第 9 章　时间序列分析和预测

内容提要

9.1 时间序列及其分解

9.2 时间序列的描述性分析

9.3 预测方法的选择

9.4 平稳序列的预测

9.5 趋势型序列的预测

9.6 复合型序列的分解预测

9.1 时间序列及其分解

- 定义与分类
- 时间序列的成分

时间序列的定义与分类

定义

时间序列，是指同一现象在不同时间上的相继观察值排列而成的序列。

表示

Y_t(其中：$t = 1,2,\cdots$，T)

分类

- 平稳序列（stationary series）
  - 不存在趋势、季节、周期性因素
  - **只包含随机因素**
- 非平稳序列（non-stationary series）
  - **包含趋势、季节或周期性因素**
  - 常见的类型
    - 有趋势的序列（线性趋势、非线性趋势）
    - 有趋势和季节性的序列
    - 复合型序列

时间序列的成分

趋势(Trend)

持续向上或持续下降的状态或规律

季节性(Seasonality)

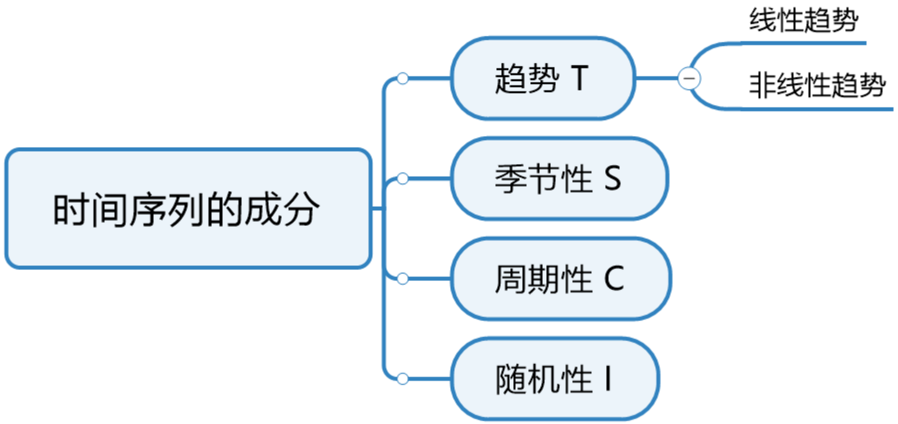
也称季节变动(Seasonal fluctuation)，时间序列在一年内重复出现的周期性波动

周期性(Cyclity)

也称循环波动(Cyclical fluctuation)，围绕长期趋势的一种波浪形或振荡式变动

随机性(Random)

