcal{X})$ is called a set-valued random variable if for each open subset $O$ of $\mathcal{X}$, $F^{-1}(O) \in \mathcal{A}$, where $F^{-1}(O) = \{\omega \in \Omega : F(\omega) \cap O \neq \emptyset\}$ and $\emptyset$ is the empty set. Two set-valued random variables are identical if $F_1(\omega) = F_2(\omega)$ for almost every $\omega \in \Omega \ (a.s. (Pr))$. Let $\mathbf{U}[\Omega, \mathbf{K}_{cc}(\mathcal{X})]$ denote the family of set-valued random variables taking values in $\mathbf{K}_{cc}(\mathcal{X})$. The $D_p$ metric of two set-valued random variables $F_1, F_2 \in \mathbf{U}[\Omega, \mathbf{K}_{cc}(\mathcal{X})]$ is defined by

$$D_p(F_1, F_2) = \left( \mathbf{E} \left( d_p^p(F_1, F_2) \right) \right)^{1/p}.$$  

(6)

**Example 2.** Let $\mathcal{X} = \mathbb{R}^1$ and $F_1(\omega) = [a_i(\omega), b_i(\omega)]$ represent random intervals, where $a_i(\omega)$ and $b_i(\omega)$ are two random variables and $a_i(\omega) \leq b_i(\omega), i = 1, 2$. By the definition of $D_p$ metric and Eq. (5), we have

$$D_p(F_1, F_2) = \left( \mathbf{E}[|a_2(\omega) - a_1(\omega)|^p + |b_2(\omega) - b_1(\omega)|^p] \right)^{1/p}.$$  

(7)

Let $\mathcal{L}^p[\Omega, \mathbf{K}_{cc}(\mathcal{X})] = \{F \in \mathbf{U}[\Omega, \mathbf{K}_{cc}(\mathcal{X})] : E\|F\|_{d_p}^p < +\infty\}$, where $\|F\|_{d_p} = d_p([0], F)$. For any $p \geq 1$, $(\mathcal{L}^p[\Omega, \mathbf{K}_{cc}(\mathbb{R}^d)], D_p)$ is a complete metric space [32].

2.3. Variance and covariance of set-valued random variables

The expectation of set-valued random variable $F$ was defined by Aumann [3] as

$$E(F) = \left\{ \int_{\Omega} fd\Pr : f \in S_F \right\},$$

(8) where $S_F = \{f : f(\omega) \in F(\omega) \ a.s. (Pr), \text{ and } f \text{ is integrable} \}$ is called the selection set of the set-valued random variable $F$. The properties of the expectation of set-valued random variables have been discussed in [20].

However, the subtraction of sets is difficult to define, since the space of subsets of $\mathcal{X}$ is not linear with respect to the addition and scalar multiplication. Therefore, extending the definitions of variance and covariance to the set-valued case is not a trivial task. Yang and Li [32] proposed to define the variance and covariance via the $D_p$ metric on $\mathbf{U}[\Omega, \mathbf{K}_{cc}(\mathbb{R}^d)]$, based on the fact that the support functions of sets are subtractive. Here, we recall the corresponding definitions.
Definition 1. The variance of a set-valued random variable $F \in \mathcal{U} [\Omega, \mathcal{K}_e (\mathcal{X})]$ is defined as

$$\text{var}(F) = D(F, E(F)) = E \left( \int_{S^e} (s(x^e, F) - s(x^e, E(F)))^2 d\mu \right).$$

(9)

The covariance of two set-valued random variables $F_1, F_2 \in \mathcal{U} [\Omega, \mathcal{K}_e (\mathcal{X})]$ is defined as

$$\text{cov}(F_1, F_2) = E \left( \int_{S^e} (s(x^e, F_1) - s(x^e, E(F_1))) (s(x^e, F_2) - s(x^e, E(F_2))) d\mu \right).$$

(10)

The correlation coefficient of $F_1$ and $F_2$, denoted by $\rho(F_1, F_2)$, is defined as

$$\rho(F_1, F_2) = \frac{\text{cov}(F_1, F_2)}{\sqrt{\text{var}(F_1) \cdot \text{var}(F_2)}}.$$  

(11)

Theorem 2 (32). The variance $\text{var}(F)$ of $F \in \mathcal{U} [\Omega, \mathcal{K}_e (\mathcal{X})]$ has the following properties:

1. $\text{var}(C) = 0$ for any constant $C \in \mathcal{K}_e (\mathcal{X})$.
2. $\text{var}(aF) = a^2 \text{var}(F)$ for any $a \geq 0$.
3. $\text{var}(F_1 + F_2) = \text{var}(F_1) + 2\text{cov}(F_1, F_2) + \text{Var}(F_2)$.
4. (Chebyshev Inequality) $\Pr(d_2(F, E(F)) \geq \varepsilon) \leq \frac{\text{var}(F)}{\varepsilon^2}$, for any $\varepsilon > 0$.

Theorem 3. Let $X_t(\omega) = [a_t(\omega), b_t(\omega)] = (c_t(\omega); r_t(\omega))$ and $X_{t+1}(\omega) = [a_{t+1}(\omega), b_{t+1}(\omega)] = (c_{t+1}(\omega); r_{t+1}(\omega))$ be two interval-valued random variables, where $c_t(\omega) = (a_t(\omega) + b_t(\omega))/2$ and $r_t(\omega) = (b_t(\omega) - a_t(\omega))/2$ are the center and radius of $X_t(\omega), i = 1, 2$. Then,

$$\text{cov}(X_1(\omega), X_2(\omega)) = \text{cov}(a_1(\omega), a_2(\omega)) + \text{cov}(b_1(\omega), b_2(\omega))$$

(12)

and

$$\text{cov}(X_1(\omega), X_2(\omega)) = 2\text{cov}(c_1(\omega), c_2(\omega)) + 2\text{cov}(r_1(\omega), r_2(\omega)).$$

(13)

3. Definition and properties of stationary set-valued time series

In this section, the stationary set-valued time series is defined based on the covariance introduced in Section 2. Here we denote $\mathbb{Z}^+ = \{0, 1, 2, \ldots \}$ and $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots \}$.

Definition 2. A set-valued time series $\{X_t : t \in \mathbb{Z}^+ \}$ is a sequence of set-valued random variables taking values in $\mathcal{K}_e (\mathcal{X})$. In particular, if $\mathcal{X} = \mathbb{R}^1$, $\{X_t \}$ is called interval-valued time series, denoted as $\{X_t \} = \{X_t = [a_t, b_t] = (c_t, r_t) : t \in \mathbb{Z}^+ \}$.

