

气候时间序列的季节调整*

项 静 恬

史久恩 任福民

(中国科学院应用数学研究所, 北京 100080)

(北京气象学院, 100081)

提 要

本文提供了一种对具有一年一度周期规律的气候时间序列进行季节调整的数据处理方法, 给出了该方法的基本思想、主要功能、数学工具及效果检验的评价标准, 最后还提供了相应的实例分析。

关键词: 季节调整; 对称滑动平均; 显著性检验。

一、前 言

以月份或季度为记录单位的气候要素资料, 大都呈现出一年一度的周期变化。在对长期气候规律进行研究的过程中, 这种季节性的起伏往往影响或混杂了数据蕴含的本质规律, 对深入研究和解释气候现象的实质带来一定的困难。为此, 在对气候数据进一步研究分析之前, 有必要运用一定的数学手段对之进行预处理, 以剔除上述这种一年一度的周期性(即季节性)影响, 这种处理手段, 我们称为季节调整。

季节调整方法产生于本世纪初, 是 1919 年由 W. M. Person 提出, 几十年来不断充实和发展。在经济数据的分析处理中已成为一个有成效的基本工具, 1985 年后, 由美国普查局编制的季节调整 X-11 程序传入我国, 在研究我国经济波动方面也很起作用。气候数据具有与经济数据类似的季节性特点, 将此方法借鉴来研究气候规律, 相信也是令人感兴趣的。

季节调整方法的主要手段是利用对称滑动平均为数学工具, 对数据结构进行分解和描述。方法的产生和发展一直以经验为基础而不是从理论出发, 因此在运用此方法时, 要注重掌握方法的基本思想, 并结合经验和具体数据规律来提高应用的成效。

二、气候序列的结构特点

研究气候序列的规律, 不难发现许多序列具有如下一种加法模式的结构特点:

$$x_t = T_t + S_t + E_t \quad (1)$$

式中 T_t 是 x_t 中所含的趋势性成分, 反映了序列长期性的本质规律, 如增长、下降、平稳及非季节性的周期等; S_t 是 x_t 中含有的季节性成分, 显示序列一年一度的周期变化, 即

1993 年 3 月 27 日收到, 6 月 8 日收到修改稿。

* 本工作得到北京市自然科学基金的部分资助。

序列的季节性; E_t 用来表示序列的随机成分,它是 x_t 中除去以上两部分后的剩余,其特点是围绕中心值变动,变化没有一定的规律。气象要素中不可控因素导致的波动,常是 E_t 的组成部分。

季节调整的过程就是将数据序列分解成以上各部分,然后分离季节因子对序列规律的影响。去掉季节因子的序列我们称之为调过序列,上述分离序列中季节因子的过程,狭义上也称为季节调整。

三、季节调整的方法和原理

行之有效的季节调整方法有三种:X-11 方法,X-11 ARIMA 法和贝叶斯季节调整法。

1 X-11 方法^[1]

X-11 方法是美国国势调查局于 1975 年 10 月发表的一种精细的季节调整法,对于加法模式(1)的季节调整,分以下四个步骤来实现:

- (1) 用一个估计 T_t 的函数 F_T 来计算 T_t 的估值

$$\hat{T}_t = F_T(x_t) \quad t = 1, 2, \dots, N$$

- (2) 从 x_t 中除掉 T_t , 得

$$\widehat{S_t + E_t} = x_t - \hat{T}_t \quad t = 1, 2, \dots, N$$

- (3) 用一个估计 S_t 的函数 F_S 来计算 S_t 的估值

$$\hat{S}_t = F_S(\widehat{S_t + E_t}) \quad t = 1, 2, \dots, N$$

- (4) 最后得 E_t 的估值

$$\hat{E}_t = \widehat{S_t + E_t} - \hat{S}_t \quad t = 1, 2, \dots, N$$

函数 F_T 和 F_S 的数学形式是滑动平均,主要是对称滑动平均。有关内容将在下文仔细讨论。X-11 方法是应用最广泛的一种季节调整法。

2 X-11 ARIMA 法

X-11 法虽因它的精细性受到广泛应用,但也存在一些由于移动平均法而导致的缺点,主要表现在以下几个方面:

- (1) 移动平均法会造成序列两端产生欠值项,项数取得越多,两端欠值项就越多。
- (2) 移动平均项数取得不同,会得到不同的季节调整结果。
- (3) 为了明确分离数据结构,希望移动平均项数多取一些,但这又导致数据过分光滑化,或可能产生原序列没有的变动。