也称不规则波动(Irregular variations)，除去趋势、周期性和季节性之后的偶然性波动

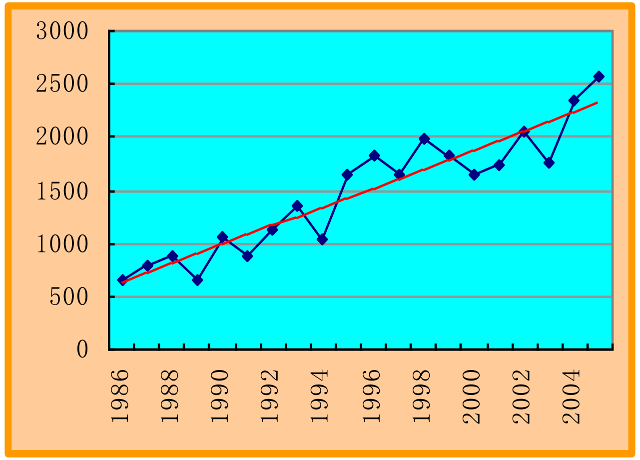


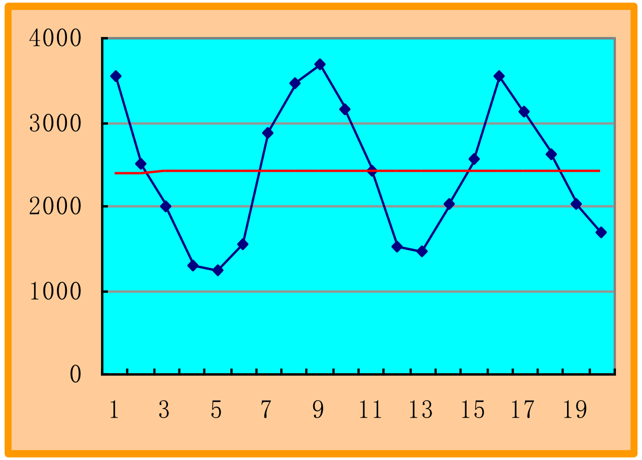
时间序列的分解（乘法模型）

$$Y_t = T_t \times S_t \times C_t \times I_t$$

含有不同成分的时间序列

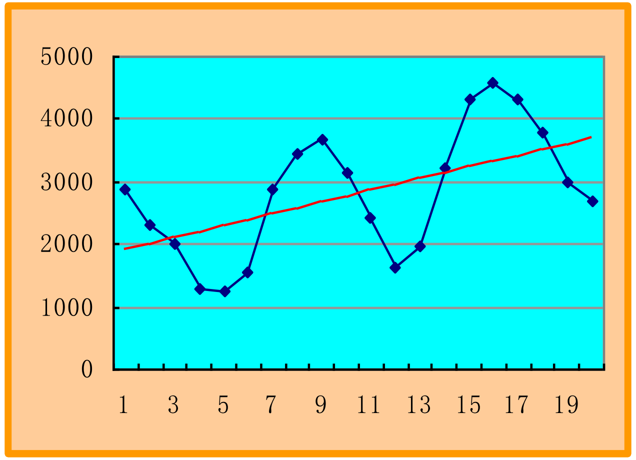
趋势

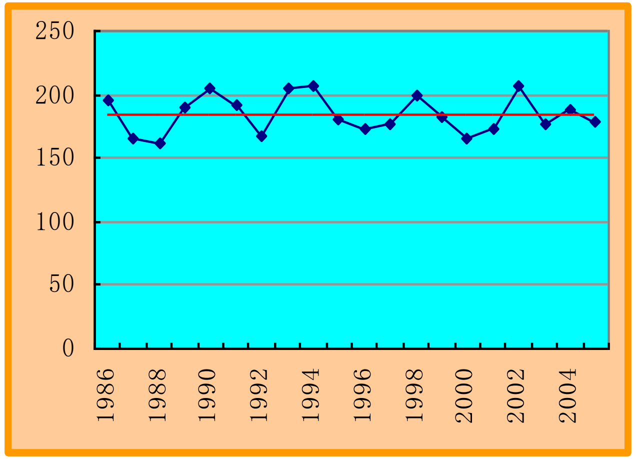




季节

趋势与季节





随机（平稳序列）

9.2 时间序列的描述性分析

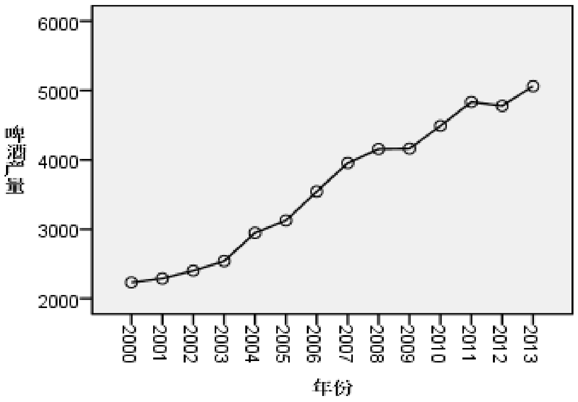
- 图形描述
- 增长率分析

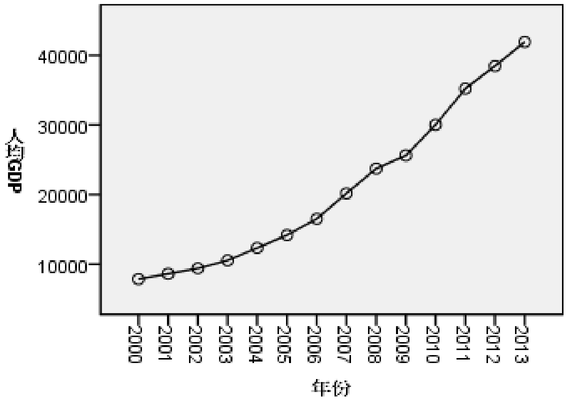
时间序列的描述性分析：图形描述

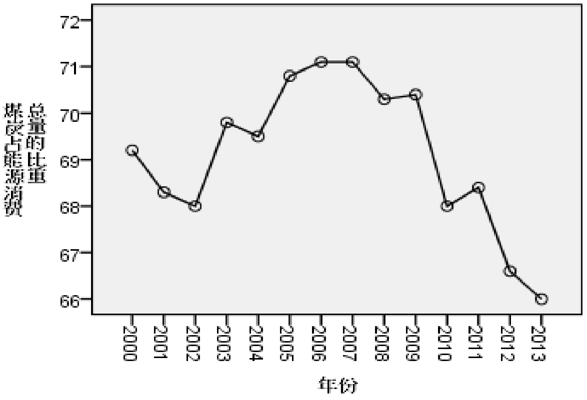
例子

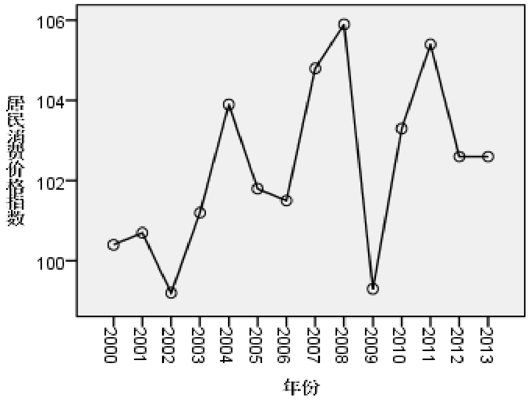
2000—2013年我国啤酒产量、人均DGP、煤炭占能源消费的比重和居民消费价格指数（CPI）的时间序列。

