

集值和区间值的多元时间序列

钟 钰, 李寿梅, 章 磊
(北京工业大学应用数理学院, 北京 100124)

摘 要: 研究集值多元时间序列的一些初步的理论, 这为拓广经典的多元时间序列模型提供了理论基础. 首先, 基于集值理论, 进一步提出集值向量、集值随机向量的定义, 并给出集值随机向量的期望向量、交叉协方差阵与交叉相关阵的定义与性质. 然后, 在此基础上, 给出集值多元时间序列的定义, 并研究关于集值多元时间序列的平稳性, 期望向量、交叉协方差阵和交叉相关阵的定义及性质, 讨论平稳的集值多元时间序列的最优线性预测问题. 最后, 在集值多元时间序列的基础之上, 讨论区间值多元时间序列, 并建立区间值多元自回归模型. 模拟研究与实证分析验证了该模型与所提出方法的合理性.

关键词: 集值随机向量; 集值多元时间序列; 交叉协方差阵; 交叉相关阵; 平稳性; 区间值多元自回归模型

中图分类号: O 211.9

文献标志码: A

文章编号: 0254-0037(2020)02-0208-09

doi: 10.11936/bjutxb2019030021

Set-valued and Interval-valued Multivariate Time Series

ZHONG Yu, LI Shoumei, ZHANG Lei

(College of Applied Sciences, Beijing University of Technology, Beijing 100124, China)

Abstract: Some preliminaries for set-valued multivariate time series were studied, which provides a theoretical basis for extending the classical multivariate time series models. First, this paper introduced the definitions of set-valued vectors and set-valued random vectors based on the set-valued random variable theory. The definitions of expectation vectors, cross covariance matrixes and cross correlation matrixes were also given in this paper. Then, the set-valued multivariate time series was defined and the stationary property, expectation vectors, cross covariance matrixes and cross correlation matrixes of set-valued multivariate time series were discussed. Furthermore, the optimal linear forecast of stationary set-valued multivariate time series was studied. Finally, the interval-valued multivariate time series were discussed based on set-valued multivariate time series, and the interval-valued vector autoregressive model was built. Simulation study and real data analysis show the efficiency of the proposed model and approach.

Key words: set-valued random vector; set-valued multivariate time series; cross-covariance matrix; cross-correlation matrix; stationary; interval-valued multivariate autoregressive model

时间序列分析的研究有着悠久的历史以及大量的学术成果, 并在很多领域中都有着广泛的应用,

例如 Box 等^[1]、Tsay^[2]、Brockwell 等^[3]的专著及其所引用的参考文献. 经典时间序列模型研究的对象

收稿日期: 2019-03-25

基金项目: 国家自然科学基金资助项目(11571024)

作者简介: 钟 钰(1989—), 女, 博士研究生, 主要从事集值、区间值的统计模型及其应用方面的研究, E-mail: emmaestchong@emails.bjut.edu.cn

通信作者: 李寿梅(1963—), 女, 教授, 主要从事随机理论及其应用方面的研究, E-mail: lishoumei2011@163.com

通常是随着时间推移而形成的随机变量或随机向量序列,通过建立时间序列模型来分析这些序列的动态规律.

然而,随着技术的发展,人们收集到的数据变得更加复杂,经典的时间序列模型不再适用于一些较为复杂的数据.例如,在分析股票数据时通常选取每个交易日的收盘价作为研究对象.然而,当今的技术使得股票数据刷新频率很高,收盘价并不足以反映当天股票价格的变化.比如说,某一个交易日里某一支股票价格所有的观测值为 $x_1, x_2, \dots, x_n$, 这里 n 的值很大,而收盘价只是集合 $\{x_1, x_2, \dots, x_n\}$ 中的一个元素.为能更准确地研究股票数据,有一种方法是取每个交易日里股票价格的最高价和最低价,将其作为一个区间的2个端点,即研究对象是

$$X_t = [\min\{x_1, x_2, \dots, x_n\}, \max\{x_1, x_2, \dots, x_n\}]$$

将区间 X_t 视为一个数据单元来研究问题.可以看出,区间 X_t 不仅包含了收盘价的信息,而且还有当天股价变动的情况,这无疑是一个好的处理方法.通过区间值或集值数据来描述一类复杂数据,是一种分析复杂数据的手段.另一方面,由于系统的复杂性,使得到的数据是某个范围而不是一个确定的实数值.比如让专家估计明年中国的经济增长率,专家们往往会给出 $3\% \sim 6\%$, $3.5\% \sim 5.7\%$ 这样的数据,显然用区间来刻画这类数据更加贴切.引入区间值和集值数据,不仅是对观测数据的一种处理方式,而且还是对收集到的数据一种客观并且准确的体现.

事实上,区间是集合的一个特例,区间值随机变量的相关性质都可以从集值随机变量的性质中推导.自从20世纪中叶,以经济学诺贝尔奖获得者 Aumann 和 Debreu 为代表的一批研究者在研究市场均衡等经济问题时引入了集值映射的概念及方法,集值随机理论有了较为完善的发展. Aumann^[4] 通过集值随机变量的选择集给出了集值随机变量期望的定义,这里集值随机变量的期望是选择集中元素的期望构成的集合. Hiai 等^[5] 定义了集值随机变量的条件期望. Lyashenko^[6] 讨论了欧式空间中集值随机变量的性质,提出了集值随机变量方差的定义,并给出了集值高斯随机变量的定义及表达定理. Vitale^[7] 研究了集合的 L_p 距离的性质. 需注意的是,相较于集合的 Hausdorff 距离, L_p 距离具有非常好的数学性质,这使得集值和区间值的统计模型大多数都建立在 L_p 距离之上. Diamond^[8] 讨论了区间值仿射函数的最小二乘估计方法. Li 等^[9-10] 研究了

集值鞅的收敛性问题. Li 等^[11] 的书中系统地总结了集值随机变量的空间、期望与条件期望,强大数定律与中心极限定理等理论. Gil 等^[12] 推广了 L_p 距离. Yang 等^[13] 通过 L_p 距离定义了集值随机变量的协方差和相关系数,并讨论了方差、协方差和相关系数的性质. 张文修等^[14] 的书中阐述了集值随机过程的一系列理论.

作为集合最简单的特例,区间值时间序列被广泛地讨论.有一类文献是将区间的2个端点或中心半径看作2个实值随机变量,然后通过区间的2个端点或中心半径来给出区间值时间序列模型,例如 Wang 等^[15]、González-Rivera 等^[16]、唐娜娜^[17]等.而另一类文献则是将区间值数据看成一个整体,使用集值随机理论来分析区间值数据,并建立区间值时间序列模型,例如王洵^[18]、Wang 等^[19]、章磊^[20]等.

区间值时间序列可以看作集值时间序列的一个特例,区间值时间序列的相关理论可以通过集值时间序列推出,而对集值时间序列的讨论建立在集值随机过程理论的框架下. 张文修等^[14] 的专著中,系统地介绍了集值随机过程的一系列理论. Wang 等^[21] 定义了严平稳的集值随机过程并讨论了它的一系列性质.然而,虽然在严平稳性的条件下,集值随机过程有许多好的性质,但在实际应用中难以使用.因此 Wang 等^[19] 通过集值随机变量的期望^[4] 和协方差^[13] 定义了集值时间序列的弱平稳性,并进一步讨论了弱平稳条件下集值时间序列统计量的估计及其渐近性质.

本文在 Wang 等^[19]、章磊^[20] 研究的基础之上,进一步讨论了集值随机向量和集值多元时间序列的相关定义及性质.首先,给出集值向量与集值随机向量的相关理论;然后,讨论集值多元时间序列;第三,给出集值多元时间序列在区间值情形下的应用;第四,给出区间值多元自回归模型的一个实证分析;最后,简要地给出结论.

