中级应用统计学

时间序列分析

张晨峰

华东理工大学商学院

2015年10月30日

主要内容

- 简介
- 时间序列的经济指标
- 经济时间序列的分解、平滑和季节调整
- 平稳ARMA过程

7.1 简介

时间序列

一个时间序列是依时间顺序生成的观测值的集合。时间序列可表示为如下形式：

$$\{y_1, y_2, ..., y_T\}$$

其中，下标t表示时间。

时间序列分类

- 针对时间点的变量（存量）
- 针对时间段的变量（流量）

7.1 简介

时间序列分析

- 时间序列分析是一种动态数据处理的统计方法。该方法基于随机过程理论和数理统计学方法，研究随机数据序列所遵从的统计规律，以用于解决实际问题。
- 它包括一般统计分析(如自相关分析，谱分析等),统计模型的建立与推断，以及关于时间序列的最优预测、控制与滤波等内容。
- 时间序列分析实际上是对离散指标的随机过程的统计分析，所以又可看作是随机过程统计分析的一个组成部分。

随机变量的无条件均值

随机变量Y_t的无条件均值为:

$$E(Y_t) \equiv \int_{-\infty}^{\infty} y_t f_{Y_t}(y_t) dy_t$$

随机变量的方差

随机变量Y_t的方差为:

$$\gamma_{0t} \equiv E(Y_t - \mu_t)^2$$

7.1 简介

自协方差

Y_t的第j个自协方差为：

$$\gamma_{jt} \equiv E(Y_t - \mu_t)(Y_{t-j} - \mu_{t-j})$$

自相关系数

$$\rho_j \equiv \frac{\gamma_j}{\gamma_0}$$

平稳性

如果随机过程Y_t的均值μ_t和自协方差γ_{jt}都不取决于t，则称随机过程Y_t是协方差平稳的或弱平稳的：

$$E(Y_t) = \mu$$

$$E(Y_t - \mu)(Y_{t-j} - \mu) = \gamma_j$$

7.2 时间序列的经济指标

发展速度

发展速度是以相除方法计算的动态比较指标，其计算公式为：**发展速度=某指标报告期数值/该指标基期数值**。发展速度一般用百分数表示，当比例数较大时，则用倍数表示较为合适。例：某地固定资产投资1994年为366亿元，1993年为328亿元，1994年与1993年比，$366 \div 328 = 1.12$，这就是发展速度，用百分数表示为112%，用倍数表示则是1.12倍。

增长速度

增长速度也称为增长率，是以相减和相除结合计算的动态比较指标，其计算公式为：**增长速度=(某指标报告期数值-该指标基期数值)/该指标基期数值**。如上例中1994年比1993年的增长速度为:$(366 - 328) \div 328 = 0.12$，用百分数表示则为12%。由上可知：增长速度=发展速度-1(或100%)。

指数

表示价格、数量或价值相对于一个基础时期的变化数。

简单指数

$$P = \frac{p_t}{p_0}$$

表 18—1　　3 型本森自动订书机的价格用 3 中不同基期构造的指数

年份	订书机的价格(美元)	价格指数(1990 年=100)	价格指数(1990—1991 年=100)	价格指数(1990—1992 年=100)
1985 年	18	90.0	$\frac{18}{21}\times100=85.7$	$\frac{18}{21.67}\times100=83.1$
1990 年	20	100.0	$\frac{20}{21}\times100=95.2$	$\frac{20}{21.67}\times100=92.3$
1991 年	22	110.0	$\frac{22}{21}\times100=104.8$	$\frac{22}{21.67}\times100=101.5$
1992 年	23	115.0	$\frac{23}{21}\times100=109.5$	$\frac{23}{21.67}\times100=106.1$
2001 年	38	190.0	$\frac{38}{21}\times100=181.0$	$\frac{38}{21.67}\times100=175.4$

未加权指数

价格指数的简单平均：$P = \frac{\sum P_i}{n}$

表 18—2　　2001 年食品价格指数的计算（1995 年=100）

项目	1995 年的价格（美元）	2001 年的价格（美元）	简单指数
1 磅纯面包	0.77	0.89	115.6
1 打鸡蛋	1.85	1.84	99.5
1 加仑纯牛奶	0.88	1.01	114.8
1 磅红金香苹果	1.46	1.56	106.8
12 盎司浓缩橙汁	1.58	1.70	107.6
1 磅 100%碾磨焙炒咖啡	4.40	4.62	105.0
总计	10.94	11.62	

