DOI:10.16016/j.1000-5404.2002.08.037
第24卷第8期 第三军医大学学报 Vol. 24, No. 8
2002年8月 ACTA ACADEMIAE MEDICINAE MILITARIS TERTIAE Aug. 2002

文章编号:1000-5404(2002)08-0955-03

论著

时间序列资料 ARIMA 季节乘积模型及其应用

张 蔚[1],张彦琦[1],杨 旭[2] ([1]第三军医大学高原军事医学系卫生统计学教研室,重庆 400038;[2]重庆电力教育培训中心,重庆 400053)

提 要:目的 用 ARIMA 季节乘积模型(p,d,q)(P,D,Q)s 对季节性时间序列资料建模并预测,并与指数平滑法进行比较,考察 ARIMA 乘积模型的预测效果。方法 用 Box-Ljung 统计量评价 ARIMA 模型的拟和度,用平均预测相对误差作为预测效果的评价指标。结果 对所分析的季节性时间序列建立了乘积 ARIMA(0,1,1)×(0,1,1)$_{12}$模型,平均预测相对误差为 4.89%,指数平滑法的平均预测相对误差为 8.14%。结论 对所分析的时间序列,ARIMA 乘积模型的预测效果优于指数平滑法。

关键词:时间序列;ARIMA 模型
中图法分类号:R195.1;R197.3 文献标识码:A

Model of multiple seasonal ARIMA and its application to data in time series

ZHANG Wei, ZHANG Yan-qi, YANG Xu (Department of Medical Statistics, Third Military Medical University, Chongqing 400038, China)

Abstract: **Objective** To establish a model of multiple seasonal autoregressive integrated moving average (ARIMA) (p, d, q) (P, D, Q)s on time-serial data, predict for the seasonal time series, and evaluate its predictive effect. **Methods** Statistics of Box-Ljung was used to evaluate the degree of fitness of ARIMA model. The predictive results were compared with the outcomes obtained by the method of exponential smoothing. And the average relative errors of predict were used as indexes to evaluate the predict effect. **Results** Model of multiple seasonal ARIMA (0, 1, 1)×(0, 1, 1)$_{12}$ was established for the time series needing analyzing with an average relative error of 4.89%. And the average relative error of exponential smoothing was 8.17%. **Conclusion** For the time serial data which need be analyzed, model of multiple seasonal ARIMA is superior to the method of exponential smoothing with better predict effect.

Key words: time series; model of autoregressive integrated moving average

20 世纪 60 年代,美国学者 Box 和英国统计学者 Jenkins 提出了一整套关于时间序列分析、预测和控制的方法,被称为 Box-Jenkins 建模方法[1]。求和自回归移动平均(Autoregressive integrated moving average, ARIMA)模型是其中重要而基本的模型之一,在各个方面的应用十分流行。针对有季节变动规律的单因素时间序列(如医院门诊量),在本研究中,我们尝试用混合有季节性和非季节性成分的 ARIMA 乘积模型(p, d, q)(P, D, Q)s 对门诊量序列进行分析和预测。

1 建模方法

设 $X=(x_1, x_2, \cdots, x_t, \cdots, x_n)$ 为一个在等间隔时间 1, 2, …, t, …, n 采集得到的时间序列,e 为误差序列。引入后移算子 B,则有 $B^m x_t = x_{t-m}$,m 为时间跨度;以及向后差分算子 ∇^d,则有 $\nabla x_t = x_t - x_{t-1} = (1-B)x_t$,$d$ 为差分的次数。

1.1 非季节性 ARIMA(p, d, q)模型

ARIMA 建模方法是以序列的平稳性为前提的,若序列没有恒定的均值和方差,则需要对序列进行变换和处理使之平稳之后再进行 ARIMA 分析。非季节性 ARIMA(p, d, q)模型的一般形式为

$$\phi(B)\nabla^d x_t = \theta(B)e_t$$

其中,$\phi(B)=1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p$ 为自回归算子,P 为自回归阶数,$\phi_1, \phi_2, \cdots, \phi_p$ 为自回归部分参数。$\theta(B)=1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q$ 为移动平均算子,q 为移动平均阶数,$\theta_1, \theta_2, \cdots, \theta_q$ 为移动平均部分参数。