1 集值向量与集值随机向量

1.1 集合与集值向量

记 $\mathbb{R}^d$ 为 d -维欧式空间, $\mathbb{R}^d$ 中所有非空紧凸集为 $K(\mathbb{R}^d)$. 任意2个集合 A, B 的加法与数乘运算定义为

$$A + B = \{a + b; a \in A, b \in B\}$$

$$\lambda A = \{\lambda a; a \in A\}, \forall \lambda \in \mathbb{R}$$

Vitale^[7] 通过支撑函数定义了 $K(\mathbb{R}^d)$ 中2个元素间的 L_p 距离, $1 \leq p < \infty$. 集合 $A \in K(\mathbb{R}^d)$ 的支撑

函数定义为

$$s(\mathbf{x}, A) = \sup_{a \in A} \langle \mathbf{x}, \mathbf{a} \rangle, \mathbf{x} \in S$$

式中: $\langle \cdot, \cdot \rangle$ 为 $\mathbb{R}^d$ 中 2 个向量的内积; S 为 $\mathbb{R}^d$ 中的单位球面. 集合 A 和 B 间的 L_p 距离定义为

$$L_p(A, B) = \left[\int_S |s(\mathbf{x}, A) - s(\mathbf{x}, B)|^p d\mu \right]^{\frac{1}{p}}$$

不失一般性, 本文接下来取 $p = 2$. 进一步地, 由 Yang 等^[13] 可知空间 $(K(\mathbb{R}^d), L_2)$ 是完备可分的.

称由 k 个 $K(\mathbb{R}^d)$ 中的集合 $A_1, A_2, \dots, A_k$ 构成的集值向量 $\mathbf{A} = (A_1, A_2, \dots, A_k)^T$ 为 k 维集值向量. 记 k 维集值向量族为 $K^k(\mathbb{R}^d)$. 需注意的是, 集值向量 $\mathbf{A}$ 是 $\mathbb{R}^{kd}$ 中一个 k 维的超立方体. 由于假定了集值向量中每个元素都是 $K^k(\mathbb{R}^d)$ 中的集合, 因此, 可以推出关系式 $K^k(\mathbb{R}^d) \subseteq K(\mathbb{R}^{kd})$ 成立.

假定 $k \leq m$, 任意的 k 维集值向量和 m 维集值向量 $\mathbf{A} = (A_1, A_2, \dots, A_k)^T, \mathbf{B} = (B_1, B_2, \dots, B_m)^T$ 间的加法运算定义为

$$\mathbf{A} + \mathbf{B} = (A_1 + B_1, \dots, A_k + B_k, B_{k+1}, \dots, B_m)^T$$

进一步地, k 维集值向量 $\mathbf{A}$ 与 $p \times k$ 的实值矩阵 $\mathbf{A} = (\lambda_{i,j})_{i=1,2,\dots,p,j=1,2,\dots,k}$ 间的矩阵乘法运算定义为

$$\mathbf{A}\mathbf{A} = \left(\sum_{j=1}^k \lambda_{1,j} A_j, \dots, \sum_{j=1}^k \lambda_{p,j} A_j \right)^T$$

式中: $\mathbf{A}\mathbf{A}$ 是一个 p 维集值向量. 特别地, 当 $p = 1$ 时, $\mathbf{A} = (\lambda_1, \lambda_2, \dots, \lambda_k)$ 是一个 $1 \times k$ 的实值行向量,

乘积 $\mathbf{A}\mathbf{A} = \sum_{j=1}^k \lambda_j A_j$ 是一个 $K(\mathbb{R}^d)$ 中的集合.

1.2 集值随机变量

令 $(\Omega, \mathcal{A}, P)$ 是一个完备的概率空间. 称集值映射 $X(\omega): \Omega \rightarrow K(\mathbb{R}^d)$ 为一个集值随机变量, 如果对于任意开集 C , 都有 $X^{-1}(C) \in \mathcal{A}$, 式中 $X^{-1}(C) = \{\omega \in \Omega; X(\omega) \cap C \neq \emptyset\}$.

记 $u[\Omega, K(\mathbb{R}^d)]$ 为 $K(\mathbb{R}^d)$ 中的集值随机变量族. 2 个集值随机变量 X_1, X_2 间的 D_2 距离定义为

$$D_2(X_1, X_2) = [E L_2^2(X_1, X_2)]^{\frac{1}{2}}$$

接下来可以通过 Aumann^[4] 积分, 给出集值随机变量 X 期望的定义

$$E[X] = \int_{\Omega} X dP = \left\{ \int_{\Omega} f dP; f \in S_X^1 \right\}$$

式中 S_X^1 是集值随机变量 X 的选择集, 即

$$S_X^1 = \{f \in L^1[\Omega, \mathbb{R}^d]; f(\omega) \in X(\omega) a. e. (P)\}$$

式中 $L^1[\Omega, \mathbb{R}^d]$ 为一阶矩存在的实值随机向量族.

注意到 $u[\Omega, K(\mathbb{R}^d)]$ 并非是一个线性空间, 这使得许多通过线性空间的特性讨论的统计问题无法平行地推广至集值的情形, 例如并不能直接定义集

值随机变量的方差和协方差, 然而, 可以通过 D_2 距离来刻画集值随机变量的离散程度. Yang 等^[15] 讨论了通过 D_2 距离定义的集值随机变量方差的性质, 并给出了集值随机变量协方差和相关系数的定义.

对于 $u[\Omega, K(\mathbb{R}^d)]$ 中的元素, 进一步地记

$$\mathcal{L}^2[\Omega, K(\mathbb{R}^d)] = \{X; D_2^2(X, \{0\}) < \infty\}$$

为所有 D_2 距离存在的集值随机变量族. 集值随机变量 $X_1, X_2 \in \mathcal{L}^2[\Omega, K(\mathbb{R}^d)]$ 的方差、协方差和相关系数分别定义为

$$\begin{aligned} \text{Var}(X_1) &= D_2^2(X_1, E[X_1]) = \\ &E \left(\int_S [s(x, X_1) - s(x, E[X_1])]^2 d\mu \right) \\ \text{Cov}(X_1, X_2) &= E \left(\int_S (s(x, X_1) - \right. \\ &\left. s(x, E[X_1]))(s(x, X_2) - s(x, E[X_2])) d\mu \right) \\ \text{Cor}(X_1, X_2) &= \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1) \text{Var}(X_2)}} \end{aligned}$$

式中 $\text{Var}(X_1) \text{Var}(X_2) \neq 0$.

为便于接下来的讨论, 下面给出集值随机变量协方差的一些性质.

命题 1 集值随机变量 X_1, X_2, X_3 的协方差具有下面几个性质:

1) $\text{Cov}(C, X_1) = 0$, 式中 C 是一个常集合.

2) $\text{Cov}(\lambda X_1, X_2) = \lambda \text{Cov}(X_1, X_2)$, 式中 $\lambda \geq 0$ 是一个常数.

3) $\text{Cov}(X_1 + X_2, X_3) = \text{Cov}(X_1, X_3) + \text{Cov}(X_2, X_3)$.

命题 1 的证明以及关于集值随机变量方差、协方差的更多性质见 Yang 等^[13]、Wang 等^[19].

1.3 集值随机向量

考虑完备的概率乘积空间

$$(\Omega_1, \mathcal{A}_1, P_1) \times \dots \times (\Omega_k, \mathcal{A}_k, P_k)$$

称向量 $\mathbf{X} = (X_1, X_2, \dots, X_k)^T$ 为 k 维集值随机向量, 如果 $X_1, X_2, \dots, X_k$ 是 $K(\mathbb{R}^d)$ 中的集值随机变量. 记 $u[\Omega_1 \times \Omega_2 \times \dots \times \Omega_k, K^k(\mathbb{R}^d)]$ 为 k 维集值随机向量族.

k 维集值随机向量 $\mathbf{X}$ 的期望 $E[\mathbf{X}]$ 定义为

$$E[\mathbf{X}] = (E[X_1], E[X_2], \dots, E[X_k])^T$$

式中 $E[X_1], E[X_2], \dots, E[X_k]$ 为 $\mathbf{X}$ 的不同分量在不同的概率测度 $P_1, P_2, \dots, P_k$ 下的期望. 然而为了简化记号, 并不对不同概率测度下的期望加以记号区分, 下面的协方差阵和交叉协方差阵等记号也同样.

