广西科学院学报 2007,23(3):153~156
Journal of Guangxi Academy of Sciences Vol. 23, No. 3 August 2007

桂林市旅游人数的时间序列预测模型*

The Model of Time Series Prediction of the Number of the Tourists in Guilin

何荣国,邓国和
HE Rong-guo, DENG Guo-he

(广西师范大学数学科学学院,广西桂林 541004)
(College of Mathematics, Guangxi Normal University, Guilin, Guangxi, 541004, China)

摘要:采用一阶自然对数差分和一阶季节差分来数学处理桂林市月旅游人数时间序列的季节性和波动性趋势,并依据1999年1月～2006年8月桂林市月旅游人数数据,建立桂林市旅游人数的时间序列预测模型,并将该模型与实际数据进行拟合和预测,结果表明该模型与实际数据的拟合性好,预测得到的数据与实际数据误差较小,可以实际用来预测未来日旅游人数的基本趋势,为管理和市场决策提供参考.

关键词:预测模型 季节因素 趋势因素
中图法分类号:O29;F59 **文献标识码**:A **文章编号**:1002-7378(2007)03-0153-04

Abstract: This paper makes mathematical treatment of the seasonal and volatility trend of the time series of the number of tourists per month in Guilin by using one order natural logarithm difference and one order season difference, establishes the model of time series prediction of the number of tourists in Guilin based on the number of tourists of Guilin from January 1999 to August 2006, and combines the model with the actual data. The results indicates that the model fits well with the actual data, with small difference between the data obtained from the forecast and the actual data. Consequently, this model can be used to predict the basic trend of the number of tourists per month in Guilin and provide a reference to the management and the market.

Key words: prediction model, seasonal factors, trend factors

桂林一贯是国内重点扶持的国际旅游城市,桂林旅游业在全区和全国旅游业中有着重要地位,它对桂林市的社会发展、经济建设和环境保护等各方面都会产生巨大影响.随着近年来旅游业规模的扩大和效益的提高,旅游业的管理和决策工作也增加了许多困难.要保证旅游业的持续稳步发展,必须以科学的方法来管理、分析、预测旅游市场的规律.

旅游行业的特点决定了以月为统计单位的旅游人数时间序列中会存在明显的季节性及波动性趋势.与年度数据不同,月度数据构成的时间序列往往会在正常年度中表现出有规律的周期性变化,这种变化被称为季节变动[1].季节变动的存在,一定程度上会掩盖时间序列短期的基本变动,可能会造成深入研究和正确解释时间序列变动规律的困难,为此必须进行季节调整[1].季节调整就是通过数学方法把原始时间序列中隐含的季节性和偶然性因素剔除.本文采用一阶自然对数差分和一阶季节差分来数学处理月旅游人数时间序列的季节调整和变动趋势分析,建立乘积季节模型.该模型实际数据拟合和预测的结果表明该模型合理.

1 乘积季节模型

通常,在一个序列中,若经过S个基本时间间隔后呈现出相似性,我们就说序列表现出以S为周期的周期特性.具有周期特性的序列就称为季节时间序列.S为周期长度,一个周期所包含的时间点称为周期点.对于月份资料来说,基本时间间隔为1个月,周期为12($S=12$),同一个周期内有12个周期点.

称模型:$\Phi(B)\nabla^d X_t=\Theta(B)a_t$为ARIMA($n,d,m$)(求和自回归滑动平均)模型,它是非平稳时间序

收稿日期:2006-12-12
作者简介:何荣国(1982-),男,硕士研究生,主要从事金融工程、金融统计研究工作.
*国家自然科学基金项目(40675023)和广西师范大学博士基金项目(2006E24)资助.

列模型. 这里 B 为后移算子, $\Phi(B)=1-\varphi_1B-\varphi_2B^2-\Lambda-\varphi_nB^n$, $\Theta(B)=1-\theta_1B-\theta_2B^2-\Lambda-\theta_mB^m$, n,m,d 分别表示自回归阶数、差分阶数、移动平均阶数. 季节模型的 ARMA 形式为

$$U(B^S)\nabla_S^D X_t=V(B^S)e_t, \quad (1)$$

其中 $U(B^S)=1-u_1B^S-u_2B^{2S}-\Lambda-u_pB^{pS}$; $V(B^S)=1-v_1B^S-v_2B^{2S}-\Lambda-v_qB^{qS}$.