为了克服上述缺点并解决欠值项的补缺,将 X-11 与 ARIMA 模型结合,产生了 X-11 ARIMA 方法,具体步骤为

- (1) 对序列 $\{x_t\}$ 模拟 ARIMA 模型^[2]。
- (2) 用 ARIMA 模型将 $\{x_t\}$ 后端延长一年(即 12 个月)。

(3) 对延长了的序列用 $X-11$ 法作季节调整。

3. 贝叶斯季节调整法

贝叶斯季节调整法是日本统计学家 H. Akaike 于 1980 年提出的。目的在于分离序列中趋势项和季节项。记 $\beta = \{T_N, T_{N-1}, \dots, T_1; S_N, S_{N-1}, \dots, S_1\}$, 在满足若干先验条件的前提下, 贝叶斯调整的基本方法就是实现对于下式的极小化, 并由此获得 β 的估计值:

$$L(\beta) = \sum_{i=1}^N \left\{ E_i^2 + d^2 \left[(s^2 \nabla^m T_i)^2 + \nabla_p S_i + Z^2 \sum_{j=0}^{p-1} S_{i-j} \right] \right\}$$

式中 d, s, z 为超参数, 可选为适当常数, ∇, ∇_p 为差分和 p 阶季节差分算子, m, n 是适当的差分和季节差分方次。

贝叶斯调整效果胜于 $X-11$ 法, 但计算工具较复杂, 计算量也较大。

四、季节调整的数学工具——对称滑动平均

对称滑动平均是用来估计趋势 T_t 和季节因子 S_t 的有效工具。

设序列为 $\{x_t\}$, $t = 1, 2, \dots, N$, 我们定义

$$Y_{\frac{t-1}{s}+t} = \frac{1}{s} \sum_{i=0}^{s-1} X_{t+i}, \quad t = 1, 2, \dots, N-s+1 \quad (2)$$

为跨度 s 的对称滑动平均, 记为 MA_s 。由此可将 Y_t 与 x_t 之间用如下关系式联结:

$$Y_{t+\frac{s-1}{2}} = MA_s(x_t) \quad (3)$$

注意等式右边实际表示了有 $x_t, x_{t+1}, \dots, x_{t+s-1}$ 共 s 个数参加运算。由于输出值 Y 的下标 $t + (s-1)/2$ 正好为参加运算的 s 个时刻的中心, 故称该运算为对称滑动平均或中心滑动平均。当 s 为奇数时, 中心值为整数; s 为偶数时, 中心值非整。将这两种情形分别记为 MA_{2m+1} 和 MA_{2m} , 显然 MA_{2m+1} 比 MA_{2m} 在实际应用中更方便, 意义也更明显, 为此有对之专门讨论的必要。

为了叙述方便, 我们将原序列 $\{x_t\}$ 称为输入, 其对应的滑动平均称为输出, 不难看出, MA_{2m+1} 具有如下三个性质:

(1) MA_{2m+1} 对直线序列“透明”, 即若输入 $\{x_t\}$, $t = 1, 2, \dots, N$ 满足 $x_t = a + bt$, 则输出 $\{y_t\}$ 的对应时刻值与 x_t 相等, 即

$$y_{m+t} = MA_{2m+1}(x_t), \quad t = 1, 2, \dots, N-2m$$

(2) 若输入是一个周期为 $2m+1$ 的序列, 满足

$$\frac{1}{2m+1} \sum_{i=0}^{2m} x_{t+i} = \mu \quad (\text{常数}), \quad t = 1, 2, \dots, N-2m,$$

则输出是一个常数序列

$$y_{m+t} = MA_{2m+1}(x_t) = \mu, \quad t = 1, 2, \dots, N-2m$$

(3) MA_{2m+1} 对白噪声序列有一定的消除作用, 即若 $\{x_t\}$ 为独立同分布序列, 方差

$$\text{var} x_t = \sigma^2, \text{ 则 } \text{var} y_t = \frac{1}{2m+1} \sigma^2.$$

以上三个性质均可根据(2)式进行直接验证。由这三个性质可见, 对称滑动平均可作为消除季节因素的有效工具。现举例形象地加以说明:

例1 设 $x_t = T_t + S_t, t = 1, 2, \dots, 15$

$$T_t = 1 + 0.5t,$$

$$S_t = \begin{cases} 0 & t = 5r \\ 4 & t = 5r + 1 \\ -0.5 & t = 5r + 2 \\ -1.5 & t = 5r + 3 \\ -2 & t = 5r + 4 \end{cases}$$

