法律声明

- 本课件包括：演示文稿，示例，代码，题库，视频和声音等，小象学院拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意，我们将保留一切通过法律手段追究违反者的权利。

- 课程详情请咨询
  - 微信公众号：小象
  - 新浪微博：ChinaHadoop

第五章 非平稳序列的随机分析

主讲教师 周仕君

本章结构

- 差分运算
- ARIMA模型
- Auto-Regressive模型
- 异方差的性质
- 方差齐性变化
- 条件异方差模型

5.1 差分运算

- 差分运算的实质
- 差分方式的选择
- 过差分

差分运算的实质

- 差分方法是一种非常简便、有效的确定性信息提取方法
- Cramer分解定理在理论上保证了适当阶数的差分一定可以充分提取确定性信息
- 差分运算的实质是使用自回归的方式提取确定性信息

$$\nabla^d x_t = (1-B)^d x_t = \sum_{i=0}^{d} (-1)^i C_d^i x_{t-i}$$

差分方式的选择

- 序列蕴含着显著的线性趋势，一阶差分就可以实现趋势平稳
- 序列蕴含着曲线趋势，通常低阶（二阶或三阶）差分就可以提取出曲线趋势的影响
- 对于蕴含着固定周期的序列进行步长为周期长度的差分运算，通常可以较好地提取周期信息

例5.1

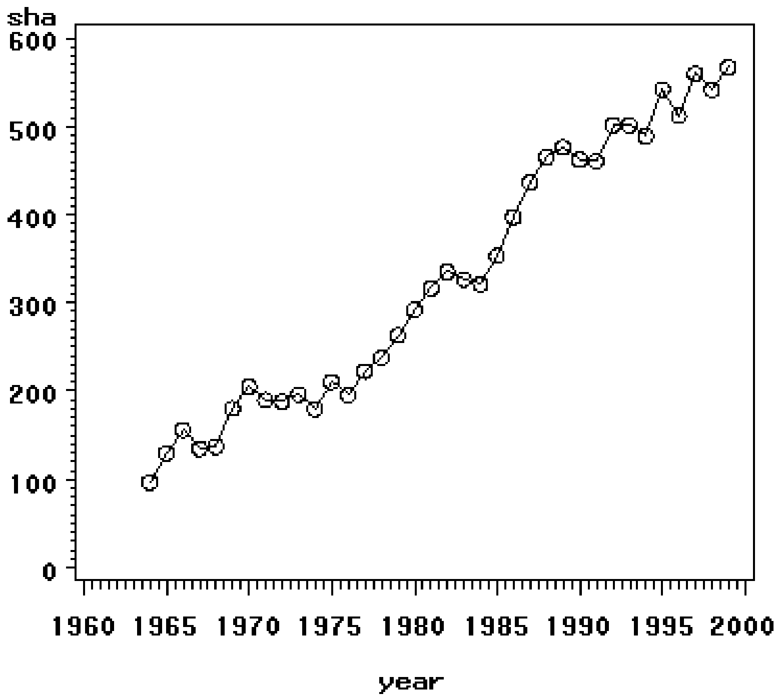
【例1.1】1964年——1999年中国纱年产量序列蕴含着一个近似线性的递增趋势。对该序列进行一阶差分运算