如果 k 维集值随机向量 $\mathbf{X}_1 = (X_{1,1}, X_{2,1}, \dots, X_{k,1})^T$, 和 $\mathbf{X}_2 = (X_{1,2}, X_{2,2}, \dots, X_{k,2})^T$ 的每一个分量都有 $X_{i,j} \in \mathcal{L}[\Omega_{i,j}, K(\mathbb{R}^d)]$, $i = 1, 2, j = 1, 2, \dots, k$, 那么 $\mathbf{X}_1$ 的协方差阵为

$$\text{Cov}(\mathbf{X}_1) =$$

$$\begin{pmatrix} \text{Var}(X_{1,1}) & \text{Cov}(X_{1,1}, X_{2,1}) & \cdots & \text{Cov}(X_{1,1}, X_{k,1}) \\ \text{Cov}(X_{2,1}, X_{1,1}) & \text{Var}(X_{2,1}) & \cdots & \text{Cov}(X_{2,1}, X_{k,1}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_{k,1}, X_{1,1}) & \text{Cov}(X_{k,1}, X_{2,1}) & \cdots & \text{Var}(X_{k,1}) \end{pmatrix}$$

$\mathbf{X}_1$ 和 $\mathbf{X}_2$ 的交叉协方差阵为

$$\text{Cov}(\mathbf{X}_1, \mathbf{X}_2) =$$

$$\begin{pmatrix} \text{Cov}(X_{1,1}, X_{1,2}) & \text{Cov}(X_{1,1}, X_{2,2}) & \cdots & \text{Cov}(X_{1,1}, X_{k,2}) \\ \text{Cov}(X_{2,1}, X_{1,2}) & \text{Cov}(X_{2,1}, X_{2,2}) & \cdots & \text{Cov}(X_{2,1}, X_{k,2}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_{k,1}, X_{1,2}) & \text{Cov}(X_{k,1}, X_{2,2}) & \cdots & \text{Cov}(X_{k,1}, X_{k,2}) \end{pmatrix}$$

$\mathbf{X}_1$ 和 $\mathbf{X}_2$ 的交叉相关阵为

$$\text{Cor}(\mathbf{X}_1, \mathbf{X}_2) = \mathbf{D}_1^{-1} \text{Cov}(\mathbf{X}_1, \mathbf{X}_2) \mathbf{D}_2^{-1}$$

式中矩阵 $\mathbf{D}_1 = \text{diag}(\sqrt{\text{Var}(X_{1,1})}, \dots, \sqrt{\text{Var}(X_{k,1})})$ 和 $\mathbf{D}_2 = \text{diag}(\sqrt{\text{Var}(X_{1,2})}, \dots, \sqrt{\text{Var}(X_{k,2})})$ 分别为由 $\mathbf{X}_1$ 和 $\mathbf{X}_2$ 的每个分量的标准差组成的对角矩阵.

命题 2 k 维集值随机向量 $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ 的交叉协方差阵具有下面几个性质:

$$1) \text{Cov}(\mathbf{X}_1, \mathbf{X}_2) = \text{Cov}(\mathbf{X}_2, \mathbf{X}_1)^T.$$

2) $\text{Cov}(\mathbf{C}, \mathbf{X}_1) = \mathbf{0}_{k \times k}$, 式中 $\mathbf{C} = (c_1, c_2, \dots, c_k)^T$ 是一个常集值向量, $\mathbf{0}_{k \times k}$ 是 $k \times k$ 的零矩阵.

3) 下面关系式成立

$$\begin{aligned} \text{Cov}(\mathbf{X}_1 + \mathbf{X}_2, \mathbf{X}_3) = \\ \text{Cov}(\mathbf{X}_1, \mathbf{X}_3) + \text{Cov}(\mathbf{X}_2, \mathbf{X}_3) \end{aligned}$$

4) $\text{Cov}(\mathbf{P}\mathbf{X}_1, \mathbf{Q}\mathbf{X}_2) = \mathbf{P}\text{Cov}(\mathbf{X}_1, \mathbf{X}_2)\mathbf{Q}^T$, 式中 $\mathbf{P} = (p_{i,j})_{i=1, \dots, p, j=1, \dots, k}$, $\mathbf{Q} = (q_{i,j})_{i=1, \dots, p, j=1, \dots, k}$ 是 2 个所有元素都非负的真实矩阵.

证明:

1) 由下面的表达式

$$\begin{aligned} \text{Cov}(\mathbf{X}_2, \mathbf{X}_1) = \\ \begin{pmatrix} \text{Cov}(X_{1,2}, X_{1,1}) & \text{Cov}(X_{1,2}, X_{2,1}) & \cdots & \text{Cov}(X_{1,2}, X_{k,1}) \\ \text{Cov}(X_{2,2}, X_{1,1}) & \text{Cov}(X_{2,2}, X_{2,1}) & \cdots & \text{Cov}(X_{2,2}, X_{k,1}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_{k,2}, X_{1,1}) & \text{Cov}(X_{k,2}, X_{2,1}) & \cdots & \text{Cov}(X_{k,2}, X_{k,1}) \end{pmatrix} \end{aligned}$$

可得 1).

2) 由命题 1, 可以推出

$$\begin{aligned} \text{Cov}(\mathbf{C}, \mathbf{X}_1) = (\text{Cov}(c_i, X_{j,1}))_{i,j=1, \dots, k} = (\mathbf{0})_{i,j=1, \dots, k} \\ \text{即为 2).} \end{aligned}$$

3) 由命题 1, 可以推出

$$\begin{aligned} \text{Cov}(\mathbf{X}_1 + \mathbf{X}_2, \mathbf{X}_3) = \\ (\text{Cov}(X_{i,1} + X_{i,2}, X_{j,3}))_{i,j=1, \dots, k} = \\ (\text{Cov}(X_{i,1}, X_{j,3}) + \text{Cov}(X_{i,2}, X_{j,3}))_{i,j=1, \dots, k} = \\ \text{Cov}(\mathbf{X}_1, \mathbf{X}_3) + \text{Cov}(\mathbf{X}_2, \mathbf{X}_3) \end{aligned}$$

即为 3).

4) 由命题 1, 可以推出

$$\begin{aligned} \text{Cov}(\mathbf{P}\mathbf{X}_1, \mathbf{Q}\mathbf{X}_2) = \\ \left(\text{Cov} \left(\sum_{m=1}^k p_{i,m} X_{m,1}, \sum_{m=1}^k q_{j,l} X_{l,2} \right) \right)_{i,j=1, \dots, p, j=1, \dots, q} = \\ \left(\sum_{m=1}^k \sum_{l=1}^k p_{i,m} q_{j,l} \text{Cov}(X_{m,1}, X_{l,2}) \right)_{i,j=1, \dots, p, j=1, \dots, q} = \\ \mathbf{P}\text{Cov}(\mathbf{X}_1, \mathbf{X}_2)\mathbf{Q}^T \end{aligned}$$

即为 4).

2 集值多元时间序列

2.1 集值多元时间序列

Wang 等^[19]在集值随机变量及其期望、方差、协方差等统计量的定义的基础上, 定义了集值时间序列, 并给出集值时间序列的平稳性、自协方差函数和自相关函数的定义. 在 Wang 等^[19]的基础上, 进一步给出集值多元时间序列的定义, 并讨论其性质.

对于集值多元时间序列的研究, 不能将其视为一个更高维空间中的集值时间序列. 记 $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_k$. 虽然一个 $u[\Omega, K^k(\mathbb{R}^d)]$ 中的元素同时也是 $u[\Omega, K(\mathbb{R}^{kd})]$ 中的集值随机变量, 但直接将其视为一个集值随机变量来做进一步的讨论, 这样就忽略了集值随机向量的几何信息. 因此应对其特别对待, 下面给出集值多元时间序列的定义.

定义 1 称由时间 $\{t = 1, 2, \dots, T\}$ 生成的 k 维集值随机向量序列 $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_T \in u[\Omega, K^k(\mathbb{R}^d)]$ 为一个 k 维集值多元时间序列.