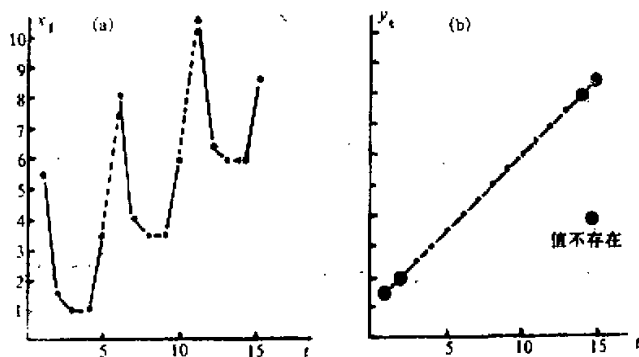


图1 滑动平均前(a)后(b)的序列形状

式中 r 为非负整数, 显然 S_t 是一个周期为 5 的序列。对 x_t 作 5 阶滑动平均, 则

$$\begin{aligned} y_{t+1} &= MA_5(x_t) = MA_5(T_t) + MA_5(S_t) \\ &= [1 + 0.5(t+2)] + \frac{1}{5} [0 + 4 + (-0.5) + (-1.5) + (-2)] \\ &= 2 + 0.5t, \quad t = 1, 2, \dots, 11. \end{aligned}$$

图 1a, b 分别表示滑动平均前后的序列形状。

由以上讨论可见, 用 MA_{2m+1} 可消除序列奇数周期, 然而气候序列周期常为偶数, 为克服这一困难, 可采用如下的高阶对称滑动平均:

$$\begin{aligned} MA_{s_1}(MA_{s_2}(x_t)) &= MA_{s_1}(y_{\frac{s_1-1}{2}+t}) = Z_{\frac{s_1-1}{2}+\frac{s_2-1}{2}+t}, \\ & \quad t = 1, 2, \dots, N - s_1 - s_2 + 2. \end{aligned} \quad (4)$$

此运算的输出量 Z_* 即为关于原序列 $\{x_t\}$ 的二阶对称滑动平均, 记为 $MA_{s_1 \times s_2}$ 。类似可定义 3 阶、4 阶对称滑动平均。(4)式中 Z_* 的中心时刻为

$$\frac{1}{2}(s_2 - 1) + \frac{1}{2}(s_1 - 1) + t = \frac{1}{2}(s_2 + s_1) + t - 1, \text{ 因此, 当 } s_1 \text{ 和 } s_2 \text{ 同为奇数}$$

或偶数时,二阶滑动平均中心时刻可取整数,且参加滑动平均的 x_i 为奇数项。因此可借用二阶滑动平均来解决消除偶周期季节波动的问题。显然,若 $\{x_i\}$ 周期为 $2m$, 取 $s_2 = 2n$, $s_1 = 2m$, 则

$$Z_{\frac{i_2-1}{2}+\frac{i_1-1}{2}+i} = MA_{i_2}(MA_{i_1}(x_i)) = MA_{i_2}(\mu) = \mu.$$

$$i = 1, 2, \dots, N - s_1 - s_2 + 2$$

$$\text{式中 } \mu = \frac{1}{s_1} \sum_{i=0}^{2m-1} x_i.$$

五、气候序列季节调整的工具

本节将提供三种最常用的对称滑动平均工具。

1. $MA_{2 \times 12}$

由定义知, $Z_{t+i} = \frac{1}{24} \left(\sum_{i=0}^{11} x_{t+i} + \sum_{i=1}^{12} x_{t+i} \right) = MA_{2 \times 12}(x_t)$, 用 $MA_{2 \times 12}$ 可消除月资料中的季节因子, 得到趋势项的估计值。从以下例子可看到其对于季节因子的消除效果:

例2 $x_t = T_t + S_t + E_t$,

$$T_t = 95 + 5t, S_t = 50 \sin \frac{\pi t}{6}, E_t \text{ 是 } [-10, 10] \text{ 上均匀分布的随机数, } t = 1, 2, \dots,$$

24.

表1列出了原序列及其分解出的各成分值, 也列出了各列经 $MA_{2 \times 12}$ 作用后的输出结果(表中关于 T_t 和 E_t 经 $MA_{2 \times 12}$ 作用后的端部丢失值均用零值代表, 今后再作专门讨论)。从滑动平均结果来看, 对于本例给出的理想情况, 效果是相当不错的。随机误差的范围经过滑动平均, 由 $[-10, 10]$ 缩小到 $[-1, 1]$, 因而使得对趋势的估计受随机因子影响的程度大为减小。

2. $MA_{3 \times 3}$, $MA_{3 \times 5}$

$MA_{3 \times 3}$ 用来估计序列的季节因子, 其计算公式为

$$Z_{2+i} = MA_3(MA_3(x_t)) = \frac{1}{3} (y_{i+i} + y_{2+i} + y_{3+i})$$

$$= \frac{1}{9} (x_t + 2x_{t+1} + 3x_{t+2} + 2x_{t+3} + x_{t+4})$$

我们要从序列 $x_t = S_t + E_t$ (趋势项 T_t 已由前法剔除) 中估计各月的季节因子。以估计1月份季节因子为例, 考虑每年1月的数据

$$x_{12i+1} = S_{12i+1} + E_{12i+1}$$

用 x'_t, S'_t, E'_t 代表原序列中1月份相应数据, 并做如下滑动平均:

表 1 $MA_{2 \times 12}$ 的模拟效果 ($x_t = T_t + S_t + E_t$)

t	x_t	T_t	S_t	E_t	$MA_{2 \times 12}(x_t)$	$MA_{2 \times 12}(T_t)$	$MA_{2 \times 12}(S_t)$	$MA_{2 \times 12}(E_t)$
1	128.48	100.00	25.00	3.48	0.00	0.00	0.00	0.00
2	121.30	105.00	43.00	3.00	0.00	0.00	0.00	0.00
3	159.13	110.00	50.00	-0.87	0.00	0.00	0.00	0.00
4	167.93	115.00	43.00	9.63	0.00	0.00	0.00	0.00
5	147.73	120.00	25.00	2.73	0.00	0.00	0.00	0.00
6	117.60	125.00	0.00	-7.40	0.00	0.00	0.00	0.00
7	102.49	130.00	-25.00	-2.51	130.06	130.00	0.00	0.06
8	93.71	135.00	-43.00	2.01	134.99	134.00	0.00	-0.01
9	83.11	140.00	-50.00	-6.89	140.41	140.00	0.00	0.41
10	110.97	145.00	-43.00	9.27	145.20	145.00	0.00	0.20
11	120.12	150.00	-25.00	-4.88	149.38	150.00	0.00	-0.62
12	152.07	155.00	0.00	-2.93	154.57	155.00	0.00	-0.43
13	180.70	160.00	25.00	-4.30	160.02	160.00	0.00	0.02
14	217.40	165.00	43.00	9.10	164.71	165.00	0.00	-0.29
15	223.05	170.00	50.00	3.05	170.00	170.00	0.00	0.00
16	219.04	175.00	43.30	0.74	175.14	175.00	0.00	0.14
17	196.73	180.00	25.00	-8.27	179.89	179.00	0.00	-0.11
18	193.34	185.00	0.00	8.34	185.00	185.00	0.00	0.00
19	157.50	190.00	-25.00	-7.50	0.00	0.00	0.00	0.00
20	151.31	195.00	-43.00	-0.39	0.00	0.00	0.00	0.00
21	152.41	200.00	-50.00	2.41	0.00	0.00	0.00	0.00
22	165.13	205.00	-43.00	3.43	0.00	0.00	0.00	0.00
23	179.87	210.00	-25.00	-5.13	0.00	0.00	0.00	0.00
24	214.98	215.00	0.00	-0.03	0.00	0.00	0.00	0.00

$$MA_{2 \times 12}(S'_t + E'_t) = MA_{2 \times 12}(S'_t) + MA_{2 \times 12}(E'_t)$$

理想状况下, $MA_{2 \times 12}(E'_t)$ 较小, 而 S'_t 因同为 1 月数据变化不大, 故输出量 $MA_{2 \times 12}(S'_t + E'_t)$ 即可作为 S'_t 的估计。

之所以用 $MA_{2 \times 12}$ 而不用 MA_{12} , 是因为前者比后者更能减少随机因子的影响, 也更精确地估计 S'_t 。

上述关于 $MA_{2 \times 12}$ 的讨论和结果可以平行地用于 $MA_{4 \times 12}$, 且 $MA_{4 \times 12}$ 对随机因子有更强的剔除功能。

3. Henderson 滑动平均 MA_H

MA_H 用于从序列 $x_t = T_t + E_t$ 中估计出趋势项 T_t , 其形式为奇数项滑动平均

$\sum_{i=-m}^m w_i x_{t+i}$, 根据序列中随机波动程度的不同, 可从 MA_{H9} , MA_{H11} 和 MA_{H21} 中选择合

适者作为 T_t 的估计工具。这三种 MA_H 分别对应于 $m = 4, 6, 11$ 时的 Henderson 滑动平均, 表 2 中列出了与之相应的滑动平均系数。

该方法中系数 w_i 是通过将 $Z_t = \sum_{i=-m}^m w_i x_{t+i}$ 作三次差分后, 对方差 $\text{var}(\nabla^3 Z_t)$ 极

表 2 Herderson 滑动平均系数 w_i

i	0	± 1	± 2	± 3	± 4	± 5
MA_{H9}	0.330	0.267	0.119	-0.010	-0.041	
MA_{H11}	0.240	0.214	0.147	0.066	0.000	-0.028
MA_{H13}	0.148	0.138	0.122	0.097	0.068	0.039
i	± 6	± 7	± 8	± 9	± 10	± 11
MA_{H11}						
MA_{H13}	-0.019					
MA_{H15}	0.013	-0.005	-0.015	-0.016	-0.011	-0.004

小化而求得,详细内容见文献[3].