Definition 3. A set-valued time series $\{X_t : t \in \mathbb{Z}^+ \}$ is stationary if it satisfies

1. $\forall t \in \mathbb{Z}^+, E(X_t) = A$ where $A$ is a constant set.
2. $\forall t, s \in \mathbb{Z}^+, \text{cov}(X_t, X_s) = \gamma(t - s)$.

The set $A$ and the series $\{\gamma(k) : k \in \mathbb{Z} \}$ are called expectation and auto-covariance function of the stationary set-valued time series $\{X_t \}$, respectively. Furthermore, the auto-correlation function $\{\rho(k) \}$ of $\{X_t \}$ is defined by $\rho(k) = \gamma(k)/\gamma(0)$.

The following theorem summarizes properties of the auto-covariance function. The proof can be found in Appendix B.

Theorem 4. The auto-covariance function $\{\gamma(t) \}$ has the following properties:

1. $\gamma(k) = \gamma(-k)$ for all $k \in \mathbb{Z}$.
2. $|\gamma(k)| \leq \gamma(0)$ for all $k \in \mathbb{Z}$. 

(3) For all \( n \in \mathbb{Z}^+ \), auto-covariance matrix

\[
\Gamma_n = (\gamma(i - j))_{i,j=1}^n = \begin{pmatrix}
\gamma(0) & \gamma(1) & \cdots & \gamma(n - 1) \\
\gamma(1) & \gamma(0) & \cdots & \gamma(n - 2) \\
\vdots & \vdots & \ddots & \vdots \\
\gamma(n - 1) & \gamma(n - 2) & \cdots & \gamma(0)
\end{pmatrix}
\]  

(14)

is positive semi-definite.

(4) If \( \gamma(0) > 0 \) and \( \gamma(k) \to 0 \) as \( k \to \infty \), \( \Gamma_n \) is a positive definite matrix for all \( n \in \mathbb{Z}^+ \).

**Definition 4.** Two set-valued time series \( \{X_t\} \) and \( \{Y_t\} \) are said to be uncorrelated if for any \( s, t \in \mathbb{Z}^+ \), \( \text{cov}(X_s, Y_t) = 0 \).

**Theorem 5.** Let \( \{X_t\} \) and \( \{Y_t\} \) be two stationary set-valued time series with expectations \( \mu_X \) and \( \mu_Y \) and auto-covariance functions \( \gamma_X(k) \) and \( \gamma_Y(k) \). If \( \{X_t\} \) and \( \{Y_t\} \) are uncorrelated, then \( \{Z_t\} = \{X_t + Y_t\} \) is also stationary, and its auto-covariance function is

\[
\gamma_Z(k) = \gamma_X(k) + \gamma_Y(k).
\]  

(15)

This theorem can be immediately justified by applying Theorem 2 (3).

### 4. Parameter estimation of stationary set-valued time series

In this section, the estimation methods of the expectation and auto-covariance function of stationary set-valued time series and their properties are presented.

**4.1. Estimation of expectation \( A \)**

Assume that \( \{X_1, \ldots, X_n\} \) is a sample of stationary set-valued time series \( \{X_t : t \in \mathbb{Z}^+\} \) with expectation \( A \). The moment estimator of \( A \) is given by

\[
\bar{X} = \frac{X_1 + \cdots + X_n}{n}.
\]  

(16)

This is an unbiased estimator of \( A \), since the Aumann integral \([3]\) guarantees that

\[
E(\bar{X}) = \frac{E(X_1) + \cdots + E(X_n)}{n} = A.
\]  

(17)

Furthermore, the following theorem indicates that \( \bar{X} \) is a consistent estimator of \( A \), i.e., for any \( \epsilon > 0 \), \( \Pr(d_2(\bar{X}, A) \geq \epsilon) \to 0 \) as \( n \to \infty \). The proof of this theorem can be found in Appendix C.

**Theorem 6.** Let \( \{X_t\} \) be a stationary set-valued time series with expectation \( A \) and auto-covariance function \( \gamma(k) \). If \( \gamma(k) \to 0 \) as \( k \to \infty \), \( \bar{X} \) is a consistent estimator of \( A \).

**4.2. Estimation of auto-covariance function \( \gamma(k) \)**

The sample auto-covariance and auto-correlation functions are defined by

\[
\hat{\gamma}(k) = \frac{1}{n} \sum_{i=1}^{n-k} \int_{S^2} (s(y^*, X_{i+k}) - s(y^*, \bar{X}))(s(y^*, X_i) - s(y^*, \bar{X}))d\mu
\]  

(18)

and

\[
\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)}
\]  

(19)

For each \( n \geq 1 \), the \( n \)-dimensional sample auto-covariance and auto-correlation matrices are defined by

\[
\hat{\Gamma}_n = (\hat{\gamma}(i - j))_{i,j=1}^n = \begin{pmatrix}
\hat{\gamma}(0) & \hat{\gamma}(1) & \cdots & \hat{\gamma}(n - 1) \\
\hat{\gamma}(1) & \hat{\gamma}(0) & \cdots & \hat{\gamma}(n - 2) \\
\vdots & \vdots & \ddots & \vdots \\
\hat{\gamma}(n - 1) & \hat{\gamma}(n - 2) & \cdots & \hat{\gamma}(0)
\end{pmatrix}
\]  

(20)
and
\[ \hat{\gamma}_n = \frac{\hat{\gamma}_n}{\gamma(0)} \]  
(21)
providing \( \gamma(0) > 0 \). Eqs. (20) and (21) have the following properties, which are proved in Appendix D.

**Theorem 7.** (1) \( \hat{\gamma}_k \) is positive semi-definite matrices for all \( k \leq n \) and \( k \in \mathbb{Z}^+ \).
(2) If \( \hat{\gamma}(0) > 0 \), \( \Gamma_k \) and \( R_k \) are positive definite matrices for all \( k \leq n \) and \( k \in \mathbb{Z}^+ \).

Actually, both \( \hat{\gamma}(k) \) and \( \hat{\gamma}(k) \) are not unbiased estimators even if \( 1/n \) in Eq. (18) is replaced by \( 1/(n-k) \). However, the following theorem shows that they are approximately unbiased for a large sample size. The proof can be found in Appendix E.