定义 2 如果对于 $\forall t = 1, 2, \dots, T$, 一个 k 维集值多元时间序列 $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_T$, 都有: 1) $E[\mathbf{X}_t] = \mathbf{X}_0$ 是一个常的集值向量, 2) $\text{Cov}[\mathbf{X}_t] = E[(\mathbf{X}_t - \mathbf{X}_0)(\mathbf{X}_t - \mathbf{X}_0)^T]$ 是 $k \times k$ 的常实值正定矩阵, 那么称这个 k 维集值多元时间序列 $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_T$ 是弱平稳的.

由定义 2 可知, 如果一个集值多元时间序列的前两阶矩不随时间变化而变化, 那么这个序列就是弱平稳的. 由于严平稳性的定义过于严格, 就使得在实际应用中很难找到满足严平稳定义的序列, 因此在这里不再讨论集值多元时间序列严平稳性的定义. 进一步地, 将弱平稳简称为平稳.

下面引入 k 维集值多元时间序列的交叉协方差阵的定义.

定义 3 滞后为 l 的 k 维集值多元时间序列的交叉协方差阵定义为

$$\boldsymbol{\Gamma}_l \triangleq \text{Cov}(\mathbf{X}_t, \mathbf{X}_{t-l})$$

滞后为 l 的交叉相关阵定义为

$$\boldsymbol{\gamma}_l \triangleq \text{Cov}(\mathbf{X}_t, \mathbf{X}_{t-l}) = \mathbf{D}_t^{-1} \boldsymbol{\Gamma}_l \mathbf{D}_{t-l}^{-1}$$

式中 $\mathbf{D}_t = \text{diag}(\sqrt{\text{Var}(X_{1,t})}, \dots, \sqrt{\text{Var}(X_{k,t})})$ 为由 $\mathbf{X}_t$ 的每个分量的标准差组成的对角矩阵.

命题 3 如果 k 维集值多元时间序列 $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_T$ 是平稳的, 那么其交叉协方差阵和交叉相关阵具有下面几个性质:

$$1) \boldsymbol{\Gamma}_l = \boldsymbol{\Gamma}_{-l}^T, \boldsymbol{\gamma}_l = \boldsymbol{\gamma}_{-l}^T.$$

$$2) \boldsymbol{\gamma}_l = \mathbf{D}_l^{-1} \boldsymbol{\Gamma}_l \mathbf{D}_l^{-1}.$$

3) 对于任意的 $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_T \in \mathbb{R}^k, T \in \{1, 2, \dots\}$, 都有 $\sum_{l=1}^T \sum_{m=1}^T \boldsymbol{\lambda}_l^T \boldsymbol{\Gamma}_{l-m} \boldsymbol{\lambda}_m \geq 0$.

证明:

1) 由平稳性可以推出

$$\begin{aligned} \boldsymbol{\Gamma}_l &= (\text{Cov}(X_{i,t}, X_{j,t-l}))_{i,j=1,\dots,k} = \\ &= (\text{Cov}(X_{i,t+l}, X_{j,t}))_{i,j=1,\dots,k} = \\ &= (\text{Cov}(X_{i,t}, X_{j,t+l}))_{i,j=1,\dots,k}^T = \boldsymbol{\Gamma}_{-l}^T \end{aligned}$$

同理可以推出 $\boldsymbol{\gamma}_l = \boldsymbol{\gamma}_{-l}^T$ 成立.

2) 由平稳性可知 $\mathbf{D}_t = \mathbf{D}_{t-l}$, 可推出 2).

3) 由平稳性, 可以由下面的计算

$$\begin{aligned} &\sum_{l=1}^T \sum_{m=1}^T \boldsymbol{\lambda}_l^T \boldsymbol{\Gamma}_{l-m} \boldsymbol{\lambda}_m = \\ &= \sum_{l=1}^T \sum_{m=1}^T \sum_{i=1}^k \sum_{j=1}^k \lambda_{j,l} \lambda_{i,m} \text{Cov}(X_{i,t-m}, X_{j,t-l}) = \\ &= \text{E} \left(\sum_{l=1}^T \sum_{m=1}^T \sum_{i=1}^k \sum_{j=1}^k \lambda_{j,l} \lambda_{i,m} \int_S (s(\mathbf{x}, X_{i,t-m}) - s(\mathbf{x}, X_{i,0})) (s(\mathbf{x}, X_{j,t-l}) - s(\mathbf{x}, X_{j,0})) d\mu \right) = \\ &= \text{E} \int_S \left[\sum_{m=1}^T \sum_{i=1}^k \lambda_{i,m} (s(\mathbf{x}, X_{i,t-m}) - s(\mathbf{x}, X_{i,0})) \right]^2 d\mu \geq 0 \end{aligned}$$

证得 3).

2.2 平稳的集值多元时间序列统计量的估计

下面讨论集值多元时间序列期望向量, 交叉协方差阵及交叉相关阵的估计. 假定 $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_T$ 是平稳的 k 维集值多元时间序列的观测值.

期望向量 $\mathbf{X}_0$ 的矩估计为

$$\begin{aligned} \hat{\mathbf{X}}_0 &\triangleq (\hat{X}_{1,0}, \hat{X}_{2,0}, \dots, \hat{X}_{k,0})^T = \\ &= \frac{1}{T} (\mathbf{X}_1 + \mathbf{X}_2 + \dots + \mathbf{X}_T) = \end{aligned}$$

$$\left(\frac{1}{T} \sum_{t=1}^T X_{1,t}, \dots, \frac{1}{T} \sum_{t=1}^T X_{k,t} \right)^T$$

对于 $l=0, 1, \dots, L, L \ll T$, 滞后为 l 的交叉协方差阵的矩估计为

$$\hat{\boldsymbol{\Gamma}}_l = (\hat{\text{Cov}}(X_{i,t}, X_{j,t-l}))_{i,j=1,\dots,k}$$

式中

$$\begin{aligned} \hat{\text{Cov}}(X_{i,t}, X_{j,t-l}) &= \\ &= \frac{1}{T-1} \sum_{t=l+1}^T \int_S (s(\mathbf{x}, X_{i,t}) - s(\mathbf{x}, \hat{X}_{i,0})) \cdot \\ &\quad (s(\mathbf{x}, X_{j,t-l}) - s(\mathbf{x}, \hat{X}_{j,0})) d\mu \end{aligned}$$

滞后为 l 的交叉相关阵的矩估计为

$$\hat{\boldsymbol{\gamma}}_l = \hat{\mathbf{D}}^{-1} \hat{\boldsymbol{\Gamma}}_l \hat{\mathbf{D}}^{-1}$$

式中 $\hat{\mathbf{D}} = \text{diag}(\sqrt{\hat{\text{Var}}(X_{1,t})}, \dots, \sqrt{\hat{\text{Var}}(X_{k,t})})$, 以及 $\hat{\text{Var}}(X_{i,t}) = \frac{1}{T-1} \sum_{t=1}^T \int_S (s(\mathbf{x}, X_{i,t}) - s(\mathbf{x}, \hat{X}_{i,0}))^2 d\mu$.

进一步地, 可通过下面的定理得出估计 $\hat{\mathbf{X}}_0$ 和 $\hat{\boldsymbol{\Gamma}}_l$ 的渐近性质.

定理 期望向量和交叉协方差阵的估计 $\hat{\mathbf{X}}_0$ 和 $\hat{\boldsymbol{\Gamma}}_l$ 具有下面的渐近性质:

1) $\hat{\mathbf{X}}_0$ 是一个无偏的相合估计, 即 $\text{E}[\hat{\mathbf{X}}_0] = \mathbf{X}_0$ 及 $\lim_{T \rightarrow \infty} \hat{\mathbf{X}}_0 = \mathbf{X}$ 成立.

2) 如果当 $l \rightarrow \infty$ 时, $\boldsymbol{\Gamma}_l \rightarrow \mathbf{0}_{k \times k}$, 那么 $\hat{\boldsymbol{\Gamma}}_l$ 是一个渐近无偏估计, 即 $\lim_{T \rightarrow \infty} \text{E}(\hat{\boldsymbol{\Gamma}}_l) = \boldsymbol{\Gamma}_l$ 成立.

