

第 1 章

导 论

1.1 概 念

时间序列的常见数据,在经济领域,人们会观测每日股票收盘价、每月价格指数、产品的年销售量等;在气象上,人们会观测每天的最高温度和最低温度、年降水与干旱指数、每小时的风速等;在农业上,人们会记录每年作物和牲畜产量、土壤侵蚀、出口销售等。作为计量经济学的一个分支,时间序列分析在计量经济学中占有重要地位,进行时间序列分析的目的主要有两个方面:一是认识产生时间序列的随机机制,即建立数据生成过程的模型;二是基于序列的历史数据,对序列未来的可能值给出预测或预报。

1.1.1 时间序列

时间序列是指被观察到的、依时间为序排列的数据序列。它的特点如下。

(1) 现实的、真实的一组数据,而不是实验数据。时间序列是反映某一现象的统计数据,它包含了现象的变化规律。

(2) 动态数据。时间序列数据反映了现象变化的动态性,对其进行分析可以描述现象的长期、短期变动规律。

时间序列所蕴含的规律会表现出各种特征,这也就是人们对它进行分析的意义所在,下面举几个典型的时间序列数据。

例 1.1.1 1820—1869 年的太阳黑子数据,见图 1.1.1。该图中,横轴是时间指标 t (在这里的 t 以年为单位),纵轴表示在 t 年太阳黑子个数的观测值,这种图被称为时间序列图。由图 1.1.1 可知,该数据具有一定周期性规律,周期大约为 11 年,且波动幅度较大。

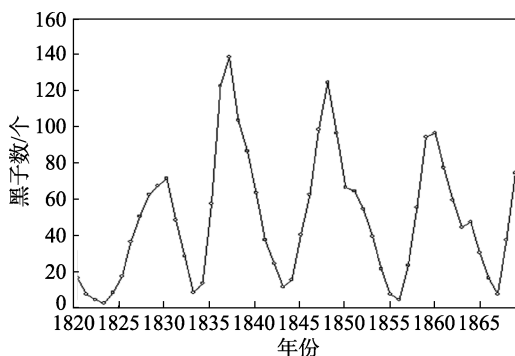


图 1.1.1 太阳黑子数据

例 1.1.2 国内生产总值 (GDP) 是对一国 (地区) 经济在核算期内所有常住单位

生产的最终产品总量的度量,常被看成反映一个国家(地区)经济状况的重要指标。本例给出了我国 1978—2007 年 GDP 数据(单位:亿元)的时间序列图(图 1.1.2)。该数据具有明显的趋势性。

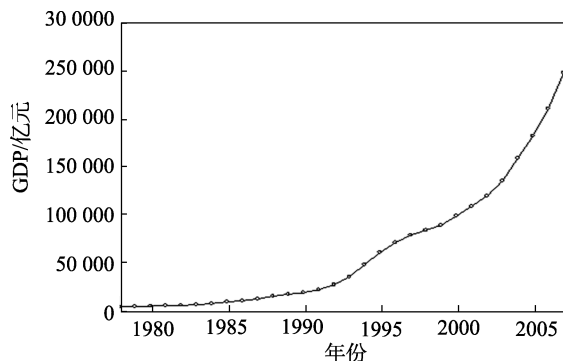


图 1.1.2 我国 GDP 数据

例 1.1.3 1992 年第 1 季度—2008 年第 1 季度我国 GDP 季度数据,见图 1.1.3。该数据既有明显周期性,也有趋势性。季度 GDP 受季节影响,呈现明显的波动特征;同时,随着时间推移,数值逐渐增加而呈现趋势性,而在图 1.1.2 的年度 GDP 数据中就看不到季节性,因为年度数据是对季度数据的总和,在加总中季度数据的差异被抹杀掉了。

例 1.1.4 1990 年 12 月 19 日—2008 年 4 月 18 日上证 A 股指数日数据(除去节假日,共 4386 条数据),见图 1.1.4。该数据属于高频数据,呈现了随机波动的杂乱特征。