加权指数——拉氏价格指数

$$P = \frac{\sum p_t q_0}{\sum p_0 q_0} \times 100$$

加权指数——帕氏价格指数

$$P = \frac{\sum p_t q_t}{\sum p_0 q_t} \times 100$$

加权指数——费舍尔理想指数

拉式指数易于过分加重价格上涨的商品的价值，而帕氏指数又易于过于加重价格下跌的商品的价值，因此提出了费舍尔理想指数。

表 18—3 食品价格的拉氏指数和帕氏指数的计算（1995 年=100）

项目	1995 年价格（美元）	1995 年数量	2001 年价格（美元）	2001 年数量
1 磅纯面包	0.77	50	0.89	55
1 打鸡蛋	1.85	26	1.84	20
1 加仑纯牛奶	0.88	102	1.01	130
1 磅红金香苹果	1.46	30	1.56	40
12 盎司浓缩橙汁	1.58	40	1.70	41
1 磅 100%碾磨焙烘咖啡	4.40	12	4.62	12

消费价格指数（CPI）

它度量了一个时期到另一个时期一篮子商品和服务的市场价格的变化。目前，在中国国家统计局制定的调查方案中，CPI的调查内容按用途划分为8个大类，包括食品、烟酒、衣着、家庭设备用品及维修服务、医疗保健和个人用品、交通和通信、娱乐教育文化用品及服务、居住。根据全国城乡13万户居民家庭消费支出调查资料中的项目以及居民消费习惯，又具体确定了262个基本分类。按照国家统计局统计调查制度的规定，在计算CPI时采用拉氏公式。

7.2 时间序列的经济指标

同比价格指数

同比价格指数一般是指当年某月与上年同月相比较计算的价格指数。

环比价格指数

环比价格指数一般是指当年某月与上月相比较计算的价格指数。

实际收入和名义收入

实际收入=100×名义收入/CPI

表 18—6　1982—1984 年和现年实际收入的计算

年份	名义收入（美元）	消费价格指数（1982—1984 年=100）	实际收入（美元）	实际收入的计算（美元）
1982—1984 年	20 000	100	20 000	$\frac{20\ 000}{100}\times(100)$
2002 年	40 000	200	20 000	$\frac{40\ 000}{200}\times(100)$

7.3 经济时间序列的分解、平滑和季节调整

时间序列的组成

一个时间序列包括4个部分：趋势项(T)、循环波动(C)、季节波动(S)和不规则波动(I)。

长期趋势

时间序列中平稳的长期走势。

循环波动

时间序列中超过一年的上升和下降波动。

季节波动

时间序列中随季节变化的波动模式。

不规则波动

时间序列中突发波动或无法预测的波动。

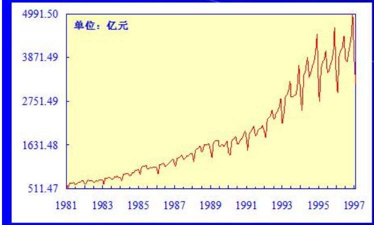


图1 我国工业总产值的时间序列 Y 图形

图2 工业总产值的趋势·循环要素 TC 图形

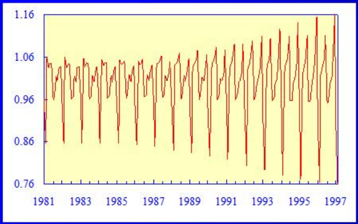


图3 工业总产值的季节变动要素 S 图形

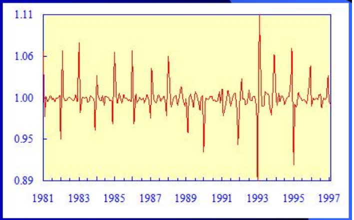


图4 工业总产值的不规则要素 I 图形

时间序列分解

时间序列的加法模型：

$$Y = T + C + S + I$$

时间序列的乘法模型：

$$Y = T \cdot C \cdot S \cdot I$$

线性时间趋势

$$Y_t = \alpha + \beta t + u$$

指数时间趋势

$$log(Y_t) = \alpha + \beta t + u$$

移动平均方法

移动平均方法不仅可以用于对时间序列进行平滑以看到该时序的趋势，还是度量季节波动的基本工具。

简单的移动平均

$$MA_t = \frac{1}{2k+1}\sum_{i=-k}^{k} y_{t+i}$$

中心化的移动平均

如果项数是偶数，常常进行所谓的“移动平均的中心化”。

表 19—4　　3 年移动平均和 5 年移动平均

年份	产量	3 年移动总和	3 年移动平均	5 年移动总和	5 年移动平均
1983	5				
1984	6	19	6.3		
1985	8	24	8.0	34	6.8
1986	10	23	7.7	32	6.4
1987	5	18	6.0	33	6.6
1988	3	15	5.0	35	7.0
1989	7	20	6.7	37	7.4
1990	10	29	9.7	43	8.6
1991	12	33	11.0	49	9.8
1992	11	32	10.7	55	11.0
1993	9	33	11.0	60	12.0
1994	13	37	12.3	66	13.2
1995	15	46	15.3	70	14.0
1996	18	48	16.0	72	14.4
1997	15	44	14.7	73	14.6
1998	11	40	13.3	75	15.0
1999	14	42	14.0	79	15.8
2000	17	53	17.7		
2001	22				