	A	B	C	D	E
1	年份	啤酒产量（万千升）	人均GDP（元）	煤炭占能源消费总量的比重(%)	居民消费价格指数（上年=100）
2	2000	2231.3	7857.7	69.2	100.4
3	2001	2288.9	8621.7	68.3	100.7
4	2002	2402.7	9398.1	68.0	99.2
5	2003	2540.5	10542.0	69.8	101.2
6	2004	2948.6	12335.6	69.5	103.9
7	2005	3126.1	14185.4	70.8	101.8
8	2006	3543.6	16499.7	71.1	101.5
9	2007	3954.1	20169.5	71.1	104.8
10	2008	4156.9	23707.7	70.3	105.9
11	2009	4162.2	25607.5	70.4	99.3
12	2010	4490.2	30015.0	68.0	103.3
13	2011	4834.5	35197.8	68.4	105.4
14	2012	4778.6	38459.5	66.6	102.6
15	2013	5061.5	41907.6	66.0	102.6









时间序列的描述性分析：增长率分析

定义

也称增长速度，是指报告期观察值与基期观察值之比减1后的结果，用%表示。

分类

- **环比增长率、定基增长率**
- **一般增长率、平均增长率和年化增长率**

环比增长率与定基增长率

- **环比增长率**

  报告期水平与**前一期**水平之比减1

$$G_{\mathrm{t}} = \frac{Y_{\mathrm{t}}}{Y_{t-1}} - 1 \qquad (t = 1,2,\cdots,n)$$