证明:

1) 对于任意的 $i=1, \dots, k$, 有

$$\text{E}[\hat{X}_{i,0}] = \text{E} \left[\frac{1}{T} \sum_{t=1}^T X_{i,t} \right] = \frac{1}{T} \sum_{t=1}^T \text{E}[X_{i,t}] = X_{i,0}$$

对于 $\forall \varepsilon > 0, T \rightarrow \infty$, 由 Wang 等^[19] 中定理 1 和定理 6 可知关系式

$$P\{L_2(\hat{X}_{i,0}, X_{i,0}) \geq \varepsilon\} \leq \frac{\text{Var}(\hat{X}_{i,0})}{\varepsilon^2} \rightarrow 0$$

成立, 由此可证得 1) 成立.

2) 只需证明

$$\lim_{T \rightarrow \infty} \text{E}(\hat{\text{Cov}}(X_{i,t}, X_{j,t-l})) = \text{Cov}(X_{i,t}, X_{j,t-l})$$

注意到关系式

$$\begin{aligned} &(s(\mathbf{x}, X_{i,t}) - s(\mathbf{x}, \hat{X}_{i,0})) (s(\mathbf{x}, X_{j,t-l}) - s(\mathbf{x}, \hat{X}_{j,0})) = \\ &= (s(\mathbf{x}, X_{i,t}) - s(\mathbf{x}, X_{i,0})) (s(\mathbf{x}, X_{j,t-l}) - s(\mathbf{x}, X_{j,0})) + \\ &\quad (s(\mathbf{x}, X_{i,t}) - s(\mathbf{x}, X_{i,0})) (s(\mathbf{x}, X_{j,0}) - s(\mathbf{x}, \hat{X}_{j,0})) + \\ &\quad (s(\mathbf{x}, X_{i,0}) - s(\mathbf{x}, \hat{X}_{i,0})) (s(\mathbf{x}, X_{j,t-l}) - s(\mathbf{x}, X_{j,0})) + \\ &\quad (s(\mathbf{x}, X_{i,0}) - s(\mathbf{x}, \hat{X}_{i,0})) (s(\mathbf{x}, X_{j,t-l}) - s(\mathbf{x}, \hat{X}_{j,0})) \end{aligned}$$

成立, 进而由集值协方差的定义可以得出关系式

$$\begin{aligned} \hat{\text{Cov}}(X_{i,t}, X_{j,t-l}) &= \text{Cov}(X_{i,t}, X_{j,t-l}) + \text{Cov}(X_{i,t}, \hat{X}_{j,0}) + \\ &\quad \text{Cov}(\hat{X}_{i,0}, X_{j,t-l}) + \text{Cov}(\hat{X}_{i,0}, \hat{X}_{j,0}) \end{aligned}$$

又当 $l \rightarrow \infty$ 时, 有 $\text{Cov}(X_{i,t}, X_{j,t-l}) = 0$, 再结合期望向量估计 $\hat{\mathbf{X}}_0$ 的表达式, 可以推出

$$\begin{aligned} & \lim_{T \rightarrow \infty} \text{E}(\hat{\text{Cov}}(X_{i,t}, X_{j,t-l})) = \\ & \lim_{T \rightarrow \infty} \frac{1}{T-1} \sum_{t=l+1}^T \text{E}[\text{Cov}(X_{i,t}, X_{j,t-l}) + \\ & \text{Cov}(X_{i,t}, \hat{X}_{j,0}) + \text{Cov}(\hat{X}_{i,0}, X_{j,t-l}) + \text{Cov}(\hat{X}_{i,0}, \hat{X}_{j,0})] = \\ & \text{Cov}(X_{i,t}, X_{j,t-l}) + \\ & \lim_{T \rightarrow \infty} \frac{1}{T-1} \sum_{t=l+1}^T \frac{1}{T} \sum_{s=1}^T \text{Cov}(X_{i,t}, X_{j,s}) + \\ & \lim_{T \rightarrow \infty} \frac{1}{T-1} \sum_{t=l+1}^T \frac{1}{T} \sum_{s=1}^T \text{Cov}(X_{i,s}, X_{j,t-l}) + \\ & \lim_{T \rightarrow \infty} \frac{1}{T-1} \sum_{t=l+1}^T \frac{1}{T^2} \sum_{s=1}^T \sum_{m=1}^T \text{Cov}(X_{i,s}, X_{j,m}) = \\ & \text{Cov}(X_{i,t}, X_{j,t-l}) \end{aligned}$$

渐近无偏性 2) 得证.

2.3 平稳的集值多元时间序列的最优线性预测

下面讨论给定平稳的 k 维集值多元时间序列观测值 $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_T$ 后, 如何求得超前 h 步的最优线性预测问题, $h = 1, 2, \dots$. 首先, 定义一个用于衡量预测质量的准则, 那么使得这个准则达到最优的 $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_T$ 的某一个线性组合, 即为最优线性预测.

记 $\tilde{\mathbf{X}}_{T+h} = (\tilde{X}_{1,T+h}, \dots, \tilde{X}_{k,T+h})^T$ 为任意超前 h 步预测, $\mathbf{X}_{T+h} = (X_{1,T+h}, \dots, X_{k,T+h})^T$ 为 $T+h$ 时刻的真实观测值. 定义均方预测误差为如下的 $k \times k$ 矩阵

$$\text{MSE}(\tilde{\mathbf{X}}_{T+h}) =$$

$$\text{E}[(L_2(\tilde{X}_{1,T+h}, X_{1,T+h}), \dots, L_2(\tilde{X}_{k,T+h}, X_{k,T+h}))^T \cdot (L_2(\tilde{X}_{1,T+h}, X_{1,T+h}), \dots, L_2(\tilde{X}_{k,T+h}, X_{k,T+h}))]$$

那么称

$$\hat{\mathbf{X}}_{T+h} = \mathbf{A}_0 + \mathbf{a}_1 \mathbf{X}_T + \dots + \mathbf{a}_T \mathbf{X}_1$$

为最优线性预测, 式中 $\mathbf{A}_0 = (A_{1,0}, A_{2,0}, \dots, A_{k,0})^T$ 为 k 维集值向量截距, $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_T$ 为 $k \times k$ 的实值系数矩阵, 如果 $\hat{\mathbf{X}}_{T+h}$ 使得均方预测误差矩阵达到最小, 即

$$\text{MSE}(\hat{\mathbf{X}}_{T+h}) \leq \text{MSE}(\tilde{\mathbf{X}}_{T+h})$$

3 区间值多元时间序列

3.1 区间与区间值随机变量

当 $d=1$ 时, 将 $\mathbb{R}^1$ 简写为 $\mathbb{R}$, 那么 $K(\mathbb{R})$ 为 $\mathbb{R}$ 中所有非空有界闭区间族, 即

$K(\mathbb{R}) = \{A = [\underline{a}, \bar{a}] : -\infty < \underline{a} \leq \bar{a} < \infty, \underline{a}, \bar{a} \in \mathbb{R}\}$ 式中 $\underline{a}$ 和 $\bar{a}$ 分别为区间 A 的左端点和右端点. 还将区间记为 $A = [\underline{a}, \bar{a}] = (a^c; a^r)$, 式中 $a^c = (\bar{a} + \underline{a})/2$, $a^r = (\bar{a} - \underline{a})/2$ 分别是区间 A 的中心和半径.

对于任意 2 个闭区间 $A = [\underline{a}, \bar{a}] = (a^c; a^r)$ 和 $B = [\underline{b}, \bar{b}] = (b^c; b^r)$, 加法和数乘运算可以写为

$$\begin{aligned} A + B &= [\underline{a} + \underline{b}, \bar{a} + \bar{b}] = (a^c + b^c; a^r + b^r) \\ \lambda A &= \begin{cases} [\lambda \underline{a}, \lambda \bar{a}], & \lambda \geq 0 \\ [\lambda \bar{a}, \lambda \underline{a}], & \lambda < 0 \end{cases} = (\lambda a^c; |\lambda| a^r) \end{aligned}$$

进一步地, 区间 $A = [\underline{a}, \bar{a}]$ 和 $B = [\underline{b}, \bar{b}]$ 间的 L_2 距离为

$$\begin{aligned} L_2(A, B) &= [(\underline{b} - \underline{a})^2 + (\bar{b} - \bar{a})^2]^{\frac{1}{2}} = \\ &[2(b^c - a^c)^2 + 2(b^r - a^r)^2]^{\frac{1}{2}} \end{aligned}$$

由集值随机变量的等价定义 (例如, 见 Li 等^[11] 的定理 1.2.2), 可以推出, 对于 $X = [\underline{x}, \bar{x}] = (x^c; x^r)$, 如果 $\underline{x}$ 和 $\bar{x}$ 或 x^c 和 x^r 是实值的随机变量, 那么 X 是区间值随机变量.

