



# 欧洲足球五大联赛进球的时间序列模型 构建及预测

全 涛

**摘要:** 通过文献资料、数理统计、对比分析等研究方法对欧洲足球五大联赛2000—2001赛季至2016—2017赛季的场均进球进行时间序列分析,构建了ARIMA模型及结果预测。研究表明:(1)欧洲五大联赛的进球特征表现出单一赛事不同时期的波动性和不同赛事之间的横向差异性。(2)欧洲五大联赛的模型分别是德甲为ARIMA(0,1,3),法甲为ARIMA(0,1,0),西甲为ARIMA(1,1,3),意甲为ARIMA(0,1,2),英超为ARIMA(1,1,2)。(3)通过模型的构建可用于短期预测,进球序列整体呈波动状态,拟合值和观测值曲线在整个区间整体拟合情况良好。

**关键词:** 欧洲足球五大联赛;进球;时间序列分析;ARIMA模型

中图分类号:G843 文献标志码:A 文章编号:1006-1207(2019)06-0071-09

DOI:10.12064/ssr.20190611

## Construction and Prediction of Goal Time Series Model in the Big Five European Football Leagues

QUAN Tao

(School of Sports and Health, LinYi University, LinYi 276000, China)

**Abstract:** On the basis of literature review, mathematical statistics, comparative analysis and other research methods, the time series analysis of the average goals in each game of the big five European football leagues during 2000/2001—2016/2017 was conducted; ARIMA model was created and results were predicted. The research shows that: (1) The goals of the big five European football leagues are characterized by fluctuations of a single game at different periods and the lateral differences between different games. (2) The models of the big five European football leagues are Bundesliga: ARIMA (0,1,3), Ligue 1: ARIMA (0,1,0), La Liga: ARIMA (1,1,3), Serie A: ARIMA (0,1,2), Premier League: ARIMA (1,1,2). (3) The construction of the model can be used for short-term prediction. The goal sequence is in a fluctuation state on the whole, and the value curves of fitting and observation fit well in the whole interval.

**Key Words:** the big five European football leagues; goal; Time Series Analysis; ARIMA Model

## 0 前言

足球在《运动训练学》项群训练理论中属于技能主导类同场对抗性运动项目,其评定方法属于命中类项目<sup>[1]</sup>,进球的结果直接影响着比赛成绩的好坏<sup>[2]</sup>。在现代足球的发展历史上,欧洲足球五大联赛代表世界足坛的顶尖水平,引领世界足球运动的发展方向。进球的特征和规律,从一定程度上体现职业联赛足球技战术打法体系的形成与革新,同时反映国家

地区的足球水平和民族精神的文化特色。

近些年,统计学方法在足球相关研究中的运用逐渐增多,从简单的计量统计,到卡方检验、t检验,再到复杂的回归分析、因子分析<sup>[3]</sup>。时间序列分析是数理统计中的一个重要分支,主要通过随机过程理论和数理统计方法来揭示数据的特点和规律,构建不同的模型和未来预测,以解决社会生活中的实际问题。通过中国知网关于“时间序列分析”的计量可视

收稿日期:2019-06-03

基金项目:2019年教育部人文社会科学研究青年基金项目(19YJC890034)。

作者简介:全涛,男,博士,副教授。主要研究方向:足球教学与训练。E-mail:979162005@qq.com。

作者单位:临沂大学 体育与健康学院,山东 临沂 276000。



化分析结果表明:(1) 发文量在近 20 年出现逐年递增的趋势,并在近 10 年保持稳定的状态;(2) 主题分布在时间序列、数学模型、ARIMA 模型、实证分析、小波分析等;(3) 学科分类主要集中在数量经济、数学、金融、计算机、地球物理、测绘、环境、临床医学、公共卫生与预防医学、国民经济、控制工程、大气、交通运输、区域经济与农业经济上。总之,时间序列分析已经广泛运用于各个领域,吴怀宇认为它在经济领域的研究和应用一直很活跃,并扩展到社会、气象、水利、交通、信息、农业、工业、教育等领域<sup>[4]</sup>。张美英和何杰强调时间序列的研究在理论及应用中都取得了极其丰硕的成果,但还有待于完善其理论及开辟新的应用领域,对于实际数据来说,没有最好的模型,只有最适合的模型,新时间序列模型的建立仍然是今后学者要继续研究的问题<sup>[5]</sup>。在体育方面,张小龙从理论的角度讨论了其可行性,通过实例分析进一步证实了在体育中应用时间序列分析的优越性<sup>[6]</sup>;任波和戴俊通过时间序列计量分析我国城乡居民消费支出与体育产业发展的互动关系<sup>[7]</sup>;辛驰和赵雪晴通过 SAS 和 E-view 对 NBA 球员林书豪进行了实例分析,建立时间序列模型并定量评价<sup>[8]</sup>;陈颇等人先后运用了时间序列分析方法,研究中国体育事业财政投入与经济增长的关系<sup>[9-10]</sup>。总之,时间序列分析具有足球赛事数据研究的可行性,既可以通过数据直观分析其特点,又可以丰富足球运动训练理论体系。

本文按照时间序列分析的方法对欧洲足球五大联赛进行多项赛事的模型构建与趋势预测,客观探索高水平足球比赛进球的特点和规律。

## 1 研究对象与方法

### 1.1 研究对象

欧洲足球五大联赛 2000—2001 赛季至 2016—2017 赛季的场均进球。

### 1.2 研究方法

#### 1.2.1 文献资料法

通过电子资源数据库查阅了有关时间序列分析、足球比赛进球分析的论文 30 余篇,阅读学习了《时间序列数据分析》《世界优秀统计工具 SPSS—高级篇》《应用统计学》相关教程与学术著作<sup>[11-15]</sup>,并在足球数据库网站上查找了关于欧洲足球五大联赛进球的数据统计与评论文章。