- **定基增长率**

  报告期水平与**某一固定期**水平之比减1

$$G_{\mathrm{t}} = \frac{Y_{\mathrm{t}}}{Y_0} - 1 \qquad (t = 1,2,\cdots,n)$$

9.3 预测方法的选择

- 选择预测方法
- 预测方法的评估

选择预测方法

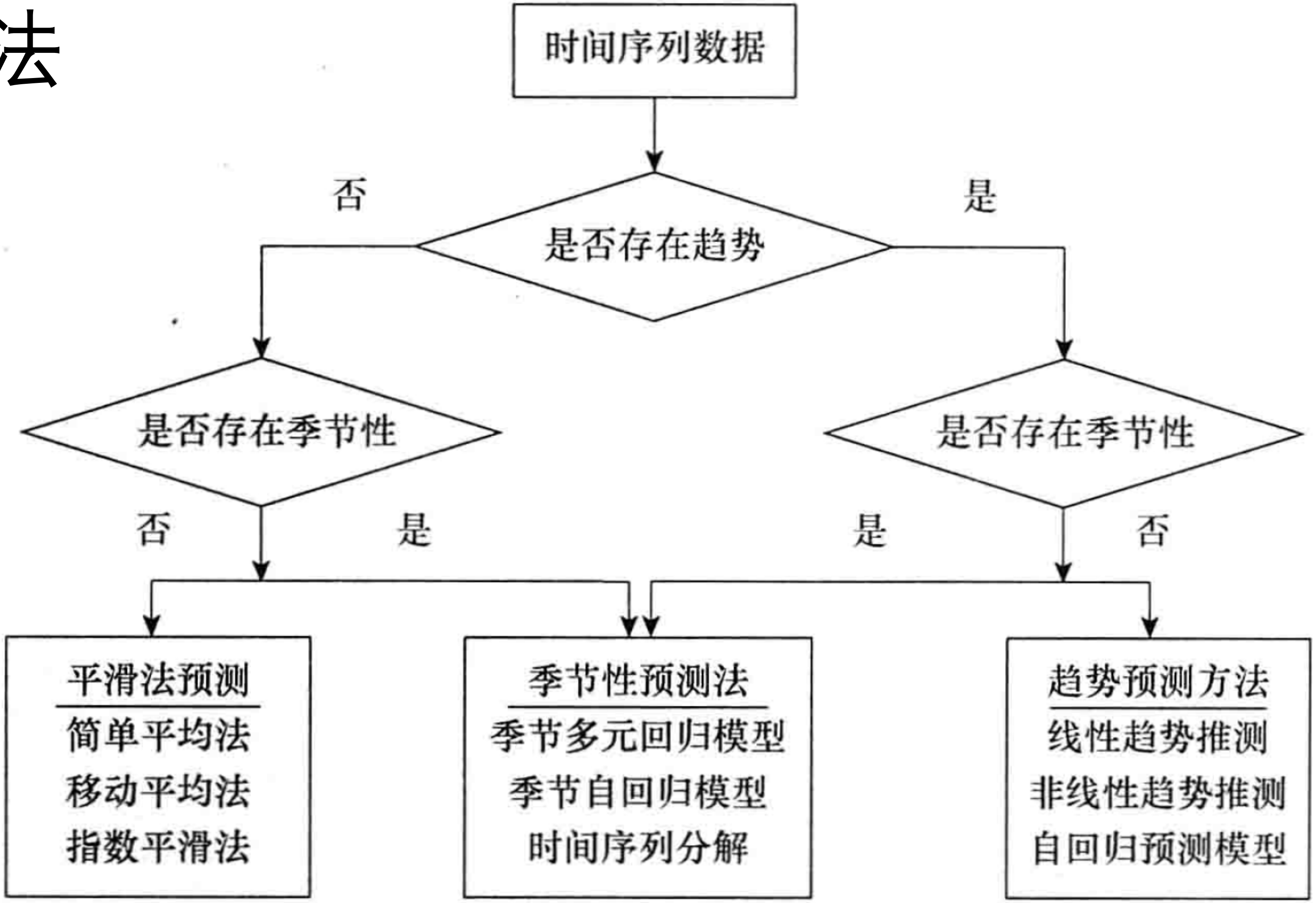


图 13—9　时间序列的类型和预测方法的选择

预测方法的评估

评估原理

最小化预测误差，即最小化预测值与实际值的差距。

评估方法

- 平均误差
- 平均绝对误差
- **均方误差**
- 平均百分比误差
- 平均绝对百分比误差

公式

- 平均误差 $$ME = \frac{\sum_{i=1}^{n}(Y_i - F_i)}{n}$$
- 平均绝对误差 $$MAD = \frac{\sum_{i=1}^{n}|Y_i - F_i|}{n}$$
- 均方误差 $$MSE = \frac{\sum_{i=1}^{n}(Y_i - F_i)^2}{n}$$
- 平均百分比误差 $$MPE = \frac{\sum\left(\frac{Y_i - F_i}{Y_i} \times 100\right)}{n}$$
- 平均绝对百分比误差 $$MAPE = \frac{\sum_{i=1}^{n}\left(\frac{|Y_i - F_i|}{Y_i} \times 100\right)}{n}$$

9.4 平稳序列的预测

- 简单平均法
- 移动平均法
- 指数平滑法

简单平均法

原理

根据过去已有的t期观察值来预测下一期的数值。

方法

- t+1期预测：$F_{t+1}=\frac{1}{t}(Y_1+Y_2+\cdots+Y_t)=\frac{1}{t}\sum_{i=1}^{t}Y_i$
- t+1期预测误差：$e_{t+1}=Y_{t+1}-F_{t+1}$
- t+2期预测：$F_{t+2}=\frac{1}{t+1}(Y_1+Y_2+\cdots+Y_t+Y_{t+1})=\frac{1}{t+1}\sum_{i=1}^{t+1}Y_i$
- t+2期预测误差：$e_{t+2}=Y_{t+2}-F_{t+2}$

以此类推……

特点

- 适合平稳时间序列的预测
- 远期数值与近期数据具有相等权重
- 当存在趋势或季节因素时，预测结果不准确

移动平均法

原理

通过对时间序列逐期递移求得一系列平均数作为预测值。

分类

- **简单移动平均法**
- **加权移动平均法**

简单移动平均法

- 方法：将最近k期数据平均作为下一期的预测值
- 公式

设移动间隔为k$(0 < k < t)$, 则t期的移动平均值为：

$$\bar{Y}_t = \frac{Y_{t-k+1} + Y_{t-k+2} + \cdots + Y_{t-1} + Y_t}{k}$$

t+1期预测为：$F_{t+1} = \bar{Y}_t = \frac{Y_{t-k+1} + Y_{t-k+2} + \cdots + Y_{t-1} + Y_t}{k}$

t+2期预测为：$F_{t+2} = \bar{Y}_{t+1} = \frac{Y_{t-k+2} + Y_{t-k+3} + \cdots + Y_t + Y_{t+1}}{k}$