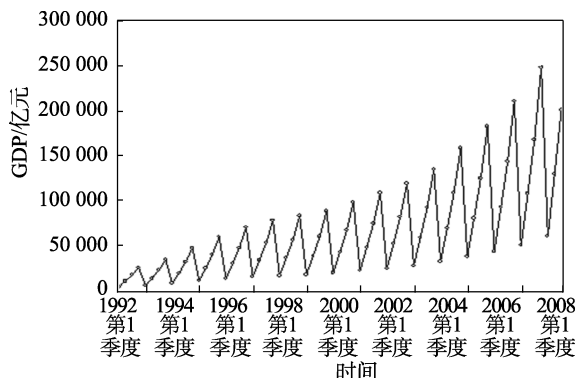


图 1.1.3 我国 GDP 季度数据

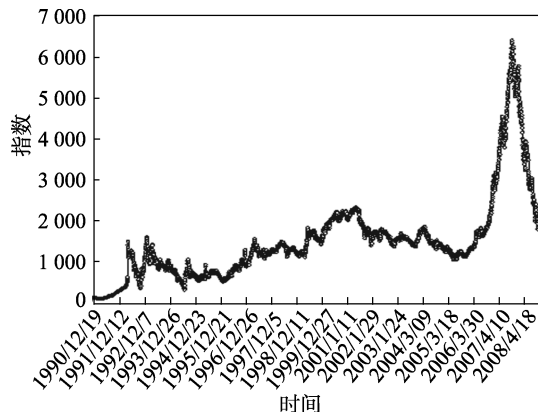


图 1.1.4 上证 A 股指数日数据

例 1.1.5 在时间序列的理论研究中,人们经常会用到模拟数据,图 1.1.5 是根据 $y_t = 0.1 + y_{t-1} + u_t$ 得到的模拟数据图,该数据的趋势中有随机波动,是两个特征的结合。

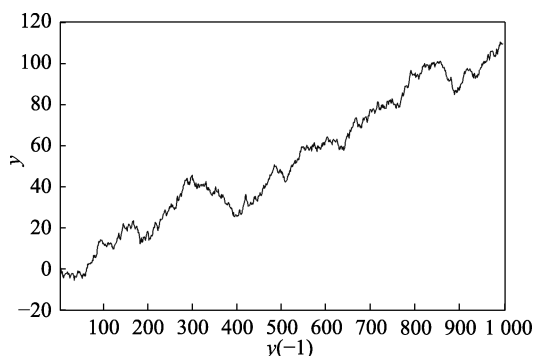


图 1.1.5 单位根模拟数据

1.1.2 时间序列分析

时间序列分析是一种根据动态数据揭示系统动态结构和规律的统计方法,其基本思想是根据系统有限长度的运行记录,建立能够比较精确地反映序列中所包含的动态依存关系的数学模型,并借以对系统的未来进行预报。具体来说,时间序列分析的本质是根据现有的样本数据所包含的相关性建立数学模型,并将模型的规律外推后进行预测。

时间序列分析的思想与经典的计量经济学的建模思想是不同的。经典的计量经济学的模型属于结构式因果模型,也就是说,要解释一个现象的变动规律时,先要在理论上说明哪些因素是影响它变动的因素,将各因素量化后再进行建模分析。例如,人们熟悉的消费函数模型可以解释消费的变动,其中,收入是人们考虑的主要因素,它的图形如图 1.1.6 所示,可以发现两变量有协同变动的趋势,而图 1.1.7 中,以收入为横坐标、以支出为纵坐标绘制的散点图显示,两者存在明显的线性关系。

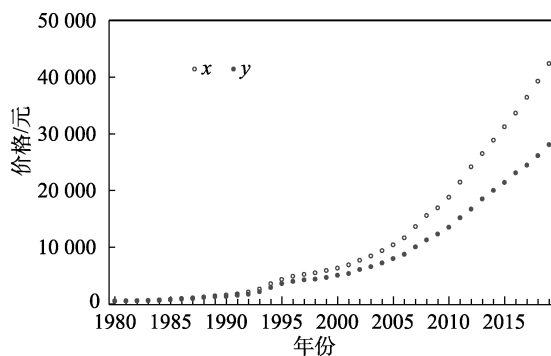


图 1.1.6 收入(x)和支出(y)

单变量时间序列建模思想与经典计量经济学建模思想更是不同的。因为,不是所有的现象变动都可以找到影响因素,另外,有一些影响因素也是不能被量化的。例如,股票价格变动受企业业绩、政策、投资人心理预期、随机因素等的影响,这里有些因素(如