**Theorem 8.** If \( \gamma(h) \rightarrow 0 \) as \( h \rightarrow \infty \), then for any \( h \), \( \hat{\gamma}(h) \) is an asymptotically unbiased estimator of \( \gamma(h) \), i.e.
\[ \lim_{n \rightarrow \infty} E(\hat{\gamma}(h)) = \gamma(h). \]  
(22)

**Example 3.** Consider the interval-valued case. The observations of an interval-valued time series are denoted by \( X_i = [a_i, b_i], i = 1, \ldots, n \). Then, the estimates of the expectation and auto-covariance function in Eqs. (16) and (18) can be explicitly given by
\[ \hat{\mathbf{X}} = [\hat{a}, \hat{b}] = \left[ \frac{1}{n} \sum_{i=1}^{n} a_i, \frac{1}{n} \sum_{i=1}^{n} b_i \right] \]  
(23)
and
\[ \hat{\gamma}(k) = \frac{1}{n} \sum_{i=1}^{n-|k|} (a_{i+|k|} - \hat{a})(a_i - \hat{a}) + (b_{i+|k|} - \hat{b})(b_i - \hat{b}). \]  
(24)

If we assume that the left- and right-endpoints of the interval-valued series follow two stationary Gaussian random processes: \( a_i \) and \( b_i \) follow Gaussian distributions \( \mathcal{N}(\mu_a, \sigma_a^2) \) and \( \mathcal{N}(\mu_b, \sigma_b^2) \), and have auto-covariance functions \( \gamma_a(k) \) and \( \gamma_b(k) \). Then, the MLE of the expectations and auto-covariance functions are
\[ \hat{\mu}_a = \hat{a}, \quad \hat{\gamma}_a(k) = \frac{1}{n} \sum_{i=1}^{n-|k|} (a_i - \hat{a})(a_{i+|k|} - \hat{a}). \]  
(25)
\[ \hat{\mu}_b = \hat{b}, \quad \hat{\gamma}_b(k) = \frac{1}{n} \sum_{i=1}^{n-|k|} (b_i - \hat{b})(b_{i+|k|} - \hat{b}). \]  
(26)

Then, the auto-covariance function of \( X_i \) is obtained from Theorem 3 (1):
\[ \text{cov}(X_i, X_{i+k}) = \frac{1}{n} \sum_{i=1}^{n-|k|} (a_{i+|k|} - \hat{a})(a_i - \hat{a}) + (b_{i+|k|} - \hat{b})(b_i - \hat{b}), \]  
(27)
which is identical to the moment estimate obtained from Eq. (24).

## 5. Stationary interval-valued time series

In this section, we consider the interval-valued time series, i.e., \( X = \mathbb{R}^1 \). In this case, \( K_{ex}(\mathbb{R}^1) = \{[a, b] : -\infty < a \leq b < \infty \} = \{(c, r) : -\infty < c < \infty, r \geq 0\} \), where \( c = (a + b)/2 \) and \( r = (b - a)/2 \) are the center and radius of interval, respectively. In the classical time series models, the distance between the actual and forecast values is used to describe the efficiency of forecast. In this work, we propose to use the \( D_2 \) metric to quantify the forecast error of interval. In Sections 5.1 and 5.2, the optimal linear and binary linear forecasts are investigated respectively. Then, a special kind of interval-valued time series, known as interval-valued autoregression model, is studied in Section 5.3.

### 5.1. Optimal linear forecast

In this section, we predict \( X_{t+h}, h \in \mathbb{Z}^+ \) using sample \( X_1, \ldots, X_n \) from a stationary interval-valued time series \( \{X_t, t \in \mathbb{Z}^+\} \) with expectation \( A = [a, b] \) and auto-covariance \( \gamma(h) \). Our purpose is to find the optimal linear combination of

\[ X_{t+h} = a \eta_t + b \rho_t, \]
${X_n, \ldots, X_1}$ in the sense of minimum forecast error. Assume that

$$k_1, \ldots, k_n \geq 0,$$

then the optimal linear predictor $P_n X_{n+h}$ has the form

$$P_n X_{n+h} = K_0 + k_1 X_1 + \cdots + k_n X_n = \left[ a_{K_0} + \sum_{i=1}^{n} k_i a_{X_i}, b_{K_0} + \sum_{i=1}^{n} k_i b_{X_i} \right],$$

where $K_0 = [a_{K_0}, b_{K_0}]$ and $X_i = [a_{X_i}, b_{X_i}], i = 1, 2, \ldots, n$. Now, we determine the coefficients $K_0, k_1, k_2, \ldots, k_n$ by minimizing the $D_2$-metric forecast error

$$S(a_{K_0}, b_{K_0}, k_1, \ldots, k_n) = (D_2(X_{n+h}, P_n X_{n+h}))^2$$

subject to $k_1, \ldots, k_n \geq 0$. Since $S(a_{K_0}, b_{K_0}, k_1, \ldots, k_n)$ is a convex quadratic function of $a_{K_0}, b_{K_0}, k_1, \ldots, k_n$ defined on the convex cone, there exists a unique value, say $(\tilde{a}_{K_0}, \tilde{b}_{K_0}, \tilde{k}_1, \ldots, \tilde{k}_n)$, of $(a_{K_0}, b_{K_0}, k_1, \ldots, k_n)$ that minimizes $S$ and satisfies $k_1, \ldots, k_n \geq 0$ and

$$\frac{\partial S}{\partial a_{K_0}} = 0, \quad \frac{\partial S}{\partial b_{K_0}} = 0.$$

Eq. (31) is equivalent to

$$E \left( a_{X_{n+h}} - \tilde{a}_{K_0} - \sum_{i=1}^{n} \tilde{k}_i a_{X_i} \right) = 0,$$

$$E \left( b_{X_{n+h}} - \tilde{b}_{K_0} - \sum_{i=1}^{n} \tilde{k}_i b_{X_i} \right) = 0,$$

which reduces to

$$\begin{cases}
\tilde{a}_{K_0} = a \left( 1 - \sum_{i=1}^{n} \tilde{k}_i \right) \\
\tilde{b}_{K_0} = b \left( 1 - \sum_{i=1}^{n} \tilde{k}_i \right)
\end{cases}.$$

Note that $\tilde{a}_{K_0} \leq \tilde{b}_{K_0}$ if $\sum_{i=1}^{n} \tilde{k}_i \leq 1$. The optimal linear forecast has the following properties. Their proofs can be found in Appendix F.