7.3 经济时间序列的分解、平滑和季节调整

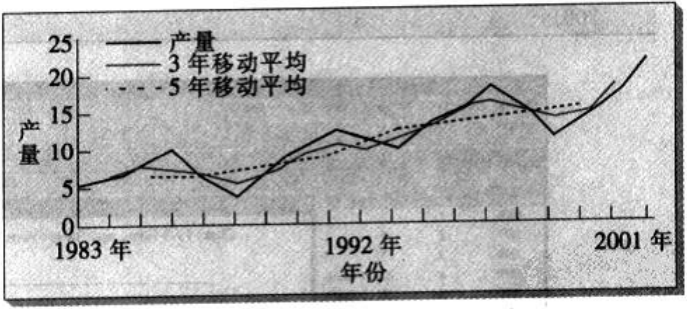


图 19—6　3 年移动平均和 5 年移动平均

表 19—5 **4 年移动平均值**

年份	产量 Y（美元）	4 年移动总和（美元）	4 年移动平均（美元）	中心化的 4 年移动平均
1993	8			
1994	11			
		42 (8+11+9+14)	10.50 (42/4)	
1995	9			10.625
		43 (11+9+14+9)	10.75 (43/4)	
1996	14			10.625
		42	10.50	
1997	9			10.625
		43	10.75	
1998	10			10.000
		37	9.25	
1999	10			9.625
		40	10.00	
2000	8			
2001	12			

季节调整

季节调整就是从时间序列中去除季节变动要素，从而显示出时间序列的趋势循环分量。常见的季节调整方法有：

- X11季节调整方法
- Census X12季节调整方法
- 移动平均比率方法
- Tramo/Seats方法

X11季节调整方法

X-11方法是基于移动平均法的季节调整方法。在计算过程中可根据数据中的随机因素大小，采用不同长度的移动平均，随机因素越大，移动平均长度越大。X-11方法是通过几次迭代来进行分解的，每一次对组成因子的估算都进一步精化。

Census X12季节调整方法

美国商务部国势普查局的X12季节调整程序是在X11方法的基础上发展而来的，包括X11季节调整方法的全部功能，并对X11方法进行了重要改进。

移动平均比率方法

- 对时间序列Y_t进行中心化移动平均，得到序列TCI_t
- 计算S序列：$S_t = Y_t / TCI_t$
- 根据调整因子进行季节调整
- 计算季节调整的最终结果

中心移动平均和消除季节

t	年	季节	销售额	中心移动平均	比值	季节指数	消除季节后的销售额
1	1	1	6.7			0.765	8.76
2	1	2	4.6			0.575	8.00
3	1	3	10.0	8.475	1.180	1.141	8.76
4	1	4	12.7	8.450	1.503	1.519	8.36
5	2	1	6.5	8.425	0.772	0.765	8.50
6	2	2	4.6	8.513	0.540	0.575	8.00
7	2	3	9.8	8.675	1.130	1.141	8.59
8	2	4	13.6	8.775	1.550	1.519	8.95
9	3	1	6.9	8.900	0.775	0.765	9.02
10	3	2	5.0	9.038	0.553	0.575	8.70
11	3	3	10.4	9.113	1.141	1.141	9.11
12	3	4	14.1	9.188	1.535	1.519	9.28
13	4	1	7.0	9.300	0.753	0.765	9.15
14	4	2	5.5	9.463	0.581	0.575	9.57
15	4	3	10.8	9.588	1.126	1.141	9.46
16	4	4	15.0	9.625	1.558	1.519	9.88
17	5	1	7.1	9.688	0.733	0.765	9.28
18	5	2	5.7	9.663	0.590	0.575	9.92
19	5	3	11.1	9.713	1.143	1.141	9.72
20	5	4	14.5	9.888	1.466	1.519	9.55
21	6	1	8.0	9.988	0.801	0.765	10.46
22	6	2	6.2	10.075	0.615	0.575	10.79
23	6	3	11.4			1.141	9.99
24	6	4	14.9			1.519	9.81

季节指数的计算：

	1	2	3	4
1			1.180	1.503
2	0.772	0.540	1.130	1.550
3	0.775	0.553	1.141	1.535
4	0.753	0.581	1.126	1.558
5	0.733	0.590	1.143	1.466
6	0.801	0.615		