选择 MA_H 估计趋势项的步骤如下 ($x_i = T_i + E_i$):

(1) 用 MA_{H11} 对 T_i 作初步估计,得 $\hat{T}_i = MA_{H11}(x_i)$,

(2) 估计随机项 E_i : $\hat{E}_i = x_i - \hat{T}_i$,

(3) 计算 $\bar{E}/\bar{T}$, $\bar{E}$ 和 $\bar{T}$ 分别表示 E_i 和 T_i 月与月(或季与季)间平均绝对变化率。根据比率大小选择不同的 MA_H 来估计 T_i 和 E_i 。选择规则列于表 3 (计算时不参考序列两端的丢失值)。

表 3 MA_H 的选择标准

$\bar{E}/\bar{T}$	0.00—0.99	1.00—3.49	4.00 以上
选择的 MA_H	MA_{H1}	MA_{H13}	MA_{H23}

六、丢失值的估计和非对称滑动平均

从对称滑动平均的计算公式(2)可见,当 $s \geq 2$ 时, $MA_s(x_i)$ 在时刻 $1 \sim \left[\frac{s}{2}\right]$ 和 $\left[N - \frac{s}{2} + 1\right] \sim N$ 上无对应值。这种现象称为滑动平均的端值丢失。被丢失值有些是我们很感兴趣的(如对近期趋势的估计),因此也必须对它们进行估计。

1. 趋势丢失值的估计

用对称的 MA_H 估计出趋势项 T_i 后,对于丢失的端值,用表 5 中相同 m 值的对应分表给出非对称滑动平均估值。

例如,若非端值有 $\hat{T}_i = MA_{H5}(x_i)$,则查表 5A 可得端值估计为

$$\hat{T}_N = -0.156x_{N-4} - 0.034x_{N-3} + 0.185x_{N-2} + 0.424x_{N-1} + 0.581x_N.$$

2. 季节因子丢失值的估计

对季节因子丢失值需分两种情形估计:

(1) 序列仅含季节项和随机项, 即 $x_t = S_t + E_t$, 则季节因子 S_t 的估计值用 $\hat{S}_t = MA(x_t)$ 来求得, 丢失值的估计方法与趋势项类同. 具体系数查表 4. 例如 $\hat{S}_t = MA_t(x_t)$, 则由表 4A 得: $\hat{S}_N = 0.39x_{N-1} + 0.61x_N$.

(2) 序列含趋势项, 即 $x_t = T_t + S_t + E_t$. 季节因子和随机因子由 $\widehat{S_t + E_t} = x_t - MA(x_t)$ 来估计, 则 $\widehat{S_t + E_t}$ 的丢失值由 $MA(x_t)$ 造成, 估计步骤如下:

1) 对 S_t 作估计: $\hat{S}_t = MA(\widehat{S_t + E_t})$,

2) 查表 5, 估计 1) 中相应的 $MA(\widehat{S_t + E_t})$ 的丢失值.

3) 用外推一年的办法估计由 $MA(x_t)$ 造成的丢失值, 即

若 S_i 是第一年某月(季)的丢失值, 则令

$$S_i = S_{i+12} \text{ (或 } S_{i+4})$$

若 S_j 是最后一年某月(季)的丢失值, 则令

$$S_j = S_{j-12} \text{ (或 } S_{j-4})$$

表 4 季节因子滑动平均权重系数表

(表中 N 表示序列的最后一年, $N+1$ 所对应的权重数值表示预报值)

A. MA_1

对应年份	加权年份 N-2	N-1	N
N+1	-0.167	0.419	0.749
N	0	0.390	0.610
N-1	0.333	0.333	0.333

B. $MA_{3 \times 3}$

对应年份	N-4	加权年份 N-3	N-2	N-1	N
N+1	0	-0.056	0.148	0.426	0.481
N	0	0	0.185	0.407	0.407
N-1	0	0.111	0.259	0.370	0.259
N-2	0.111	0.222	0.333	0.222	0.222

C. $MA_{3 \times 5}$

对应年份	N-6	加权年份 N-5	N-4	N-3	N-2	N-1	N
N+1	0	0	-0.034	0.134	0.300	0.300	0.300
N	0	0	0	0.150	0.283	0.283	0.283
N-1	0	0	0.067	0.183	0.250	0.250	0.250
N-2	0	0.067	0.133	0.217	0.217	0.217	0.150
N-3	0.067	0.133	0.200	0.200	0.200	0.133	0.067