#### 1.2.2 数理统计法

通过 500 彩票网<sup>[16]</sup>和足球之夜数据库<sup>[17]</sup>收集了

欧洲足球五大联赛 2000—2001 赛季至 2016—2017 赛季共 17 年的年场均进球数据,并逐一核对,确定原始数据库。利用 SPSS22.0 软件在计算机上对数据进行了时间序列分析,构建欧洲五大联赛的进球 ARIMA 模型并预测。

#### 1.2.3 对比分析法

在数理统计的基础上,利用比较分析、逻辑归纳等对欧洲五大联赛的进球特征进行分析。

## 2 研究结果

### 2.1 数据集的预处理

首先把数据录入到 SPSS 中(见图 1),共有 1 个字符串型“赛季”、5 个数值型“进球”变量,其中数值保留两位小数。

赛季	英超进球	西甲进球	意甲进球	德甲进球	法甲进球	YEAR	DATE
00-01	2.61	2.88	2.76	2.94	2.48	2000	2000
01-02	2.63	2.52	2.63	2.92	2.33	2001	2001
02-03	2.61	2.68	2.58	2.68	2.20	2002	2002
03-04	2.66	2.67	2.66	2.96	2.33	2003	2003
04-05	2.56	2.58	2.53	2.91	2.17	2004	2004
05-06	2.48	2.47	2.61	2.81	2.13	2005	2005
06-07	2.45	2.48	2.55	2.74	2.25	2006	2006
07-08	2.64	2.69	2.55	2.81	2.28	2007	2007
08-09	2.48	2.90	2.60	2.92	2.26	2008	2008
09-10	2.77	2.71	2.61	2.83	2.44	2009	2009
10-11	2.80	2.74	2.51	2.92	2.34	2010	2010
11-12	2.81	2.76	2.56	2.86	2.52	2011	2011
12-13	2.80	2.87	2.64	2.93	2.54	2012	2012
13-14	2.77	2.75	2.72	3.16	2.46	2013	2013
14-15	2.57	2.66	2.69	2.75	2.49	2014	2014
15-16	2.70	2.74	2.58	2.83	2.53	2015	2015
16-17	2.80	2.94	2.96	2.87	2.61	2016	2016

图 1 欧洲五大联赛进球数据集(2000/2001—2016/2017)

Figure 1 Goal Data of the Big Five European Football Leagues (2000/2001-2016/2017)

由于欧洲足球五大联赛受到地中海气候或海洋性气候的影响,赛制实行跨年度举行,如“2001—2002 赛季意大利足球甲级联赛”这一称谓,因此软件系统或者分析中的“2000 年”等同于“2000—2001 赛季”,其他依次类推。

#### 2.1.1 离群点的检验

离群点是指一个时间序列中,远离序列一般水平的极端大值和极端小值。而形成离群点的系统外部干扰是多种多样的,可能是采样中的误差,也可能是被研究现象本身由于受各种偶然非正常的因素影响而引起的<sup>[18]</sup>。为此,运用 SPSS 对五大联赛 17 个赛季的场均进球进行了探索性分析,用箱图检验离群点,以便提高构建模型的准确性(图 2)。

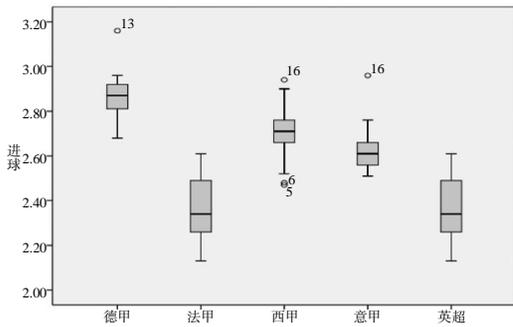


图2 进球数据箱图

Figure 2 Goal Data Boxplot

从图2可以看出,法甲、英超不存在离群点,德甲、西甲、意甲存在离群点现象。由图1数据对应后得到:德甲2013年数值3.16,西甲2005、2006年数值2.47、2.48和2016年数值2.94,意甲2016年数值2.96,共计5个数值存在不同程度的离群。

### 2.1.2 平稳性的判定

时间序列分析的前提条件是数据必须具有平稳性特征,在统计学上可以运用序列图观察和自相关、偏相关分析图判定时间序列平稳性<sup>[9]</sup>。为判断该数据是否平稳,原始数据的序列图包含5个变量。