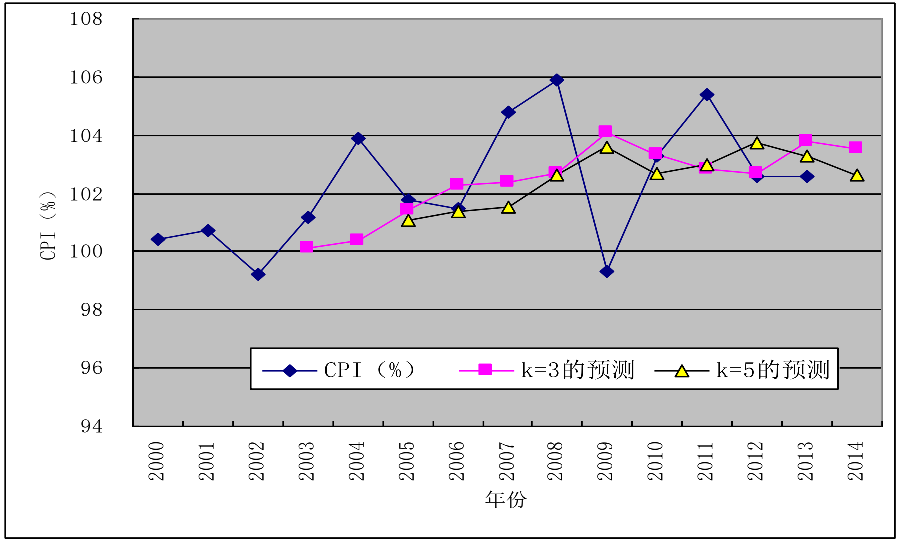
以此类推……

简单移动平均法：例子

数据

	A	B	C	D	E	F	G	H
1	年份	CPI（%）	移动平均预测 k=3	预测误差	预测误差平方	移动平均预测 k=5	预测误差	预测误差平方
2	2000	100.4						
3	2001	100.7						
4	2002	99.2						
5	2003	101.2	100.10	1.10	1.21			
6	2004	103.9	100.37	3.53	12.48			
7	2005	101.8	101.43	0.37	0.13	101.08	0.72	0.52
8	2006	101.5	102.30	-0.80	0.64	101.36	0.14	0.02
9	2007	104.8	102.40	2.40	5.76	101.52	3.28	10.76
10	2008	105.9	102.70	3.20	10.24	102.64	3.26	10.63
11	2009	99.3	104.07	-4.77	22.72	103.58	-4.28	18.32
12	2010	103.3	103.33	-0.03	0.00	102.66	0.64	0.41
13	2011	105.4	102.83	2.57	6.59	102.96	2.44	5.95
14	2012	102.6	102.67	-0.07	0.00	103.74	-1.14	1.30
15	2013	102.6	103.77	-1.17	1.36	103.30	-0.70	0.49
16	**2014**	—	**103.5**	—	—	**102.6**	—	—
17	**合计**	—	—	—	**61.14**	—	—	**48.40**