$$\nabla x_t = x_t - x_{t-1}$$

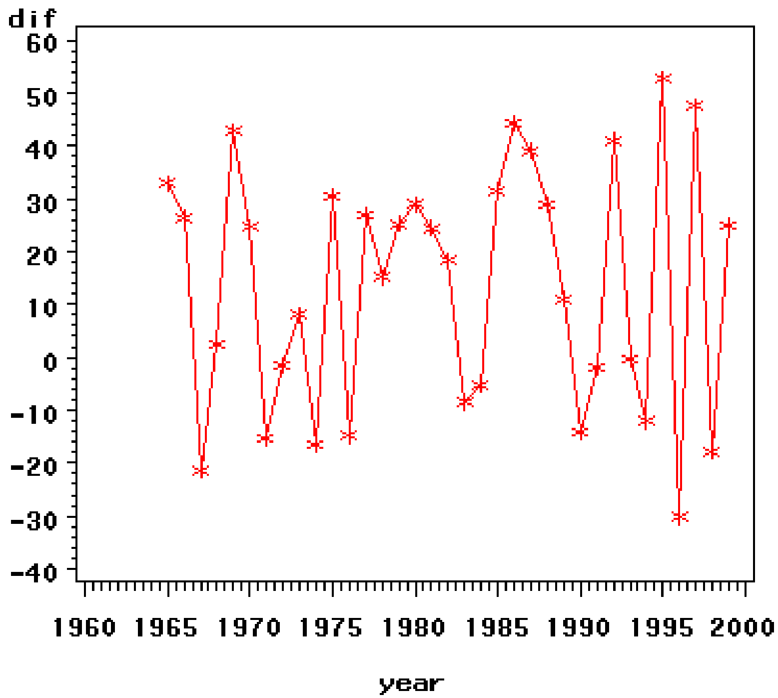
考察差分运算对该序列线性趋势信息的提取作用

差分前后时序图

- 原序列时序图

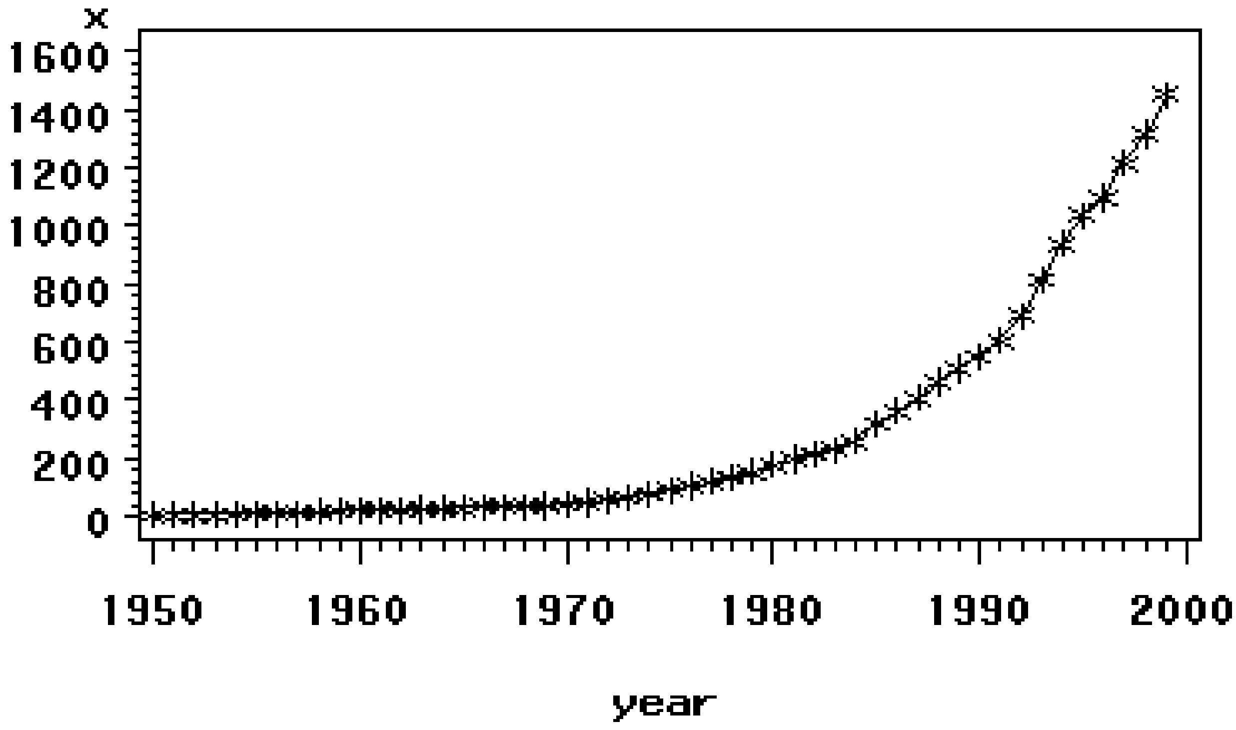


- 差分后序列时序图



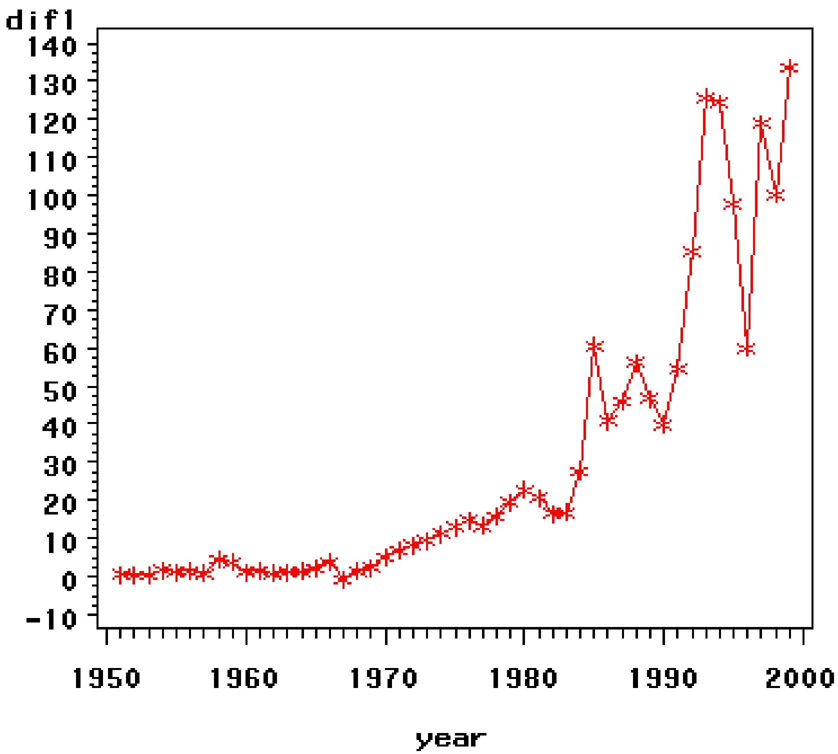
例5.2

- 尝试提取1950年——1999年北京市民用车辆拥有量序列的确定性信息

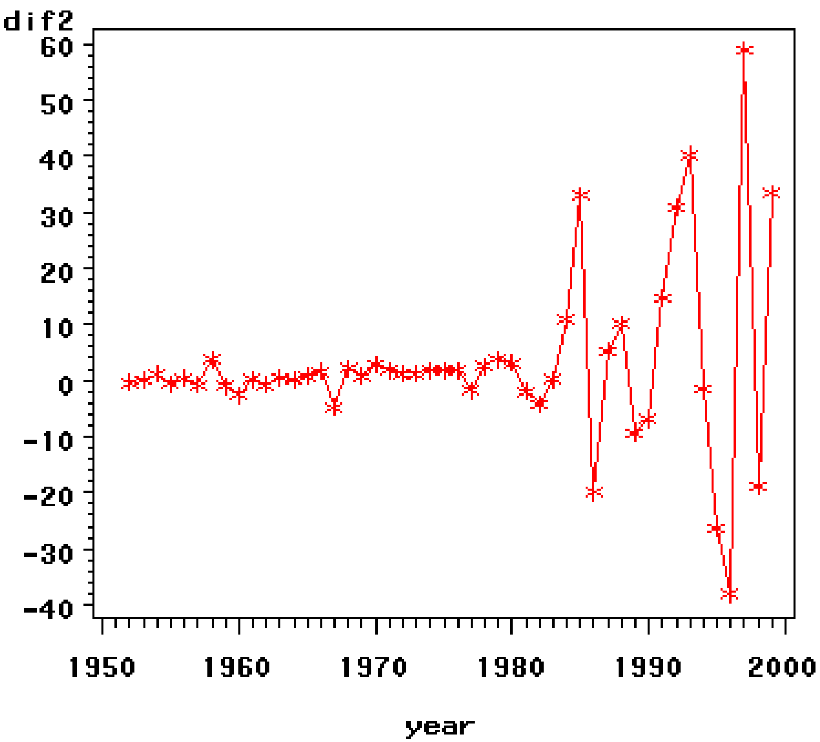


差分后序列时序图

- 一阶差分

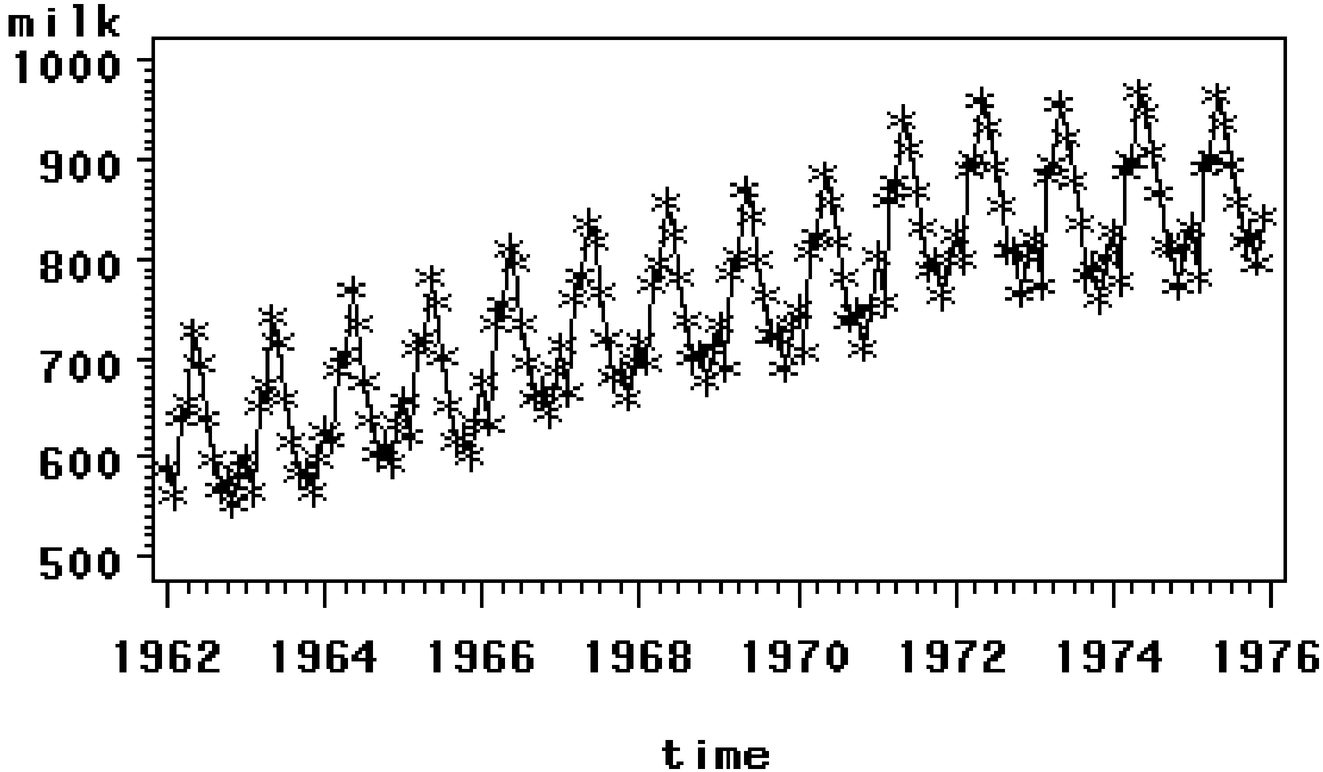


- 二阶差分



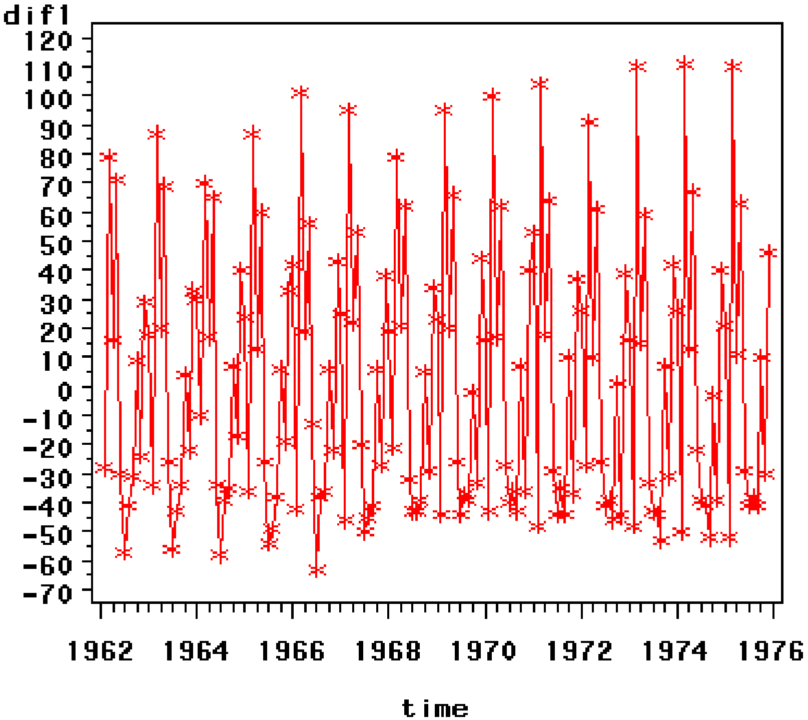
例5.3

- 差分运算提取1962年1月——1975年12月平均每头奶牛的月产奶量序列中的确定性信息

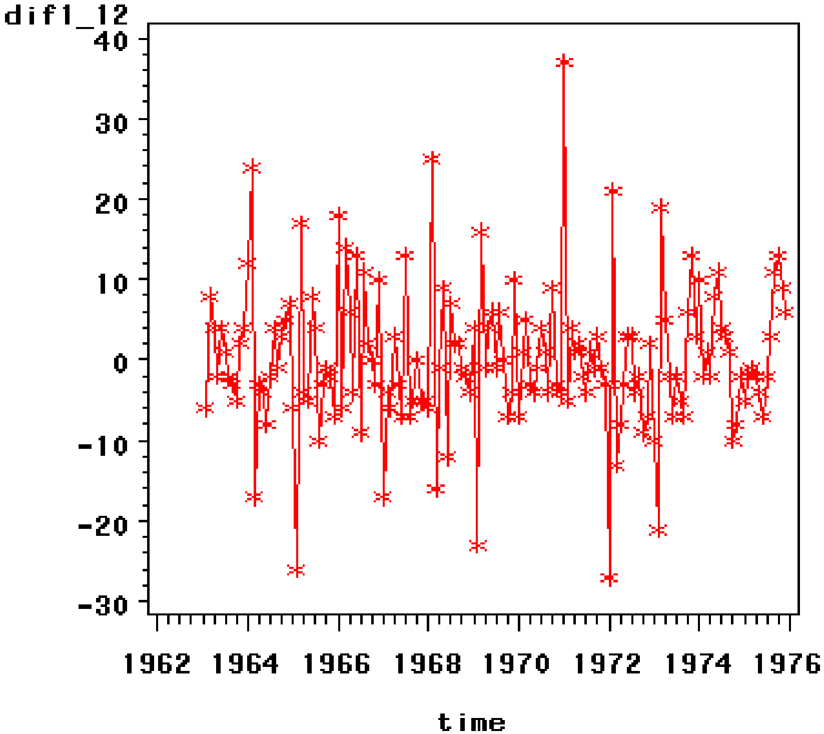


差分后序列时序图

- 一阶差分



- 1阶－12步差分



过差分

- 足够多次的差分运算可以充分地提取原序列中的非平稳确定性信息
- 但过度的差分会造成有用信息的浪费

例5.4

- 假设序列如下 $x_t = \beta_0 + \beta_1 t + a_t$
- 考察一阶差分后序列和二阶差分序列的平稳性与方差

比较

- 一阶差分
  - 平稳

$$\nabla x_t = x_t - x_{t-1}$$
$$= \beta_1 + a_t - a_{t-1}$$

  - 方差小

$$Var(\nabla x_t) = Var(a_t - a_{t-1})$$
$$= 2\sigma^2$$

- 二阶差分（过差分）
  - 平稳

$$\nabla^2 x_t = \nabla x_t - \nabla x_{t-1}$$
$$= a_t - 2a_{t-1} + a_{t-2}$$

  - 方差大

$$Var(\nabla^2 x_t) = Var(a_t - 2a_{t-1} + a_{t-2})$$
$$= 6\sigma^2$$

5.2 ARIMA模型

- ARIMA模型结构
- ARIMA模型性质
- ARIMA模型建模
- ARIMA模型预测
- 疏系数模型
- 季节模型

ARIMA模型结构

- 使用场合
  - 差分平稳序列拟合
- 模型结构

$$
\begin{cases}
\Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \\
E(\varepsilon_t) = 0, \quad Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t\varepsilon_s) = 0, s \neq t \\
Ex_s\varepsilon_t = 0, \forall s < t
\end{cases}
$$

ARIMA 模型族

- d=0

  ARIMA(p,d,q)=ARMA(p,q)

- P=0

  ARIMA(P,d,q)=IMA(d,q)

- q=0

  ARIMA(P,d,q)=ARI(p,d)

- d=1,P=q=0

  ARIMA(P,d,q)=random walk model

随机游走模型(random walk)

- 模型结构

$$\begin{cases} x_t = x_{t-1} + \varepsilon_t \\ E(\varepsilon_t) = 0, \quad Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ Ex_s \varepsilon_t = 0, \forall s < t \end{cases}$$

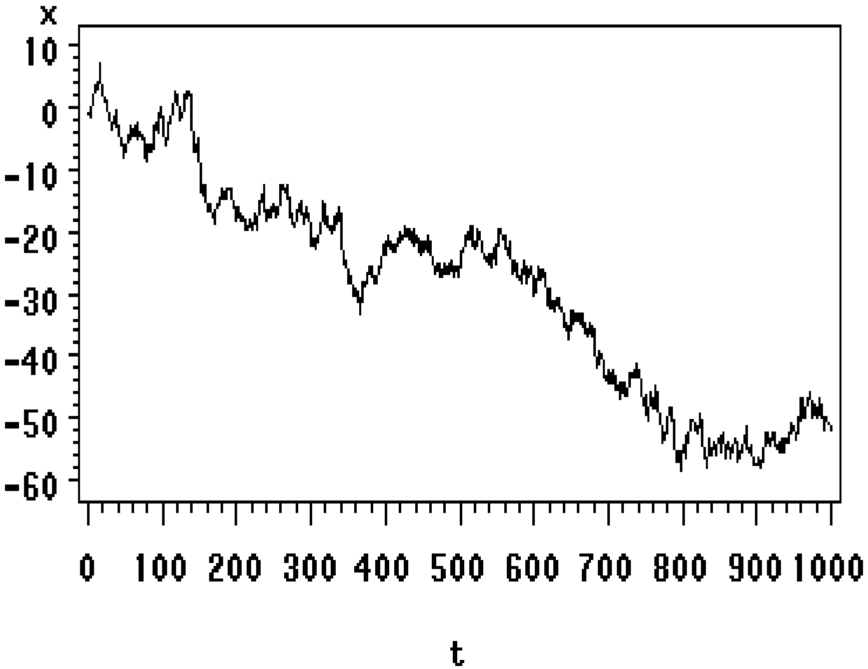
- 模型产生典故
  - Karl Pearson(1905)在《自然》杂志上提问：假如有个醉汉醉得非常严重，完全丧失方向感，把他放在荒郊野外，一段时间之后再去找他，在什么地方找到他的概率最大呢？

ARIMA模型的平稳性

- ARIMA(p,d,q)模型共有p+d个特征根，其中p个在单位圆内，d个在单位圆上。
- 所以当 $d \neq 0$ 时ARIMA(p,d,q)模型非平稳。

- 例5.5

ARIMA(0,1,0)时序图



ARIMA模型的方差齐性

- $d \neq 0$时，原序列方差非齐性

$$ARIMA(0,1,0)\text{模型}$$

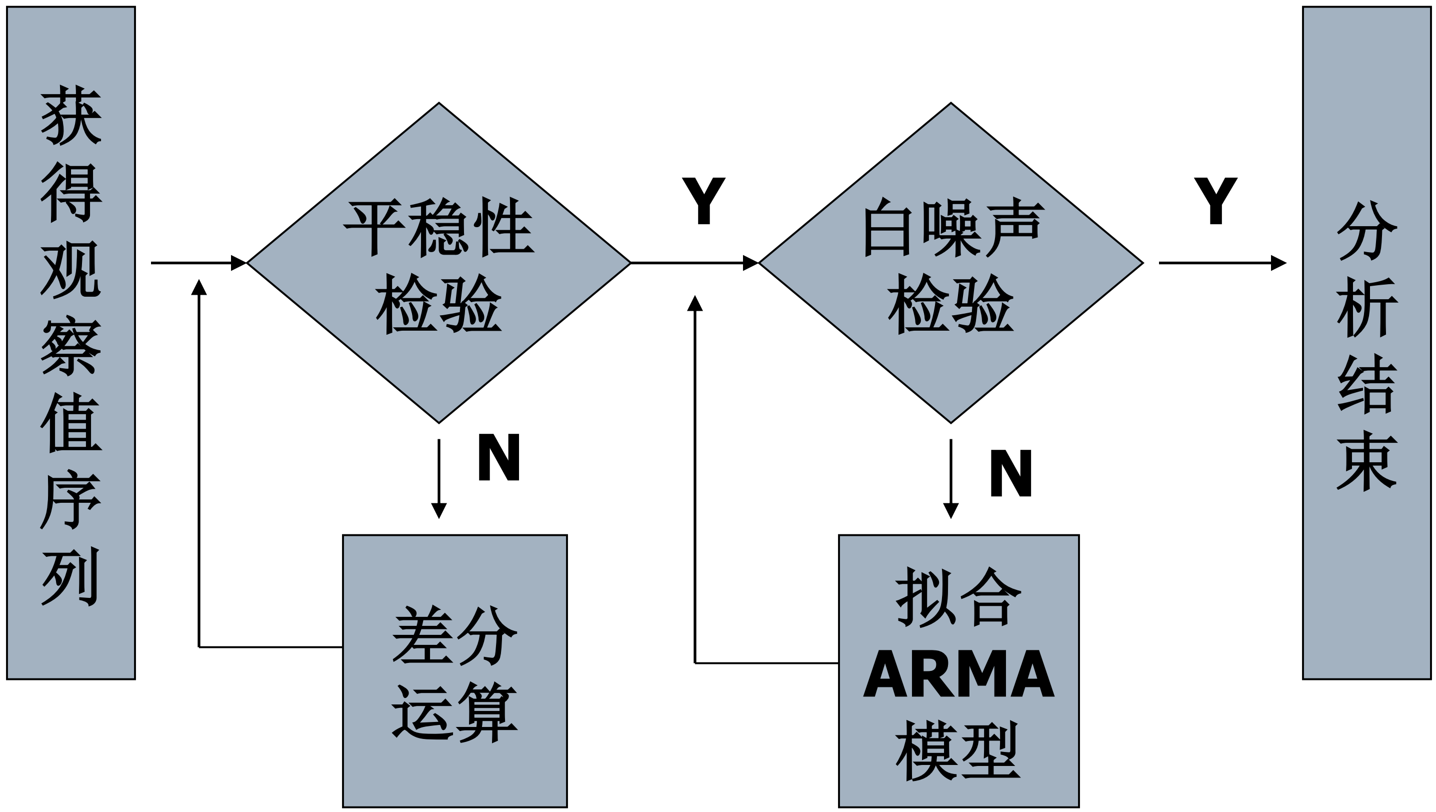
$$Var(x_t) = Var(x_0 + \varepsilon_t + \varepsilon_{t-1} + \cdots \varepsilon_1) = t\sigma_\varepsilon^2$$

- d阶差分后，差分后序列方差齐性

$$ARIMA(0,1,0)\text{模型}$$

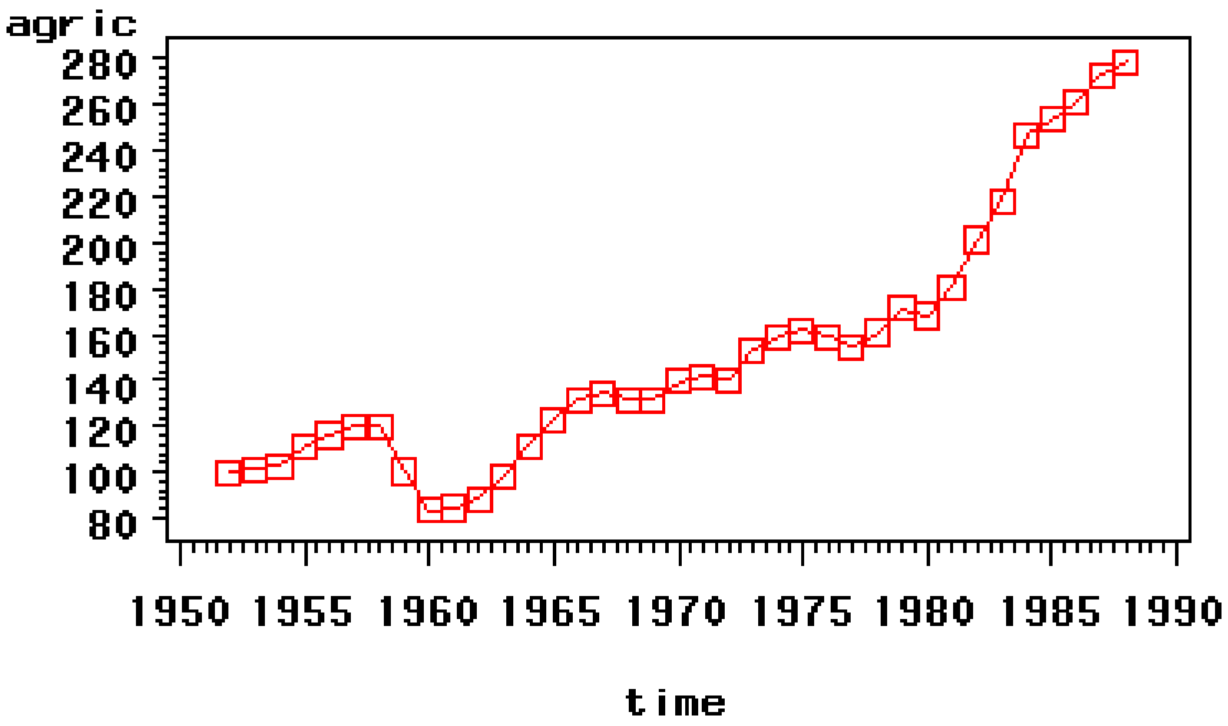
$$Var(\nabla x_t) = Var(\varepsilon_t) = \sigma_\varepsilon^2$$