从图3可以看出,从2001—2002赛季开始不同程度的出现下滑现象,2005—2006赛季逐渐回升,

在2009—2010赛季开始处于2~3年的平稳状态,之后随着球员转会自由市场和足球商业化的日益发展,五大联赛进球对比日益缩小。

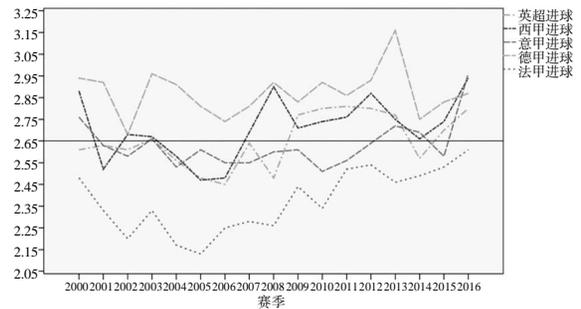


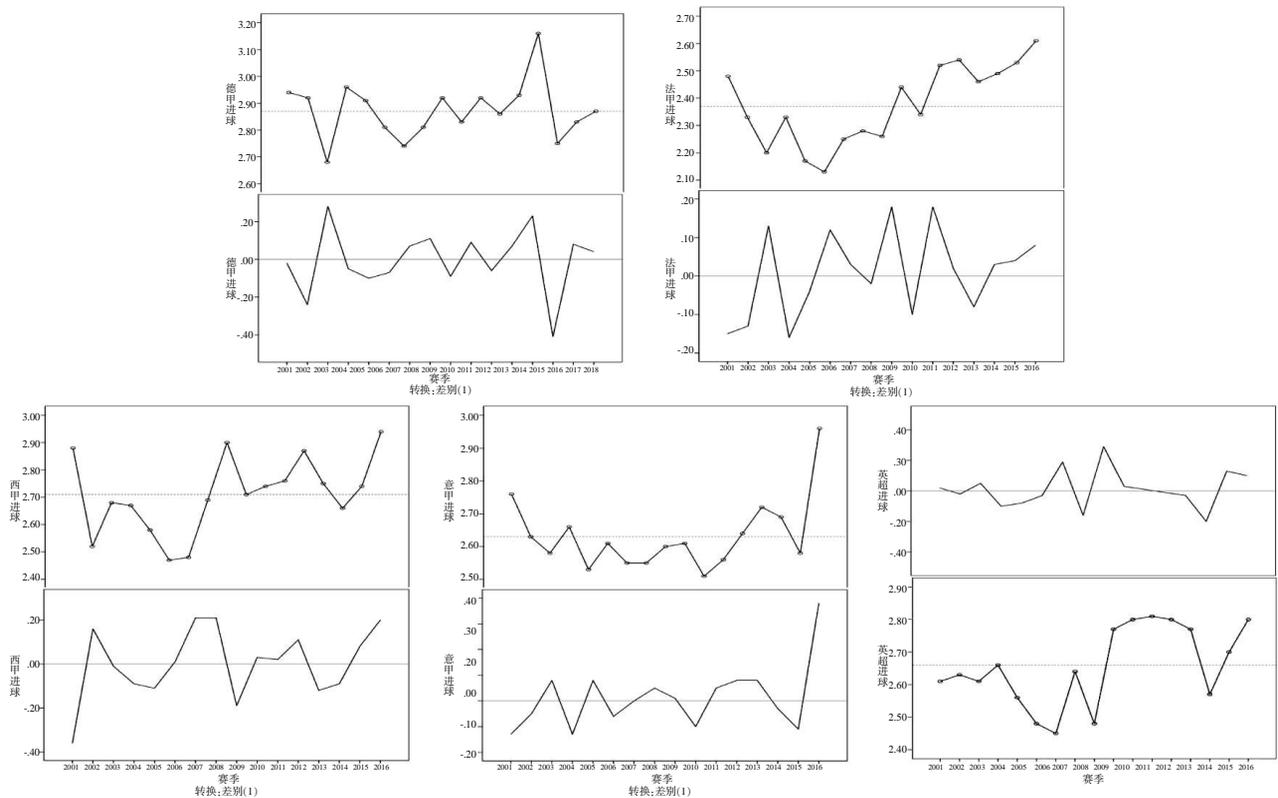
图3 欧洲足球五大联赛进球序列图

Figure 3 Goal Sequence Diagram of the Big Five European Football Leagues

## 2.2 模型识别

### 2.2.1 序列图

从原始序列图(3)来看,场均进球数随着时间的推移在起初下降后逐步上升,该序列线性特征明显。现对五大联赛的原始数据进行1阶差分,序列图(4)围绕均线上波动,状态基本平稳,差分效果良好。



注:从左到右顺序依次为德甲、法甲、西甲、意甲、英超

图4 欧洲足球五大联赛原始数据与1阶差分序列对比图

Figure 4 Comparison of the Raw Data and 1st Order Difference Sequence in the Big Five European Football Leagues