1.2 季节性 ARIMA(P, D, Q)s 模型

对于有季节性变动规律的时间序列,其季节性 ARIMA 模型为

$$\phi(B^s)\nabla_s^D x_t = \theta(B_s)e_t$$

其中,s 为季节性周期,$\Phi(B^s)=1-\Phi_1 B^s-\Phi_2 B^{2s}-\cdots-\Phi_p B^{ps}$ 为季节性自回归算子,P 为季节性自回归阶数,$\Phi_1, \Phi_2, \cdots, \Phi_P$ 为季节性自回归部分参数。$\Theta(B^s)=1-\Theta_1 B^s-\Theta_2 B^{2s}-\cdots-\Theta_Q B^{Qs}$ 为季节性移动平均算子,Q 为季节性移动平均阶数,$\Theta_1, \Theta_2, \cdots, \Theta_Q$ 为季节性移动平均部分参数。

1.3 乘积(p, d, q)(P, D, Q)s 模型

有的季节性时间序列不仅含有季节性成分,还混有非季节性成分,若单一用季节性或非季节性 ARIMA 模型进行分析,其预测效果往往不理想。季节性 ARIMA 与非季节性 ARIMA 的混合效应通常表现为相乘的形式。

$$\phi_p(B)\Phi_p(B^s)\nabla^d\nabla_s^D x_t = \theta_q(B)\Theta_Q(B^s)e_t$$

ARIMA 建模法分为三个阶段:模型识别,参数估计和诊断检验,通过这三个处理步骤的反复进行,最终确定一个用于预报或控制的最优模型。这三个阶段都需依赖计算机软件进行大量的运算。下面我们结合一个实例,对季节性时间序列资料进行预测分析。

2 应用

表 1 数据来源于某医院病案室,采集了 1995 年 1 月-2001 年 11 月间每月门诊人次,以此数据作为时间序列建立模型,对该院的门诊接诊人次进行预测。

作者简介:张 蔚(1960-),女,江苏省南京市人,实验师,主要从事卫生统计学方面的研究,发表论文 7 篇。电话:(023)68752343
收稿日期:2001-01-25;修回日期:2002-03-12

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

表 1 某医院 1995 年 1 月－2001 年 11 月门诊量(人次)

Tab 1 Amount of outpatient in a certain hospital from Jan. 1995 to Nov. 2001(person-time)

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1995	16 183	18 169	29 310	32 665	26 645	21 800	23 321	20 621	19 255	20 219	17 985	15 567
1996	14 867	13 555	18 161	19 776	17 563	16 080	18 844	14 973	15 728	17 878	19 701	19 263
1997	19 198	13 650	21 031	21 148	20 108	20 075	21 824	17 908	18 473	18 597	17 595	18 071
1998	12 542	17 109	23 003	21 315	18 044	20 969	21 400	20 266	19 796	19 798	20 816	20 078
1999	17 196	15 499	22 106	20 474	19 811	19 173	20 464	19 397	19 122	17 931	20 417	17 438
2000	15 889	14 482	22 053	20 324	22 431	23 510	22 514	23 089	20 354	22 867	22 398	19 734
2001	16 921	21 727	24 371	23 824	27 759	27 578	28 497	28 304	27 456	28 096	27 818	—

2.1 模型识别[2]