投资人心理预期)就是无法被量化的。计量经济学家认为,尽管影响现象发展的因素无法探求,但其结果之间却存在着一定的联系,可以用相应的模型表示出来,尤其在随机性现象中。这就是单变量时间序列分析建立的基础:来自一个现象的数据,由于现象发展变化具有关联性,导致前后数据之间是有联系的,它们的相关性可以用模型表示出来,就形成了时间序列模型。从性质上说,它们不是包含因果关系的模型,是将结果间的相关性体现出来的模型。

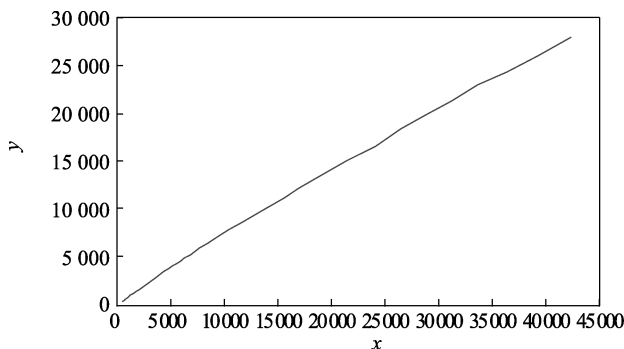


图 1.1.7 收入 (x) 和支出 (y) 线性关系

1.2 时间序列构成要素的分析

依据特征可以将时间序列分为长期趋势 (Trend, T)、季节变动 (Season, S)、循环变动 (Cycle, C) 和不规则变动 (Irregular, I), 也可以说, 一个完整的时间序列 (Y) 是由上述四个部分构成的, 它们与序列值之间遵循乘法或加法关系, 即

$$Y = TSCI$$

或

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

时间序列每一部分的性质、特征不一样, 在分析时人们通常会把各构成因素从序列值中分离出来, 进一步进行分析。

1.2.1 长期趋势分析

长期趋势是指存在于时间序列中的方向确定的运动态势, 它可以是向上的、向下的, 也可以是线性的、非线性的。之所以能呈现长期趋势, 是因为有某种确定因素持续、恒定地影响时间序列值, 数据依时间变化表现为一种确定倾向, 它按某种规则稳步地增长或下降。

长期趋势分析是把长期趋势值从时间序列值中分离出来, 具体的方法有移动平均法、模型拟合法、指数平滑法等。

这里介绍移动平均法和模型拟合法, 并比较它们的差异。

1. 移动平均法

移动平均法就是计算几项数据的平均值, 且以固定项数不断向后移动计算平均数,

利用所得到的平均数表示趋势值,对应的时期是移动项数的中间时期。移动平均法利用了平均数的性质来修正数据得到趋势值,趋势值是贯穿于数据中的不可观测的量,它引导序列值围绕它波动,因为“贯穿”的性质,所以把高低起伏的原始数据平均后,就可以得到贯穿在原始序列值中的趋势值。同时,移动平均保证了得到的趋势值能随着序列值的变化而变化。

例 1.2.1 某企业历年产品销售情况如表 1.2.1 所示,试分析其长期趋势。

将数据作图,如图 1.2.1 所示。序列值有明显的长期趋势和季节变动特征,为消除季节变动对长期趋势值的影响,取 $N=4$ 进行移动平均,结果如表 1.2.1 的第四、第五列所示。偶数项被移动平均后,平均数对应的是两期数据中间,此时,需要为相邻两项值计算平均数,该平均数就是对应时期的趋势值。例如,第一个移动平均值是 23.68,是前四期的平均值,对应的是第 2.5 期,第二个平均值是 25.45,对应的是第 3.5 期,将 23.68 与 25.45 平均后,平均值 24.56 对应的就是第一年第 3 季度,是它的趋势值。