### 2.2.2 自相关与偏自相关分析

为了更好地把握五大联赛进球特点,通过自相关系数(ACF)、偏自相关系数(PACF)分析图进一步

识别模型和目标序列定阶。运用 SPSS 对数据进行了自相关分析,输出结果如图 5。

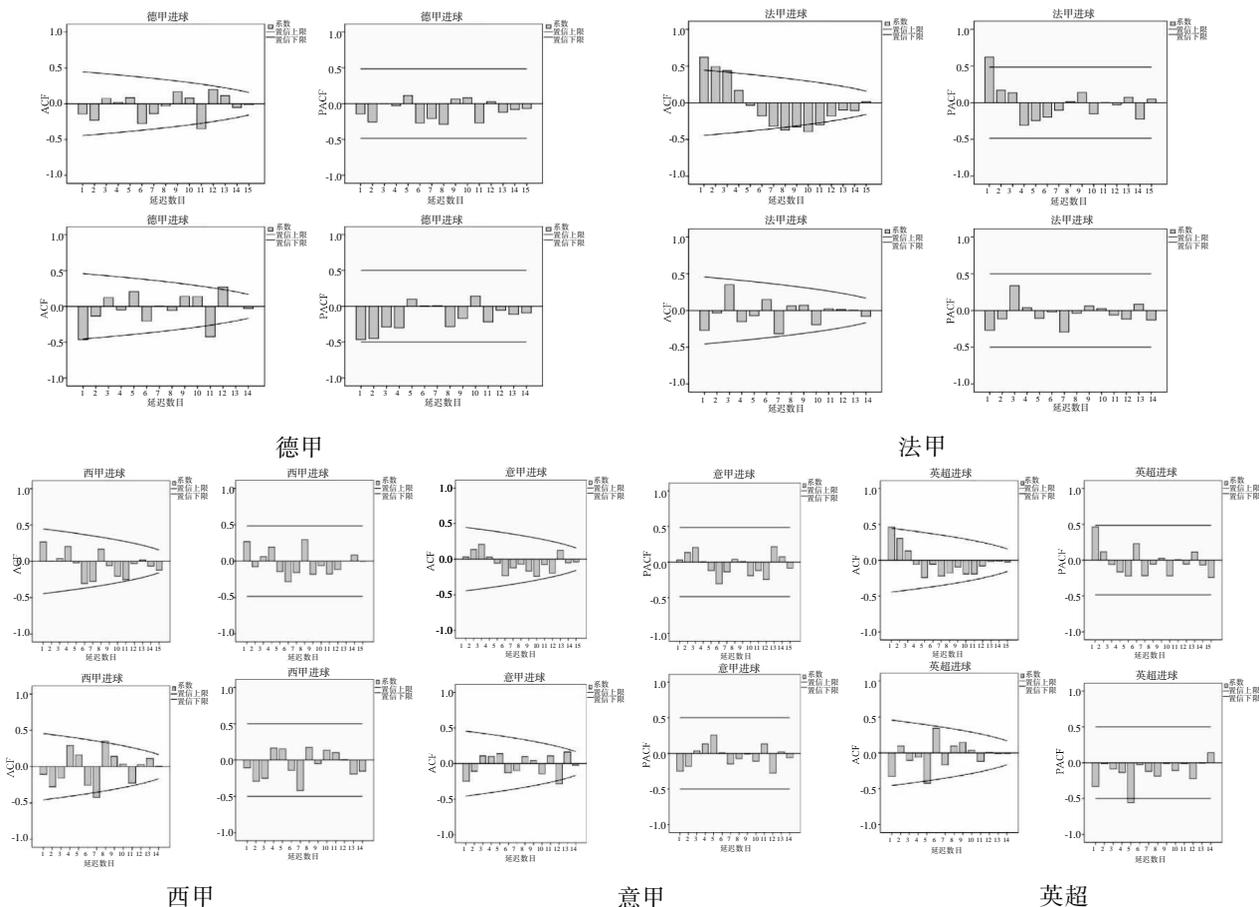


图 5 欧洲足球五大联赛进球自相关与偏自相关分析图(0,1)

Figure 5 Analysis of Autocorrelation and Partial Autocorrelation of Goals in the Big Five European Football Leagues (0,1)

总体上随着延迟数目的增加,自相关系数并没有显著的趋近于 0,且有近一半数值落在了置信区间之外,同时,偏自相关系数也逐渐减弱,说明该时间序列并非平稳。为了更好确认数据的平稳性,对原始数据进行 1 阶差分处理,得到自相关系数、偏自相关系数大多数进入置信区间。

### 2.2.3 假设检验

假设  $H_0$ : 原始时间序列  $X_t$  平稳,  $H_1$ : 序列  $X_t$  不平稳;采用显著性格水平  $\alpha=0.05$  进行检验。从表 1 来看,原始数据  $P < 0.01$ ,具有很强的显著性差异,则拒绝  $H_0$  假设接受  $H_1$  假设,证明该数列不平稳;进行 1 阶差分后,  $P > 0.05$ ,相反则接受  $H_0$  假设判定该数列为平稳序列,且数值逐渐增大总体态势为上升,与序列图的分析一致。

这里只列出国足球联赛自相关和偏相关函数的 Box-Ljung 统计量情况,其他联赛可以采用类似的方法确定其初步模型。

### 2.2.4 p,q 取值

Box-Jenkins 方法是根据时间序列模型 ACF 和 PACF 图的识别规则,建立相应的 ARIMA 模型<sup>[20]</sup>。序列 ACF、PACF 性质及其与模型参数之间的关系<sup>[21]</sup>,整理如表 2。

依据函数特征表的识别规则,对五大联赛 ACF、PACF 函数分析图进行了详细分析,统计了 1 阶差分后的 ACF 和 PACF 落在置信区外、内的数量,并对非常接近的特殊分布作了不确定、可能认定,用来初步识别模型,分布和模型情况。

由表 3 得到初步模型,其中德甲为 ARIMA (0,1,3) 或 ARIMA (2,1,3),法甲为 ARIMA (0,1,0) 或 ARIMA (0,1,1),西甲为 ARIMA (1,1,2) 或 ARIMA (1,1,3),意甲为 ARIMA (0,1,1) 或 ARIMA (0,1,2),英超为 ARIMA (1,1,1) 或 ARIMA (1,1,2)。值得注意的是,在这个过程中,充分考虑了离群点(值)的影响,运用加权的方法将特定的时间点作为离群值来建模。



表 1 法甲进球自相关和偏相关函数对比表(原始与 1 阶差分)

Table I Comparison of the Autocorrelation and Partial Correlation Functions of the French Armor Goals (Raw and 1st Order Difference)