由 SPSS 8.0 Graphs→Sequence 绘制序列图，见图 1，可发现数据有明显的季节变动规律。逐年的门诊量呈现升高的趋势，且年内也呈现波动，每年 1 月或 2 月门诊量最低，4 月达到高峰。因此先提取序列的季节性成分，对序列作周期为 12 的季节性差分，并绘制序列的自相关函数图(ACF)和偏自相关函数图(PACF)(Graphs→Time Series→Auto correlations)，见图 2A、B，在 ACF 图中，时点 12 处有一个低峰，PACF 图中，在时点 12、24、36 处偏自相关函数呈现指数衰减的趋势，用文献[1]介绍的方法，构建季节性 ARIMA 模型为$(0,1,1)_{12}$。因为我们不仅关心各年相应月份门诊量的变化，还关心同一年中各月份门诊量的变化，因此作季节性 ARIMA 模型 $(0,1,1)_{12}$ 残差序列的 ACF 和 PACF 图，见图 2C、D，从中识别非季节成分，拟合非季节 ARIMA 模型为$(0,1,1)$。故而最终得到的乘积 ARIMA 模型为$(0,1,1)\times(0,1,1)_{12}$，即：$\nabla\nabla_{12}x_t=(1-\theta B)(1-\Theta_1 B^{12})e_t$

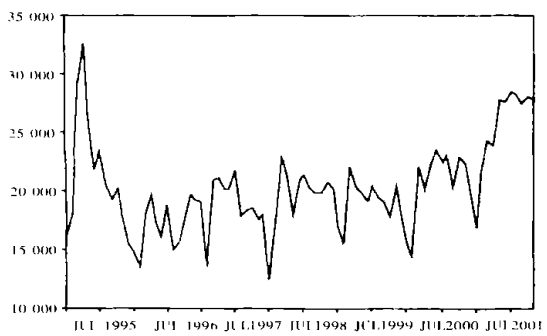
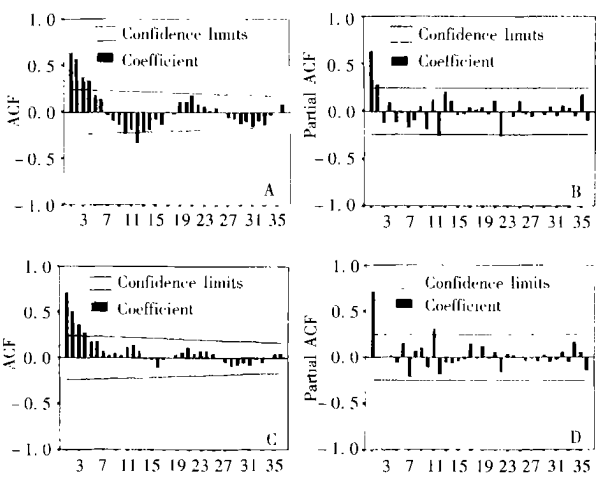


图 1 某医院 7 年来门诊量趋势图

Fig 1 Trendency curve of amount of outpatient of the hospital in 7 years



A: ACF of outpatient; B: PACF of outpatient; C: ACF of errors; D: PACF of errors

图 2 门诊量及残差的 ACF 和 PACF 图

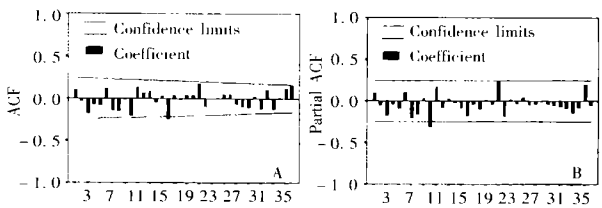
Fig 2 ACF and PACF of outpatient amount and errors of ARIMA$(0,1,1)_{12}$

2.2 参数估计

由 Statistics→Time Series→ARIMA 估计参数得 $\theta_1=0.3616$，$\Theta_1=0.5063$，且都具有显著水平。

2.3 诊断检验

对乘积模型$(0,1,1)\times(0,1,1)_{12}$残差序列作 ACF 和 PACF 图，见如图 3，得到残差的自相关函数和偏自相关函数均在可信限以内，同时 Box-Ljung 统计量无显著性，说明残差是随机分布的，因此认为模型的识别是合适的。