在随机季节模型中, 由于序列 e_t 不是独立的, 因此, 不妨假设 e_t 适合一个 ARIMA(n,d,m), 则

$$\Phi(B)\nabla^d e_t=\Theta(B)a_t, \quad (2)$$

其中 a_t 为白噪声序列, 且 $\Phi(B)=1-\varphi_1B-\varphi_2B^2-\Lambda-\varphi_nB^n$, $\Theta(B)=1-\theta_1B-\theta_2B^2-\Lambda-\theta_mB^m$. 结合(1)、(2)式有

$$\Phi(B)\bigcup(B^S)\nabla^d\nabla_S^D X_t=V(B^S)\Theta(B)a_t. \quad (3)$$

这里 $\Phi(B)\nabla^d X_t$ 仅表示同一周期内不同周期点的相关性, 而 $U(B^S)\nabla_S^D$ 则描述不同周期的同一周期点上的相关性. 模型(3)刻画了这两者因素的作用, 它是随机季节模型与 ARIAM 模型的结合式, 故称为乘积季节模型, 其阶数用 $(n,d,m)\times(p,D,q)_s$ 表示. 我们知道, 一个 n 阶自回归模型 AR(n) 的偏自相关函数(PACF)在滞后大于 n 时是截尾的. 然而, 对移动平均模型 MA(m) 而言, 其自相关函数(ACF)具有超出 m 步的截尾性.

2 桂林市旅游人数时序模型的建立

2.1 数据来源及预处理

表1数据来源于各年的《桂林经济社会统计年鉴》[2](其中, 2006年数据来源于桂林市统计局), 记 X_t' 为月旅游人数.

要建立时间序列模型, 首先要检验序列的平稳性. 为此, 描绘出表1中92个数据 X_t' 折线(如图1). 从表1和图1可以看出: 从1999年1月～2003年2月, X_t' 总体呈现平稳的上升趋势和季节波动. 但由于受 SARS 的影响, 从2003年4月开始出现大幅度下降, 直到2004年6月份才开始基本恢复到原来的发展水平. 为了消除 SARS 对 X_t' 的影响(异常值), 需对2003年4月～2004年5月旅游人数月数据进行修正(用线性插值法), 则得到了新月旅游人数序列(记为 X_t), 图2即为修正后的数据.

其次, 对修正后数据列进行模型初步识别, 即求出 X_t 的自相关系数(见表2和图3). 从表2可以看出: 前36期的自相关系数在12的整数倍上一直较大, 这说明序列 X_t 确实存在周期为12的季节性. 此外, 自相关系数除了在12的整数倍的点以外, 数值并不明显地增大, 但也没有明显地快速衰减, 这从图2也同样可以看出. 因此, 可以认为 X_t 存在明显的增长趋势, 应对 X_t 序列进行一阶自然对数逐期差分, 即 $(1-B)\ln X_t=Y_t$.

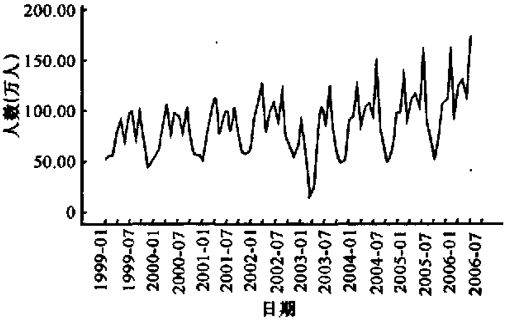

图1 1999年1月～2006年8月桂林市旅游人数折线

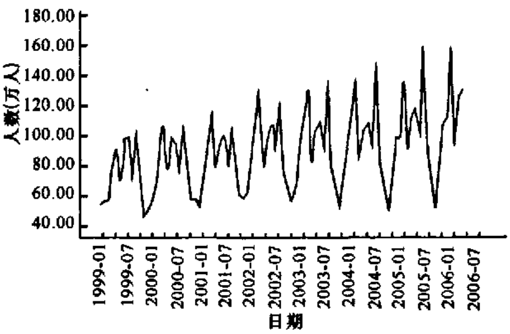

图2 修正SARS影响月旅游人数

表1 1999年1月～2006年8月桂林市月旅游人数数据

年份	人数(万人)											
	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
1999	53.58	56.29	56.81	81.41	94.05	69.49	99.94	101.33	72.18	102.09	67.52	43.84
2000	51.42	55.84	63.99	90.66	109.83	76.59	100.10	97.39	77.82	106.82	76.14	56.77
2001	57.39	51.41	82.49	99.59	116.16	79.05	94.78	102.89	80.01	107.36	77.55	60.55
2002	58.42	60.20	93.28	106.12	130.69	79.50	102.15	110.88	89.89	122.93	76.25	65.49
2003	55.40	65.08	92.69	66.05	15.14	25.61	79.72	105.68	84.23	128.58	76.26	59.85
2004	49.22	51.97	93.07	95.29	127.42	85.53	106.10	110.22	94.99	148.90	82.77	65.95
2005	48.84	63.39	100.12	99.37	139.54	91.00	113.88	120.39	102.85	160.79	90.93	73.98
2006	50.80	72.86	109.29	113.89	160.05	97.50	129.05	134.11				