**Theorem 9.** The optimal linear predictor $P_n X_{n+h}$ is unique and satisfies

1. $E(P_n X_{n+h}) = E(X_{n+h})$.
2. $D_2^2(P_n X_{n+h}, X_{n+h}) = \gamma(0) - 2\tilde{K}_{n}^\top \gamma_{n}(h) + \tilde{K}_{n}^\top \Gamma_{n}^\top n_{n} \Gamma_{n} \tilde{K}_{n}$, in which $K_{n} = (\tilde{k}_1, \ldots, \tilde{k}_n)^\top$ and $\gamma_{n}(h) = (\gamma(h + n - 1), \ldots, \gamma(h))^\top$.
3. $\text{cov}(P_n X_{n+h}) = D_2^2(P_n X_{n+h}, E(P_n X_{n+h})) = \tilde{K}_{n}^\top \Gamma_{n} \tilde{K}_{n}$.

5.2. Optimal binary linear forecast

In this section, we propose another forecast method: optimal binary linear forecast, i.e., forecasting $X_{n+h}, h \in \mathbb{Z}^+$ through binary linear combination [30] of $X_n, \ldots, X_1$ which minimizes the $D_2$ error. The binary linear predictor is defined by

$$P_n X_{n+h} = K_0 + (k_1, i_1) X_1 + \cdots + (k_n, i_n) X_n = \left( c_{K_0} + \sum_{i=1}^{n} k_i c_{X_i}, r_{K_0} + \sum_{i=1}^{n} l_i r_{X_i} \right),$$

where $K_0 = (c_{K_0}, r_{K_0})$ and $X_i = (c_{X_i}, r_{X_i})$. Then, the coefficients $c_{K_0}, r_{K_0}, k_1, \ldots, k_n, l_1, \ldots, l_n$ are decided by minimizing

$$S(c_{K_0}, r_{K_0}, k_1, \ldots, k_n, l_1, \ldots, l_n) = D_2^2(X_{n+h}, P_n X_{n+h})$$

$$= 2E \left( \left( c_{X_{n+h}} - c_{K_0} - \sum_{i=1}^{n} k_i c_{X_i} \right)^2 + \left( r_{X_{n+h}} - r_{K_0} - \sum_{i=1}^{n} l_i r_{X_i} \right)^2 \right).$$
Let
\[ \frac{\partial S}{\partial c_{k_0}} = 0, \quad \frac{\partial S}{\partial r_{k_0}} = 0, \quad \frac{\partial S}{\partial k_j} = 0, \quad \frac{\partial S}{\partial l_j} = 0, \quad j = 1, 2, \ldots, n, \] (37)
that is,
\[ E\left( c_{x_{n+h}} - c_{k_0} - \sum_{i=1}^{n} k_i c_{x_i} \right) = 0, \] (38)
\[ E\left( r_{x_{n+h}} - r_{k_0} - \sum_{i=1}^{n} l_i r_{x_i} \right) = 0, \] (39)
\[ E\left( (c_{x_{n+h}} - c_{k_0} - \sum_{i=1}^{n} k_i c_{x_i}) c_{x_i} \right) = 0. \] (40)
\[ E\left( (r_{x_{n+h}} - r_{k_0} - \sum_{i=1}^{n} l_i r_{x_i}) r_{x_i} \right) = 0. \] (41)

Then, we obtain
\[
\begin{align*}
c_{k_0} &= c \left( 1 - \sum_{i=1}^{n} k_i \right), \\
r_{k_0} &= r \left( 1 - \sum_{i=1}^{n} l_i \right), \\
\gamma_{cn}(h) &= \Gamma_{cn} \mathbf{K}_n, \\
\gamma_{rn}(h) &= \Gamma_{rn} \mathbf{L}_n,
\end{align*}
\] (42)
where
\[
\mathbf{K}_n = (k_1, \ldots, k_n)^\top, \quad \mathbf{L}_n = (l_1, \ldots, l_n)^\top,
\] (43)
\[
\Gamma_{cn} = (\gamma_c (i-j))_{i,j=1}^{n}, \quad \Gamma_{rn} = (\gamma_r (i-j))_{i,j=1}^{n},
\] (44)
\[
\gamma_{cn}(h) = (\gamma_c (h + n - 1), \ldots, \gamma_c (h))^\top, \quad \gamma_{rn}(h) = (\gamma_r (h + n - 1), \ldots, \gamma_r (h))^\top.
\] (45) (46)

Since \( S(c_{k_0}, r_{k_0}, k_1, \ldots, k_n, l_1, \ldots, l_n) \) is a quadratic function and bounded below by zero, \( c_{k_0}, r_{k_0}, k_1, \ldots, k_n, l_1, \ldots, l_n \) minimizing \( S \) are obtained by Eq. (42). Then, the optimal binary linear predictor is
\[
P_n x_{n+h} = \left( c \left( 1 - \sum_{i=1}^{n} k_i \right) + \sum_{i=1}^{n} k_i c_{x_i}; r \left( 1 - \sum_{i=1}^{n} l_i \right) + \sum_{i=1}^{n} l_i r_{x_i} \right),
\] (47)
where \( k_i \) and \( l_i \) are decided by the last two equations of Eq. (42).

The uniqueness of the optimal binary linear predictor can be guaranteed by the fact that Eq. (36) is a strictly convex function.

5.3. Interval-valued autoregression model

In this section, the interval-valued time series model is investigated in the Box–Jenkins framework. More specifically, the interval autoregression (IAR) model is considered. Wang and Li [29] first introduced this model and used a minimum error of history data forecast to estimate the model parameters. This paper considers this model in the framework of set-valued time series and estimates the parameters using a statistical moment method. A \( p \)-order IAR model, denoted as IAR(\( p \)), is defined by
\[
X_t = A + \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \epsilon_t, \quad t \in \mathbb{Z}^+,
\] (48)
in which, \( \{X_t = [a_t, b_t] = (c_t; r_t) : t \in \mathbb{Z}^+ \} \) is a sequence of intervals, \( A = [a_0, b_0] = (c_0; r_0) \) is a constant interval, \( \alpha_1, \ldots, \alpha_p (> 0) \) are autoregression coefficients, and \( \{\epsilon_t, t \in \mathbb{Z}^+ \} \) is a series of uncorrelated and identically distributed random intervals satisfying that any \( \epsilon_t \) is uncorrelated with \( X_{t-1}, \ldots, X_1. \)
Assume that \{X_t\} is a stationary set-valued time series satisfying the model (48) and the roots of the polynomial \(y^p - \sum_{i=1}^{p} \alpha_i y^{p-i} = 0\) lie within the unit circle. The auto-covariance functions are obtained from Theorem 2 by
\[
\begin{pmatrix}
\gamma_1 \\
\vdots \\
\gamma_p
\end{pmatrix} = \begin{pmatrix}
\text{cov}(X_p, X_{p+1}) \\
\vdots \\
\text{cov}(X_p, X_{2p})
\end{pmatrix} = \alpha_1 \begin{pmatrix}
\text{cov}(X_p, X_p) \\
\vdots \\
\text{cov}(X_p, X_{p-1})
\end{pmatrix} + \cdots + \alpha_p \begin{pmatrix}
\text{cov}(X_p, X_{p}) \\
\vdots \\
\text{cov}(X_p, X_{p-1})
\end{pmatrix} = I_p \alpha,
\]
which is known as the Yule–Walker equation in the classical time series theory. Then, the autoregression coefficients \(\alpha = (\alpha_1, \ldots, \alpha_p)^T\) can be estimated by
\[
\hat{\alpha} = \left( \hat{r}_p^T \hat{r}_p \right)^{-1} \hat{r}_p \hat{y}_p,
\]
where \(\hat{r}_p = (\hat{\gamma}(i-j))_{i,j=1}^p\) and \(\hat{y}_p = (\hat{\gamma}(1), \ldots, \hat{\gamma}(p))^T\) are obtained using the auto-covariance estimation method in Section 4.2 and observed data \(X_1, \ldots, X_n\). By taking expectation on both sides of Eq. (48) and replacing \(\alpha_1, \ldots, \alpha_p\) by their estimates \(\hat{\alpha}_1, \ldots, \hat{\alpha}_p\), the estimate \(\hat{A}\) of the constant interval \(A\) can be obtained.