滞后	自相关	标准误差 <sup>a</sup>	Box-Ljung 统计量			滞后	自相关	标准误差 <sup>a</sup>	Box-Ljung 统计量		
			值	df	Sig. <sup>b</sup>				值	df	Sig. <sup>b</sup>
1	0.622	0.223	7.801	1	0.005	1	-0.270	0.228	1.397	1	0.237
2	0.491	0.215	12.989	2	0.002	2	-0.032	0.220	1.418	2	0.492
3	0.439	0.208	17.434	3	0.001	3	0.350	0.212	4.125	3	0.248
4	0.169	0.201	18.143	4	0.001	4	-0.151	0.204	4.676	4	0.322
5	-0.032	0.193	18.170	5	0.003	5	-0.070	0.195	4.804	5	0.440
6	-0.178	0.185	19.101	6	0.004	6	0.147	0.186	5.428	6	0.490
7	-0.321	0.176	22.439	7	0.002	7	-0.317	0.177	8.651	7	0.279
8	-0.373	0.167	27.419	8	0.001	8	0.064	0.167	8.798	8	0.360
9	-0.329	0.157	31.795	9	0.000	9	0.073	0.156	9.015	9	0.436
10	-0.394	0.147	38.972	10	0.000	10	-0.195	0.144	10.835	10	0.371
11	-0.299	0.136	43.782	11	0.000	11	0.022	0.132	10.863	11	0.455
12	-0.181	0.124	45.903	12	0.000	12	0.014	0.118	10.877	12	0.539
13	-0.099	0.111	46.695	13	0.000	13	0.005	0.102	10.880	13	0.621
14	-0.105	0.096	47.888	14	0.000	14	-0.079	0.083	11.786	14	0.623

注:a 表示假定的基础过程是独立性(白噪音);b 表示基于渐近卡方近似

表 2 零均值平稳序列的自相关函数和偏自相关函数特征表

Table II Characteristics of Autocorrelation Function and Partial Autocorrelation Function of Zero-mean Stationary Series

模型	AR(p)	MA(q)	ARMA(p,q)
自相关函数	拖尾	截尾	拖尾
偏自相关函数	截尾	拖尾	拖尾

表 3 欧洲五大联赛进球 ACF、PACF 分布和 p,d,q 取值一览表(1 阶)

Table III List of ACF, PACF Distribution and p, d, q Values of Goals in the Big Five European Football Leagues (1st Order)

联赛	自相关		偏自相关		接近(不确定)	p,d,q 取值
	外部	内部	外部	内部		
德甲*	3	其他	0	全部	PACF:2	(0,1,3)/(2,1,3)
法甲	0	全部	0	全部	ACF:1	(0,1,0)/(0,1,1)
西甲*	2	其他	0	全部	ACF:1,PACF:1	(0,1,2)/(1,1,2)/(1,1,3)
意甲*	1	其他	0	全部	ACF:1	(0,1,1)/(0,1,2)
英超	1	其他	1	其他	ACF:1	(1,1,1)/(1,1,2)

注:\* 表示该联赛存在离群点(值)

### 2.3 模型构建

运用 SPSS22.0 软件计算,得到五大联赛的模型结果分析见表 4。

### 2.4 模型预测

模型的适用性检验按白噪音独立性检验准则,其基本思想是:若由估计模型拟合的残差纯粹由干

扰产生,则该模型是适用的,可用于外推预测;否则,估计模型不合适<sup>[21]</sup>。为此,对进球残差自相关和偏相关进行了检验,结果如图 6。

通过建立 ARIMA 模型对欧洲足球各大联赛的均进球数据进行拟合,并对 2017-2020 赛季的进球数值进行了结果预测,形成模型拟合及预测图(图 7)。

同时,根据命令输出了 2017—2020 年的预测值、UCL、LCL 值详见表 5,以供参考。



表 4 欧洲足球五大联赛进球模型参数表 (d=1)

Table IV Goal Model Parameter Table of the Big Five European Football Leagues (d=1)

联赛	模型	平稳性(R方)	离群值	选项	估计	SE	t	Sig.		
德甲	ARIMA(0,1,3)	0.707	2013 (1)	常数	0.000	0.003	0.043	0.966		
				MA 滞后 1	1.420	32.458	0.044	0.966		
				滞后 2	0.071	15.042	0.005	0.996		
				滞后 3	-0.536	17.365	-0.031	0.976		
	ARIMA(2,1,3)	0.715	2013 (1)	常数	0.000	0.003	0.173	0.866		
				AR 滞后 1	-0.320	1.013	-0.315	0.760		
				滞后 2	0.148	1.095	0.135	0.896		
				MA 滞后 1	0.971	1259.963	0.001	0.999		
				滞后 2	0.971	1067.638	0.001	0.999		
				滞后 3	-1.000	467.993	-0.002	0.998		
法甲	ARIMA(0,1,0)	0.423	0	常数	0.008	0.028	0.290	0.776		
	ARIMA(0,1,1)	0.468	0	常数	0.010	0.020	0.529	0.605		
				MA 滞后 1	0.309	0.268	1.154	0.268		
西甲	ARIMA(1,1,2)	0.404	2005	常数	0.007	0.009	0.748	0.474		
			2006	AR 滞后 1	-0.388	1.399	-0.277	0.788		
			2016	MA 滞后 1	0.841	9.167	0.092	0.929		
			(3)	滞后 2	0.148	2.640	0.056	0.956		
			ARIMA(1,1,3)	0.362	2005	常数	0.005	0.013	0.405	0.696
					2006	AR 滞后 1	-0.988	0.201	-4.920	0.001
	2016	MA 滞后 1			-0.929	101.946	-0.009	0.993		
	(3)	滞后 2			0.941	191.280	0.005	0.996		
				滞后 3	0.986	96.872	0.010	0.992		
	意甲	ARIMA(0,1,1)	0.639	2016 (1)	常数	-0.004	0.011	-0.403	0.693	
				MA 滞后 1	0.522	0.255	2.046	0.062		
ARIMA(0,1,2)		0.658	2016 (1)	常数	-0.002	0.006	-0.420	0.682		
				MA 滞后 1	0.592	35.477	0.017	0.987		
				滞后 2	0.406	14.605	0.028	0.978		
英超	ARIMA(1,1,1)	0.265	0	常数	0.012	0.008	1.479	0.163		
				AR 滞后 1	0.417	0.455	0.917	0.376		
				MA 滞后 2	1.000	192.498	0.005	0.996		
	ARIMA(1,1,2)	0.173	0	常数	0.010	0.022	0.440	0.668		
				AR 滞后 1	-1.000	0.098	-10.186	0.000		
				MA 滞后 1	-0.679	0.903	-0.752	0.467		
			滞后 2	0.308	0.474	0.650	0.528			