表2 X_t 的自相关函数值

滞后	样本自相关函数值											
	1	2	3	4	5	6	7	8	9	10	11	12
1~12	0.399	0.223	−0.030	−0.228	−0.013	−0.113	0.008	−0.243	0.006	0.231	0.328	0.779
13~24	0.285	0.137	−0.084	−0.282	−0.043	−0.126	−0.034	−0.253	−0.041	0.156	0.239	0.608
25~36	0.214	0.083	−0.080	−0.244	−0.072	−0.137	−0.068	−0.247	−0.055	0.097	0.170	0.462

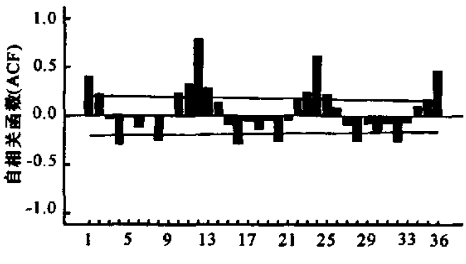

图3 X_t 的自相关函数

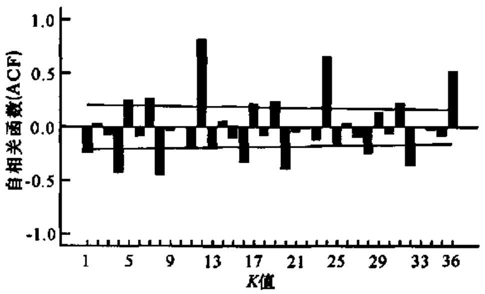

图4 Y_t 的自相关函数

现计算序列 Y_t 的自相关系数(如图 4),可以看出滞后期 $k=12n, n=1,2,3,4$ 时,自相关系数明显较大,这表明应该对序列 Y_t 进行 12 阶季节差分,即建立关系式:$(1-B^{12})Y_t=(1-B^{12})(1-B)\ln X_t=W_t$. 于是进一步分析 W_t 序列的自相关和偏自相关系数(如图5和图6). 从图5和图6可以看出,此时序列 W_t 基本趋于平稳.

2.2 模型定价

经过数据预处理后,序列增长趋势和季节性基本消除,故 $d=D=1$. 由乘积季节模型,可以选用 W_t 序列模型为 $(n,d,m)\times(p,D,q)_s$. 从图 6 可知,模型的自回归项的阶数 $n=2$ 或者 $n=4$ 是比较合适的,同样地,移动平均项的阶数 $m=1$ 合适. 此时可供选择的 (n,m) 组合有(2,1),(4,1). 由于 $k=12$ 时,样本的自相关系数和偏自相关系数显著不为 0,故 $p=q=1$. 下面检验 $(2,1,1)\times(1,1,1)_{12}$ 和 $(4,1,1)\times(1,1,1)_{12}$ 模型,为此计算它们的参数和相关性(如表 3).

从表3可知,这两个模型都满足平稳性要求,根据 AIC 和 BIC 准则,第1个模型的 AIC 和 BIC 都比第2个模型的 AIC 和 BIC 小,所以选用模型 $(2,1,1)\times(1,1,1)_{12}$ 较为合理. 于是桂林市月旅游人数的时间序列模型为:

$$(1-\varphi_1B-\varphi_2B^2)(1-u_1B^{12})(1-B)(1-B^{12})\ln X_t = (1-v_1B^{12})(1-\theta_1B)a_t. \quad (4)$$

2.3 模型(4)的参数估计

用 SPSS 软件建立模型(4)的参数值与检验结果如表4.

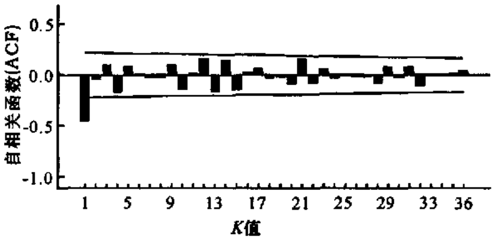

图5 W_t 的自相关函数

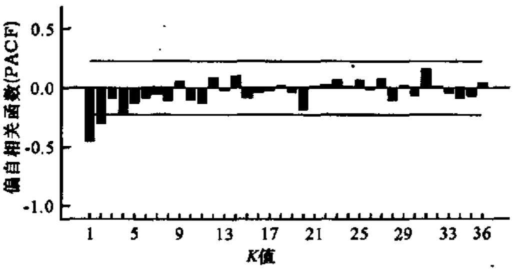

图6 W_t 的偏自相关函数

表3 各模型参数估计与检验结果

(n,m)	φ_1	φ_2	φ_3	φ_4	θ_1	u_1	v_1	R^2	AIC	BIC
(2,1)	0.118	0.045			0.874	−0.102	−0.303	0.370	−194.669	−180.452
(4,1)	0.139	0.055	0.118	−0.102	0.891	−0.176	−0.404	0.350	−192.145	−173.189