图形



指数平滑法

原理

通过对过去观察值加权平均进行预测。

特点

- 加权平均的特殊形式
- 观察值越久远，权重越低
- 权重呈现指数下降

分类

- 一次指数平滑法
- 二次指数平滑法
- 三次指数平滑法

一次指数平滑法

- 公式：$F_{t+1} = \alpha Y_t + (1-\alpha)F_t$

  其中：α为平滑系数， $0 < \alpha < 1$

- 第1期预测值：$F_1 = Y_1$
- 第2期预测值：

$$F_2 = \alpha Y_1 + (1-\alpha)F_1 = \alpha Y_1 + (1-\alpha)Y_1 = Y_1$$

- 第3期预测值：

$$F_3 = \alpha Y_2 + (1-\alpha)F_2 = \alpha Y_2 + (1-\alpha)Y_1$$

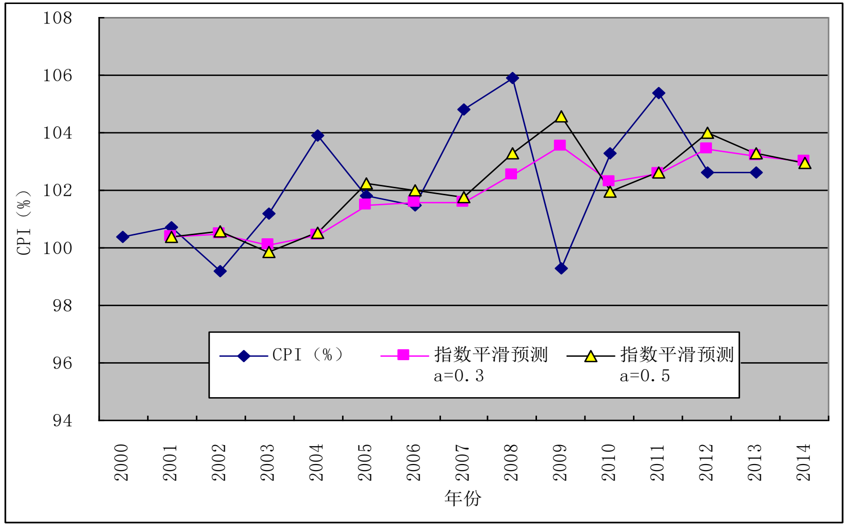
以此类推……

一次指数平滑法：例子

数据

	A	B	C	D	E	F	G	H
1	年份	CPI（%）	指数平滑预测 α=0.3	预测误差	预测误差平方	指数平滑预测 α=0.5	预测误差	预测误差平方
2	2000	100.4						
3	2001	100.7	100.40	0.30	0.09	100.40	0.30	0.09
4	2002	99.2	100.49	-1.29	1.66	100.55	-1.35	1.82
5	2003	101.2	100.10	1.10	1.20	99.88	1.33	1.76
6	2004	103.9	100.43	3.47	12.03	100.54	3.36	11.31
7	2005	101.8	101.47	0.33	0.11	102.22	-0.42	0.18
8	2006	101.5	101.57	-0.07	0.01	102.01	-0.51	0.26
9	2007	104.8	101.55	3.25	10.57	101.75	3.05	9.27
10	2008	105.9	102.52	3.38	11.39	103.28	2.62	6.88
11	2009	99.3	103.54	-4.24	17.95	104.59	-5.29	27.97
12	2010	103.3	102.27	1.03	1.07	101.94	1.36	1.84
13	2011	105.4	102.58	2.82	7.97	102.62	2.78	7.72
14	2012	102.6	103.42	-0.82	0.68	104.01	-1.41	1.99
15	2013	102.6	103.18	-0.58	0.33	103.31	-0.71	0.50
16	2014	—	103.00	—	—	102.95	—	—
17	合计	—	—	—	65.06	—	—	71.57

图形



9.5 趋势型序列的预测

- 线性趋势预测
- 非线性趋势预测

线性趋势预测

线性趋势

现象随着时间的推移而呈现出稳定增长或下降的线性变化规律

预测原理

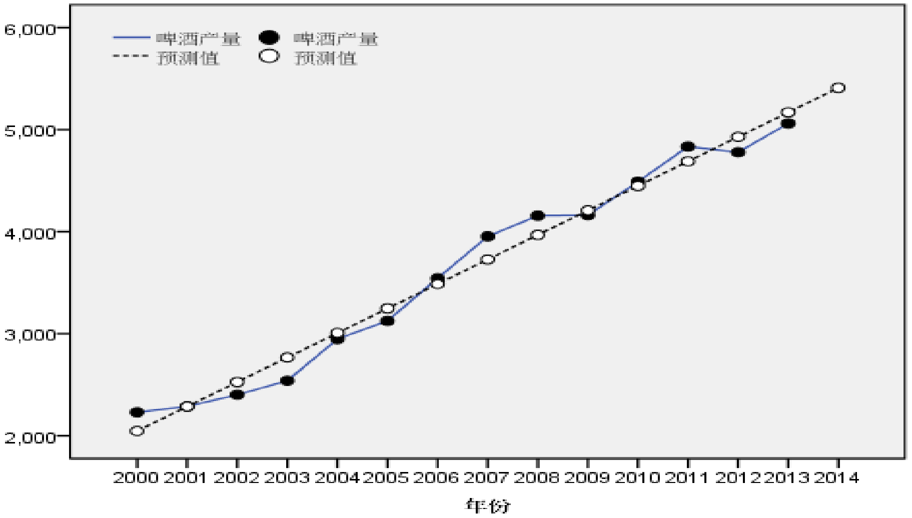
通过**一元线性回归方程**进行预测。

公式

$$\hat{Y}_t = b_0 + b_1 t$$

例子

根据右表中的啤酒产量数据，进行线性趋势预测。



	年份	时间代码	啤酒产量	预测值	残差
1	2000	1	2231.3	2047.35	183.95
2	2001	2	2288.9	2287.53	1.37
3	2002	3	2402.7	2527.72	-125.02
4	2003	4	2540.5	2767.90	-227.40
5	2004	5	2948.6	3008.09	-59.49
6	2005	6	3126.1	3248.27	-122.17
7	2006	7	3543.6	3488.46	55.14
8	2007	8	3954.1	3728.64	225.46
9	2008	9	4156.9	3968.83	188.07
10	2009	10	4162.2	4209.01	-46.81
11	2010	11	4490.2	4449.20	41.00
12	2011	12	4834.5	4689.38	145.12
13	2012	13	4778.6	4929.57	-150.97
14	2013	14	5061.5	5169.75	-108.25
15	2014	15	.	5409.94	.