ARIMA模型建模步骤

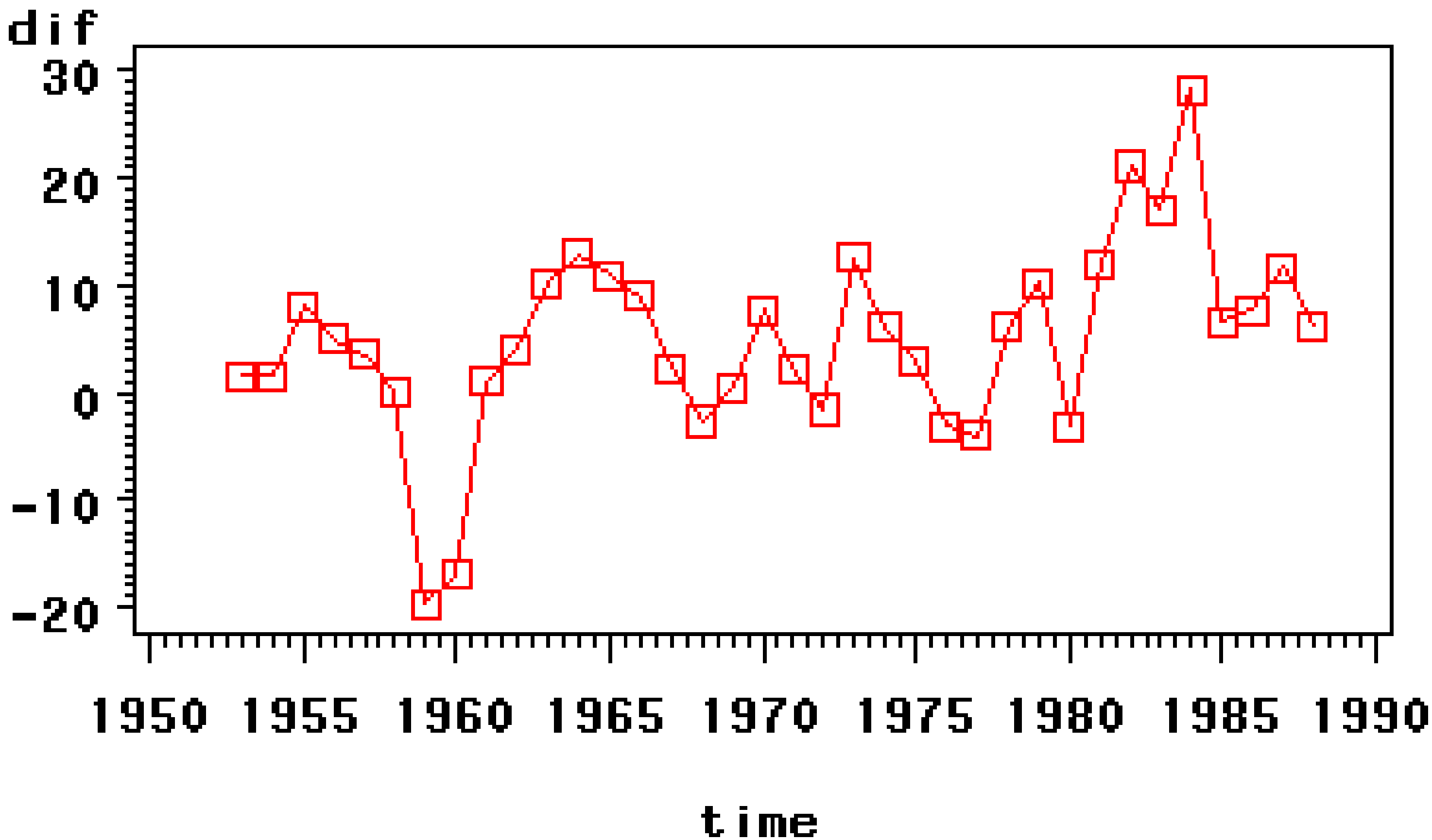


例5.6

- 对1952年——1988年中国农业实际国民收入指数序列建模



一阶差分序列时序图



dif
30
20
10
0
-10
-20
1950
1955
1960
1965
1970
1975
1980
1985
1990
time

一阶差分序列自相关图

```
                                      Autocorrelations

Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

  0     78.239167        1.00000     |                    |********************|
  1     42.075733        0.53778     |             .      |***********         |
  2     16.246605        0.20765     |            .       |****   .            |
  3      7.058588        0.09022     |           .        |**      .           |
  4    -11.132207        -.14228     |           .     ***|        .           |
  5     -7.917076        -.10119     |           .      **|        .           |
  6     -9.245185        -.11817     |           .      **|        .           |
  7    -11.564313        -.14781     |           .     ***|        .           |
  8      7.108735        0.09086     |           .        |**      .           |
  9     12.965116        0.16571     |           .        |***     .           |
 10     -0.909105        -.01162     |           .        |        .           |
 11     -2.455085        -.03138     |           .       *|        .           |
 12     -3.501852        -.04476     |           .       *|        .           |
 13     -6.583063        -.08414     |           .      **|        .           |
 14     -7.883765        -.10076     |           .      **|        .           |
 15     -4.783310        -.06114     |           .       *|        .           |
 16      2.087515        0.02668     |           .        |*       .           |
 17     12.894776        0.16481     |           .        |***     .           |
 18     15.631250        0.19979     |           .        |****    .           |

                            "." marks two standard errors
```

一阶差分后序列白噪声检验

延迟阶数	χ^2统计量	P值
6	15.33	0.0178
12	18.33	0.1060
18	24.66	0.1344

拟合ARMA模型

偏自相关图

```
                           Partial Autocorrelations

     Lag     Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

       1         0.53778    |             .      |***********         |
       2        -0.11474    |             .    **|      .             |
       3         0.03912    |             .      |*     .             |
       4        -0.27003    |             . *****|      .             |
       5         0.16219    |             .      |***   .             |
       6        -0.17787    |             .  ****|      .             |
       7         0.03051    |             .      |*     .             |
       8         0.21004    |             .      |****  .             |
       9         0.02902    |             .      |*     .             |
      10        -0.25795    |             . *****|      .             |
      11         0.04421    |             .      |*     .             |
      12         0.04346    |             .      |*     .             |
      13        -0.03857    |             .     *|      .             |
      14        -0.15591    |             .   ***|      .             |
      15         0.21892    |             .      |****  .             |
      16         0.00855    |             .      |      .             |
      17         0.05496    |             .      |*     .             |
      18         0.01825    |             .      |      .             |
```

建模

- 定阶
  - ARIMA(0,1,1)
- 参数估计

$$(1-B)x_t = 4.99661 + (1 + 0.70766B)\varepsilon_t$$

$$Var(\varepsilon_t) = 56.48763$$

- 模型检验
  - 模型显著
  - 参数显著

ARIMA模型预测

- 原则
  - 最小均方误差预测原理
- Green函数递推公式

$$
\begin{cases}
\psi_1 = \phi_1 - \theta_1 \\
\psi_2 = \phi_1 \psi_1 + \phi_2 - \theta_2 \\
\vdots \\
\psi_j = \phi_1 \psi_{j-1} + \cdots + \phi_{p+d} \psi_{j-p-d} - \theta_j
\end{cases}
$$

预测值

$$x_{t+l} = (\varepsilon_{t+l} + \psi_1\varepsilon_{t+l-1} + \cdots + \psi_{l-1}\varepsilon_{t+1}) + (\psi_l\varepsilon_t + \psi_{l+1}\varepsilon_{t-1} + \cdots)$$

$$e_t(l) \qquad\qquad \hat{x}_t(l)$$

$$E[e_t(l)] = 0$$

$$Var[e_t(l)] = (1 + \psi_1^2 + \cdots + \psi_{l-1}^2)\sigma_\varepsilon^2$$

例5.7

- 已知ARIMA(1,1,1)模型为

$$(1-0.8B)(1-B)x_t = (1-0.6B)\varepsilon_t$$

且 $x_{t-1} = 4.5$ $x_t = 5.3$ $\varepsilon_t = 0.8$ $\sigma_\varepsilon^2 = 1$

- 求 x_{t+3} 的95%的置信区间

预测值

- 等价形式

$$(1-1.8B+0.8B^2)x_t=(1-0.6B)\varepsilon_t$$

$$x_t=1.8x_{t-1}-0.8x_{t-2}+\varepsilon_t-0.6\varepsilon_{t-1}$$

- 计算预测值

$$\hat{x}_t(1)=1.8x_t-0.8x_{t-1}-0.6\varepsilon_t=5.46$$

$$\hat{x}_t(2)=1.8\hat{x}_t(1)-0.8x_t=5.59$$

$$\hat{x}_t(3)=1.8\hat{x}_t(2)-0.8\hat{x}_t(1)=5.69$$

计算置信区间

- Green函数值

$$\begin{cases} \psi_1 = 1.8 - 0.6 = 1.2 \\ \psi_2 = 1.8\psi_1 - 0.8 = 1.36 \end{cases}$$

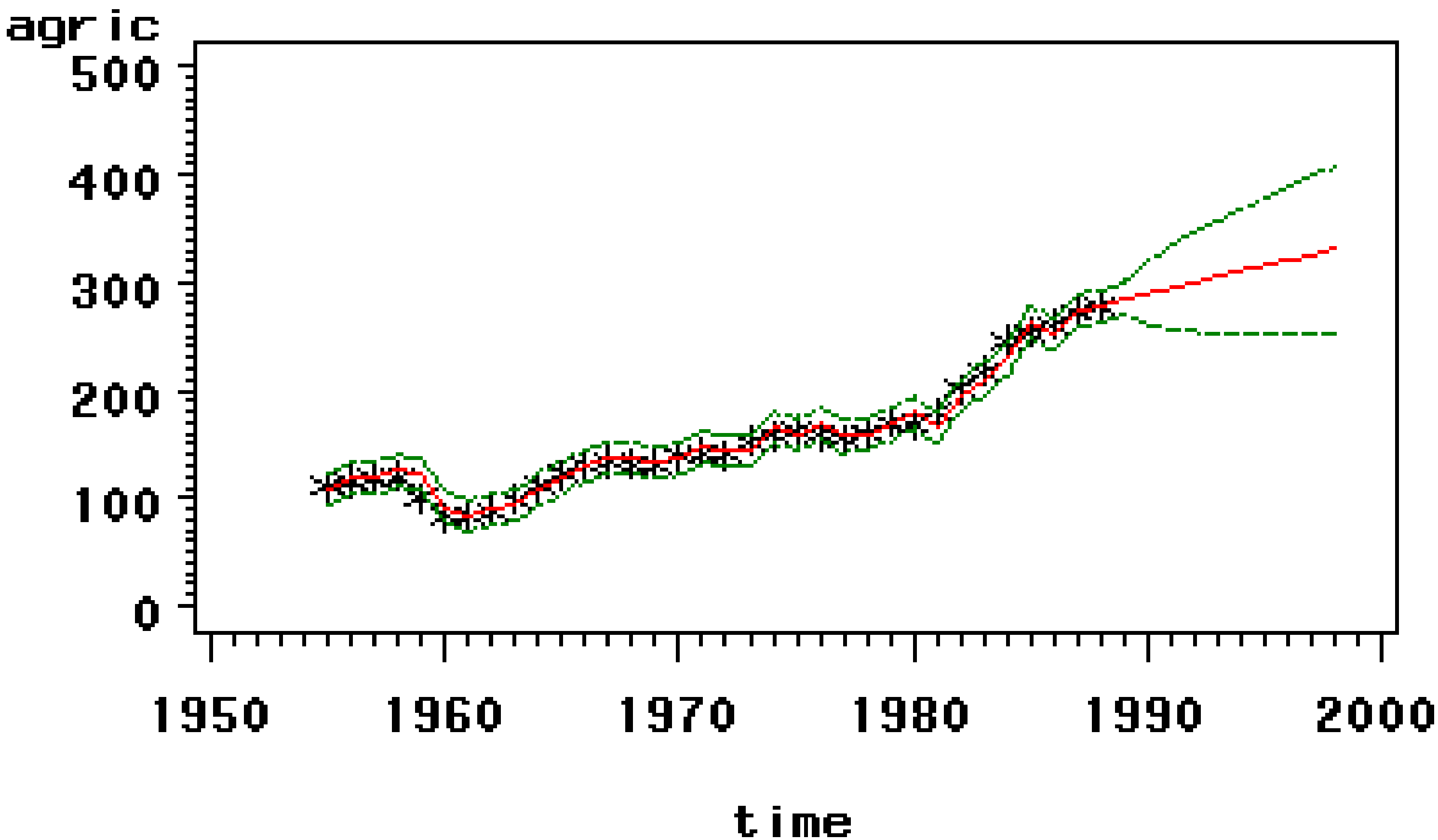
- 方差 $Var[e(3)] = (1 + \psi_1^2 + \psi_2^2)\sigma_\varepsilon^2 = 4.2896$

- 95%置信区间

$$(\hat{x}_t(3) - 1.96\sqrt{Var(e(3))}\ ,\ \hat{x}_t(3) + 1.96\sqrt{Var(e(3))}\,)$$

$$\Rightarrow (1.63, 9.75)$$

例5.6续：对中国农业实际国民收入指数序列做为期10年的预测



疏系数模型

- ARIMA(p,d,q)模型是指d阶差分后自相关最高阶数为p，移动平均最高阶数为q的模型，通常它包含p+q个独立的未知系数：

$$\phi_1,\cdots,\phi_p,\theta_1,\cdots,\theta_q$$

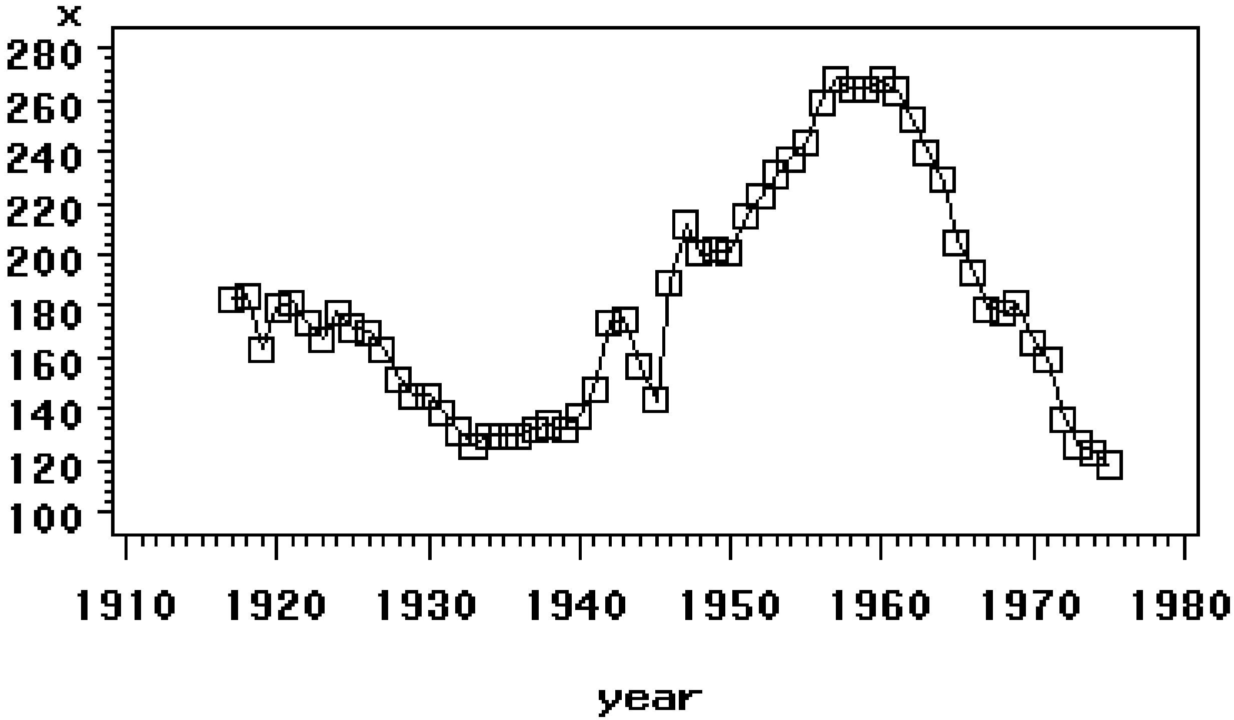
- 如果该模型中有部分自相关系数 $\phi_j, 1 \le j < p$ 或部分移动平滑系数 $\theta_k, 1 \le k < q$ 为零，即原模型中有部分系数省缺了，那么该模型称为疏系数模型。