Then, \(X_{n+1}\) can be forecasted by
\[
\hat{X}_{n+1} = \hat{A} + \hat{\alpha}^T X + E(\epsilon_{n+1}),
\]
where \(X = (X_n, \ldots, X_{n-p+1})^T\).

Note that the proposed parameter estimation and forecast methods are independent of the distribution of random error interval \(\epsilon\). The robustness of the proposed method is justified by the estimation efficiency of moments shown in Section 4.2. Furthermore, if the distribution would be assumed or estimated, a common procedure to decide the order \(p\) in the classical autoregression model could be generalized to the interval case. In particular, the Akaike information criterion (AIC) [1] is defined as
\[
\text{AIC}(p) = 2p - \ln L,
\]
where \(L\) is the likelihood function of the model depending on the observed data and the estimated parameters. The optimal order of the model is determined by minimizing Eq. (52) with respect to \(p\). In general, a model with a higher order can better fit the observed data. However, in order to avoid overfitting, AIC decides the order via a trade-off between the goodness of fit (assessed by the likelihood function) and the complexity of the model (represented by a penalty term, i.e., the first term of the right side of Eq. (52)). Let us assume further that the center and radius series are uncorrelated and their distributions are explicitly given. Then, the goodness of fit of the model can be justified if the residuals follow the assumed distributions, which are defined as
\[
\epsilon_\tau = \hat{c}_0 - \sum_{i=1}^{p} \hat{\alpha}_i \epsilon_{\tau-i}, \quad \tau = p + 1, \ldots, n,
\]
and
\[
\epsilon_{\tau} = \hat{r}_0 - \sum_{i=1}^{p} |\hat{\alpha}_i| \epsilon_{\tau-i}, \quad \tau = p + 1, \ldots, n.
\]
Here, \(\hat{c}_0\) and \(\hat{r}_0\) are the center and radius of the estimate \(\hat{A}\).

6. Numerical studies

In this section, the usefulness of the proposed model is justified via a simulation experiment and a real data example.

6.1. Simulations experiment

In this subsection, we consider a 2-order IAR model
\[
X_t = A + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \epsilon_t, \quad t = 3, \ldots, n,
\]
where \(\epsilon_t\) is a series of independent random intervals whose center \(c_\epsilon\) and radius \(r_\epsilon\) follow Gaussian distribution \(\mathcal{N}(0, \sigma^2)\) and exponential distribution \(\exp(\lambda)\) respectively. Here we let \(\alpha_1 = 0.6, \alpha_2 = 0.2,\) and \(A = [0.27, 0.33]\). Fig. 1 shows a realization of \(\{X_t\}\) with a length of time \(T = 100\). The error parameters are \(\sigma = 0.3\) and \(\lambda = 20\). Then, the sample mean \(\bar{X}\) and auto-covariance function \(\hat{\gamma}(k), k = 0, 1, \ldots, 98\), can be estimated via Eqs. (23) and (24), respectively. Fig. 2 displays an estimate of \(\hat{\gamma}(k), k = 0, 1, \ldots, 98\), which indicates that as \(k\) increases, the sample auto-covariance function \(\hat{\gamma}(k)\) decreases and tends to 0. This result is consistent with the model assumption Eq. (55). Then, the model parameters \(\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2)^T\) and \(\hat{A}\) can be estimated by
\[
\hat{\alpha} = \left( \hat{r}_2^T \hat{r}_2 \right)^{-1} \hat{r}_2 \hat{y}_2
\]
Fig. 1. A realization of interval-valued time series \( \{X_t, t = 1, \ldots, 100\} \).

Fig. 2. Estimates of \( \gamma(k) \), \( k = 0, 1, \ldots, 98 \).

Table 1

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \hat{\alpha}_1 )</th>
<th>( \hat{\alpha}_1 )’s error</th>
<th>( \hat{\alpha}_2 )</th>
<th>( \hat{\alpha}_2 )’s error</th>
<th>( \hat{A} )</th>
<th>( \hat{A} )’s error</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.646</td>
<td>0.046</td>
<td>0.100</td>
<td>0.100</td>
<td>[0.274, 0.378]</td>
<td>0.048</td>
</tr>
<tr>
<td>500</td>
<td>0.632</td>
<td>0.032</td>
<td>0.156</td>
<td>0.044</td>
<td>[0.288, 0.353]</td>
<td>0.029</td>
</tr>
<tr>
<td>1000</td>
<td>0.615</td>
<td>0.015</td>
<td>0.189</td>
<td>0.011</td>
<td>[0.270, 0.326]</td>
<td>0.004</td>
</tr>
</tbody>
</table>

Table 1 shows the parameter estimates using three sets of data \( \{X_t, t = 1, \ldots, n\} \) obtained from Eq. (55) with \( n = 100, 500, 1000 \) respectively. Both the estimates and their errors are displayed in Table 1, in which the error of the estimated constant interval \( \hat{A} \) is defined by its \( d_2 \)-distance to the true value \( A \). The results justify the robustness of the proposed method: as the sample size increases, all the three estimates converge to their true values.