$C_1 MA_{H11}$

对应月份	加权月份 N-22	N-21	N-20	N-19	N-18	N-17	N-16	N-15	N-14	N-13	N-12	N-11
N	0	0	0	0	0	0	0	0	0	0	0	-0.077
N-1	0	0	0	0	0	0	0	0	0	0	-0.046	-0.041
N-2	0	0	0	0	0	0	0	0	0	-0.022	-0.025	-0.025
N-3	0	0	0	0	0	0	0	0	-0.008	-0.014	-0.018	-0.015
N-4	0	0	0	0	0	0	0	-0.001	-0.008	-0.013	-0.012	-0.003
N-5	0	0	0	0	0	0	0.003	-0.006	-0.011	-0.011	-0.002	0.015
N-6	0	0	0	0	0	0.002	-0.006	-0.012	-0.011	-0.003	0.015	0.039
N-7	0	0	0	0	0.001	-0.007	-0.013	-0.011	-0.003	0.015	0.039	0.069
N-8	0	0	0	-0.002	-0.007	-0.013	-0.013	-0.003	0.014	0.039	0.068	0.097
N-9	0	0	-0.003	-0.010	-0.015	-0.014	-0.005	0.014	0.040	0.060	0.097	0.122
N-10	0	-0.004	-0.011	-0.016	-0.015	-0.005	0.013	0.039	0.068	0.097	0.122	0.138
N-11	-0.004	-0.011	-0.016	-0.015	-0.005	-0.013	0.039	0.068	0.097	0.122	0.138	0.148

 $C_1 MA_{H11}$

对应月份	加权月份 N-10	N-9	N-8	N-7	N-6	N-5	N-4	N-3	N-2	N-1	N
N	-0.064	-0.049	-0.028	0.002	0.039	0.084	0.133	0.182	0.227	0.263	0.288
N-1	-0.035	-0.024	-0.004	0.025	0.061	0.101	0.141	0.176	0.203	0.219	0.224
N-2	-0.019	-0.005	0.018	0.049	0.082	0.116	0.146	0.166	0.177	0.176	0.166
N-3	-0.004	0.015	0.042	0.073	0.103	0.129	0.147	0.154	0.150	0.134	0.112
N-4	0.015	0.040	0.068	0.098	0.121	0.137	0.142	0.136	0.119	0.195	0.066
N-5	0.039	0.067	0.095	0.119	0.134	0.139	0.131	0.114	0.088	0.059	0.027
N-6	0.068	0.096	0.118	0.134	0.138	0.132	0.114	0.089	0.059	0.027	0.001
N-7	0.096	0.120	0.135	0.140	0.133	0.116	0.090	0.060	0.031	0.005	-0.015
N-8	0.120	0.137	0.140	0.136	0.118	0.094	0.064	0.034	0.008	-0.010	-0.021
N-9	0.138	0.143	0.166	0.120	0.095	0.067	0.037	0.011	-0.007	-0.017	-0.019
N-10	0.144	0.138	0.122	0.097	0.068	0.039	0.013	-0.005	-0.015	-0.016	-0.011
N-11	0.138	0.122	0.097	0.068	0.039	0.013	-0.005	-0.015	-0.016	-0.011	-0.004

七、季节调整效果的检验和评价

季节调整效果的好坏,与数据本身的结构及选用的季节调整工具都有很大关系,因此有必要提供一套评价标准和检验方法,以便提高季节调整的效果。

1. 季节因子的显著性检验

我们希望通过季节调整能得到一个较为稳定的季节因子序列 $\{S_t\}$ 。且不同月份的季节因子有较显著的差异,为达到这一目的,可借用单因素方差分析,以检验各月季节因子是否存在显著差别,显著性水平建议选 1%。

2. 不同跨度的百分变化率

设 $\{x_t\}$, $t = 1, 2, \dots, N$ 为气候序列,定义

$$P(K) = \frac{1}{N-K} \sum_{i=1}^{N-K} \frac{x_{i+K} - x_i}{x_i} \times 100\% \quad (5)$$

为跨度为 K 的平均百分变化率

$$P_+(K) = \frac{1}{N-K} \sum_{i=1}^{N-K} \left| \frac{x_{i+K} - x_i}{x_i} \right| \times 100\% \quad (6)$$

为跨度为 K 的平均绝对百分变化率, K 的取值为 1, 2, ..., 12。

用此量检验季节调整效果时有以下几点参考评价标准(以下用 x, E, S, T 分别代表原序列、随机项、周期项、趋势项,下同):