A: ACF of errors; B: PACF of errors

图 3 ARIMA$(0,1,1)\times(0,1,1)_{12}$的 ACF 和 PACF 残差图

Fig 3 ACF and PACF errors of ARIMA$(0,1,1)\times(0,1,1)_{12}$

2.4 预测

用建立的乘积 ARIMA$(0,1,1)\times(0,1,1)_{12}$模型对 2001 年 7 月～11 月的门诊量进行预测，得到预测数据如表 2。

表 2 ARIMA$(0,1,1)(0,1,1)_{12}$模型预测结果

Tab 2 Predictive results of ARIMA$(0,1,1)(0,1,1)_{12}$

Month(2001)	Series value	Predicted value	Error	Relative error
7	28 497	27 503	994	0.0349
8	28 304	26 882	1 422	0.0502
9	27 456	25 458	1 998	0.0728
10	28 096	26 512	1 584	0.0564
11	27 818	26 979	839	0.0302
		Average relative error		0.0489

平均相对误 $\bar{e}=\frac{1}{n}\sum_{i=1}^{n}\frac{|x_t-\hat{x}_t|}{x_t}=0.0489$。可见预测效果良好。

3 比较

有季节规律的时间序列用线性或非线性回归模型进行预测显然是牵强的，这里就不再详述。这里用指数平滑的方法进行建模[3]，得模型的常规参数 $\alpha=0.50$，趋势参数 $\gamma=0.00$，周期参数 $\delta=0.00$。把其预测结果与乘积 *ARIMA* 模型的预测结果进行比较，见表 3。

表 3 乘积 ARIMA$(0,1,1)(0,1,1)_{12}$模型与指数平滑模型预测结果比较

Tab 3 Comparison on predict results by ARIMA$(0,1,1)(0,1,1)_{12}$ and exponential smoothing

Month(2001)	Series value	Predicted value		Relative error	
		ARIMA	Exponential smoothing	ARIMA	Exponential smoothing
7	28 497	27 503	27 713	0.0349	0.0275
8	28 304	26 882	25 248	0.0502	0.1080
9	27 456	25 458	24 444	0.0728	0.1097
10	28 096	26 512	25 364	0.0564	0.0972
11	27 818	26 979	26 027	0.0302	0.0644
			Average relative error	0.0489	0.0814

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

4 讨论

ARIMA 乘积模型是针对有季节性变动的时间序列提出的建模方法。它对每一季节周期中相同时间点的序列值进行分析，提取季节趋势；并且针对每个季节周期内部序列值的变化提取非季节性成分，确立最优模型。指数平滑法是对近期序列值赋予大的权重，来估计序列的平均水平、趋势性和周期性。通过比较，可以看出 ARIMA 乘积模型的预测平均误差为 4.89%，而指数平滑法为 8.14%，ARIMA 乘积模型的预测效果优于指数平滑法。本研究只利用一例时间序列比较二者的预测效果，在实际应用中，由于数据的复杂性，可选择多种方法建模，确立最优预测模型。

ARIMA 建模法的最大特点之一就是其对序列平稳型的要求，在应用模型以前，对序列进行预处理，以保证序列均值和方差的平稳性。在本例中，我们用一阶季节性差分和一阶非季节性差分的方法，用组合算子$(1-B)(1-B^{12})\chi_t$ 对原序列变换，实现平稳化。差分往往要求大样本的特性，即至少 7 个季节周期的序列值。但在更多的实际应用中，我们仍需要不断地改进模型的估计方法，主要是寻找对序列进行平稳性处理的方法，以提高预测模型的精度。

参考文献：

[1] Geoge E. P. Box 著. 顾岚主译. 范金城校译. 时间序列分析预测与控制[M]. 北京：中国统计出版社，1997. 101－135.
[2] 章扬熙. 医学统计预测[M]. 北京：中国科技技术出版社，1995. 64－90.
[3] 徐国强，胡清友. 统计预测和决策[M]. 上海：上海财经大学出版社，2001. 113～131.