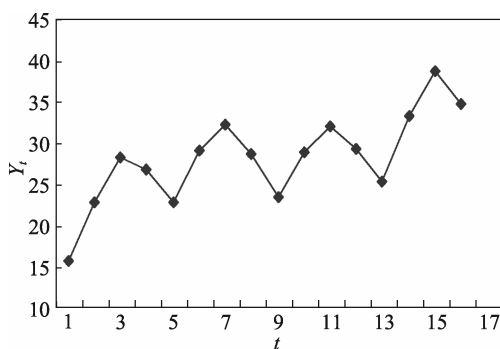


图 1.2.1 某企业历年产品销售

需要注意的是:①移动项数为偶数时,要进行二次平均。②对于有季节性波动的序列,移动的项数要与季节性波动的周期一致,这是为了消除季节性对趋势值的影响,如果移动平均项数与季节周期不一致,季节性的高低起伏变化会对趋势值计算有影响;而如果两者一致,移动平均的将一直都是一个周期内的数值,只是可能是跨年度的,这样可以消除季节性的影响。

2. 模型拟合法

模型拟合法是通过模型拟合得到趋势值,通常拟合的是直线模型。设有时间序列 $\{Y_t\}$ ($t=1,2,\dots,n$), 对其拟合直线方程为

$$Y_t = a + bt(t=1,2,\dots,n) \quad (1.2.1)$$

$\hat{Y}_t$ 为 Y_t 的估计值或趋势值。通常用最小二乘法 (OLS) 估计 a 、 b , OLS 估计参数的原理是时间序列值与它们各自的估计值之间的离差平方和最小,也就是满足

$$Q = \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 = \min \quad (1.2.2)$$

$$Q = \sum_{t=1}^n (Y_t - a - bt)^2 = \min \quad (1.2.3)$$

为使剩余平方和最小, 有

$$\frac{\partial Q}{\partial a} = 2 \sum_{t=1}^n (Y_t - a - bt)(-1) = 0 \quad (1.2.4)$$

$$\frac{\partial Q}{\partial b} = 2 \sum_{t=1}^n (Y_t - a - bt)(-t) = 0 \quad (1.2.5)$$

简化后为

$$\begin{cases} \sum_{t=1}^n Y_t = na + b \sum_{t=1}^n t \\ \sum_{t=1}^n Y_t \cdot t = a \sum_{t=1}^n t + b \sum_{t=1}^n t^2 \end{cases} \quad (1.2.6)$$

利用 OLS 对例 1.2.1 数据拟合直线。先对时间序列中时间赋值, 如表 1.2.1 所示, 代入式 (1.2.6) 有

$$\begin{cases} 456.5 = 16a + b \times 136 \\ 4170.2 = 136a + b \times 1496 \end{cases} \quad (1.2.7)$$

解方程得 $a=21.31$, $b=0.85$, 则

$$\hat{Y}_t = 21.31 + 0.85t \quad (1.2.8)$$

将时间项代入方程, 可得 $\hat{Y}_t$, 也就是趋势值 T , 结果如表 1.2.1 所示。移动平均法和直线模型得到的趋势值如图 1.2.2 所示。从结果及图形可以看出两种方法的区别, 直线模型得到的趋势值每期以固定的数值增加, 而移动平均法得到的趋势值则是波动的, 将随数据的变化而变化。实际经济数据很少有直线性质的趋势值, 所以, 移动平均法得到的趋势值应用最广, 例如, 它是 X-11 季节调整方法的基础。

表 1.2.1 长期趋势分析

年	季度	销售量 (Y)	移动平均数 $N=4$	二次移动平均数 (T)	t	t ²	t Y _t	T
第一年	1	16.0	—	—	1	1	16	22.16
	2	23.0	—	—	2	4	46	23.01
	3	28.6	23.68	24.56	3	9	85.8	23.86
	4	27.1	25.45	26.24	4	16	108.4	24.71
第二年	1	23.1	27.03	27.53	5	25	115.5	25.56
	2	29.3	28.03	28.25	6	36	175.8	26.41
	3	32.6	28.48	28.54	7	49	228.2	27.26
	4	28.9	28.6	28.58	8	64	231.2	28.11
第三年	1	23.6	28.55	28.53	9	81	212.4	28.96
	2	29.1	28.5	28.59	10	100	291	29.81
	3	32.4	28.68	28.91	11	121	356.4	30.66
	4	29.6	29.15	29.73	12	144	355.2	31.51
第四年	1	25.5	30.3	31.13	13	169	331.5	32.36
	2	33.7	31.95	32.63	14	196	471.8	33.21
	3	39.0	33.3	—	15	225	585	34.06
	4	35.0	—	—	16	256	560	34.91