均值：0.767　0.576　1.144　1.522　4.009

调整的均值：0.765　0.575　1.141　1.519　4.000

趋势分解

将趋势和循环要素进行分解的方法有多种方法，比较常用的方法有回归分析方法、移动平均法、阶段平均法(phase average，PA方法)、HP滤波方法和频谱滤波方法（frequency (band-pass) filer， BP滤波）。

HP滤波法

HP滤波法方法在Hodrick和Prescott(1980)分析战后美国经济周期的论文中首次使用。设

$$Y_t = Y_t^T + Y_t^C$$

时间序列Y_t中的趋势Y_t^T被定义为下面最小化问题的解：

$$min \sum_{t=1}^{T} \{(Y_t - Y_t^T)^2 + \lambda[c(L)Y_t^T]^2\}$$

差分方程

一阶差分方程：$y_t = \rho y_{t-1} + \epsilon_t$

二阶差分方程：$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \epsilon_t$

滞后算子

$$Lx_t \equiv x_{t-1}$$

$$L^2 x_t \equiv x_{t-2}$$

$$(1 - \rho L) y_t = \epsilon_t$$

$$(1 - \rho_1 L - \rho_2 L^2) y_t = \epsilon_t$$

7.4 平稳ARMA模型

平稳性：一阶差分方程

$$|\rho| < 1$$

平稳性：二阶差分方程

将$(1-\rho_1 L-\rho_2 L^2)$分解成

$$(1-\rho_1 L-\rho_2 L^2) = (1-\lambda_1 L)(1-\lambda_2 L)$$

如果λ_1和λ_2的绝对值都小于1，则该时间序列是稳定的。

移动平均过程(MA)

$$Y_t = \mu + \epsilon_t + \theta \epsilon_{t-1}$$

$$Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$$

自回归过程(AR)

$$Y_t = c + \rho Y_{t-1} + \epsilon_t = (c + \epsilon_t) + \rho(c + \epsilon_{t-1}) + \ldots = [c/(1-\rho)] + \epsilon_t + \rho \epsilon_{t-1} + \rho^2 \epsilon_{t-2} + \ldots$$

$$Y_t = c + \rho_1 Y_{t-1} + \rho_2 Y_{t-2} + \ldots + \rho_p Y_{t-p} + \epsilon_t$$

7.4 平稳ARMA模型

移动平均过程(MA)

$$Y_t = \mu + \epsilon_t + \theta\epsilon_{t-1}$$

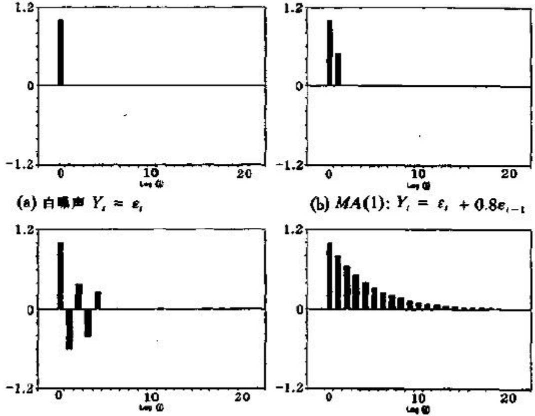
$$Y_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \ldots + \theta_q\epsilon_{t-q}$$

自回归过程(AR)

$$Y_t = c+\rho Y_{t-1}+\epsilon_t = (c+\epsilon_t)+\rho(c+\epsilon_{t-1})+\ldots = [c/(1-\rho)]+\epsilon_t+\rho\epsilon_{t-1}+\rho^2\epsilon_{t-2}+\ldots$$

$$Y_t = c + \rho_1 Y_{t-1} + \rho_2 Y_{t-2} + \ldots + \rho_p Y_{t-p} + \epsilon_t$$

7.4 平稳ARMA模型



(a) 白噪声 $Y_t = \varepsilon_t$

(b) $MA(1)$: $Y_t = \varepsilon_t + 0.8\varepsilon_{t-1}$

(c) $MA(4)$: $Y_t = \varepsilon_t - 0.6\varepsilon_{t-1} + 0.3\varepsilon_{t-2} - 0.5\varepsilon_{t-3} + 0.5\varepsilon_{t-4}$

(d) $AR(1)$: $Y_t = 0.8Y_{t-1} + \varepsilon_t$

自回归综合移动平均过程$ARMA(p, q)$

$$Y_t = c + \rho_1 Y_{t-1} + \rho_2 Y_{t-2} + ... + \rho_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + ... + \theta_q \epsilon_{t-q}$$

$$(1 - \rho_1 L - \rho_2 L^2 - ... - \rho_p L^p) Y_t = c + (1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q) \epsilon_t$$