In order to decide the optimal order of the IAR model, the parameter estimation procedure is repeated with assumed order \( p = 1, \ldots, 5 \). Here the simulated data are generated with a sample size \( n = 500 \), the error parameters are \( \sigma = 0.003 \)

and

\[
\hat{A} = \left(1 - \hat{\alpha}_1 - \hat{\alpha}_2\right) \hat{a}_X + \frac{1}{\lambda} \left(1 - \hat{\alpha}_1 - \hat{\alpha}_2\right) \hat{b}_X - \frac{1}{\lambda},
\]  

where \( \hat{a}_X \) and \( \hat{b}_X \) are the left- and right-endpoints of the sample mean \( \bar{X} \). In order to investigate the convergence of the parameter estimates, the above procedure is repeated with different sample size. Table 1 shows the parameter estimates using three sets of data \( \{X_t, t = 1, \ldots, n\} \) obtained from Eq. (55) with \( n = 100, 500, 1000 \) respectively. Both the estimates and their errors are displayed in Table 1, in which the error of the estimated constant interval \( \hat{A} \) is defined by its \( d_2 \)-distance to the true value \( A \). The results justify the robustness of the proposed method: as the sample size increases, all the three estimates converge to their true values.

In order to decide the optimal order of the IAR model, the parameter estimation procedure is repeated with assumed order \( p = 1, \ldots, 5 \). Here the simulated data are generated with a sample size \( n = 500 \), the error parameters are \( \sigma = 0.003 \)
and $\lambda = 20$. Since the center and radius of the random error intervals follow a Gaussian and an exponential distribution respectively, the log-likelihood function of the observations $X_t = (c_t; r_t)$, $t = p + 1, \ldots, n$, given the parameter estimates is

$$
\ln L = \ln L_c + \ln L_r,
$$

in which $L_c$ and $L_r$ are the probability density functions (PDFs) of the residual sequences of center and radius ($\epsilon_{ct}$ and $\epsilon_{rt}$, $t = p + 1, \ldots, n$, in Eqs. (53) and (54)). Fig. 3(a) displays the log-likelihood functions with five different assumed orders $p = 1, \ldots, 5$. Obviously, the model with $p = 2$ fits the data better than $p = 1$, but a higher order does not further improve the model since the log-likelihood function has no significant change. Then, by substituting Eq. (58) back to Eq. (52), the optimal order of the IAR model can be decided via AIC. Fig. 3(b) shows the values of AIC as a function of $p$, which return an optimal order $p = 2$ (equal to the order assumed in the IAR model (55)). Finally, the PDFs of the residuals of center and radius are computed using kernel density estimation, which are shown in Fig. 4. It is clear that the estimated PDFs are almost overlapped with the actual ones, which justifies the goodness of fit of the used model.

### 6.2. Empirical study

In this subsection, an interval-valued stock price analysis is shown to present how the proposed model can be used to deal with practical problems. More specifically, daily stock price is modeled by an IAR model, the model parameters are estimated by the interval-valued data, and finally future prices are forecasted by the estimated model.

Here, we use the data of Shanghai Stock Index from 22 January to 14 April 2015 (53 trading days), which are shown in Fig. 5. Each line segment in this figure stands for the interval-valued price of each day, composed by the daily lowest and highest prices. The interval-valued price is modeled by an IAR(2) model:

$$
X_t = A + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \epsilon_t,
$$

Fig. 3. (a) likelihood function of observations and (b) AIC as a function of order $p$.

Fig. 4. PDFs of the center (a) and the radius (b) of random error interval. The curves and points represent the real and estimated PDFs respectively.
where $\epsilon_t$ follows an interval-valued random variable with expectation $\{0\}$. Then, the autoregression coefficients $\alpha = (\alpha_1, \alpha_2)^\top$ and the constant interval $A$ are estimated by

$$\hat{\alpha} = \left( \hat{I}_2^\top \hat{I}_2 \right)^{-1} \hat{I}_2 \hat{\gamma}_2 = (0.87, 0.06)^\top$$

(60)

and

$$\hat{A} = (1 - \hat{\alpha}_1 - \hat{\alpha}_2) \bar{X} = [238, 243].$$

(61)

Finally, the prices $\tilde{X}_t$ of the following 17 trading days (15 April–8 May, 2015) are forecasted by the estimated model

$$\tilde{X}_t = \hat{A} + \hat{\alpha}_1 X_{t-1} + \hat{\alpha}_2 X_{t-2}, \quad t = 54, \ldots, 70.$$  

(62)

Fig. 6 shows the forecast intervals of the 17 days (dashed line segments) and their actual values (solid line segments). It is clear that the estimated model is able to reflect the future trend of the stock index.

We remark that in this example the parameter estimation and forecast methods are based on the estimates of first and second moments which do not depend on the distribution of stock prices. In fact, estimating the distribution needs a relative high sample size, especially for stock price which may be a function of time. The proposed method provides a way to model the interval-valued time series in the case of small sample and lack of information of distribution in real applications.

7. Conclusions

The stationarity is a significant concept in the classical time series, which signifies that the expectation and covariance do not vary with respect to time. However, this definition cannot be straightforwardly generalized to the set-valued time
These second inequality holds due to Schwarz inequality. The unbiasedness and consistency of the expectation estimator and the asymptotic unbiasedness of the auto-covariance function estimator are also justified.

Interval-valued time series is a special case of set-valued time series, which can be used to deal with a wide range of practical problems, e.g., stock price analysis and weather forecast. In this paper, the optimal linear and binary linear forecast methods are presented to forecast stationary interval-valued time series. In order to further crystallize the interval-valued time series, the IAR model and its parameter estimation method are investigated. Then, a simulation experiment is carried out to justify the performance of the proposed model and methods. Finally, an empirical study of stock price analysis and forecast is introduced to present how the proposed model can be used in real applications.

The parameter estimation and forecast methods proposed in this paper do not assume the distribution of interval-valued error. The modification of the model to take into account the distribution and the corresponding MLE would be a nature extension of the current approach. In fact, Example 3 has shown that the MLE under the assumption of Gaussian-distributed interval endpoints introduces the same estimates of random interval’s expectation and auto-covariance function as the proposed method. A deeper study of MLE with other distributions would be interesting to investigate the relationship between the moment estimation and MLE in the interval-valued time series.

**Acknowledgments**

The research is financially supported by NSFC (No. 11571024 and No. 11271039). Zhongzhan Zhang’s research is also partly supported by Collaborative Innovation Center on Capital Social Construction and Social Management (No. 006000546615539). The authors would also like to acknowledge the anonymous reviewers for their valuable suggestions and comments.