注: 西甲模型 ARIMA(0,1,2)因迭代在优化期间不收敛,可能不可靠数据未列出;离群值标出具体的年(赛季),(\*)为个数,均对其进行加权处理。该数据选择 1 阶差分和无转换处理,因表格有限未列出

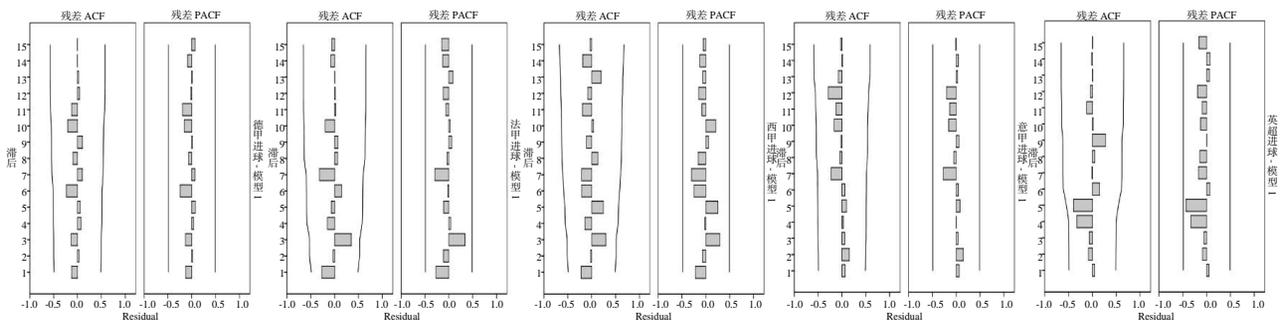


图 6 欧洲足球五大联赛进球残差自相关和偏相关检验图

Figure 6 Autocorrelation and Partial Correlation Test Card of Goal Scores in the Big Five European Football Leagues

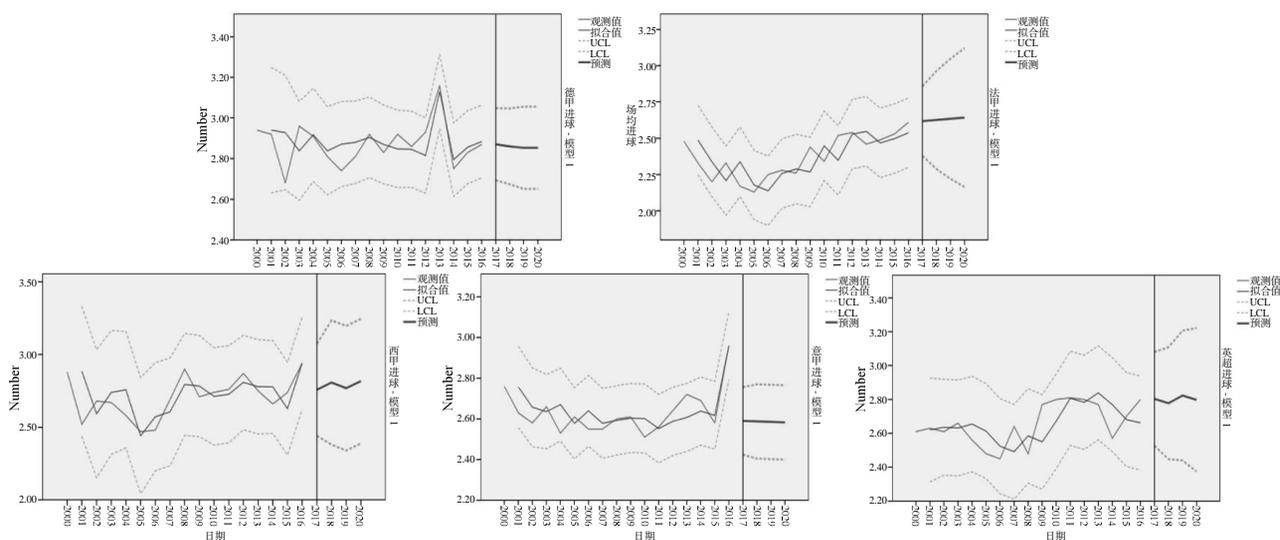


图7 欧洲足球五大联赛进球模型及预测图(2017—2020)

Figure 7 Predicted Value of the Goal Model of the Big Five European Football Leagues(2017-2020)

表5 欧洲足球五大联赛进球模型预测值(2017—2020)

Table V Predicted Value of the Goal Model of the Big Five European Football Leagues (2017-2020)