非线性趋势预测：指数曲线

指数趋势

时间序列以几何级数递增或递减的趋势

指数曲线的趋势方程

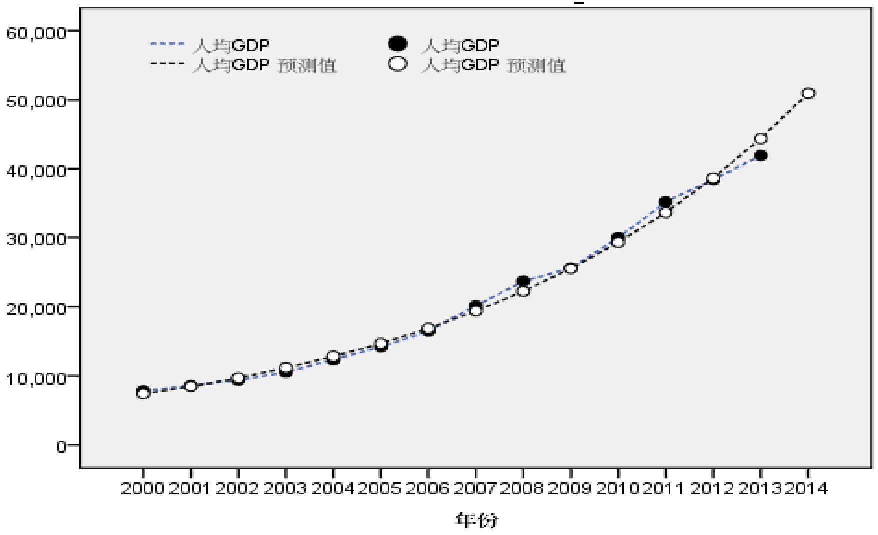
$$\hat{Y}_t = b_0 {b_1}^t$$

说明

- b_0，b_1为参数
- 若$b_1 > 1$, 增长率随着时间t的增加而增加
- 若$b_1 < 1$, 增长率随着时间t的增加而降低
- 若$b_0 > 1$，$b_1 < 1$, 趋势值逐渐降低到以0为极限

例子

根据右表中的人均GDP数据，进行指数趋势预测。



	年份	时间代码	人均GDP	预测值	残差
1	2000	1	7857.7	7377.37	480.33
2	2001	2	8621.7	8469.45	152.25
3	2002	3	9398.1	9723.20	-325.10
4	2003	4	10542.0	11162.55	-620.55
5	2004	5	12335.6	12814.96	-479.36
6	2005	6	14185.4	14711.98	-526.58
7	2006	7	16499.7	16889.83	-390.13
8	2007	8	20169.5	19390.06	779.44
9	2008	9	23707.7	22260.41	1447.29
10	2009	10	25607.5	25555.67	51.83
11	2010	11	30015.0	29338.72	676.28
12	2011	12	35197.8	33681.79	1516.01
13	2012	13	38459.5	38667.78	-208.28
14	2013	14	41907.6	44391.85	-2484.25
15	2014	15	.	50963.26	.

非线性趋势预测：多阶曲线

应用场景

时间序列呈现出几个拐点的趋势

K阶曲线函数的一般方程

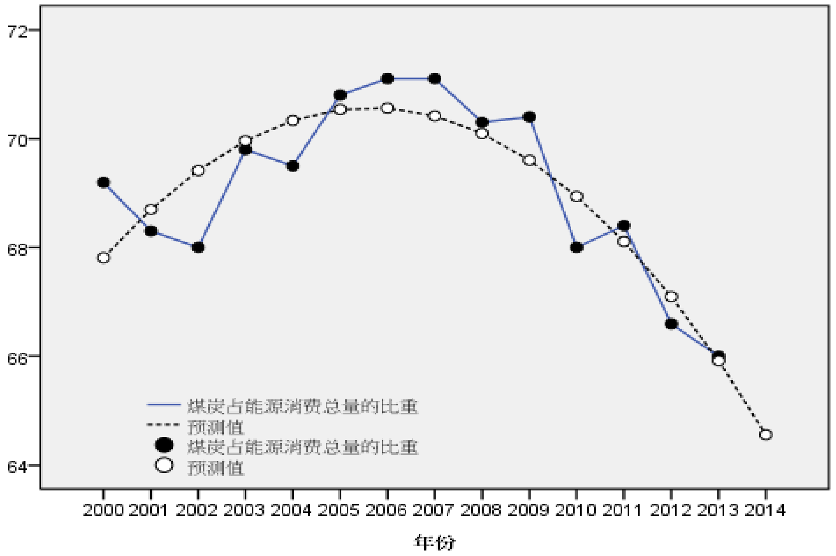
$$\hat{Y}_t = b_0 + b_1 t + b_2 t^2 + \cdots + b_k t^k$$

说明

- 当有1个拐点时，可以拟合二阶曲线；
- 当有2个拐点时，需要拟合三阶曲线；
- 当有k-1个拐点时，需要拟合k阶曲线

例子

根据右表中的煤炭占能源消费比重的数据，进行多阶曲线预测。



	年份	时间代码	煤炭占能源消费总量的比重	预测值	残差
1	2000	1	69.2	67.81	1.39
2	2001	2	68.3	68.70	-.40
3	2002	3	68.0	69.42	-1.42
4	2003	4	69.8	69.96	-.16
5	2004	5	69.5	70.33	-.83
6	2005	6	70.8	70.53	.27
7	2006	7	71.1	70.56	.54
8	2007	8	71.1	70.42	.68
9	2008	9	70.3	70.10	.20
10	2009	10	70.4	69.61	.79
11	2010	11	68.0	68.94	-.94
12	2011	12	68.4	68.11	.29
13	2012	13	66.6	67.10	-.50
14	2013	14	66.0	65.92	.08
15	2014	15	.	64.56	.