进一步地, 区间值随机变量 X 的期望可以写为

$$\text{E}[X] = [\text{E}\underline{x}, \text{E}\bar{x}] = (\text{E}x^c; \text{E}x^r)$$

$u[\Omega, K(\mathbb{R})]$ 中区间值随机变量 X_1 和 X_2 间的 D_2 距离可以写为

$$\begin{aligned} D_2(X_1, X_2) &= [\text{E}(\underline{x}_2 - \underline{x}_1)^2 + \text{E}(\bar{x}_2 - \bar{x}_1)^2]^{\frac{1}{2}} = \\ &[2\text{E}(x_2^c - x_1^c)^2 + 2\text{E}(x_2^r - x_1^r)^2]^{\frac{1}{2}} \end{aligned}$$

$\mathcal{L}^2[\Omega, K(\mathbb{R})]$ 中的区间值随机变量 X_1, X_2 的方差和协方差可以写为

$$\begin{aligned} \text{Var}(X_1) &= \text{E}(\underline{x}_1 - \text{E}\underline{x}_1)^2 + \text{E}(\bar{x}_1 - \text{E}\bar{x}_1)^2 = \\ &2\text{E}(x_1^c - \text{E}x_1^c)^2 + 2\text{E}(x_1^r - \text{E}x_1^r)^2 \\ \text{Cov}(X_1, X_2) &= \end{aligned}$$

$$\begin{aligned} &\text{E}[(x_1 - \text{E}x_1)(x_2 - \text{E}x_2) + (\bar{x}_1 - \text{E}\bar{x}_1)(\bar{x}_2 - \text{E}\bar{x}_2)] = \\ &2\text{E}[(x_1^c - \text{E}x_1^c)(x_2^c - \text{E}x_2^c)] + 2\text{E}[(x_1^r - \text{E}x_1^r)(x_2^r - \text{E}x_2^r)] \end{aligned}$$

3.2 区间值向量与区间值随机向量

由 k 个非空有界的闭区间 $A_1 = [\underline{a}_1, \bar{a}_1], \dots, A_k = [\underline{a}_k, \bar{a}_k]$ 构成了 k 维区间值向量 $\mathbf{A} = (A_1, A_2, \dots, A_k)^T$. 当 $k=2$ 时, $\mathbf{A}$ 是二维欧式空间中的矩形. 当 $k=3$ 时, $\mathbf{A}$ 是三维欧式空间中的长方体. 当 $k \geq 4$ 时, $\mathbf{A}$ 是 k 维的超立方体. 可以发现, 相较于集值向量, 区间值向量具有更加直观的几何形状.

称向量 $\mathbf{X} = (X_1, X_2, \dots, X_k)^T$ 为 k 维区间值随机向量, 如果 $X_1, X_2, \dots, X_k$ 是 $K(\mathbb{R})$ 中的区间值随机变量. 记 $u[\Omega_1 \times \Omega_2 \times \dots \times \Omega_k, K^k(\mathbb{R})]$ 为 k 维区间值随机向量族.

k 维区间值随机向量 $\mathbf{X}$ 的期望 $\text{E}[\mathbf{X}]$ 为

$$E[\mathbf{X}] = ([E\mathbf{x}_1, E\bar{\mathbf{x}}_1], \dots, [E\mathbf{x}_k, E\bar{\mathbf{x}}_k])^T =$$

$$((Ex_1^c; Ex_1^r), \dots, (Ex_k^c; Ex_k^r))^T$$

k 维区间值随机向量的协方差阵,交叉协方差阵及交叉相关阵形式与 k 维集值随机向量的相同,在此不再给出。

进一步地,注意到 k 维区间值随机向量 $\mathbf{X}$ 的任一分量 $X_i = [\underline{x}_i, \bar{x}_i]$ 都可以由其 2 个端点 $\underline{x}_i$ 和 $\bar{x}_i$ 或中心 x_i^c 半径 x_i^r 所决定,所以可以通过实值随机变量 $x_1, \bar{x}_1, \dots, x_k, \bar{x}_k$ 或 $x_1^c, x_1^r, \dots, x_k^c, x_k^r$ 来研究 $\mathbf{X}$ 的性质。需注意的是,在如此分析问题,应保持关系式 $\underline{x}_i \leq \bar{x}_i$ 或 $x_i^c \geq 0$, 对所有的 $i = 1, 2, \dots, k$ 均成立,否则区间值随机向量将无意义。

3.3 平稳的区间值多元时间序列

平稳的 k 维区间值多元时间序列由时间生成的 k 维区间值随机向量序列 $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_T$ 组成,并且其前二阶矩不随时间变化而变化。

给定平稳的 k 维区间值多元时间序列观测值 $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_T$ 后,可以通过极小化均方预测误差

$MSE(\tilde{\mathbf{X}}_{T+h})$ 求出最优线性预测

$$\hat{\mathbf{X}}_{T+h} = \mathbf{A}_0 + \mathbf{a}_1 \mathbf{X}_T + \dots + \mathbf{a}_T \mathbf{X}_1 =$$

$$\begin{pmatrix} ([\underline{a}_{1,0}, \bar{a}_{1,0}] + \sum_{t=1}^T \sum_{i=1}^k a_{1i,t} [\underline{x}_{i,T-t+1}, \bar{x}_{i,T-t+1}]) \\ \vdots \\ ([\underline{a}_{k,0}, \bar{a}_{k,0}] + \sum_{t=1}^T \sum_{i=1}^k a_{ki,t} [\underline{x}_{i,T-t+1}, \bar{x}_{i,T-t+1}]) \end{pmatrix}$$

式中: $\mathbf{A}_0 = ([\underline{a}_{1,0}, \bar{a}_{1,0}], \dots, [\underline{a}_{k,0}, \bar{a}_{k,0}])^T$ 为 k 维区间值向量截距; $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_T$ 为 $k \times k$ 的系数矩阵。

3.4 区间值多元自回归模型

p 阶的 k 维区间值多元自回归模型,简记为 IVAR(p) 模型,其表达式为

$$\mathbf{X}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{X}_{t-1} + \mathbf{\Phi}_2 \mathbf{X}_{t-2} + \dots + \mathbf{\Phi}_p \mathbf{X}_{t-p} + \mathbf{Z}_t$$

式中: $\mathbf{\Phi}_0 = ([\underline{\phi}_{1,0}, \bar{\phi}_{1,0}], \dots, [\underline{\phi}_{k,0}, \bar{\phi}_{k,0}])^T$ 为 k 维区间值向量截距; $\mathbf{\Phi}_1, \mathbf{\Phi}_2, \dots, \mathbf{\Phi}_p$ 为 $k \times k$ 的实值系数矩阵; $\mathbf{Z}_t$ 为不相关的同分布的 k 维区间值随机向量,并且 $\mathbf{Z}_t$ 与 $\mathbf{X}_{t-1}, \dots, \mathbf{X}_1$ 不相关。为讨论方便,下面进一步假设 $\mathbf{\Phi}_1, \mathbf{\Phi}_2, \dots, \mathbf{\Phi}_p$ 的所有元素皆非负。