模型		2017	2018	2019	2020
德甲进球 - 模型 ARIMA(0,1,3)	预测	2.87	2.86	2.85	2.85
	LCL-UCL	2.69-3.05	2.67-3.05	2.65-3.06	2.65-3.06
法甲进球 - 模型 ARIMA(0,1,0)	预测	2.62	2.63	2.63	2.64
	LCL-UCL	2.38-2.86	2.29-2.96	2.22-3.05	2.16-3.12
西甲进球 - 模型 ARIMA(1,1,3)	预测	2.76	2.81	2.77	2.82
	LCL-UCL	2.44-3.07	2.38-3.23	2.34-3.20	2.39-3.25
意甲进球 - 模型 ARIMA(0,1,2)	预测	2.59	2.59	2.58	2.58
	LCL-UCL	2.42-2.76	2.40-2.77	2.40-2.77	2.40-2.77
英超进球 - 模型 ARIMA(1,1,2)	预测	2.80	2.78	2.82	2.80
	LCL-UCL	2.53-3.08	2.45-3.11	2.44-3.21	2.37-3.22

注:对于每个模型,预测都在请求的预测时间段范围内的最后一个非缺失值之后开始,在所有预测值的非缺失值都可用的最后一个时间段或请求预测时间段的结束日期(以较早者为准)结束

### 3 分析与讨论

#### 3.1 欧洲足球五大联赛的进球特征

在体育运动项目的模型构建过程中,原始数据本身具有随机性,非稳定性特征比较常见,这就需要采用合适的方法进行序列分析,以保证研究结果的正确性。前期对离群点的分析和平稳性的判定非常重要,否则会直接影响模型的拟合精度,得到一些虚假的信息,对后续的时间序列分析会造成一定影响。

对“五个离群点”的进一步分析发现:德甲 2013

年场均 3.16 个进球,说明德甲保持较高的场均进球,一直是崇尚进攻的联赛。西甲 2005 和 2006 年出现的 2.47、2.48 两个最低的进球数值,原因在于那个时期西甲球员出现新老交替,进攻能力下降;意甲 2016 年的场均进球为 2.96,这与意甲改变传统保守的技战术打法有关。另一方面,从进球序列图的走势来看,表现出单一赛事不同时期的波动性和不同赛事之间的横向差异性特征。五大联赛场均进球的数据随着时间的推移上下波动,并有上升的趋势。在作时间序列分析时,首先初步判断该序列的稳定性特征,如果不稳定需要对数据进行自相关函数、偏相关



函数、特征根的检验,采取1阶或2阶差分处理,直至数据平稳,才能继续进行模型的识别,以期构建合理的模型和预测,探索欧洲足球五大联赛的进球特点和规律。路云亭把第五种足球描述为“盛传于世界各地且带有本民族文化传统意味的多元化的足球”<sup>[23]</sup>,笔者认为欧洲五大联赛具备不同的差异,反映着一个国家的民族精神和区域文化特色。如法甲最讲究“思想性”与“公平性”,其竞争力有待于提升;意甲受到经济的影响较大,表现出传统保守的特点;英超的商业化最为成熟,比赛更加注重身体对抗、比赛节奏较快,观赏性高;西甲联赛注重技术与进攻,具有很强的观赏性;德甲的民族特点突出,注重整体配合。

### 3.2 欧洲五大联赛的 ARIMA 模型

由于足球比赛中的场均进球序列不存在季节性周期趋势,这里的模型识别没有考虑季节问题。在模型识别时,则是通过序列图和自相关与偏相关的1阶差分处理,为的是保证数据的平稳性,使ACF和PACF进入95%的置信区间,进一步确定 $d=1$ ,同时为下一步的参数估计做好准备。

在这里需借助于自相关函数(ACF)和偏自相关函数(PACF)以及其相关分析图来识别时序特性,如法甲(表1)运用BOX-JENKINS预测法假设检验,采用显著性格水平0.05进行1阶的前后对比检验,统计量由小于0.01到大于0.05,表明效果良好,证明这一方法的可行性。在确定 $p, q$ 值时,需根据自相关函数(ACF)和偏自相关函数(PACF)图的拖尾、截尾两个特征,内外部分布情况,参数之间的关系3个方面,初步建立相应的ARIMA模型。运用SPSS软件的“分析—预测—创建模型”命令,自动输出的统计量和相关信息,包括平稳的(R方)、离群值、参数估计、SE、t、Sig。主要是通过比较平稳的(R方)、P值、观测值和拟合值的曲线拟合度情况,来确定各大联赛的最终进球模型。分析如下:德甲:平稳R方分别为0.715和0.707,P值分别为0.966和0.866,第一个模型的两项数值比第二模型高,所以模型定为ARIMA(0,1,3)。法甲:两个模型的常数估计值为0.008和0.010,P值分别为0.605和0.776,ARIMA(0,1,0)拟合度高于ARIMA(0,1,1),加上R方相差不大,取P值为0.776,模型确定为ARIMA(0,1,0)。西甲:因ARIMA(0,1,2)迭代在优化期间不收敛,可能不可靠数据放弃。另两个模型的平稳R方分别为0.404和0.362,P值分别为0.474和0.696,结合模型的拟合情况确定为ARIMA(1,1,3)。同样的方法推

断意甲模型为ARIMA(0,1,2),英超模型为ARIMA(1,1,2)。

### 3.3 欧洲五大联赛的短期预测

通过自相关和偏相关的残差检验表明,数值都落在95%置信区间内,与0无显著差异,模型的检验效果比较好,所以残差通过白噪声检验,由此诊断模型是可行的,可用于短期预测。

从预测图(图7)左侧可以看出,进球序列上整体上成波动状态,拟合值和观测值曲线在整个区间整体上拟合情况良好,拟合值的波动非常接近实际观察值波动性。从参考线右侧预测了2017—2020赛季的场均进球态势,其中德甲趋于缓慢下降,法甲继续稳步上升,西甲处于上下波动、上升趋势,意甲趋于稳定,英超则呈上下波动、下降趋势。当然,这只是单纯数据的预测,不排除有其他特殊因素的影响,在前期“进球数量稳定特征”<sup>[24]</sup>的研究结论有所论述,即在常态下,场均进球应该是一个比较平稳的状态,不会有很大的波动,当然不排除特殊情境(争冠、降级、出线等)的影响。总之,由于足球运动是攻守平衡的一项同场对抗项目,进球总体上应该趋于稳定状态,介于2.5~3.0之间。