疏系数模型类型

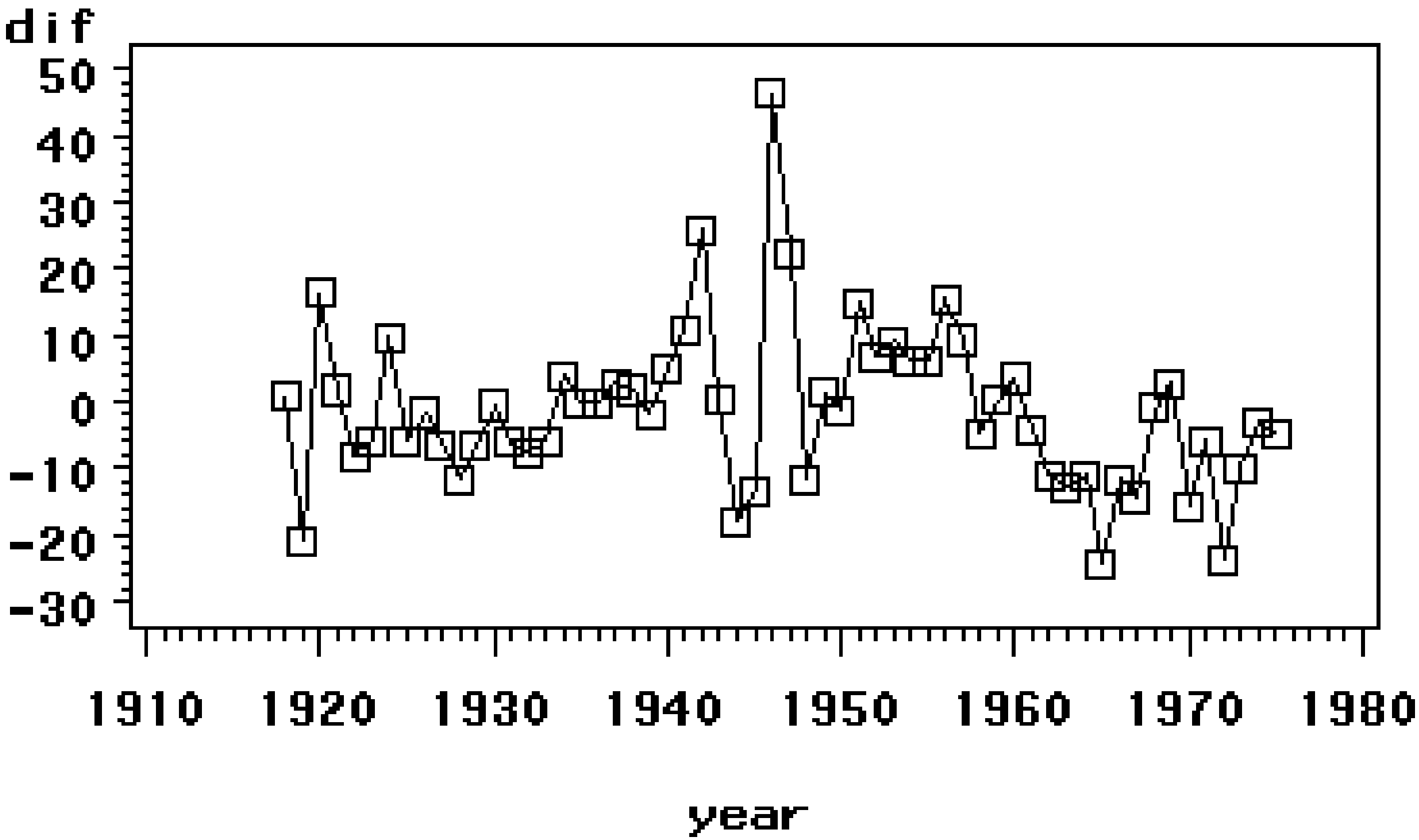
- 如果只是自相关部分有省缺系数，那么该疏系数模型可以简记为 $ARIMA((p_1,\cdots,p_m),d,q)$
  - $p_1,\cdots,p_m$ 为非零自相关系数的阶数
- 如果只是移动平滑部分有省缺系数，那么该疏系数模型可以简记为 $ARIMA(p,d,(q_1,\cdots,q_n))$
  - $q_1,\cdots,q_n$ 为非零移动平均系数的阶数
- 如果自相关和移动平滑部分都有省缺，可以简记为 $ARIMA((p_1,\cdots,p_m),d,(q_1,\cdots,q_n))$

例5.8

- 对1917年－1975年美国23岁妇女每万人生育率序列建模



一阶差分


dif
50
40
30
20
10
0
-10
-20
-30
1910
1920
1930
1940
1950
1960
1970
1980
year

自相关图

```
                                           Autocorrelations

Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

  0      145.928        1.00000     |                    |********************|
  1    41.471097        0.28419     |               .    |******              |
  2    -6.153942        -.04217     |              .    *|     .              |
  3    11.679824        0.08004     |              .     |**   .              |
  4    47.775344        0.32739     |              .     |*******             |
  5    47.552927        0.32586     |              .     |*******             |
  6    21.691431        0.14864     |             .      |***   .             |
  7     9.102516        0.06238     |             .      |*     .             |
  8     3.544593        0.02429     |             .      |      .             |
  9    12.658538        0.08674     |             .      |**    .             |
 10    13.544854        0.09282     |             .      |**    .             |
 11    -5.308811        -.03638     |             .     *|      .             |
 12   -10.154165        -.06958     |             .     *|      .             |
 13    -6.521036        -.04469     |             .     *|      .             |
 14    -7.235034        -.04958     |             .     *|      .             |
 15   -16.998688        -.11649     |             .    **|      .             |
 16   -18.817139        -.12895     |             .   ***|      .             |
 17   -10.170169        -.06969     |             .     *|      .             |
 18   -21.791201        -.14933     |             .   ***|      .             |

                            "." marks two standard errors
```

偏自相关图

```
                    Partial Autocorrelations

 Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

   1        0.28419    |              .    |******             |
   2       -0.13373    |              .  ***|    .              |
   3        0.14580    |              .    |*** .              |
   4        0.28405    |              .    |******             |
   5        0.19604    |              .    |****.              |
   6        0.07973    |              .    |**  .              |
   7        0.02965    |              .    |*   .              |
   8       -0.10604    |              .   **|    .              |
   9       -0.05098    |              .    *|    .              |
  10       -0.06054    |              .    *|    .              |
  11       -0.14640    |              .  ***|    .              |
  12       -0.05923    |              .    *|    .              |
  13       -0.05089    |              .    *|    .              |
  14       -0.06281    |              .    *|    .              |
  15       -0.07425    |              .    *|    .              |
  16       -0.03383    |              .    *|    .              |
  17        0.02752    |              .    |*   .              |
  18       -0.08911    |              .   **|    .              |
```

建模

- 定阶
  - ARIMA((1,4),1,0)
- 参数估计

$$(1-B)x_t = \frac{1}{1-0.26633B-0.33597B^4}\varepsilon_t$$

- 模型检验
  - 模型显著
  - 参数显著

季节模型

- 简单季节模型
- 乘积季节模型

简单季节模型

- 简单季节模型是指序列中的季节效应和其它效应之间是加法关系

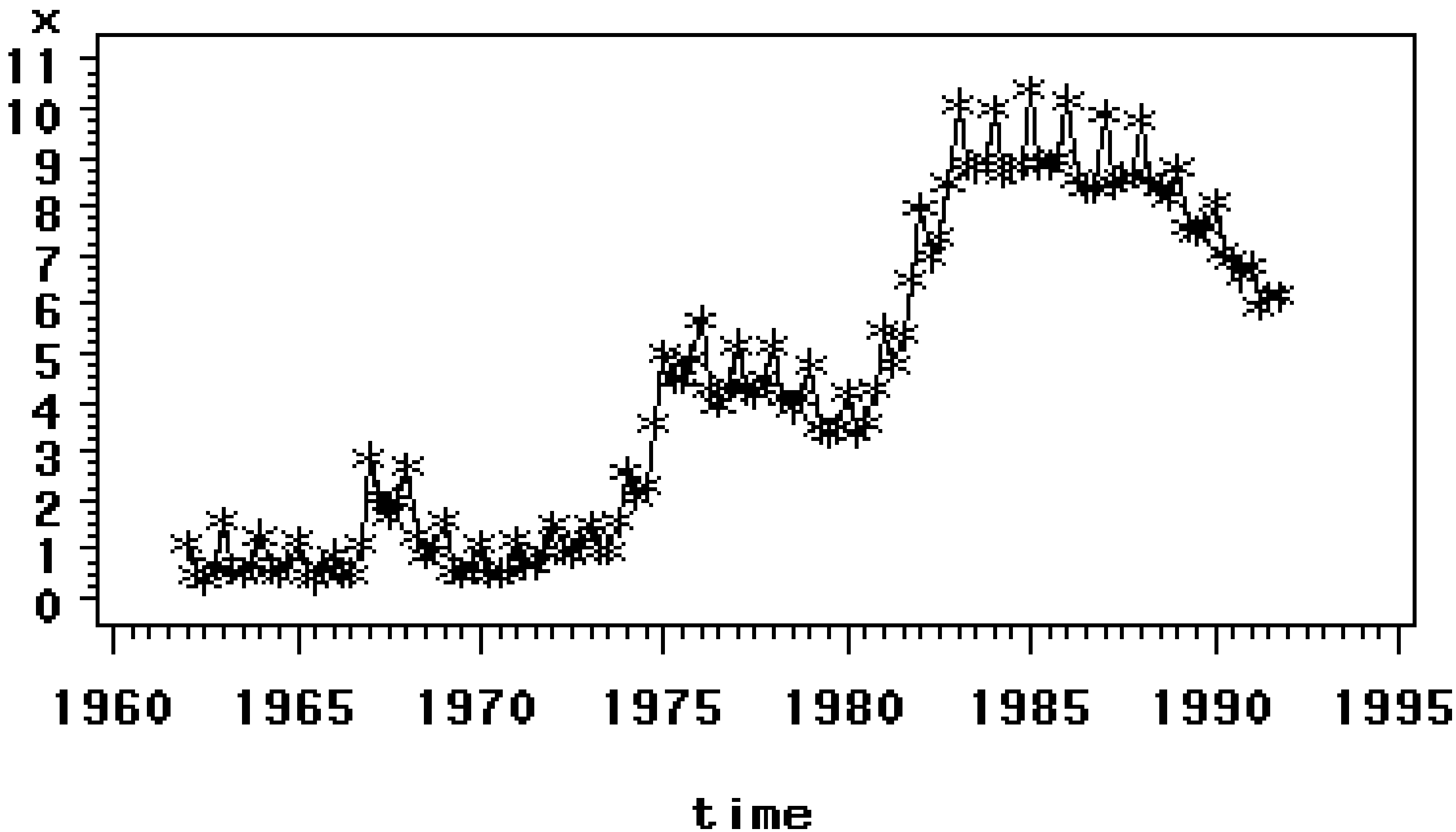
$$x_t = S_t + T_t + I_t$$

- 简单季节模型通过简单的趋势差分、季节差分之后序列即可转化为平稳，它的模型结构通常如下

$$\nabla_D \nabla^d x_t = \frac{\Theta(B)}{\Phi(B)} \varepsilon_t$$

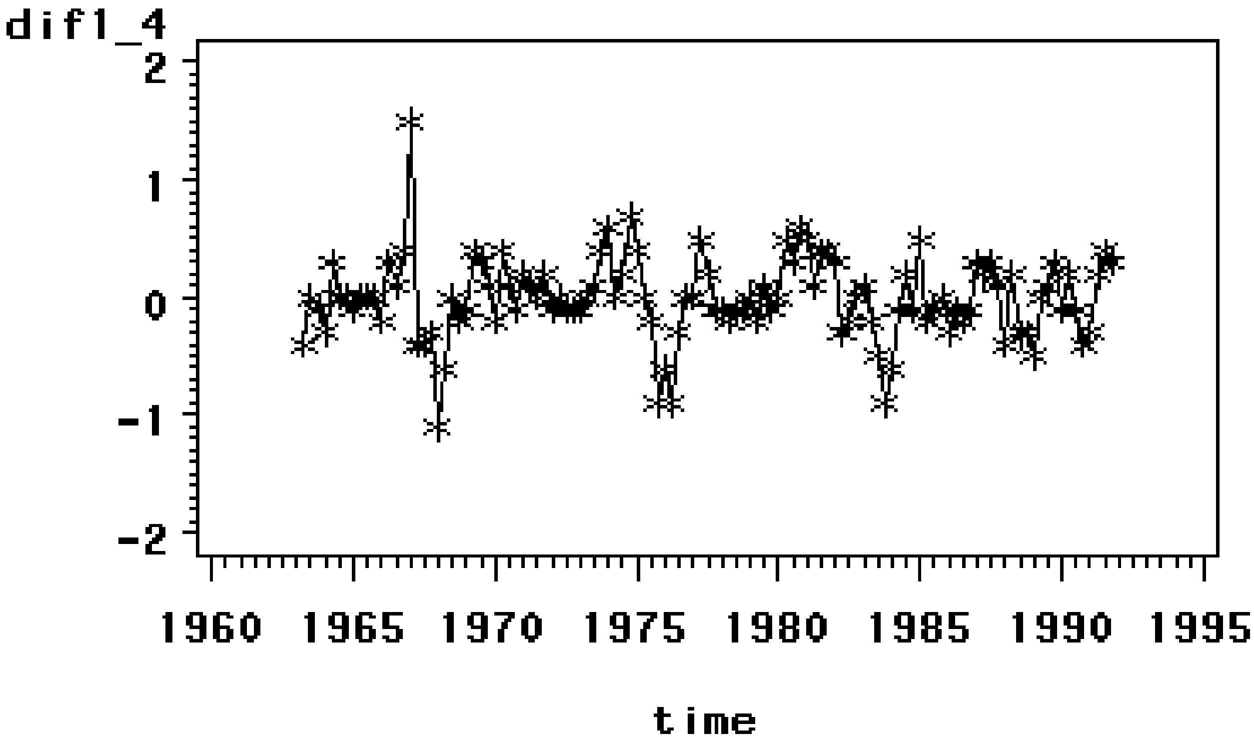
例5.9

- 拟合1962——1991年德国工人季度失业率序列



差分平稳

- 对原序列作一阶差分消除趋势，再作4步差分消除季节效应的影响，差分后序列的时序图如下



白噪声检验

延迟阶数	χ^2统计量	P值
6	43.84	<0.0001
12	51.71	<0.0001
18	54.48	<0.0001

差分后序列自相关图

```
                                   Autocorrelations

Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

  0      0.124721        1.00000    |                    |********************|
  1      0.052641        0.42207    |                .   |********            |
  2      0.023569        0.18898    |                .   |****                |
  3     0.0090404        0.07248    |                .   |*  .                |
  4     -0.029789        -.23884    |               *****|   .                |
  5     -0.029783        -.23879    |               *****|    .               |
  6     -0.022559        -.18087    |               .****|    .               |
  7     -0.024312        -.19493    |               .****|    .               |
  8     -0.011082        -.08885    |               .  **|    .               |
  9    -0.0074505        -.05974    |               .   *|    .               |
 10     0.0027160        0.02178    |               .    |    .               |
 11     0.0027155        0.02177    |               .    |    .               |
 12     -0.013672        -.10962    |               .  **|    .               |
 13    -0.0025423        -.02038    |               .    |    .               |
 14    -0.0033988        -.02725    |               .   *|    .               |
 15     -0.015005        -.12031    |               .  **|    .               |
 16     0.0021367        0.01713    |               .    |    .               |
 17     0.0070261        0.05633    |               .    |*   .               |
 18    -0.0046198        -.03704    |               .   *|    .               |

                          "." marks two standard errors
```