记 $\mathbf{\Gamma}_0 = \text{Cov}(\mathbf{X}_t)$ 为 k 维区间值多元时间序列的协方差矩阵, $\mathbf{Z}_t$ 的协方差阵为 $\mathbf{\Sigma}$ 。由 IVAR(p) 模型的表达式可以推出 IVAR(p) 序列 $\mathbf{X}_t$ 的协方差阵为

$$\mathbf{\Gamma}_0 = \mathbf{\Phi}_1 \mathbf{\Gamma}_1 + \mathbf{\Phi}_2 \mathbf{\Gamma}_2 + \dots + \mathbf{\Phi}_p \mathbf{\Gamma}_p + \mathbf{\Sigma}$$

另一方面,对于 $\forall l > 0$, 由 IVAR(p) 模型的表达式可以推出关系式

$$\mathbf{\Gamma}_l = \mathbf{\Phi}_1 \mathbf{\Gamma}_{l-1} + \mathbf{\Phi}_2 \mathbf{\Gamma}_{l-2} + \dots + \mathbf{\Phi}_p \mathbf{\Gamma}_{l-p}$$

成立。然后分别取 $l = 1, 2, \dots, p$, 联立方程组,可以得到

$$\begin{pmatrix} \mathbf{\Gamma}_1 \\ \mathbf{\Gamma}_2 \\ \vdots \\ \mathbf{\Gamma}_p \end{pmatrix} = \begin{pmatrix} \mathbf{\Gamma}_0 & \mathbf{\Gamma}_1^T & \dots & \mathbf{\Gamma}_{p-1}^T \\ \mathbf{\Gamma}_1 & \mathbf{\Gamma}_0 & \dots & \mathbf{\Gamma}_{p-2}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Gamma}_{p-1} & \mathbf{\Gamma}_{p-2} & \dots & \mathbf{\Gamma}_0 \end{pmatrix} \begin{pmatrix} \mathbf{\Phi}_1 \\ \mathbf{\Phi}_2 \\ \vdots \\ \mathbf{\Phi}_p \end{pmatrix}$$

式中使用了命题 3 的结论 $\mathbf{\Gamma}_{-l} = \mathbf{\Gamma}_l^T$ 。通过求解这个方程组,可以得到系数矩阵 $\mathbf{\Phi}_1, \mathbf{\Phi}_2, \dots, \mathbf{\Phi}_p$ 的估计,记为 $\hat{\mathbf{\Phi}}_1, \hat{\mathbf{\Phi}}_2, \dots, \hat{\mathbf{\Phi}}_p$ 。进一步地,将系数矩阵的估计代入 IVAR(p) 模型中推出维区间值向量截距 $\mathbf{\Phi}_0$ 的估计 $\hat{\mathbf{\Phi}}_0$ 。得到模型系数的估计之后,IVAR(p) 模型的一步预测可以写为

$$\tilde{\mathbf{X}}_{t+1} = \hat{\mathbf{\Phi}}_0 + \hat{\mathbf{\Phi}}_1 \mathbf{X}_t + \hat{\mathbf{\Phi}}_2 \mathbf{X}_{t-1} + \dots + \hat{\mathbf{\Phi}}_p \mathbf{X}_{t-p+1}$$

4 数值模拟与实证分析

4.1 数值模拟

考虑一个一阶的 2 维区间值多元自回归模型

$$\mathbf{X}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{X}_{t-1} + \mathbf{Z}_t, t = 2, 3, \dots, T$$

式中: $\mathbf{\Phi}_0 = ([\underline{\phi}_{1,0}, \bar{\phi}_{1,0}], [\underline{\phi}_{2,0}, \bar{\phi}_{2,0}])^T$ 为 2 维区间值向量截距; $\mathbf{\Phi}_1 = \begin{pmatrix} \phi_{11,1} & \phi_{12,1} \\ \phi_{21,1} & \phi_{22,1} \end{pmatrix}$ 为实值回归系数矩阵;

$\mathbf{Z}_t = \begin{pmatrix} Z_{1,t} \\ Z_{2,t} \end{pmatrix} = \begin{pmatrix} z_{1,t}^c; z_{1,t}^r \\ z_{2,t}^c; z_{2,t}^r \end{pmatrix}$ 为与区间值时间序列

$\mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \dots, \mathbf{X}_1$ 不相关的噪声序列,并且其分量的中心半径 $z_{1,t}^c, z_{2,t}^c$ 与 $z_{1,t}^r, z_{2,t}^r$ 分别服从正态分布 $N(0, \sigma^2)$ 与伽马分布 $Ga(\alpha, \lambda)$ 。下面取 $\underline{\phi}_{1,0} = 0.2, \bar{\phi}_{1,0} = 0.5, \underline{\phi}_{2,0} = 0.1, \bar{\phi}_{2,0} = 0.3, \phi_{11,1} = 0.4, \phi_{12,1} = 0.2, \phi_{21,1} = 0.1, \phi_{22,1} = 0.5$, 以及 $\sigma = 0.3, \alpha = 1, \lambda = 20$ 。分别取 $T = 100, 500, 1\,000$, 模拟 1 000 次,将参数估计的均方误差列入表 1 中。可以看出,随着 T 增加,各个系数估计的均方误差逐渐减小,这说明了方法具有相合性。

4.2 实证分析

本节分析 2017-06-16—2017-08-08 北京市的空气指数数据,数据收集于网站 <http://beijingair.sinaapp.com>。数据由北京市 12 个空气监测站所收集到的空气指数、空气水平、主要污染物、PM 2.5、PM 10、CO、NO₂、SO₂ 的质量浓度组成,每个观测站收集到的数据都是每小时更新 1 次。

本文的研究对象是 PM 2.5 和 PM 10 的质量浓度,取每一天里这些空气监测站收集到的 PM 2.5

表1 系数估计的均方误差

Table 1 MSEs of coefficient estimations			
系数	$T=100$	$T=500$	$T=1\,000$
$\phi_{1,0}$	0.001 7	0.001 3	0.001 1
$\overline{\phi}_{1,0}$	0.037 6	0.014 6	0.011 5
$\phi_{2,0}$	0.001 7	0.001 9	0.001 9
$\overline{\phi}_{2,0}$	0.027 8	0.019 4	0.017 1
$\phi_{11,1}$	0.008 0	0.001 4	0.000 8
$\phi_{12,1}$	0.020 3	0.011 4	0.010 6
$\phi_{21,1}$	0.017 0	0.011 8	0.010 7
$\phi_{22,1}$	0.008 0	0.001 5	0.000 7

和PM 10 质量浓度的最小值和最大值作为当天的区间值随机向量观测值的端点,并分别记每天的PM 2.5 和PM 10 的质量浓度为 $X_{1,t}$ 和 $X_{2,t}$. 为计算方便,将所有数据除以100. 图1 给出了由PM 2.5 和PM 10 的质量浓度组成的区间值二元时间序列.

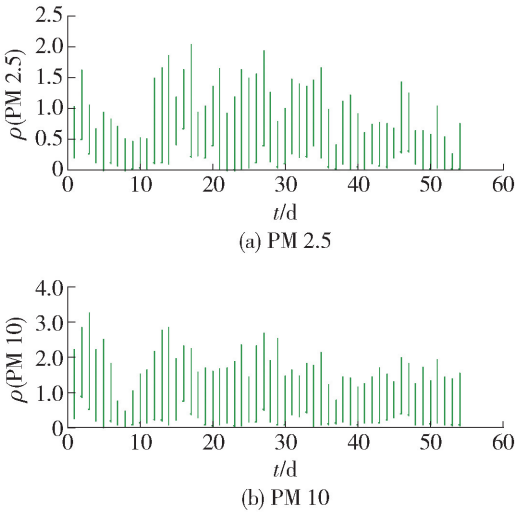


图1 2017-06-16 到2017-08-08 的北京市PM 2.5 和PM 10 质量浓度的时间序列

Fig. 1 Time series plots of Beijing's PM 2.5 and PM 10 concentration from June 16,2017 to August 8,2017

通过计算,可以得出区间值多元时间序列 $\{\mathbf{X}_t=(X_{1,t},X_{2,t})^T;t=1,2,\cdots,54\}$ 的协方差矩阵和自协方差矩阵的估计

$$\hat{\mathbf{\Gamma}}_0=\begin{pmatrix}0.21 & 0.17 \\ 0.17 & 0.35\end{pmatrix},\hat{\mathbf{\Gamma}}_1=\begin{pmatrix}0.11 & 0.10 \\ 0.10 & 0.20\end{pmatrix}$$