(1) $P_{+s}(K) \geq P_{+(T+E)}(K)$, $K = 1, 2, \dots, 12$;

(2) $P_{+E}(K)$ 对不同 K 值应无显著性差异。原因是随机项 E_t 应是零均值等方差的独立随机量;

(3) $P_{+s}(12) = \min P_{+s}(K)$ (季数据等式左边括号内为 4)。此式成立的原因是当跨度取 12 (季数据取 4) 时,相同月份(季)的 S_t 值应具有相对稳定性。变化应显著地小于 K 取其它值时的 $P_{+s}(K)$;

(4) 若季节调整所得的 $\{T_t\}$ 具有明显趋势性,则 $P_{+T}(K)$ 随 K 递增。

3. 趋势显著期 (MCD)

为了描述趋势 T 的显著性及受随机因子 E 的影响程度,定义

$$MCD = \min_{1 \leq K \leq 12} \left\{ K: \frac{P_{+E}(K)}{P_{+T}(K)} < 1 \right\} \quad (7)$$

为趋势显著期 (Month for Cyclical Dominance)。例如,当 $MCD = 3$ 时,则可以在平均的意义上认为,跨度大于或等于 3 时,就可以在调过序列 $T + E$ 中看到明显的趋势性。

本检验方法的参考评价标准为: 如果 $MCD < 6$, 则认为趋势明显, 否则认为不明显。

4. 相对贡献率

为叙述方便起见, 记

$$\bar{x}_t = P_{+s}(t), t = 1, 2, \dots, 12 \quad (8)$$

设序列结构为 $x = T + S + E$, 且各部分相互独立, 则可近似地有

$$\bar{x}_t^2 \approx \bar{T}_t^2 + \bar{S}_t^2 + \bar{E}_t^2$$

令

$$(\bar{x}_t')^2 = \bar{T}_t^2 + \bar{S}_t^2 + \bar{E}_t^2 \quad (9)$$

则

$$\bar{T}_t^2/(\bar{x}_t')^2, \bar{S}_t^2/(\bar{x}_t')^2, \bar{E}_t^2/(\bar{x}_t')^2 \quad (10)$$

可用来描述跨度为 t 时各部分变化的相对大小, 我们称(10)式的量为相对贡献率 (Relative Contribution) $(\bar{x}_t')^2/(\bar{x}_t)^2$ 可用来描述近似的程度。一个具有明显季节性的序列, S 的相对贡献率 $\bar{S}_t^2/(\bar{x}_t')^2$ 在 $t = 1$ 时会相当大, 并随 t 的增加而减小, 而 $\bar{T}_t^2/(\bar{S}_t)^2$ 则随 t 的增加而增加。 $\bar{E}_t^2/(\bar{x}_t')^2$ 一般没有较大变化。

本检验的参考评价标准为

(1) 若 $\bar{S}_1^2/(\bar{x}_1')^2 < 50\%$, 则认为季节性不明显。

(2) 若 $\bar{E}_t^2/(\bar{x}_t')^2 > 30\%$, 则认为随机部分影响过大, 季节调整无效。

(3) 若 $\bar{E}_t^2/(\bar{x}_t')^2 > 14\%$, 对 $t = 1, 2, \dots, 12$ 全成立, 则也认为季节因子影响过大, 季节调整无效。

5. 平均游程 (ADR)

平均游程 (Average Duration of Run) 按如下方法计算: 设序列 x_t 按时间顺序排列可划分为 I 个单调段 (单调上升和单调下降均为单调段), 且第 i 段中含有样本个数为 t_i , $i = 1, 2, \dots, I$ 。则该序列 $\{x_t\}$ 的 ADR 值为:

$$ADR = \frac{1}{I} \left[\sum_{i=1}^I t_i + (t_i - 1) \right]$$

例: $\{x_t\}$ 为 1.0, 2.1, 2.5, 2.3, 3.0, 3.0, 3.2, 3.0, 3.0,

$$\text{则 } ADR = \frac{1}{4} [(3-1) + 1 + 3 + 2] = 2$$

由上所示, ADR 表征一个序列中所有单调段的平均持续长度, ADR 越大, 表明序列愈平滑, 反之则变化愈剧烈, 但 ADR 仅反映序列变化频度, 并不反映变化幅度。一个无穷长的白噪声序列, 其 ADR 值为 1.5; 一个长度大于 12 的白噪声序列若其分布趋于正态, 则在 5% 的显著水平下其 ADR 有置信区间 [1.36, 1.75], 如果序列 ADR 值在此区间外, 则表明序列有一定相关性。