差分后序列偏自相关图

```
                           Partial Autocorrelations

Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

  1        0.42207     |               .   |********           |
  2        0.01319     |               .   |   .               |
  3       -0.01435     |               .   |   .               |
  4       -0.32452     |             ******|   .               |
  5       -0.03822     |               .  *|   .               |
  6       -0.01721     |               .   |   .               |
  7       -0.07052     |               .  *|   .               |
  8       -0.03361     |               .  *|   .               |
  9       -0.07938     |               . **|   .               |
 10        0.05648     |               .   |*  .               |
 11       -0.07772     |               . **|   .               |
 12       -0.19344     |               ****|   .               |
 13        0.04133     |               .   |*  .               |
 14       -0.01802     |               .   |   .               |
 15       -0.12741     |               .***|   .               |
 16        0.01261     |               .   |   .               |
 17        0.03710     |               .   |*  .               |
 18       -0.11550     |               . **|   .               |
```

模型拟合

- 定阶
  - ARIMA((1,4),(1,4),0)

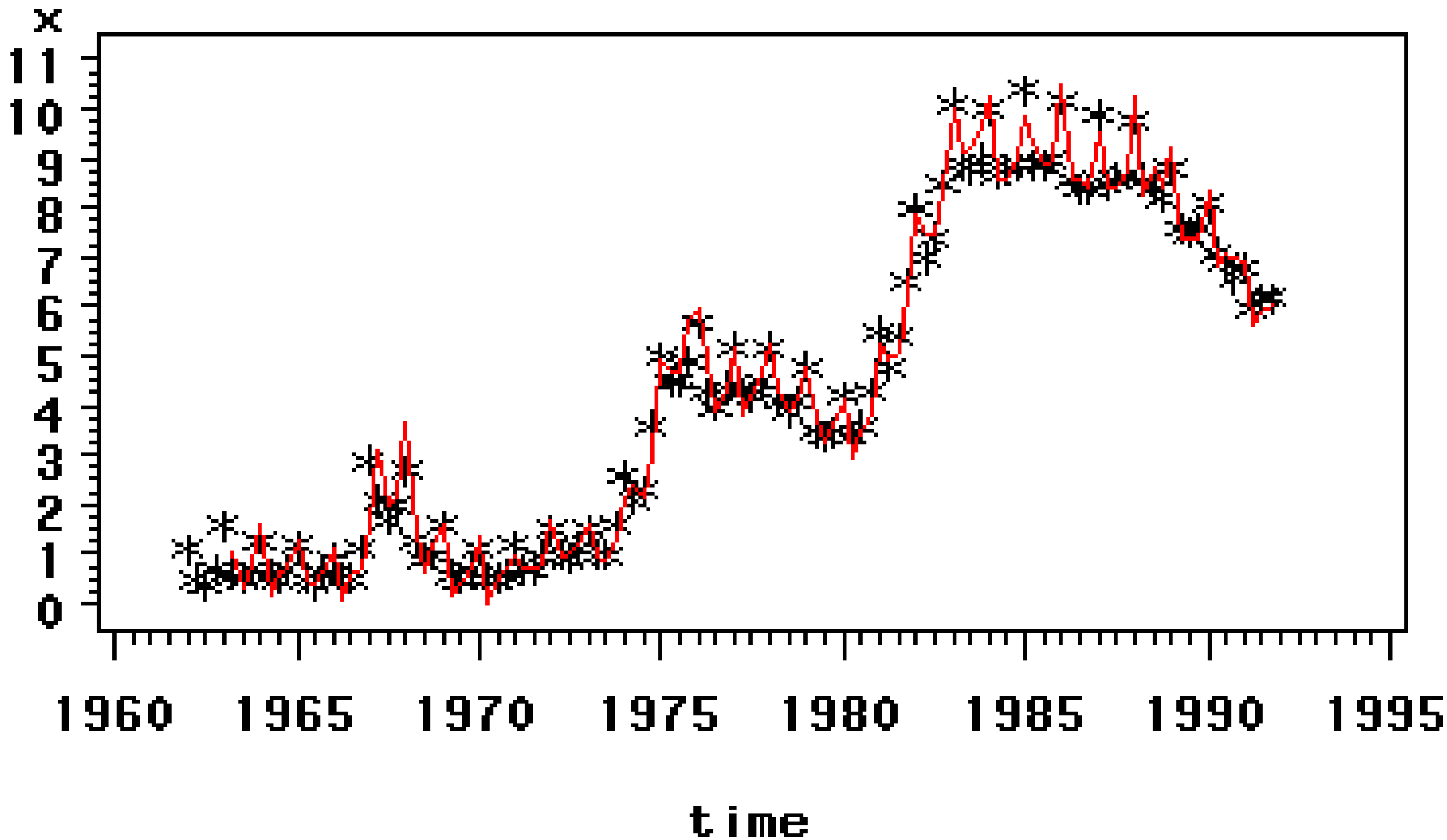
- 参数估计

$$(1-B)(1-B^4)x_t = \frac{1}{1-0.44746B+0.28132B^4}\varepsilon_t$$

模型检验

残差白噪声检验			参数显著性检验		
延迟阶数	χ^2统计量	P值	待估参数	t 统计量	P值
6	2.09	0.7191	θ_1	5.48	<0.0001
12	10.99	0.3584	θ_4	-3.41	<0.0001

拟合效果图

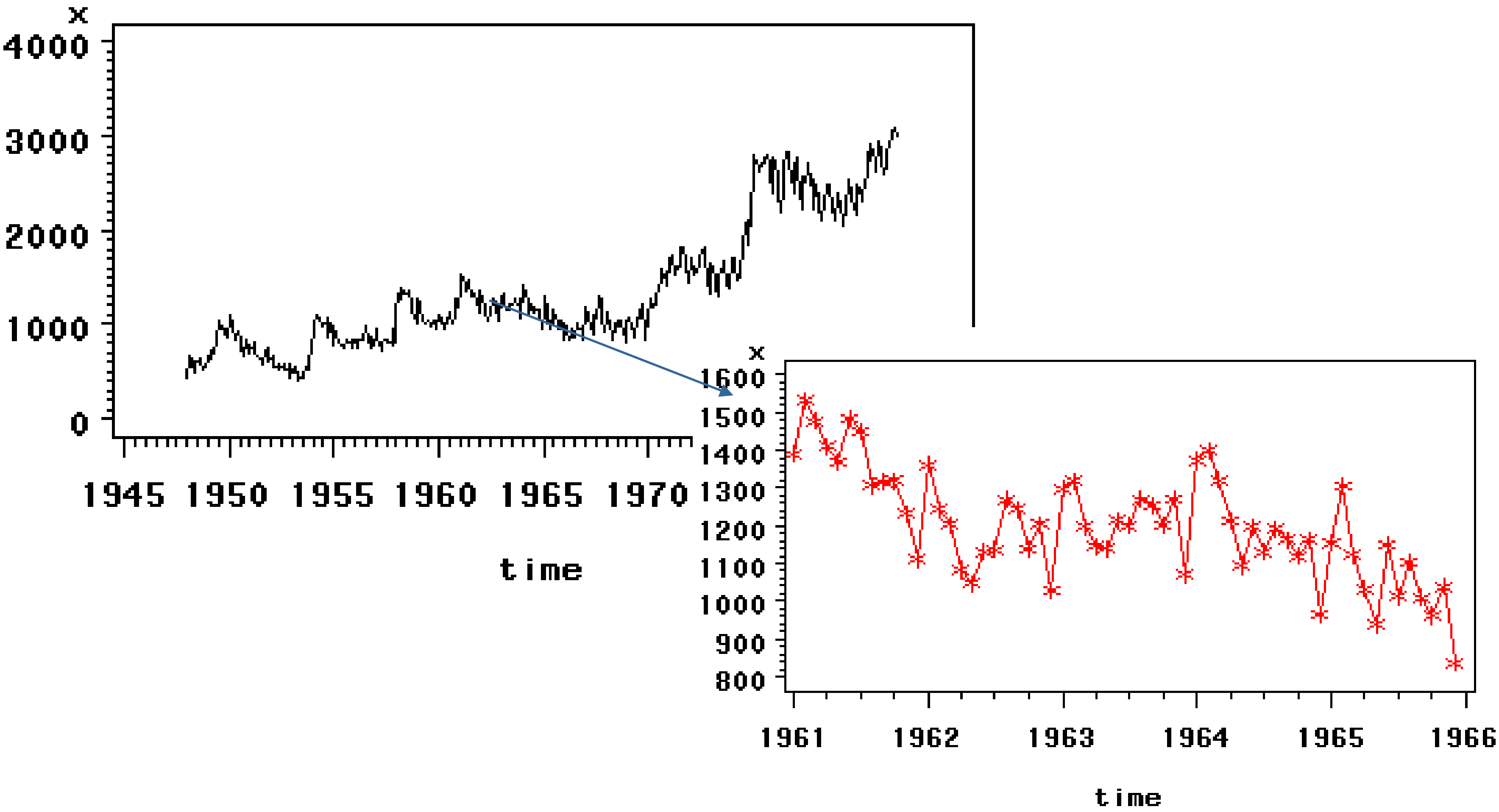

x
11
10
9
8
7
6
5
4
3
2
1
0
1960
1965
1970
1975
1980
1985
1990
1995
time

乘积季节模型

- 使用场合
  - 序列的季节效应、长期趋势效应和随机波动之间有着复杂地相互关联性，简单的季节模型不能充分地提取其中的相关关系
- 构造原理
  - 短期相关性用低阶ARMA(p,q)模型提取
  - 季节相关性用以周期步长S为单位的ARMA(P,Q)模型提取
  - 假设短期相关和季节效应之间具有乘积关系，模型结构如下

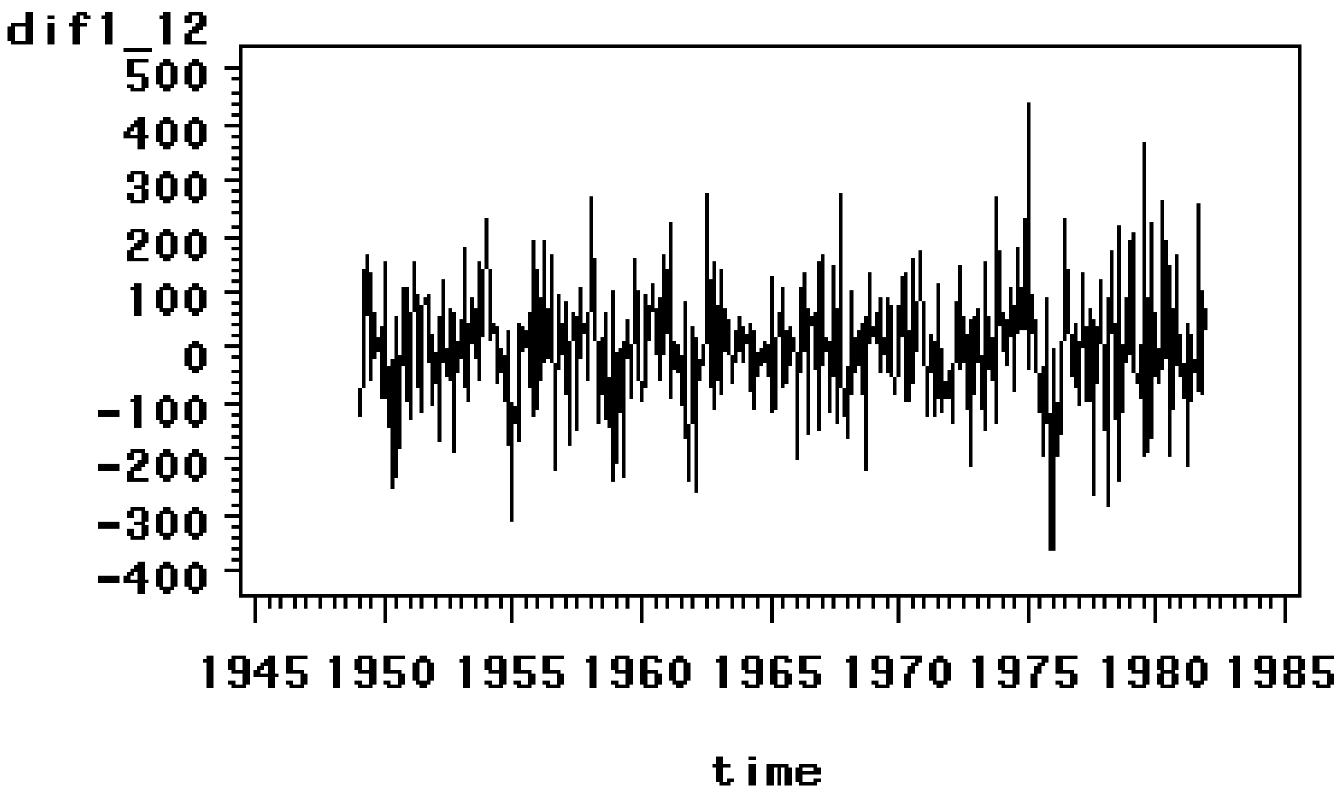
$$\nabla^d \nabla_S^D x_t = \frac{\Theta(B)}{\Phi(B)} \frac{\Theta_S(B)}{\Phi_S(B)} \varepsilon_t$$

例5.10 :拟合1948——1981年美国女性月度失业率序列



差分平稳

- 一阶、12步差分



差分后序列自相关图

```
                                           Autocorrelations

 Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

   0     12557.531        1.00000    |                    |********************|
   1     -1734.813        -.13815    |                 ***| .                  |
   2      2375.783        0.18919    |                  . |****                |
   3       279.359        0.02225    |                  . | .                  |
   4       759.808        0.06051    |                  . |*.                  |
   5       138.150        0.01100    |                  . | .                  |
   6       655.488        0.05220    |                  . |*.                  |
   7     -1028.202        -.08188    |                  **| .                  |
   8       533.734        0.04250    |                  . |*.                  |
   9      -158.605        -.01263    |                  . | .                  |
  10     -1606.237        -.12791    |                 ***| .                  |
  11       858.206        0.06834    |                  . |*.                  |
  12     -5698.457        -.45379    |           *********| .                  |
  13       521.120        0.04150    |                 .  |* .                 |
  14      -509.219        -.04055    |                 . *|  .                 |
  15     -1020.660        -.08128    |                 .**|  .                 |
  16      -730.212        -.05815    |                 . *|  .                 |
  17       429.071        0.03417    |                 .  |* .                 |
  18      -825.235        -.06572    |                 . *|  .                 |
  19       592.947        0.04722    |                 .  |* .                 |
  20      -565.282        -.04502    |                 . *|  .                 |
  21       206.681        0.01646    |                 .  |  .                 |
  22      -117.966        -.00939    |                 .  |  .                 |
  23       774.691        0.06169    |                 .  |* .                 |
  24      -929.421        -.07401    |                 . *|  .                 |

                                 "." marks two standard errors
```

差分后序列偏自相关图

```
                         Partial Autocorrelations

 Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

   1       -0.13815    |                 ***| .                  |
   2        0.17342    |                  . |***                 |
   3        0.07127    |                  . |*.                  |
   4        0.04101    |                  . |*.                  |
   5        0.00652    |                  . | .                  |
   6        0.03643    |                  . |*.                  |
   7       -0.08399    |                  **| .                  |
   8        0.00291    |                  . | .                  |
   9        0.01776    |                  . | .                  |
  10       -0.14008    |                 ***| .                  |
  11        0.04184    |                  . |*.                  |
  12       -0.42881    |           *********| .                  |
  13       -0.07313    |                  .*| .                  |
  14        0.11856    |                  . |**                  |
  15       -0.04173    |                  .*| .                  |
  16       -0.03232    |                  .*| .                  |
  17        0.04320    |                  . |*.                  |
  18        0.00766    |                  . | .                  |
  19       -0.03205    |                  .*| .                  |
  20       -0.00893    |                  . | .                  |
  21        0.03529    |                  . |*.                  |
  22       -0.12175    |                  **| .                  |
  23        0.09376    |                  . |**                  |
  24       -0.30002    |              ******| .                  |
```