## 4 研究结论

4.1 足球在《运动训练》学具有技能和命中的双重属性,决定了进球是多种技战术相互作用的结果,在现有足球规则倡导进攻的条件下,场均进球较少,有稳定增加的态势,这符合足球竞技体育项目的发展规律。

4.2 欧洲五大联赛的进球特征表现出单一赛事不同时期的波动性和不同赛事之间的横向差异性。波动性主要是随着时间的推移上下波动,并有上升的趋势;差异性反映不同国家的民族精神和区域文化特色,从进球曲线表现上,依次为:德甲、西甲、英超、意甲、法甲。

4.3 在建模的过程中,要充分考虑离群值、数据平稳性问题,严格按照时间序列分析的相关分析、假设检验、 $p$ 与 $q$ 取值等步骤建模。欧洲五大联赛的模型是德甲为ARIMA(0,1,3),法甲为ARIMA(0,1,0),西甲为ARIMA(1,1,3),意甲为ARIMA(0,1,2),英超为ARIMA(1,1,2)。

4.4 通过模型的构建可用于短期预测,进球序列上整体上成波动状态,拟合值和观测值曲线在整个区间整体上拟合情况良好。



## 参考文献:

- [1] 田麦久.《运动训练学》[M].北京:人民体育出版社,2000:39.
- [2] 全涛,孙雪翰.2008年中国足球超级联赛成绩与技术统计的因子分析[J].数理统计与管理,2011(2):370-380.
- [3] 赵陆.近五年国内足球运动科学研究现状与发展趋势[D].石家庄:河北师范大学,2012.
- [4] 吴怀宇.时间序列分析与综合[M].武汉:武汉大学出版社,2004.
- [5] 张美英,何杰.时间序列预测模型研究综述[J].数学的实践与认识,2011(18):189-194.
- [6] 张小龙.时间序列自回归(AR)模型在体育预测中的应用[J].北京体育大学学报,2010(2):86-88.
- [7] 任波,戴俊.我国城乡居民消费支出与体育产业发展互动关系的计量研究——基于2006—2015年的时间序列数据[J].天津体育学院学报,2017(1):87-92.
- [8] 辛驰,赵雪晴.基于时间序列分析的体育比赛成绩分析与预测[J].科技风,2012(16):185-186.
- [9] 张羽,杨铁黎,赵鑫.体育财政投入、体育事业发展与经济增长——基于我国时间序列数据的实证研究[J].北京体育大学学报,2015(6):12-17.
- [10] 陈颇.中国体育事业财政投入与经济增长关系的实证研究——基于1977—2010年的时间序列数据分析[J].武汉体育学院学报,2012(5):34-39.
- [11] 王黎明,王连,杨楠编著.应用时间序列分析[M].上海:复旦大学出版社,2009.09.
- [12] 肖枝洪,郭明月编著.时间序列分析与SAS应用(第2版)[M].武汉:武汉大学出版社,2012.
- [13] 张文彤主编.世界优秀统计工具 SPSS 11 统计分析教程(高级篇)[M].北京:北京希望电子出版社,2002.06.
- [14] 赵玮,温小霓.应用统计学教程(上)[M].西安:西安电子科技大学出版社,2003.
- [15] 赵玮,温小霓.应用统计学教程(下)[M].西安:西安电子科技大学出版社,2003.
- [16] 500彩票网-数据统计[DB/OL][2019-06-03].http://lian sai.500.com/zuqiu-4435/.
- [17] 足球之夜-欧洲赛事[DB/OL][2019-06-03].http://www.uhchina.com/ouzhouzuqiusaishi/.
- [18] 王振龙主编.时间序列分析[M].北京:中国统计出版社,2000:10-13.
- [19] 范涛涛,寇艳廷.时间序列分析中数据的平稳性判定研究[J].现代电子技术,2013(4):66-68,72.
- [20] 王晓鹏,曹广超,丁生喜.基于Box-Jenkins方法的青南高原降水量时间序列分析建模与预测[J].数理统计与管理,2008(4):5-10.
- [21] 安鸿志.时间序列的分析与应用[M].北京:科学出版社,1983:73-83.
- [22] 赵晓葵.基于Box-Jenkins方法的中国年度GDP时间序列分析建模与预测[J].青海师范大学学报(自然科学版),2009(3):18.
- [23] 路云亭.足球中的国民性[J].体育与科学,2015(1):28-37.
- [24] 全涛.基于时间序列分析的2016年中超联赛进球ARIMA模型研究[J].统计与管理,2018(5):86-91.
- [25] 浦义俊,戴福祥,江长东.德国足球甲级联赛的历史演进与支持系统分析[J].成都体育学院学报,2016(3):66-72.
- [26] 杨世东,段博文,杨祖辉.法国足球甲级联赛研究[J].体育文化导刊,2011(5):17-21.
- [27] 张俊斌.欧洲足球五大联赛研究[J].体育文化导刊,2012(11):66-70.
- [28] 邱希,曾庆欢.西班牙足球甲级联赛研究[J].体育文化导刊,2014(1):96-99.
- [29] 梁国强.欧洲五大联赛对足球运动发展的推动作用[J].山西师大体育学院学报,2011(3):97-100.

(责任编辑:刘畅)