**Appendix A. Proof of Theorem 3**

By the definition of covariance of set-valued random variables,

\[
\text{cov}(X_1(\omega), X_2(\omega)) = E \left( \int_{S^*} \left( s(y^*, X_1) - s(y^*, E(X_1)) \right) \left( s(y^*, X_2) - s(y^*, E(X_2)) \right) d\mu \right) \\
= E \left( (a_1 - E(a_1))(a_2 - E(a_2)) + (b_1 - E(b_1))(b_2 - E(b_2)) \right) \\
= \text{cov}(a_1, a_2) + \text{cov}(b_1, b_2).
\]  

Replacing \( a_i, b_i \) in (A.1) by \( c_i - r_i, c_i + r_i, \ i = 1, 2 \), we obtain

\[
\text{cov}(X_1(\omega), X_2(\omega)) = E \left( (c_1 - E(c_1) - (r_1 - E(r_1)))(c_2 - E(c_2) - (r_2 - E(r_2))) \right) \\
+ E \left( (c_1 - E(c_1) + (r_1 - E(r_1)))(c_2 - E(c_2) + (r_2 - E(r_2))) \right) \\
= 2\text{cov}(c_1, c_2) + 2\text{cov}(r_1, r_2).
\]

**Appendix B. Proof of Theorem 4**

(1) is obvious by the definition of auto-covariance function.

(2) The second property holds by

\[
|\gamma(k)| \leq E \int_{S^*} \left| s(y^*, X_{k+1}) - s(y^*, A) \right| \left| s(y^*, X_1) - s(y^*, A) \right| d\mu \\
\leq E \left( \left( \int_{S^*} \left| s(y^*, X_{k+1}) - s(y^*, A) \right|^2 d\mu \right)^{1/2} \left( \int_{S^*} \left| s(y^*, X_1) - s(y^*, A) \right|^2 d\mu \right)^{1/2} \right) \\
\leq \gamma(0).
\]

The second inequality holds due to Schwarz inequality.

(3) For any real vector \( \alpha = (a_1, \ldots, a_n)^T \), we have

\[
\alpha^T I_{n \alpha} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j E \left( \int_{S^*} \left( s(y^*, X_i) - s(y^*, A) \right) \left( s(y^*, X_j) - s(y^*, A) \right) d\mu \right) \\
= E \left( \int_{S^*} \left( \sum_{i=1}^{n} a_i \left( s(y^*, X_i) - s(y^*, A) \right) \right)^2 d\mu \right) \geq 0.
\]
Therefore, $\Gamma_n$ is positive semi-definite.

(4) Suppose that there exists an $n$, s.t. $\Gamma_n$ is positive definite, but $|\Gamma_{n+1}| = 0$. Therefore, for any $b = (b_1, \ldots, b_n)^T \neq 0$, $b^T \Gamma_n b > 0$ and there exists an $a = (a_1, \ldots, a_{n+1})^T \neq 0$ with $a_{n+1} \neq 0$,

$$a^T \Gamma_{n+1} a = E \left( \int_{S_n} \left( \sum_{i=1}^{n+1} a_i (s(y^n, X_i) - s(y^n, A)) \right)^2 d\mu \right) = 0. \quad (B.3)$$

Thus, $\sum_{i=1}^{n+1} a_i (s(y^n, X_i) - s(y^n, A)) = 0$, a.s., that is, $s(y^n, X_{n+1}) - s(y^n, A)$ can be represented by a linear combination of $s(y^n, X_i) - s(y^n, A)$, $i = 1, \ldots, n$. Moreover, for any $k \geq 1$, $s(y^n, X_{n+k}) - s(y^n, A)$ can be represented by the linear combination of $s(y^n, X_1), \ldots, s(y^n, X_n) - s(y^n, A)$, a.s., which means that there exists a vector $c = (c_1, \ldots, c_n)^T \neq 0$, such that

$$s(y^n, X_{n+k}) - s(y^n, A) = c^T s, \quad (B.4)$$

where $s = (s(y^n, X_1), \ldots, s(y^n, X_n) - s(y^n, A))^T$. Let $0 < \lambda_1 \leq \cdots \leq \lambda_n$ be the eigenvalues of $\Gamma_n$, then there exists an orthogonal matrix $T$, s.t., $T \Gamma_n T^T = \text{diag}(\lambda_1, \ldots, \lambda_n)$. Hence, we have

$$\gamma(0) = E \int_{S_n} (s(y^n, X_{n+k}) - s(y^n, A))^2 d\mu = E \int_{S_n} c^T c s^T s d\mu \geq \lambda_1 c^T T^T c = \lambda_1 |c|^2, \quad (B.5)$$

therefore, $|c| \leq \sqrt{\gamma(0)/\lambda_1}$. Moreover,

$$\gamma(0) = E \int_{S_n} c^T s (s(y^n, X_{n+k}) - s(y^n, A)) d\mu = c^T \begin{pmatrix} \gamma(n+k-1) \\ \gamma(n+k-2) \\ \vdots \\ \gamma(k) \end{pmatrix} \leq |c| \left( \sum_{i=1}^{n} \gamma^2(n+k-i) \right)^{\frac{1}{2}} \leq \left( \frac{\gamma(0)}{\lambda_1} \sum_{i=1}^{n} \gamma^2(n+k-i) \right)^{\frac{1}{2}} \rightarrow 0, \quad \text{as } k \rightarrow \infty, \quad (B.6)$$

which contradicts $\gamma(0) \geq \lambda_1 |c|^2$. Thus we conclude $|\Gamma_{n+1}| \neq 0$.

### Appendix C. Proof of Theorem 6

The variance of $\bar{X}$ is

$$\text{var}(\bar{X}) = D_2^2(\bar{X}, A) = E \left( \int_{S_n} (s(y^n, \bar{X}) - s(y^n, A))^2 d\mu \right)$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} E \left( \int_{S_n} (s(y^n, X_i) - s(y^n, A))(s(y^n, X_j) - s(y^n, A)) d\mu \right)$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma(i-j) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (n - |i - j|) \gamma(i-j)$$

$$= \frac{1}{n} \sum_{k=-n}^{n} \left( 1 - \frac{|k|}{n} \right) \gamma(k). \quad (C.1)$$

Therefore, if $\gamma(k) \rightarrow 0$ as $k \rightarrow \infty$, the right hand side of Eq. (C.1) converges to zero. By Chebyshev inequality (Theorem 1 (4)), $\forall \varepsilon > 0$, $\Pr(d_2(\bar{X}, A) \geq \varepsilon) \leq \text{Var}(\bar{X})/\varepsilon^2 \rightarrow 0$ as $n \rightarrow \infty$. Therefore, $\bar{X}$ is a consistent estimator of $A$. 