简单季节模型拟合结果

延迟阶数	拟合模型残差白噪声检验					
	AR(1,12)		MA(1,2,12)		ARMA((1,12),(1,12))	
	χ^2值	P值	χ^2值	P值	χ^2值	P值
6	14.58	0.0057	9.5	0.0233	15.77	0.0004
12	16.42	0.0883	14.19	0.1158	17.99	0.0213
结果	拟合模型均不显著					

乘积季节模型拟合

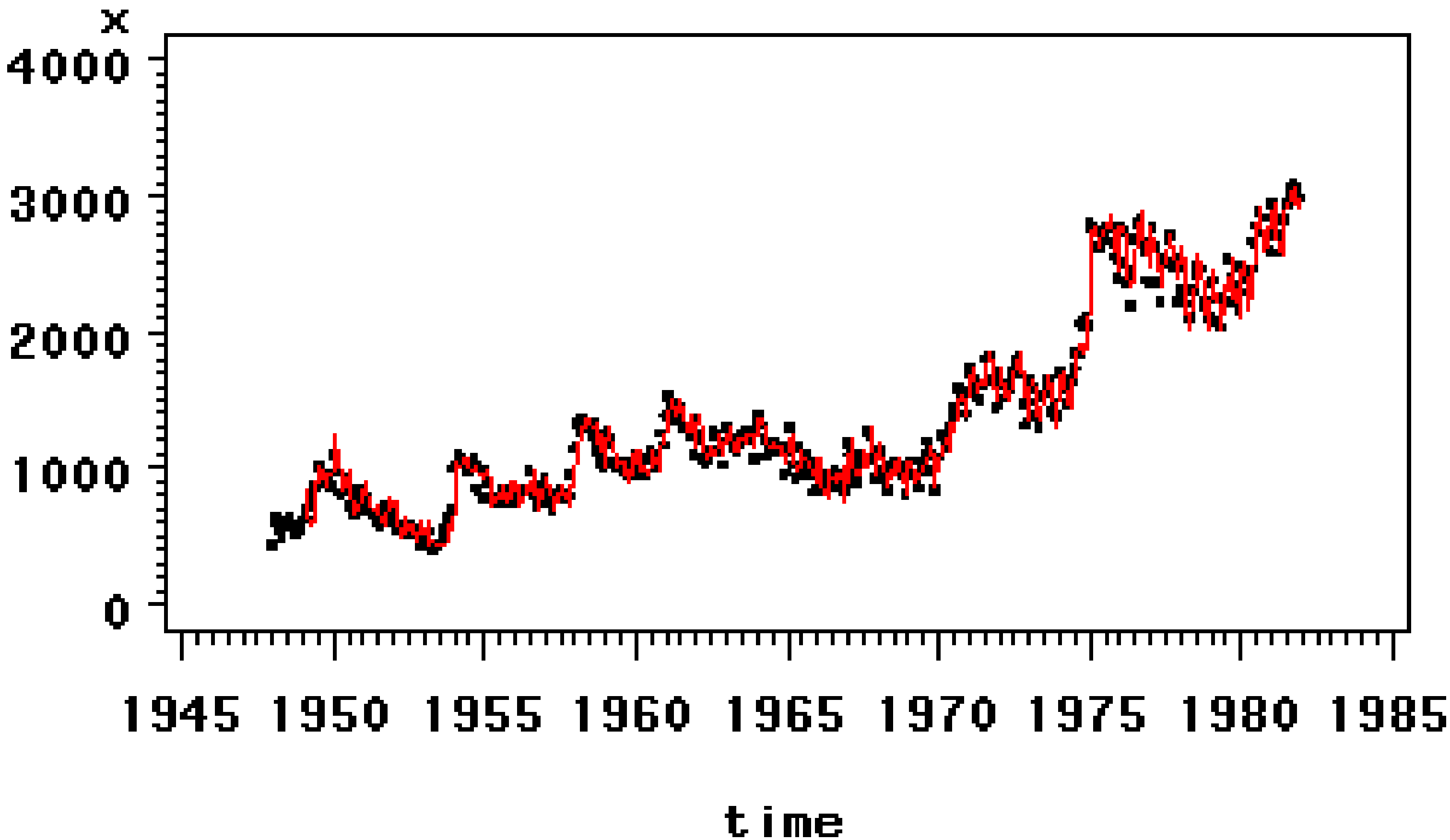
- 模型定阶
  - ARIMA$(1,1,1)\times(0,1,1)_{12}$
- 参数估计

$$\nabla\nabla_{12}x_t=\frac{1+0.66137B}{1+0.78978B}(1-0.77394B^{12})\varepsilon_t$$

模型检验

残差白噪声检验			参数显著性检验		
延迟阶数	χ^2统计量	P值	待估参数	χ^2统计量	P值
6	4.50	0.2120	θ_1	-4.66	<0.0001
12	9.42	0.4002	θ_{12}	23.03	<0.0001
18	20.58	0.1507	ϕ_1	-6.81	<0.0001
结果	模型显著		参数均显著		

乘积季节模型拟合效果图



5.3 Auto-Regressive模型

- 构造思想
  - 首先通过确定性因素分解方法提取序列中主要的确定性信息

$$x_t = T_t + S_t + \varepsilon_t$$

  - 然后对残差序列拟合自回归模型，以便充分提取相关信息

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \cdots + \phi_p \varepsilon_{t-p} + a_t$$

Auto-Regressive模型结构

$$
\begin{cases}
x_t = T_t + S_t + \varepsilon_t \\
\varepsilon_t = \phi_1 \varepsilon_{t-1} + \cdots + \phi_p \varepsilon_{t-p} + a_t \\
E(a_t) = 0, Var(a_t) = \sigma^2, Cov(a_t, a_{t-i}) = 0, \forall i \geq 1
\end{cases}
$$

对趋势效应的常用拟合方法

- 自变量为时间t的幂函数

$$T_t = \beta_0 + \beta_1 \cdot t + \cdots + \beta_k \cdot t^k + \varepsilon_t$$

- 自变量为历史观察值

$$T_t = \beta_0 + \beta_1 \cdot x_{t-1} + \cdots + \beta_k \cdot x_{t-k} + \varepsilon_t$$

对季节效应的常用拟合方法

- 给定季节指数

$$S_t = S_t'$$

- 建立季节自回归模型

$$T_t = \alpha_0 + \alpha_1 \cdot x_{t-m} + \cdots + \alpha_l \cdot x_{t-lm}$$

例5.6续

- 使用Auto-Regressive模型分析1952年－1988年中国农业实际国民收入指数序列。
- 时序图显示该序列有显著的线性递增趋势，但没有季节效应，所以考虑建立如下结构的Auto-Regressive模型

$$
\begin{cases}
x_t = T_t + \varepsilon_t \quad , \quad t = 1,2,3,\cdots \\
\varepsilon_t = \phi_1 \varepsilon_{t-1} + \cdots + \phi_p \varepsilon_{t-p} + a_t \\
E(a_t) = 0, Var(a_t) = \sigma^2, Cov(a_t, a_{t-i}) = 0, \forall i \geq 1
\end{cases}
$$

趋势拟合

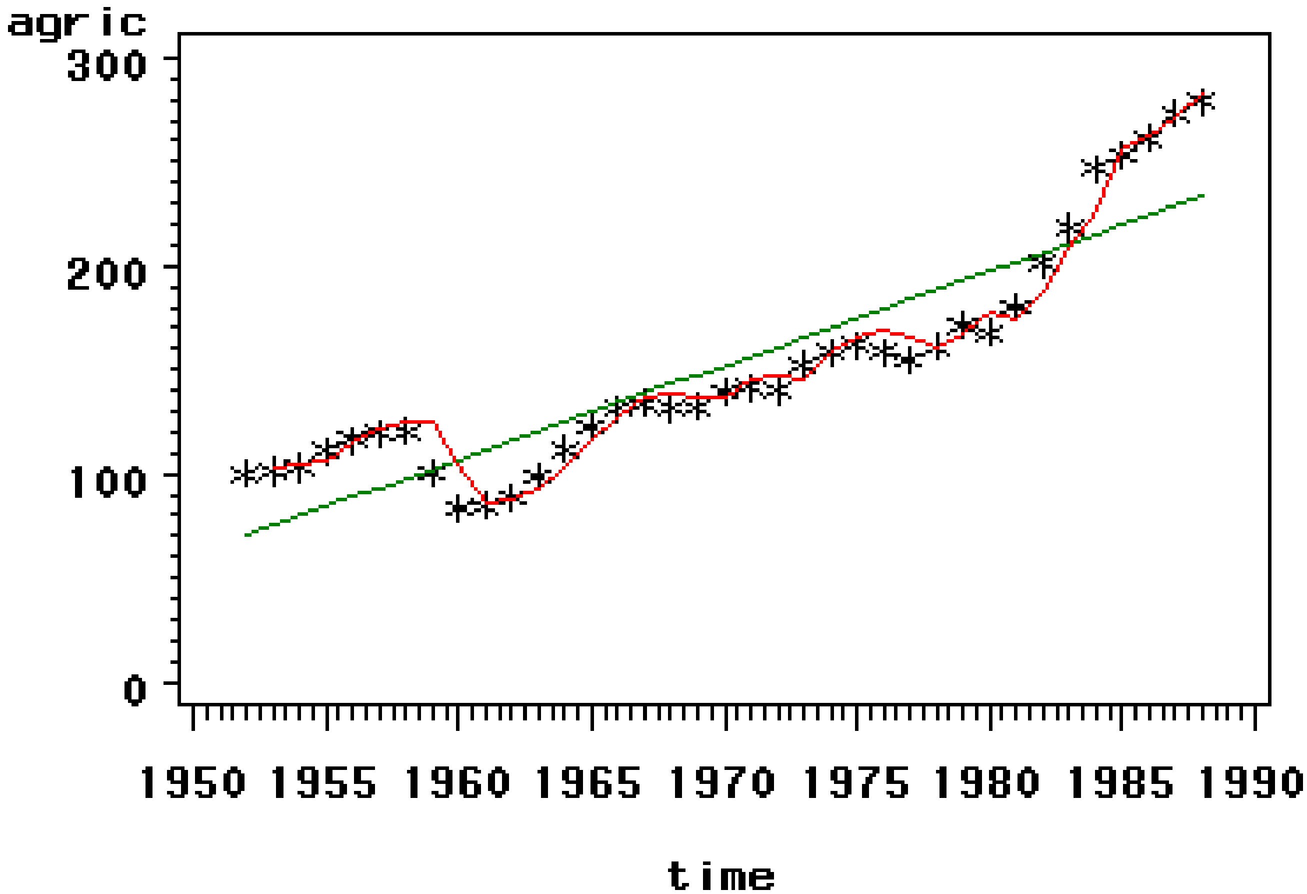
- 方法一：变量为时间t的幂函数

$$T_t = 66.1491 + 4.5158t \quad , \quad t = 1,2,3,\cdots$$

- 方法二：变量为一阶延迟序列值 x_{t-1}

$$\hat{x}_t = 1.0365 x_{t-1} \quad , \quad t = 1,2,3,\cdots$$

趋势拟合效果图



agric
300
200
100
0
1950 1955 1960 1965 1970 1975 1980 1985 1990
time

残差自相关检验

- 检验原理
  - 回归模型拟合充分，残差的性质

$$E(\varepsilon_t,\varepsilon_{t-j})=0\ ,\ \forall j\geq 1$$

  - 回归模型拟合得不充分，残差的性质

$$E(\varepsilon_t,\varepsilon_{t-j})\neq 0\ ,\ \exists\ j\geq 1$$

Durbin-Waston检验（DW检验）

- 假设条件
  - 原假设：残差序列不存在一阶自相关性

$$H_0: E(\varepsilon_t, \varepsilon_{t-1}) = 0 \Leftrightarrow H_0: \rho = 0$$

  - 备择假设：残差序列存在一阶自相关性

$$H_0: E(\varepsilon_t, \varepsilon_{t-1}) \neq 0 \Leftrightarrow H_0: \rho \neq 0$$

DW统计量

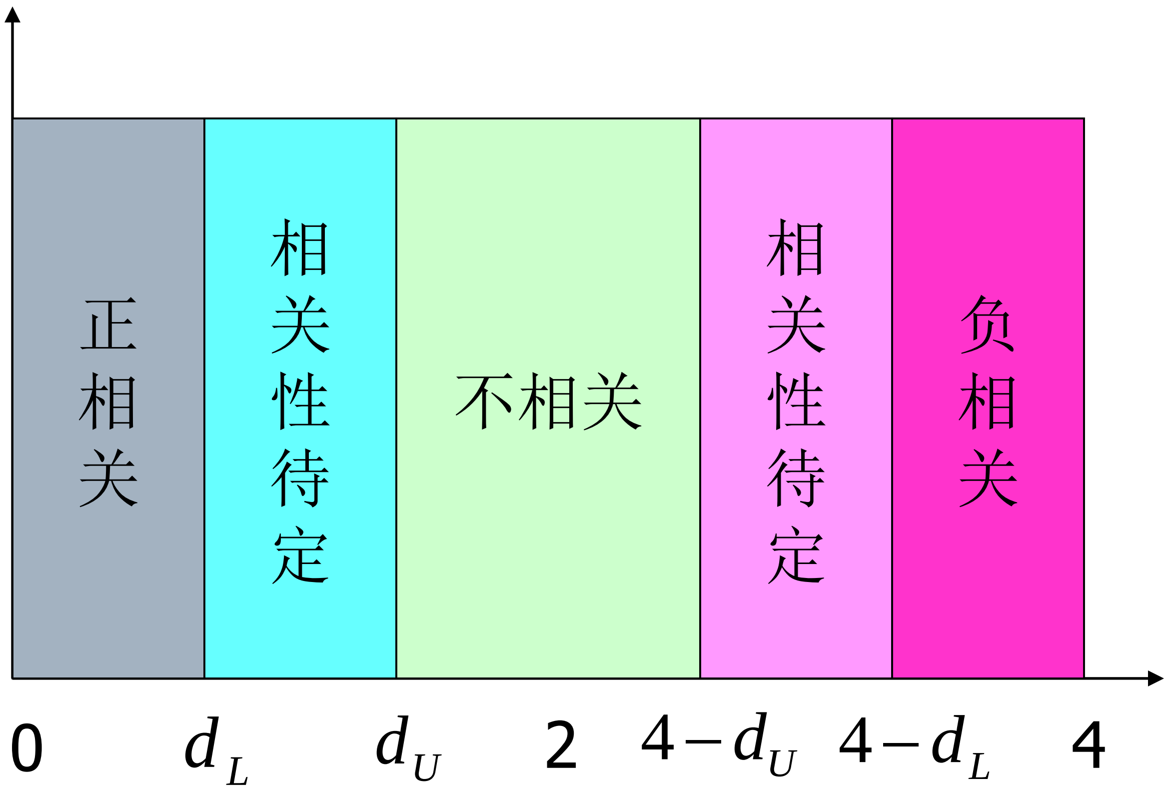
- 构造统计量

$$DW = \frac{\sum_{t=2}^{n}(\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n}\varepsilon_t^2}$$