表4 模型参数估计与检验结果

变量	参数	标准误差	t-统计量
AR(1)	0.118	0.148	0.802
AR(2)	0.045	0.140	0.324
MA(1)	0.874	0.091	9.629
SAR(1)	−0.102	0.673	−0.159
SMA(1)	−0.303	0.626	−4.483

把各参数代入模型(4),即可得到本文要建的模型:

$$(1-0.118B-0.045B^2)(1+0.102B^{12})(1-B^{12})\ln X_t=(1+0.303B^{12})(1-0.874B)a_t. \quad (5)$$

2.4 适应性检验

为了验证模型(5)的残差序列 a_t 是否为白噪声序列,我们采取估计相关系数法,通过直接观察残差序列的自相关系数(如图7),发现其自相关系数都落入了随机区间,可认为与零无明显差异,这表明序列是 a_t 独立的,因此模型(5)通过了检验.

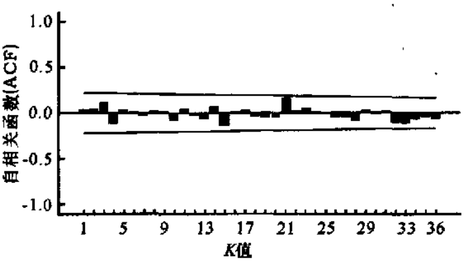

图7 残差序列 a_t 的自相关函数

2.5 模型预测

利用MATLAB软件,对原序列 X_t 进行预测并作图,通过与实际数据值对比,可以看到两者吻合较好,这表明模型(5)具有很好的拟合性(如图8).再利用模型(5)预测了2006年9月和10月的旅游人数数据分别为112.25和175.83,与最近桂林市旅游局网站(http://www.guilin.com.cn)公布的数据115.27和178.59比较,误差较小.

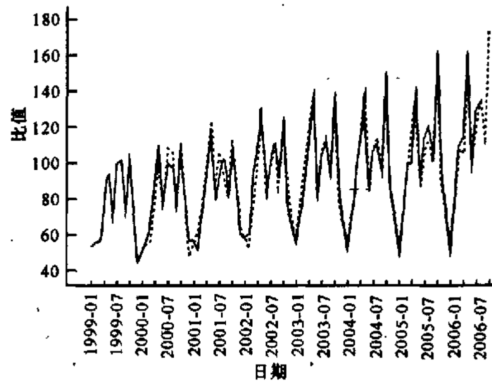

图8 X_t 的预测值与实际值的对比

3 结束语

乘积季节模型在建立月旅游人数预测模型方面是符合实际的理想模型,该模型可以为各旅游局、旅游公司、酒店等预测未来月旅游人数的基本走势和管理及市场决策提供很好的参考价值.

参考文献:

[1] 孙玉环.中国入境旅游外汇收入季节调整实证分析[J].旅游学刊,2006,21(7):29-33.

[2] 桂林市统计局.桂林经济社会统计年鉴:1999－2005[M].北京:中国统计出版社,1999-2005.

[3] 王振龙.时间序列分析[M].北京:中国统计出版社,2000.

[4] GEORGE E P BOX,GWILYM M.JENKINS,GREGORY C REINSEL.时间序列分析预测与控制[M].顾岚主,译.北京:中国统计出版社,1997.

(责任编辑:邓大玉 尹 闯)

喝咖啡结合运动可预防皮肤癌

美国新泽西州罗格斯大学(Rutgers)的研究人员研究认为,喝咖啡再结合经常做运动,能够消灭遭阳光中紫外线伤害的细胞,预防皮肤癌的发生。

研究人员对两组无毛实验鼠进行比较,其中一组老鼠喝的水里含有相当于每天一到两杯咖啡的咖啡因,它们还要踏转动轮做运动;而另外对照组的老鼠,完全不喝咖啡,或不做运动。实验结果是,只喝咖啡的老鼠,细胞凋亡增加了95%,只做运动的老鼠增加120%,而那些喝咖啡和做运动的老鼠则增加了几乎400%。该类凋亡细胞是出现异常会导致皮肤癌的细胞。

(据科学网)