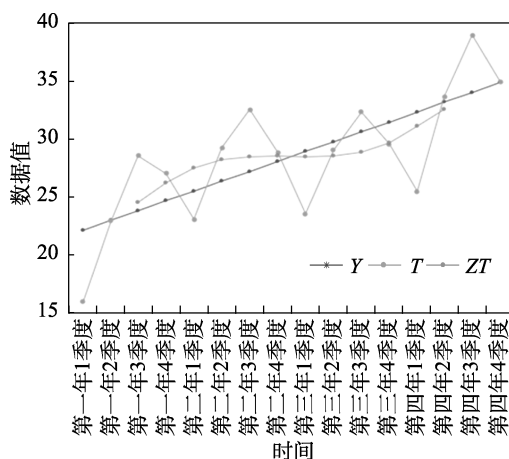


图 1.2.2 长期趋势值

注：Y—序列原始值，T—移动平均法趋势值，ZT—模型拟合趋势值

1.2.2 季节性分析

受季节更替等因素影响，序列依一个固定周期而规则性的变化被称为季节性，又称商业循环。时间序列之所以呈现季节变动特征，是因为变化是受季节的更替影响而形成的，多见于季节性商品的销售、生产等，例如，冷饮、汗衫等。近年来，季节性分析还包括节假日对序列的影响等。与一般意义上的经济周期序列不同的是，季节性序列的周期长度是固定的，因为四季的更替是固定的，且没有两个经济周期的长度是一样的。

季节性分析的方法主要是计算季节指数，季节指数的计算如式（1.2.9）所示。

$$\text{季节指数} = \frac{\text{同月（或季）平均数}}{\text{总月（或季）平均数}} \quad (1.2.9)$$

式（1.2.9）中，分子是由各月（或季）序列值计算得出的平均数，由于处于相同月或季，受季节影响的数值比较接近，所以，分子代表某月或某季的平均水平值；分母是所有序列值的平均数，消除了季节对序列值的影响。季节指数表示某月或某季相对平均值的大小，当季节指数大于1时，其被称为旺季；当季节指数小于1时，其被称为淡季。

例 1.2.2 对例 1.2.1 数据进行季节性分析，见表 1.2.2。

表 1.2.2 季节性分析

	1	2	3	4	
第一年	16.0	23.0	28.6	27.1	
第二年	23.1	29.3	32.6	28.9	
第三年	23.6	29.1	32.4	29.6	
第四年	25.5	33.7	39.0	35.0	
各年同季平均数	22.05	28.78	33.15	30.15	总平均数（28.53）
季节指数（%）	77.29	100.88	116.19	105.68	400.04
整理后的季节指数（%）	77.28	100.87	116.18	105.67	400

注：四个季节指数的合计应为400%，不满足则需要进行调整，即：调整各季节指数 = $\frac{400\%}{\text{实际合计数}} \times \text{各季节指数}$ 。

用原始数据计算季节指数会受到趋势值的影响，因此，在计算季节指数时，要先扣除趋势值的影响，结果见表 1.2.3 和表 1.2.4。

表 1.2.3 剔除长期趋势数据结果

年	季度	销售量 Y (万元)	移动平均数 T (N=4)	Y/T (%)
第一年	1	16.0	—	—
	2	23.0	—	—
	3	28.6	24.56	116.45
	4	27.1	26.24	103.28
第二年	1	23.1	27.53	83.91
	2	29.3	28.25	103.72
	3	32.6	28.54	114.233
	4	28.9	28.58	101.12
第三年	1	23.6	28.53	82.72
	2	29.1	28.59	101.78
	3	32.4	28.91	112.07
	4	29.6	29.73	99.56
第四年	1	25.5	31.13	81.91
	2	33.7	32.63	103.28
	3	39.0	—	—
	4	35.0	—	—

表 1.2.4 剔除长期趋势的季节指数

	1	2	3	4	
第一年	—	—	116.45	103.28	
第二年	83.91	103.72	114.23	101.12	
第三年	82.72	101.78	112.07	99.56	
第四年	81.91	103.28	—	—	
各年同季度平均数	82.85	102.93	114.25	101.32	总平均数 (100.34)
季节指数	82.57	102.58	113.86	100.98	400
季节指数 (%) 未去趋势	77.28	100.87	116.18	105.67	