本检验的参考评价标准是: 若 $ADR(E)$ 不在 [1.36, 1.75] 内, 则认为 $\{E_t\}$ 有相关性, 是否认为季节调整有效, 视 $\bar{E}_t^2/(\bar{x}_t')^2$ 而定。

八、气候序列分析实例

1. 计算结果

图 2 给出了北京从 1960 年 1 月至 1989 年 12 月的月降水量序列曲线图。采用 X-11 方法的加法模式进行季节调整, 计算结果如下:

(1) 计算趋势 T_t 的估值 $\hat{T}_t$, 如图 3 所示, 从图中可以看到有非季节性变化的长期趋势成分。

(2) 从原序列 x_t 中剔除 $\hat{T}_t$, 得剩余部分为 $\widehat{S_t + E_t} = x_t - \hat{T}_t, t = 1, 2, \dots, N$. 计算出季节性成分 S_t 的估值 $\hat{S}_t$, 如图 4 所示。从图可见, $\hat{S}_t$ 是一非常有规律的季节性变化曲线。

(3) 随机成分 E_t 的估值 $\hat{E}_t$ 是 x_t 中除去以上两部分后的剩余部分, 即

$$\hat{E}_t = \widehat{S_t + E_t} - \hat{S}_t = x_t - \hat{T}_t - \hat{S}_t, t = 1, 2, \dots, N \text{ (图 5)}.$$

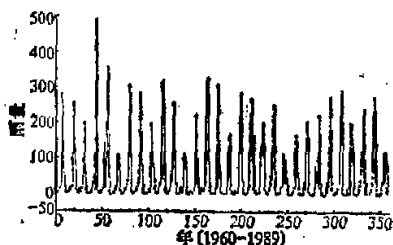


图 2 北京月降水量变化曲线
(1960 年 1 月—1989 年 12 月)

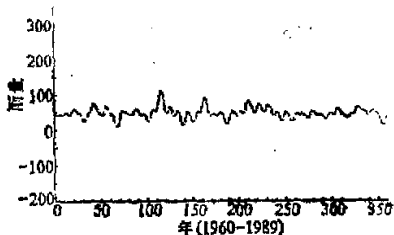


图 3 北京月降水量的趋势
估值 $\hat{T}_t$

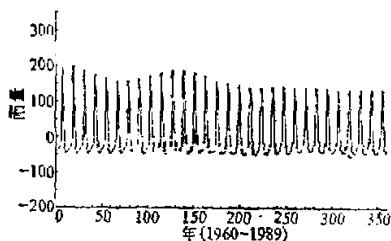


图 4 北京月降水量的季节性
变化部分 $\hat{S}_t$

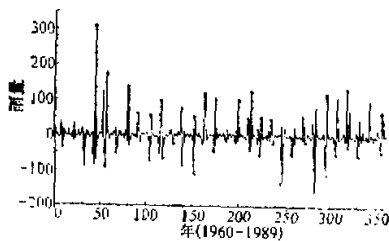


图 5 北京月降水量的随机性
变化部分 $\hat{E}_t$

2. 分析与讨论

(1) 季节调整的 X-11 方法在经济数据分析处理中十分有效。气候序列具有与经济序列类似的季节性特点, 从上面实例的数据计算结果来看, 基本上反映了气候变化规律, 说明此方法可以在气象中应用。

(2) 通过上述例子可见 X-11 方法主要是对数据主体规律的提炼和加工, 可以作为在对气候数据进行分析研究之前的预处理, 并可与其他方法结合起来, 分别深入研究气候序列的趋势影响、循环和周期影响、季节性和随机影响等的变化规律和预测问题。

(3) 本项工作是为了介绍一种新的方法, 所以借用了经济序列处理程序 X-11, 考虑到气候序列演变的特殊性, 不论在方法上或计算机软件方面都需要作进一步的探索。

参 考 文 献

- [1] 郭瑞, 项静恬, 1993, 经济时间序列的 X-11 季节调整方法, 数理统计与管理, 12, 第1—3期。
- [2] 项静恬、史久恩等, 1991, 动态和静态数据处理, 气象出版社。
- [3] Kenny, P. B. and Dubbin, J., 1982, Local Trend Estimation and Seasonal Adjustment of Economic and Social Time Series, J. Amer. Statistics Assoc., 77.

The Seasonal Adjustment Method of Climatic Time Series

Xiang Jingtian

(Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing 100080)

Shi Jiuen Ren Fumin

(Beijing Meteorological College, Beijing 100081)

Abstract

This paper aims to introduce a new method of dynamic data processing of climatic time series with a seasonal pattern. It cludes basic idea, main functions, mathematical procedures and significance testing for the method. Finally, we give a real example of analyzing the monthly rainfall in Beijing.

Key words: Seasonal adjustment; Symmetry moving average; Significance test.