（编辑　王　红）

文章编号：1000-5404(2002)08-0957-01

个案与短篇

慢性呼吸衰竭病人血循环内皮细胞变化

Changes of circulating endothelial cells in patients with chronic respiratory failure

李松柏，　骏青，唐学良　　（安阳钢铁集团公司职工医院，安阳 455004）

本研究通过研究肺心病慢性呼吸衰竭病人循环内皮细胞（CEC）变化，探讨肺动脉血栓形成的机制，为肺心病慢性呼吸衰竭病人的防治提供依据。

1 材料与方法

病例选择：选择本院住院治疗的 29 例病人，所有病人均符合 1997 年 COPD 诊断标准。其中 23 例并有肺心病。依动脉血气分析结果，将病人分为 2 组：①呼吸衰竭组（呼衰组）：16 例，男 15 例，女 1 例；年龄（64±12）岁；② 非呼吸衰竭组（对照组）：13 例，男 11 例，女 2 例；年龄（67±7）岁。

方法：①血气分析。在不吸氧情况下，用肝素化注射器抽取动脉血 2 ml。在 IL-1 306 血气分析仪上进行血气分析。② CEC 测定，按彭文鸿[1] 介绍的方法测定。③统计学处理。数据用 $\bar{x}\pm s$ 表示。两组数据差异显著性采用 t 检验。并计算 CEC 与 pH、$PaCO_2$ 和 PaO_2 的相关系数。

2 结果

两组病人动脉血气分析和循环内皮细胞计数比较，见表 1。

表 1　两组病人动脉血气和循环内皮细胞计数比较（$\bar{x}\pm s$）

	pH	$PaCO_2$(kPa)	PaO_2(kPa)	SO_2(%)	CEC(0.9 均)
呼衰组	7.30±0.11 *	7.92±2.91 **	6.71±1.24 **	78.59±11.78 **	1.98±0.79 *
对照组	7.38±0.05 *	5.47±0.72 *	10.08±1.36	94.20±2.01	1.35±0.79

*：$P<0.05$，**：$P<0.01$ 与对照组比较

CEC 与 pH、$PaCO_2$ 和 PaO_2 的相关关系：呼衰组、对照组或全组病人 CEC 与 pH、$PaCO_2$ 以及 PaO_2 均无明显相关。

3 讨论

循环内皮细胞是指在生理和病理情况下，从循环血中获得的内皮细胞。研究发现[2]，导致血管内皮损伤的因素均可使 CEC 增多。CEC 可以直接而特异地反映活体内血管内皮细胞损伤情况，而主要反映血管内皮细胞的脱落性损伤。本研究表明，慢性呼吸衰竭病人 CEC 显著多于对照组，提示慢性呼吸衰竭病人伴有明显的血管内皮细胞损伤。另外，本研究发现，CEC 与 pH、$PaCO_2$ 和 PaO_2 均无明显相关，提示该血管内皮损伤并非某个单一因素作用，而可能系缺氧、二氧化碳潴留、酸血症等综合作用所致。

当血管内皮细胞损伤、脱落，内皮下胶原组织暴露后，可引起血小板粘附、聚集；而本已处于高凝状态的血液则更易凝聚，进而引起继发性肺内多发性血栓形成，从而加重慢性呼吸衰竭患者病情。因此，临床上对 COPD 慢性呼吸衰竭病人，除应用抗生素防治感染，积极氧疗，改善通气等综合处理外，给予内皮细胞保护措施也是必要的。

关键词：呼吸衰竭；内皮细胞
中图法分类号：R563.802　　文献标识码：B

参考文献：

[1] 彭文鸿，邹霞英，毛宝龄，等. 慢性阻塞性肺疾病和肺心病患者血液动力学与循环内皮细胞变化[J]. 中华结核和呼吸杂志，1998，21(11)：678－680.
[2] 李　琦，钱桂生，陈正堂，等. 急性呼吸窘迫综合症患者的血液及肺组织中血管内皮细胞标志物的变化[J]. 中华结核和呼吸杂志，1999，22(9)：533－535.

（编辑　王　红）

作者简介：李松柏（1960－），男，河南省温县人，硕士研究生，副主任医师，主要从事肺心病呼吸衰竭、高血压病的研究，发表论文 13 篇。电话：(0372)311185

收稿日期：2002-05-18；修回日期：2002-07-18

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net