Appendix D. Proof of Theorem 7

(1) It is manifested that \( \hat{T}_k \) is positive semi-definite for all \( k < m \) if \( \hat{T}_m \) is positive semi-definite, so we only need to prove the case \( k = n \).

Let \( n \times 2n \) matrix \( T_n \) be

\[
T_n = \begin{pmatrix}
0 & \cdots & 0 & Y_1 & \cdots & Y_{n-1} & Y_n \\
0 & \cdots & 0 & Y_1 & \cdots & Y_{n-1} & 0 \\
\vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
0 & Y_1 & \cdots & Y_{n-1} & Y_n & \cdots & 0
\end{pmatrix},
\]

in which \( Y_i = s(y^*, X_i) - s(y^* - \bar{X}), i = 1, 2, \ldots, n \). Then, \( \hat{T}_n = \frac{1}{n} \int_{S^*} T_n T_n^T d\mu \). Hence, for any real \( n \times 1 \) vector \( \alpha \), we have

\[
\alpha^T \hat{T}_n \alpha = \frac{1}{n} \int_{S^*} (T_n^T \alpha)^T (T_n^T \alpha) d\mu \geq 0,
\]

which implies that \( \hat{T}_n \) is positive semi-definite.

(2) Assume that \( \hat{\gamma}(0) > 0 \) and \( \exists k \in \mathbb{Z}^+, \hat{T}_k \) is not positive definite, i.e. \( \hat{T}_k \) is singular. Therefore, there exists a nonzero vector \( \alpha \), such that

\[
\alpha^T \hat{T}_k \alpha = \frac{1}{n} \int_{S^*} \alpha^T T_k T_k^T \alpha d\mu = \frac{1}{n} \int_{S^*} (T_k^T \alpha)^T (T_k^T \alpha) d\mu = 0,
\]

which implies \( \alpha^T T_k = 0 \) a.s.(\( \mu \)), and hence the rank of \( T_k \) is almost surely less than \( k \). On the other hand, \( \hat{\gamma}(0) > 0 \) if and only if there is at least one \( Y_i \), which is nonzero with positive measure based on \( \mu \). Let \( Y_i \) be the first nonzero value of \( Y_1, \ldots, Y_k \), and consider the \( k \times k \) submatrix of \( T_k \) consisting of columns from \((i + 1)\) to \((i + k)\). Obviously, this submatrix is nonsingular with positive measure since its determinant has absolute value \( |Y_i|^k \neq 0 \), which indicates \( T_k \) has rank \( k \), a contradiction.

Appendix E. Proof of Theorem 8

Let \( E(X_1) = A \), then we have

\[
\hat{\gamma}(k) = \frac{1}{n} \sum_{i=1}^{n-k} \int_{S^*} (s(y^*, X_{i+k}) - s(y^* - \bar{X})) (s(y^*, X_i) - s(y^* - \bar{X})) d\mu
\]

\[
= \frac{1}{n} \sum_{i=1}^{n-k} \int_{S^*} (s(y^*, X_{i+k}) - s(y^*, A)) (s(y^*, X_i) - s(y^*, A)) d\mu + \frac{1}{n} \sum_{i=1}^{n-k} \int_{S^*} (s(y^*, \bar{X}) - s(y^*, A))^2 d\mu
\]

\[
= \frac{1}{n} \sum_{i=1}^{n-k} \int_{S^*} (s(y^*, \bar{X}) - s(y^*, A)) (s(y^*, X_{i+k}) + s(y^*, X_i) - 2s(y^*, A)) d\mu.
\]

Therefore, we have

\[
\lim_{n \to \infty} E(\hat{\gamma}(k)) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n-k} \gamma(k) + \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n-k} \text{var}(\bar{X}) - \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n-k} (\text{cov}(\bar{X}, X_{i+k}) + \text{cov}(\bar{X}, X_i))
\]

\[
= \gamma(k) - \lim_{n \to \infty} \frac{1}{n^2} \sum_{i=1}^{n-k} \sum_{j=1}^{n} (\gamma(j-i) + \gamma(j-i))
\]

\[
= \gamma(k) - \lim_{n \to \infty} \frac{1}{n^2} \sum_{i=1}^{n} (n - |i|) (\gamma(j-i) + \gamma(j-i))
\]

\[
= \gamma(k) - \lim_{n \to \infty} \frac{1}{n} \sum_{i=-n}^{n} \left(1 - \frac{|i|}{n}\right) (\gamma(|i|) + \gamma(i)) = \gamma(k).
\]

The last equality holds due to \( \gamma(h) \to 0 \) as \( h \to \infty \).
Appendix F. Proof of Theorem 9

By Eq. (34), it holds that

$$P_nX_{n+h} = \left[ a + \sum_{i=1}^{n} \tilde{k}_i (aX_i - a), b + \sum_{i=1}^{n} \tilde{k}_i (bX_i - b) \right]. \quad (F.1)$$

(1) The unbiasedness is immediately obtained by taking expectation on both sides of Eq. (F.1).

(2) Eq. (F.1) leads to

$$D_n^2 (P_nX_{n+h}, X_{n+h}) = E \left( aX_{n+h} - a - \sum_{i=1}^{n} \tilde{k}_i (aX_i - a) \right)^2 + E \left( bX_{n+h} - b - \sum_{i=1}^{n} \tilde{k}_i (bX_i - b) \right)^2$$

$$= \gamma (0) - 2 \sum_{i=1}^{n} \tilde{k}_i \gamma (n + h - i) + \sum_{i=1}^{n} \sum_{j=1}^{n} \tilde{k}_i \tilde{k}_j \gamma (i - j)$$

$$= \gamma (0) - 2 \mathbf{K}_n^{\top} \gamma_n (h) + \mathbf{K}_n^{\top} \Gamma_n \mathbf{K}_n. \quad (F.2)$$

(3) By Theorem 3,

$$D_n^2 (P_nX_{n+h}, E(P_nX_{n+h})) = E \left( \sum_{i=1}^{n} \tilde{k}_i (aX_i - a) \right)^2 + E \left( \sum_{i=1}^{n} \tilde{k}_i (bX_i - b) \right)^2$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \tilde{k}_i \tilde{k}_j \text{cov}(X_i, X_j) = \mathbf{K}_n^{\top} \Gamma_n \mathbf{K}_n. \quad (F.3)$$

References