9.6 复合型序列的分解预测

- 复合型序列的定义与预测
- 复合型序列的预测：例子

复合型序列的定义与预测

定义

复合型序列，是指含有趋势、季节、周期和随机成分的时间序列。

常用的分解模型

$$\hat{Y}_t = T_t \times S_t \times I_t$$

说明

- 周期因素涉及多个年度，通常不考虑

预测步骤

1. **确定并分离季节成分**
   - 计算季节指数，确定时间序列中的季节成分
   - 将季节成分从时间序列中分离出去，即用每一个观测值除以相应的季节指数，以消除季节性
2. **建立预测模型并进行预测**
   - 对消除季节成分的序列建立适当的预测模型，并根据这一模型进行预测
3. **计算出最后的预测值**
   - 用预测值乘以相应的季节指数，得到最终的预测值

复合型序列的预测：例子

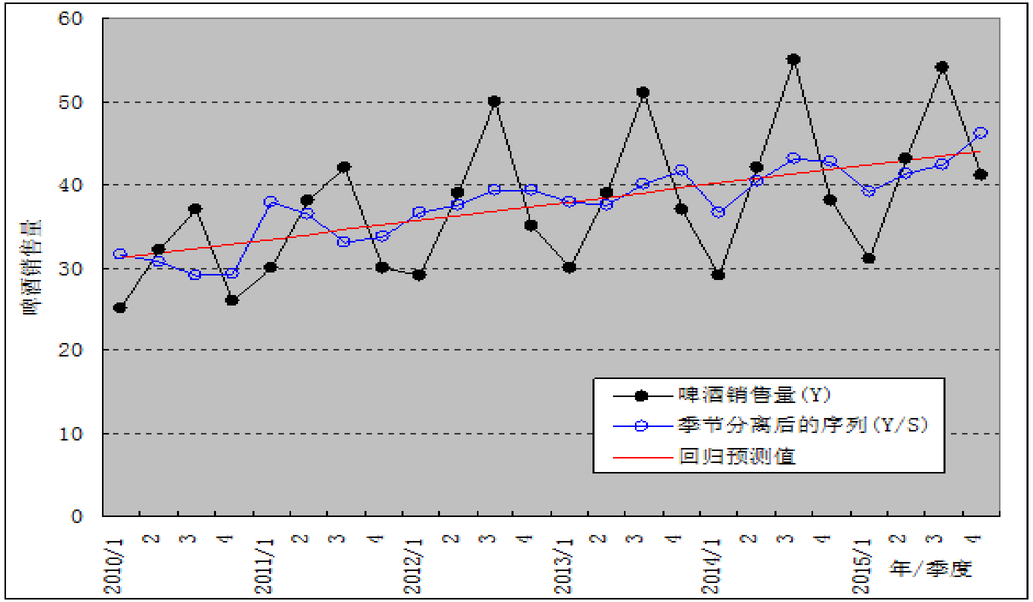
例子：下表是某家啤酒生产企业2010-2015年各个季度的啤酒销售数据。

	A	B	C	D	E
1	年份	季　度			
2		1	2	3	4
3	2010	25	32	37	26
4	2011	30	38	42	30
5	2012	29	39	50	35
6	2013	30	39	51	37
7	2014	29	42	55	38
8	2015	31	43	54	41

季节指数

	A	B	C	D	E
1	年份	季　度			
2		1	2	3	4
3	2010	—	—	1.2082	0.8125
4	2011	0.8989	1.1014	1.2043	0.8602
5	2012	0.8056	1.0365	1.3029	0.9091
6	2013	0.7767	1.0000	1.3035	0.9397
7	2014	0.7205	1.0275	1.3333	0.9129
8	2015	0.7447	1.0269	—	—
9	合计	3.9464	5.1924	6.3522	4.4344
10	平均	0.7893	1.0385	1.2704	0.8869
11	**季节指数**	**0.7922**	**1.0424**	**1.2752**	**0.8902**

季节性及其分解



本章小结

- 时间序列的分析与预测
  - 概念
    - 定义
    - 时间序列的成分
      - 趋势 T
      - 季节性 S
      - 周期性 C
      - 随机性 I
  - 分类
    - 平稳时间序列(I)
    - 非平稳序列
      - 有趋势的序列(T)
      - 有趋势和季节性的序列(T*S)
      - 复合型序列 （T*S*I)
  - 描述型分析
    - 图形描述
    - 增长率分析
      - 环比增长率
      - 定基增长率
  - 预测方法
    - 平滑预测法
      - 简单平均法
      - 移动平均法
      - 指数平滑法
    - 季节预测法
    - 趋势预测法
      - 线性趋势预测
      - 非线性趋势预测
        - 指数预测法
        - 多阶曲线预测
  - 预测方法评估
    - 平均误差
    - 平均绝对误差
    - 均方误差
    - 平均百分比误差
    - 平均绝对百分比误差