比较两个季节指数可以发现，剔除趋势值后，季节指数间的差距明显减小。

1.2.3 循环变化分析

循环变化是指序列值呈现的周期不固定的波动变化，类似经济周期的特征。对循环变化没有确定的分析方法，人们常常将之与随机性变化合在一起分析。在具体分析时，需要把趋势值和季节变动剔除，即

$$\frac{Y}{T \cdot S} = C \cdot I \tag{1.2.10}$$

对剩余的 C·I 进行移动平均，进一步消除随机波动的影响，余下的就是循环变动，结果见表 1.2.5。

表 1.2.5 循环变化结果

年份	季度	销售量 Y(万元)	T	S(%)	Y/(T·S)	移动平均数(N=3)C	随机波动项 I
2015	1	16.0	—	82.57	—	—	—
	2	23.0	—	102.58	—	—	—
	3	28.6	24.56	113.86	102.27	—	—
	4	27.1	26.24	100.98	102.28	102.06	100.22

续表

年份	季度	销售量 Y(万元)	T	S(%)	Y/(T·S)	移动平均数(N=3)C	随机波动项 I
2016	1	23.1	27.53	82.57	101.62	101.67	99.95
	2	29.3	28.25	102.58	101.11	101.02	100.09
	3	32.6	28.54	113.86	100.32	100.52	99.80
	4	28.9	28.58	100.98	100.14	100.21	99.93
2017	1	23.6	28.53	82.57	100.18	99.85	100.33
	2	29.1	28.59	102.58	99.22	99.27	99.95
	3	32.4	28.91	113.86	98.43	98.75	99.68
	4	29.6	29.73	100.98	98.60	98.75	99.85
2018	1	25.5	31.13	82.57	99.21	99.50	99.71
	2	33.7	32.63	102.58	100.68	—	—
	3	39.0	—	113.86	—	—	—
	4	35.0	—	100.98	—	—	—

1.2.4 不规则变化分析

对不规则变化分析，以前的分析认为它是由随机因素引起的，没有进一步研究的必要，但是，后来的研究发现，不规则变化序列前后值之间是有联系的，可以用模型再现它们的规律，这就是人们常说的时间序列分析，也是本书时间序列分析的对象。也就是说，对一般的时间序列序列，要剔除掉它的趋势、季节、循环部分后才能进行时间序列分析。