由上面的估计可以推出IVAR(1)模型系数的估计

$$\hat{\boldsymbol{\Phi}}_{1,0}=[0.065\ 1,0.069\ 1],\hat{\boldsymbol{\Phi}}_{2,0}=[0.482\ 0,0.736\ 4]$$
$$\hat{\boldsymbol{\phi}}_1=\begin{pmatrix}0.510\ 8 & 0.024\ 3 \\ 0.036\ 4 & 0.555\ 9\end{pmatrix}$$

进一步地做了3步预测,如图2所示,实线段是真实的观测区间,虚线段是通过IVAR(1)模型得到的预测区间.可以看出,预测区间明显地反映出了接下来的趋势.进一步地,可以计算出 $\hat{\mathbf{X}}_{T+1}$ 、 $\hat{\mathbf{X}}_{T+2}$ 和 $\hat{\mathbf{X}}_{T+3}$ 的均方预测误差为

$$\text{MSE}(\hat{\mathbf{X}}_{T+1})=\begin{pmatrix}0.049\ 4 & 0.021\ 0 \\ 0.021\ 0 & 0.009\ 0\end{pmatrix}$$
$$\text{MSE}(\hat{\mathbf{X}}_{T+2})=\begin{pmatrix}0.070\ 5 & 0.192\ 0 \\ 0.192\ 0 & 0.523\ 1\end{pmatrix}$$
$$\text{MSE}(\hat{\mathbf{X}}_{T+3})=\begin{pmatrix}0.056\ 1 & 0.029\ 1 \\ 0.029\ 1 & 0.015\ 1\end{pmatrix}$$

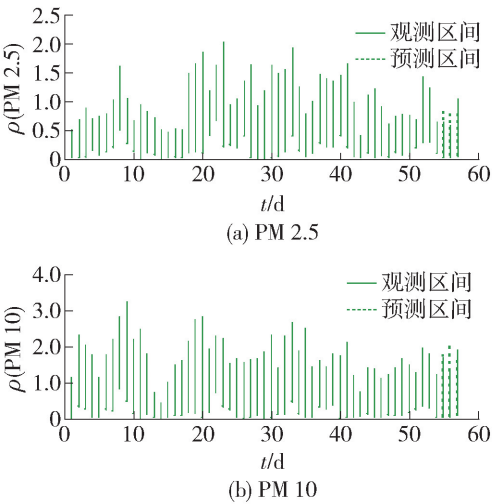


图2 北京市PM 2.5 和PM 10 的质量浓度的时间序列

Fig.2 Time series plots of Beijing's PM 2.5 and PM 10 concentration

5 结论

1) 对于实际中多元的复杂的时间序列数据,探讨多元集值时间序列基础理论,建立其时间序列模型是非常必要的. 由集值随机向量的定义可知,集值随机向量是一个具有特定几何特征的集值随机变量. 如果忽略这个向量形式而依旧将其视为集值随机变量来讨论相关的问题,将会导致信息的损失,这是一种非常粗糙的处理方式. 因此,从集值向量的定义出发,进一步定义了集值随机向量,并讨论了集值随机变量的期望向量、协方差阵、交叉协方差阵以及交叉相关阵的定义及性质.

2) 由于区间可以通过它的2个端点或中心半径来确定,因此对区间值多元时间序列的讨论是一

个相对更为具体的问题. 相较而言集合则抽象得多, 对集值多元时间序列的讨论本质上是对区间值多元时间序列的一个推广. 给出了集值多元时间序列及其平稳性的定义, 并研究了在平稳的条件下, 集值多元时间序列的期望向量、交叉协方差阵与交叉相关阵的性质, 这些统计量的估计及估计的渐近性质, 以及最优线性预测问题. 以上研究为对集值多元时间序列的进一步讨论打下了理论基础.

3) 讨论了区间值情形下的随机向量和多元时间序列, 并进一步给出区间值多元自回归模型的表达形式及估计. 最后, 通过一个实证分析展示了本文提出的模型及方法在实践中的应用.

参考文献:

- [1] BOX G E P, JENKIS G M, REINSEL G C, et al. Time series analysis: forecasting and control[M]. 5th ed. New Jersey: Wiley, 2015: 47-92.
- [2] TSAY R S. Multivariate time series analysis with R and financial applications[M]. New Jersey: Wiley, 2014: 27-89.
- [3] BROCKWELL P J, DAVIS R A. Time series: theory and methods[M]. Berlin: Springer, 2015: 77-110.
- [4] AUMANN R J. Integrals of set-valued functions[J]. Journal of Mathematical Analysis and Applications, 1965, 12(1): 1-12.
- [5] HIAI F, UMEGAKI H. Integrals, conditional expectations, and martingales of multivalued functions[J]. Journal of Multivariate Analysis, 1977, 7(1): 149-182.
- [6] LYASHENKO N N. Statistics of random compacts in Euclidean space[J]. Journal of Mathematical Sciences, 1983, 21(1): 76-92.
- [7] VITALE R A. L_p metrics for compact, convex sets[J]. Journal of Approximation Theory, 1985, 45(3): 280-287.
- [8] DIAMOND P. Least squares fitting of compact set-valued data[J]. Journal of Mathematical Analysis and Applications, 1990, 147(2): 351-362.
- [9] LI S, OGURA Y. Convergence of set valued sub- and super-martingales in the Kuratowski-Mosco sense[J]. Annals of Probability, 1998, 26(3): 1384-1402.
- [10] LI S, OGURA Y. Convergence of set-valued and fuzzy-valued martingales[J]. Fuzzy Sets and Systems, 1999, 101(3): 453-461.
- [11] LI S, OGURA Y, KREINOVICH V. Limit theorems and applications of set-valued and fuzzy set-valued random variables[M]. Dordrecht: Kluwer Academic Publishers, 2002: 1-108.
- [12] GIL M, LUBIANO M, MONTENEGRO M, et al. Least squares fitting of an affine function and strength of association for interval-valued data[J]. Metrika, 2002, 56(2): 97-111.
- [13] YANG X, LI S. The D_p -metric space of set-valued random variables and its application to covariances[J]. International Journal of Innovative Computing, Information and Control, 2005, 1(1): 73-82.
- [14] 张文修, 李寿梅, 汪振鹏, 等. 集值随机过程引论[M]. 北京: 科学出版社, 2007: 155-208.
- [15] WANG X, LI S. The interval autoregressive time series model [C] // Proceedings of IEEE International Conference on Fuzzy Systems. Washington D C: IEEE Computer Society, 2011: 2528-2533.
- [16] GONZALEZ-RIVERA G, LIN W. Constrained regression for interval-valued data[J]. Journal of Business and Economic Statistics, 2013, 31(4): 473-490.
- [17] 唐娜娜. 带约束条件的区间线性回归及自回归时间序列模型[D]. 北京: 北京工业大学, 2017.
- TANG N. Interval-valued linear regression and autoregressive time series model with constraints [D]. Beijing: Beijing University of Technology, 2017. (in Chinese)
- [18] 王洵. 平稳集值时间序列与区间值时间序列的自回归模型的研究[D]. 北京: 北京工业大学, 2011.
- WANG X. Stationary set-valued time series and interval-valued time series autoregressive model [D]. Beijing: Beijing University of Technology, 2011. (in Chinese)
- [19] WANG X, ZHANG Z, LI S. Set-valued and interval-valued stationary time series[J]. Journal of Multivariate Analysis, 2016, 145: 208-223.
- [20] 章磊. 多维区间值向量自回归时间序列模型理论研究和实证分析[D]. 北京: 北京工业大学, 2018.
- ZHANG L. Theoretical research and empirical analysis of interval-valued vector autoregressive time series models [D]. Beijing: Beijing University of Technology, 2018. (in Chinese)
- [21] WANG R, WANG Z. Set-valued stationary processes[J]. Journal of Multivariate Analysis, 1997, 63(1): 180-198.

(责任编辑 杨开英)