- DW统计量和自相关系数的关系

$$DW \cong 2(1-\rho)$$

DW统计量的判定结果



例5.6续

- 检验第一个确定性趋势模型

$$x_t = 66.1491 + 4.5158t + \varepsilon_t \ , \ t = 1,2,3,\cdots$$

残差序列的自相关性。

DW检验结果

- 检验结果

DW统计量的值	d_L	d_U	P值
0.1378	1.42	1.53	0.0001

- 检验结论
  - 检验结果显示残差序列高度正自相关。

Durbin h检验

- DW统计量的缺陷
  - 当回归因子包含延迟因变量时，残差序列的DW统计量是一个有偏统计量。在这种场合下使用DW统计量容易产生残差序列正自相关性不显著的误判
- Durbin h检验

$$Dh = DW\frac{n}{1-n\sigma_{\beta}^{2}}$$

例5.6续

- 检验第二个确定性趋势模型

$$x_t = 1.0365 x_{t-1} + \varepsilon_t \ , \ t = 1,2,3,\cdots$$

残差序列的自相关性。

Dh检验结果

- 检验结果

Dh统计量的值	P值
2.8038	0.0025

- 检验结论
  - 检验结果显示残差序列高度正自相关。

残差序列拟合

- 确定自回归模型的阶数
- 参数估计
- 模型检验

例5.6续

- 对第一个确定性趋势模型的残差序列

$$\varepsilon_t = x_t - T_t = x_t - 66.1491 - 4.5158t \,,\ t = 1,2,\cdots$$

进行拟合

残差序列自相关图

```
                            Autocorrelations

Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

    1.00000    |                    |********************|
    0.86045    |             .      |*****************   |
    0.65506    |          .         |*************       |
    0.44082    |        .           |*********  .        |
    0.21516    |       .            |****        .       |
    0.01276    |       .            |            .       |
    -.12360    |       .          **|            .       |
    -.21001    |       .        ****|            .       |
    -.20305    |       .        ****|            .       |
    -.15935    |       .         ***|            .       |

              "." marks two standard errors
```

残差序列偏自相关图

```
Partial Autocorrelations

 -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

|             .      |*****************    |
|             *******|      .              |
|             .    **|      .              |
|             .  ****|      .              |
|             .     *|      .              |
|             .      |*     .              |
|             .     *|      .              |
```

模型拟合

- 定阶
  - AR(2)
- 参数估计方法
  - 极大似然估计
- 最终拟合模型口径

$$\begin{cases} x_t = 69.1491 + 4.5158t + \varepsilon_t \\ \varepsilon_t = 1.4859\varepsilon_{t-1} - 0.5848\varepsilon_{t-2} + a_t \end{cases}$$

例5.6

- 第二个Auto－Regressive模型的拟合结果

$$\begin{cases} x_t = 1.033x_{t-1} + \varepsilon_t \\ \varepsilon_t = 0.4615\varepsilon_{t-1} + a_t \end{cases}$$

三个拟合模型的比较

模型	AIC	SBC
ARIMA(0,1,1)模型： $(1-B)x_t = 4.99661+(1+0.70766B)\varepsilon_t$	249.3305	252.4976
Auto－Regressive模型一： $\begin{cases} x_t = 69.1491+4.5158t+\varepsilon_t \\ \varepsilon_t = 1.4859\varepsilon_{t-1} - 0.5848\varepsilon_{t-2} + a_t \end{cases}$	260.8454	267.2891
Auto－Regressive模型二： $\begin{cases} x_t = 1.033x_{t-1}+\varepsilon_t \\ \varepsilon_t = 0.4615\varepsilon_{t-1} + a_t \end{cases}$	250.6317	253.7987

5.4 异方差的性质

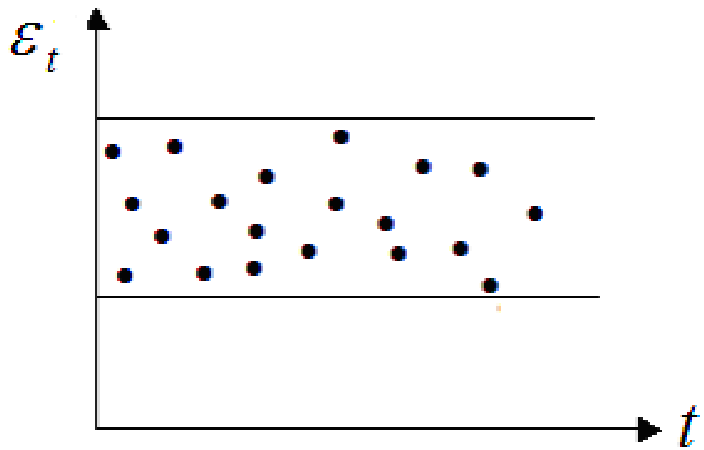
- 异方差的定义
  - 如果随机误差序列的方差会随着时间的变化而变化，这种情况被称作为异方差
- 异方差的影响 $Var(\varepsilon_t) = h(t)$
  - 忽视异方差的存在会导致残差的方差会被严重低估，继而参数显著性检验容易犯纳伪错误，这使得参数的显著性检验失去意义，最终导致模型的拟合精度受影响。

异方差直观诊断

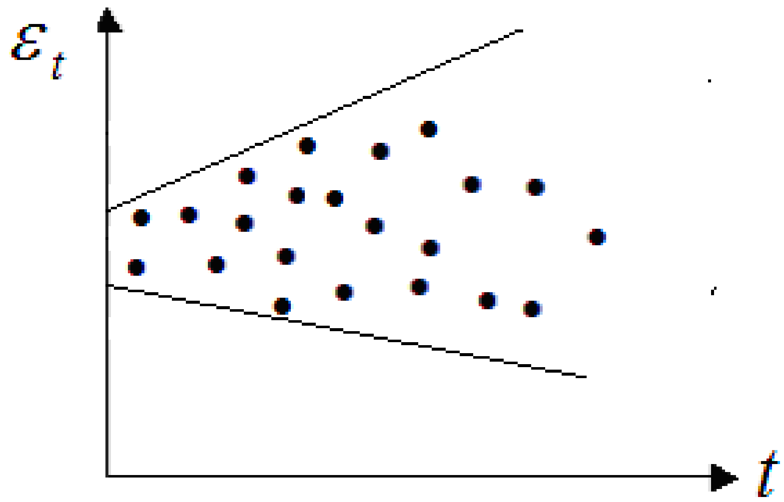
- 残差图
- 残差平方图

残差图

□ 方差齐性残差图



□ 递增型异方差残差图



残差平方图

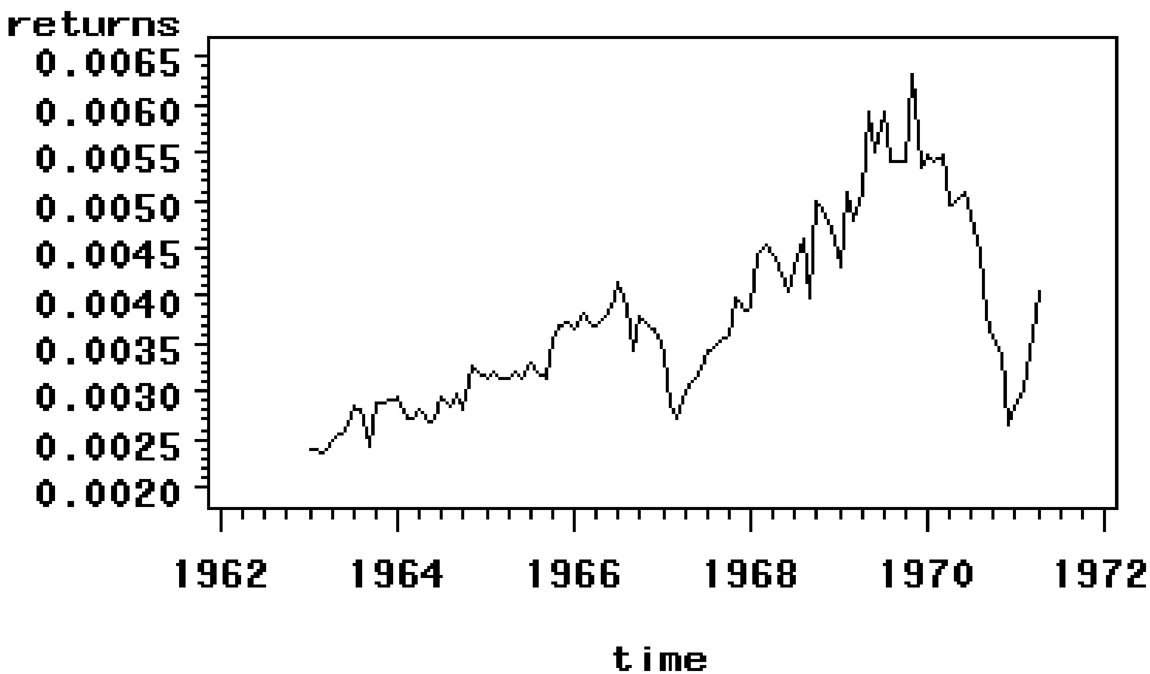
- 原理
  - 残差序列的方差实际上就是它平方的期望。

$$Var(\varepsilon_t) = E(\varepsilon_t^2)$$

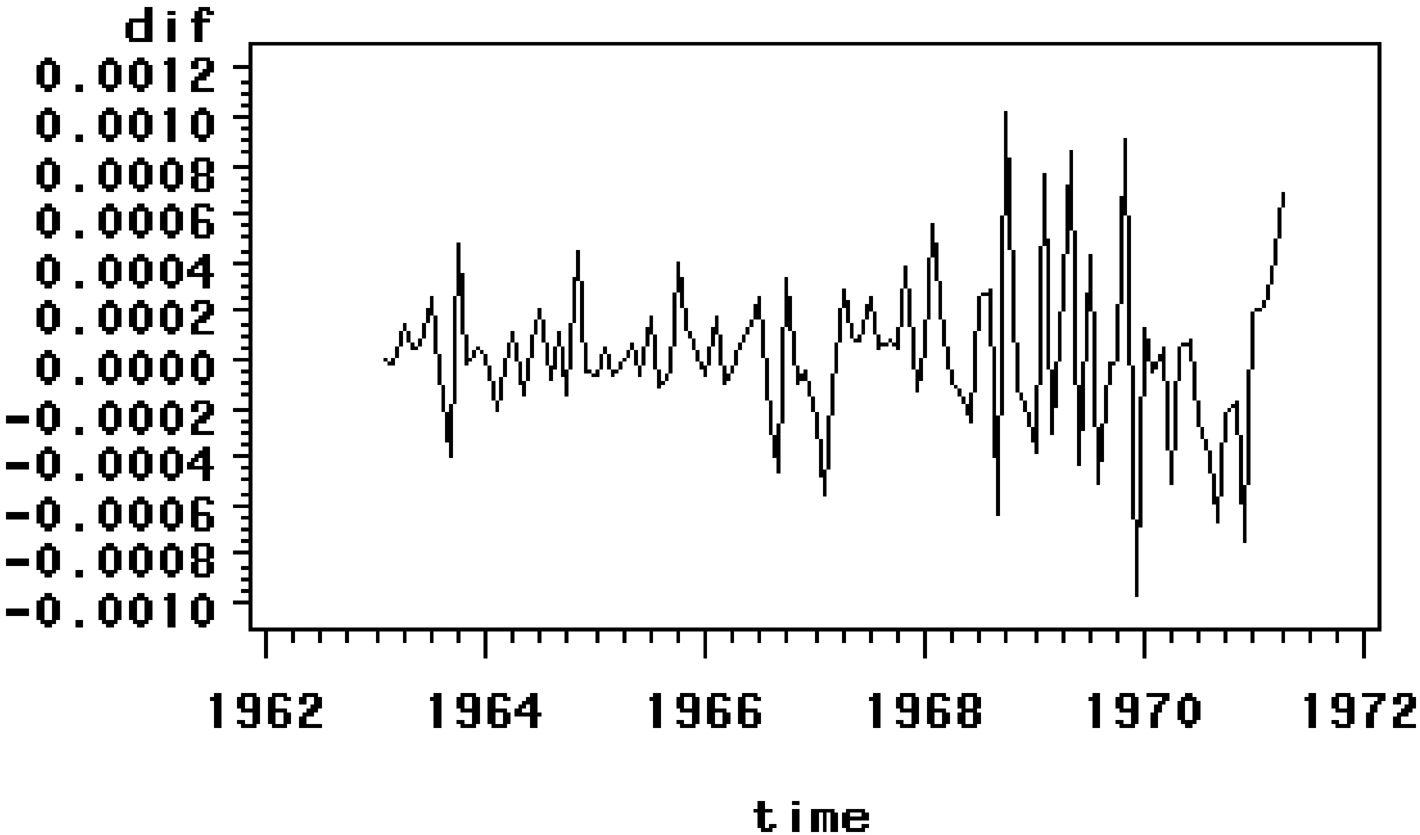
  - 所以考察残差序列是否方差齐性，主要是考察残差平方序列是否平稳

例5.11

- 直观考察美国1963年4月——1971年7月短期国库券的月度收益率序列的方差齐性。

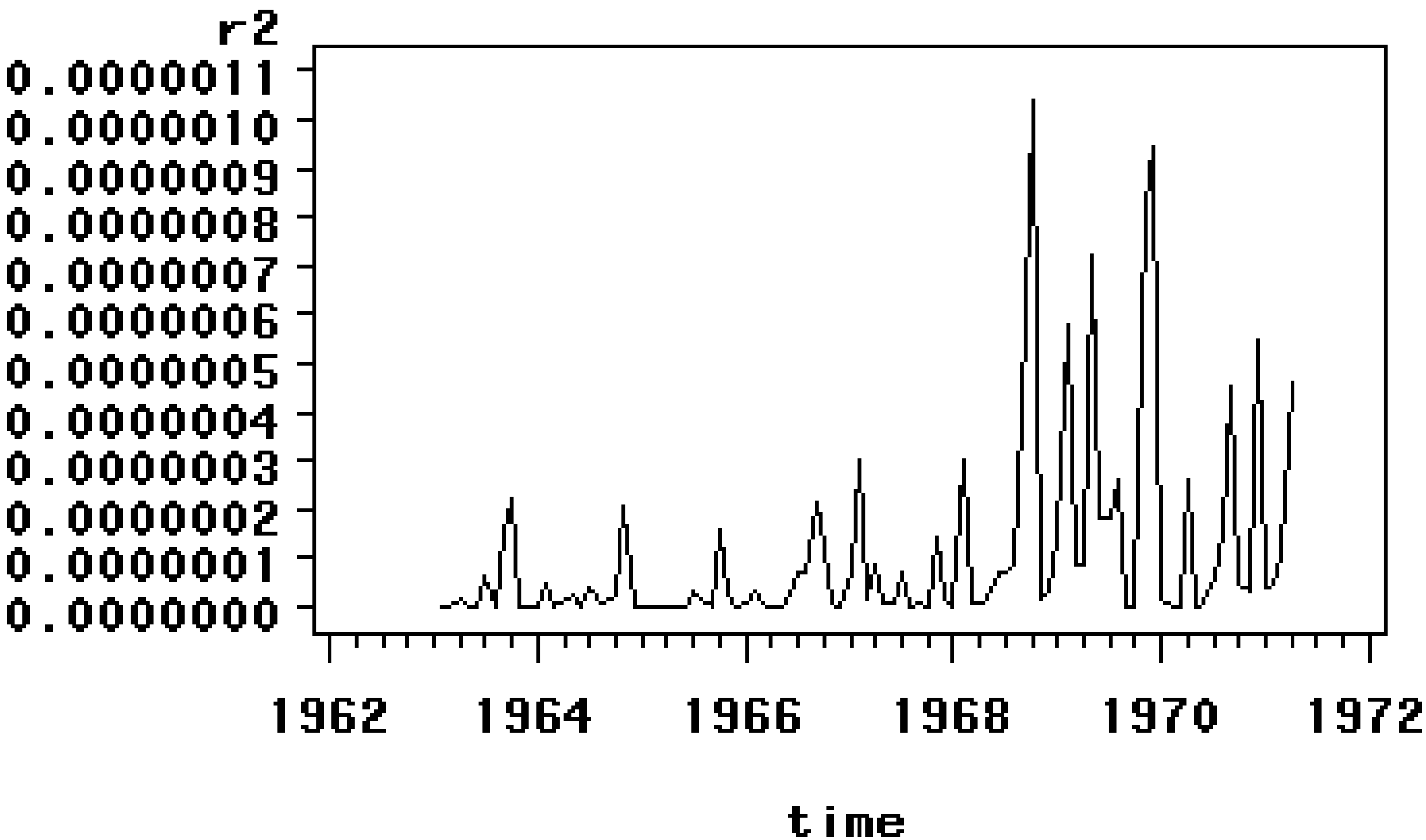


一阶差分后残差图



dif
0.0012
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
-0.0002
-0.0004
-0.0006
-0.0008
-0.0010
1962
1964
1966
1968
1970
1972
time