例 1.2.1 中时间序列的季节波动、循环波动和不规则波动结果分别如图 1.2.3、图 1.2.4 和图 1.2.5 所示。

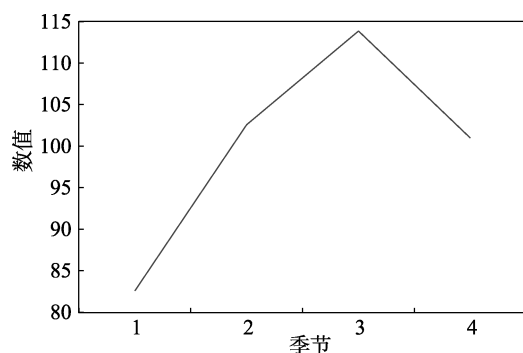


图 1.2.3 季节指数

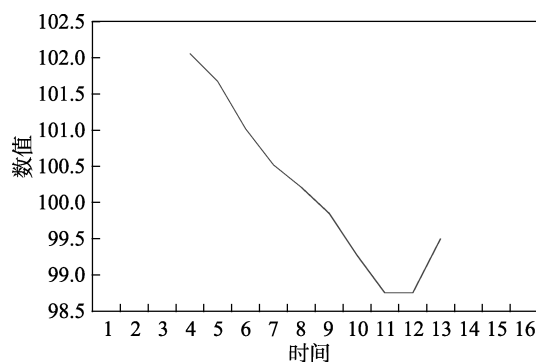


图 1.2.4 循环波动

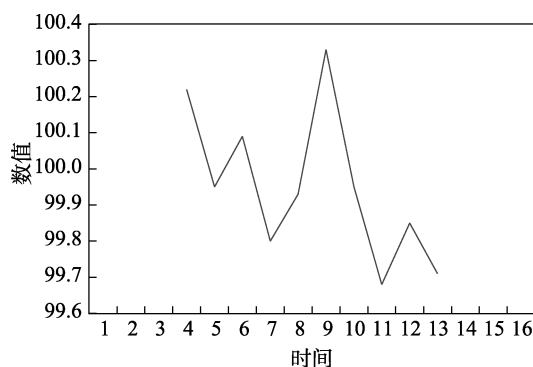


图 1.2.5 不规则变动

上述不同构成因素从时间序列分离出来的方法,从性质上说可以分成两类,见图 1.2.6。趋势变化分析、周期变化分析及循环变化分析属于确定性分析,因为这些因素变化的方向和量是确定的,而不规则变化的方向和量则是随机的。

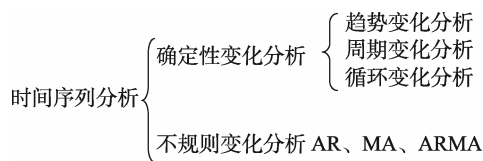


图 1.2.6 时间序列分析结构

1.3 时间序列分析发展历史

时间序列分析理论与方法是 20 世纪 40 年代分别由诺伯特·维纳和安德烈·科雷莫戈诺尔独立给出的,他们对发展时间序列的参数模型拟和推断过程作出了贡献,提供了与此相关的重要文献,促进了时间序列分析在工程领域的应用。

20 世纪 70 年代, G.P.Box 和 Jenkins 发表专著《时间序列分析: 预测和控制》,使时间序列分析的应用成为可能,通常被称为 B-J 方法。

B-J 方法假定数据是平稳的,是时间序列建模的基础。但是,“二战”后,世界经济的快速发展导致时间序列多呈现非平稳特征,20 世纪 70 年代后,非平稳时间序列的建模问题得到关注,在理论和方法上有重大突破,引领了计量经济学进入新的领域,而其中的协整理论与方法对计量经济学有里程碑式的意义和贡献(详细内容见第 9 章)。

2003 年诺贝尔经济学奖的获得者是美国经济学家罗伯特·恩格尔和英国经济学家克莱夫·格兰杰。获奖原因为“发明了处理许多经济时间序列两个关键特性的统计方法: 时间变化的变更率和非平稳性”。时间变化的变更率是指方差随时间变化而变化的频率,这主要是指恩格尔在 1982 年发表的条件异方差模型 (ARCH),最初主要用于研究英国的通货膨胀问题,后来被广泛用作金融分析的高级工具。在传统的计量经济学研究中,人们通常假定经济数据和产生这些数据的随机过程是平稳的,格兰杰的贡献主要是在非平稳过程假定下所进行的严格计量模型的建立,即非平稳时间序列的建模。

练 习 题

1. 比较古典计量经济学与时间序列分析建模思想的不同。
2. 一个时间序列 $\{Y_t\}$ 的构成因素有哪些? 它们与 $\{Y_t\}$ 的关系如何表示? 简述将每种构成因素从 Y_t 中分离出来的方法。
3. HP 滤波是从 $\{Y_t(Y_t = Y_t^T + Y_t^C)\}$ 中将其趋势成分 Y_t^T 分离出来,使下式达到最小,请将其展开并说明其中的含义,其中, $L^k Y_t = Y_{t-k}$ 。

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

$$c(L) = (L^{-1} - 1) - (1 - L)$$

4. 下表为企业的季度产值, 请利用统计软件将时间序列的四个构成要素从序列值中分离出来。

第一年		第二年		第三年		第四年	
季度	产值/百万元	季度	产值/百万元	季度	产值/百万元	季度	产值/百万元
1	36.4	1	43.1	1	40.8	1	44.6
2	43.1	2	49.5	2	46.1	2	52.5
3	48.3	3	52.1	3	50.2	3	58.1
4	46.9	4	46.1	4	48.4	4	54.6

5. 模拟生成一个容量为 48、随机独立的正态分布过程, 并绘制时间序列图, 考察其是否显示出“随机性”。

6. 模拟生成一个容量为 100、完全随机、自由度为 5 的独立 t 分布过程, 并绘制时间序列图, 考察是否显示出“随机性”和“非正态性”。



自
学
自
测



扫
描
此
码