一阶差分后残差平方图


r2
0.0000011
0.0000010
0.0000009
0.0000008
0.0000007
0.0000006
0.0000005
0.0000004
0.0000003
0.0000002
0.0000001
0.0000000
1962
1964
1966
1968
1970
1972
time

异方差处理方法

- 假如已知异方差函数具体形式，进行方差齐性变化
- 假如不知异方差函数的具体形式，拟合条件异方差模型

5.5 方差齐性变换

- 使用场合
  - 序列显示出显著的异方差性，且方差与均值之间具有某种函数关系

$$\sigma_t^2 = h(\mu_t)$$

其中：$h(\cdot)$是某个已知函数

- 处理思路
  - 尝试寻找一个转换函数$g(\cdot)$，使得经转换后的变量满足方差齐性

$$Var[g(x_t)] = \sigma^2$$

转换函数的确定原理

- 转换函数$g(x_t)$在μ_t附近作一阶泰勒展开

$$g(x_t) \cong g(\mu_t) + (x_t - \mu_t) g'(\mu_t)$$

- 求转换函数的方差

$$Var[g(x_t)] \cong Var[g(\mu_t) + (x_t - \mu_t) g'(\mu_t)]$$
$$= [g'(\mu_t)]^2 h(\mu_t)$$

- 转换函数的确定

$$g'(\mu_t) = \frac{1}{\sqrt{h(\mu_t)}}$$

常用转换函数的确定

- 假定

$$\sigma_t = \mu_t \Leftrightarrow h(\mu_t) = \mu_t^2$$

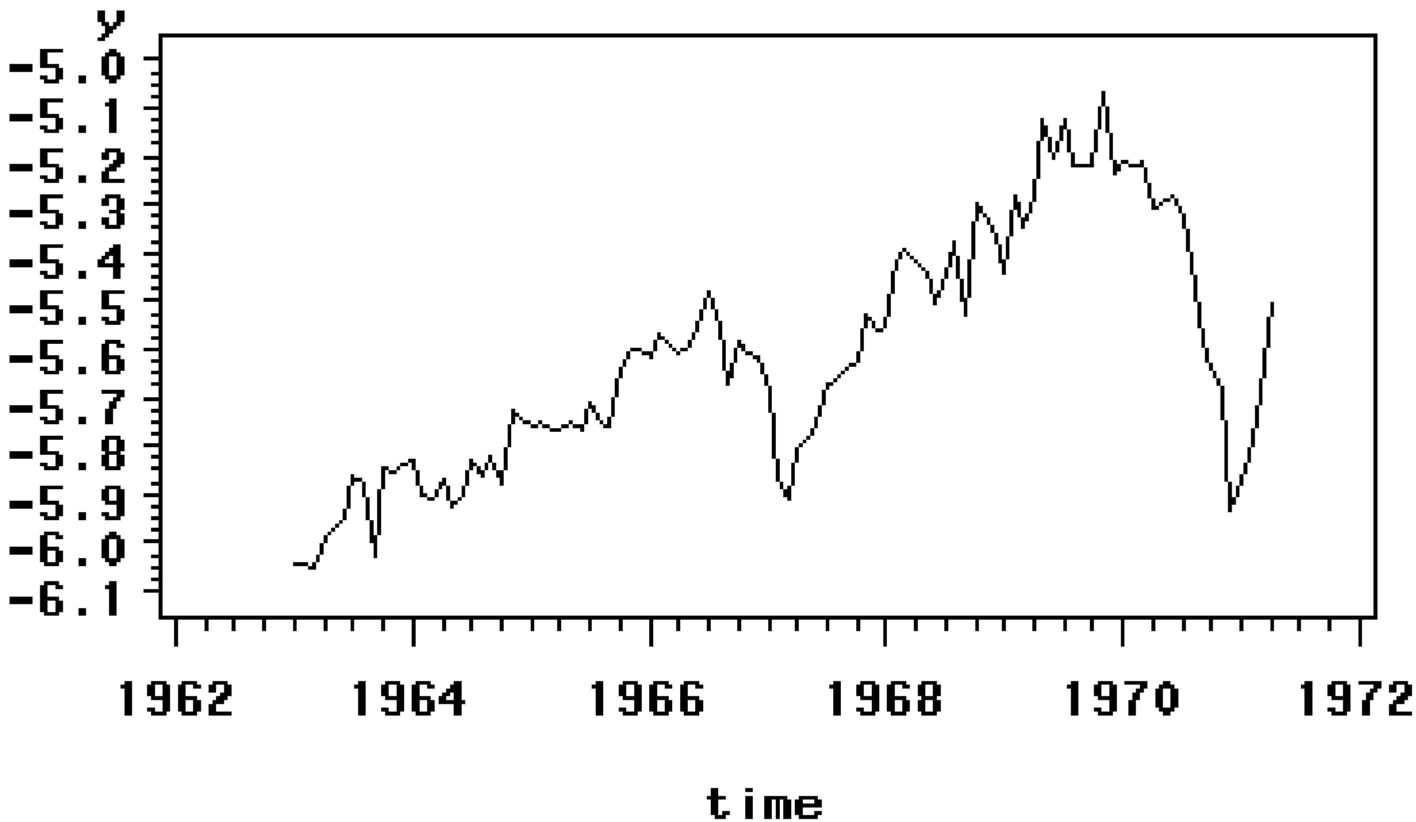
- 转换函数的确定

$$g'(\mu_t) = \frac{1}{\sqrt{h(\mu_t)}} = \frac{1}{\mu_t} \quad \Rightarrow \quad g(\mu_t) = \log(\mu_t)$$

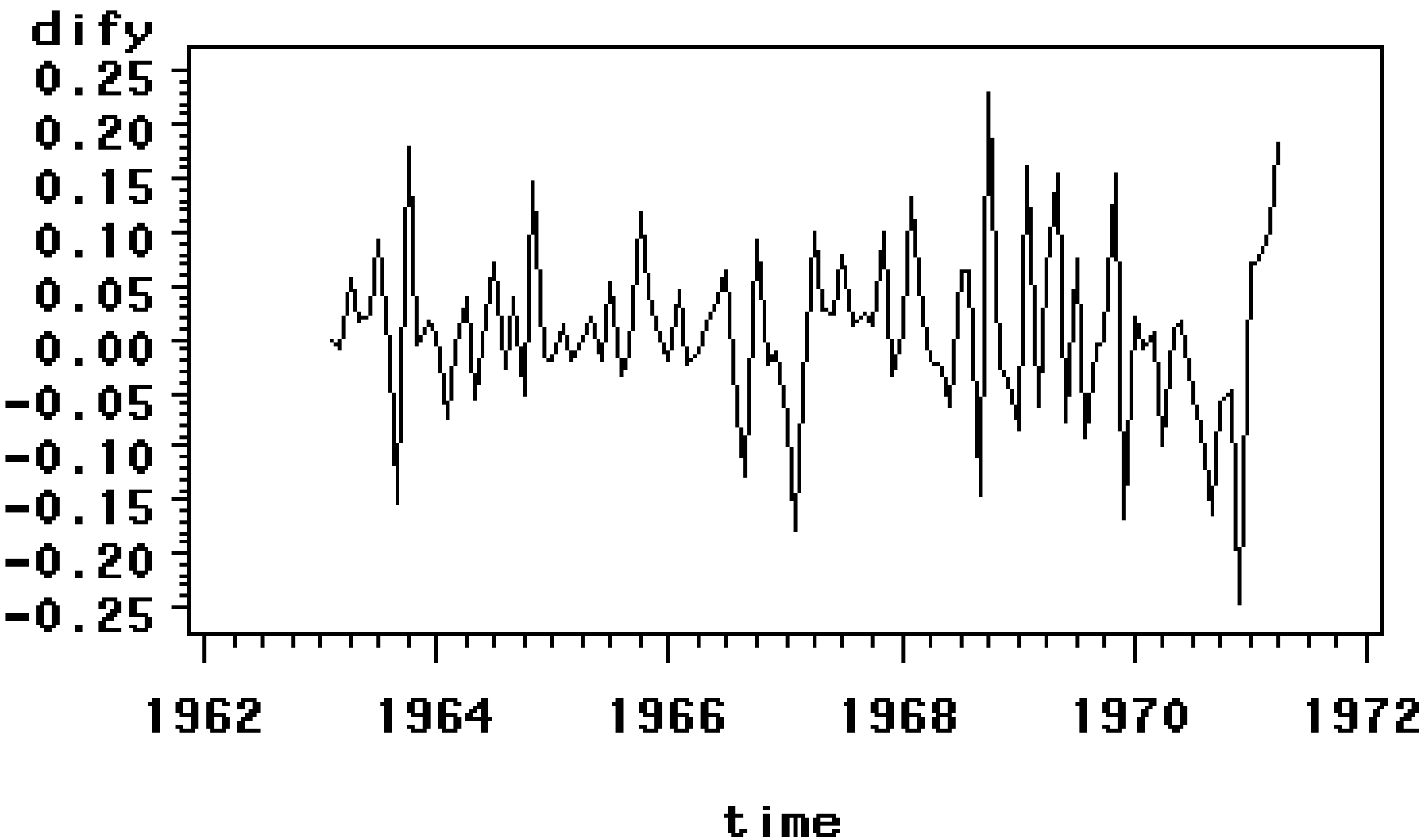
例5.11续

- 对美国1963年4月——1971年7月短期国库券的月度收益率序列使用方差齐性变换方法进行分析
- 假定$\sigma_t = x_t$
- 函数变换 $y_t = \log(x_t)$

对数序列时序图


y
-5.0
-5.1
-5.2
-5.3
-5.4
-5.5
-5.6
-5.7
-5.8
-5.9
-6.0
-6.1
1962
1964
1966
1968
1970
1972
time

一阶差分后序列图

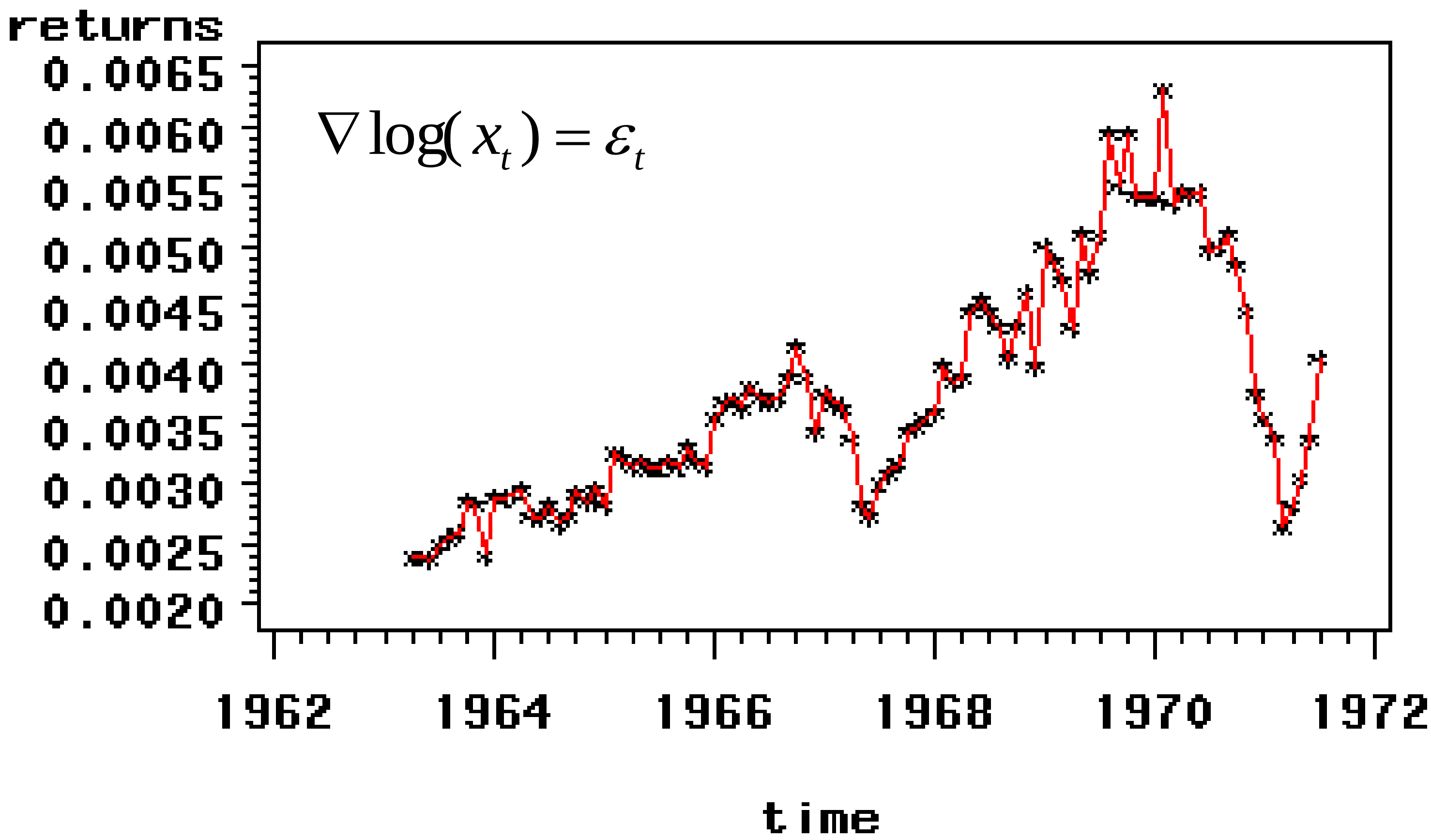


dify
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
-0.15
-0.20
-0.25
1962
1964
1966
1968
1970
1972
time

白噪声检验

延迟阶数	LB统计量	P值
6	3.58	0.7337
12	10.82	0.5441
18	21.71	0.2452

拟合模型口径及拟合效果图



5.6 条件异方差模型

- ARCH模型
- GARCH模型
- GARCH模型的变体
  - EGARCH模型
  - IGARCH模型
  - GARCH-M模型
  - AR-GARCH模型

ARCH模型

- 假定

$$\varepsilon_t / \sqrt{h_t} \sim N(0,1)$$

- 原理
  - 通过构造残差平方序列的自回归模型来拟合异方差函数

- ARCH(q)模型结构

$$\begin{cases} x_t = f(t, x_{t-1}, x_{t-2}, \cdots) + \varepsilon_t \\ \varepsilon_t = \sqrt{h_t} e_t \\ h_t = \omega + \sum_{j=1}^{q} \lambda_j \varepsilon_{t-j}^2 \end{cases}$$

GARCH 模型结构

- 使用场合
  - ARCH模型实际上适用于异方差函数短期自相关过程
  - GARCH模型实际上适用于异方差函数长期自相关过程
- 模型结构

$$\begin{cases} x_t = f(t, x_{t-1}, x_{t-2}, \cdots) + \varepsilon_t \\ \varepsilon_t = \sqrt{h_t} e_t \\ h_t = \omega + \sum_{i=1}^{p} \eta_i h_{t-i} + \sum_{j=1}^{q} \lambda_j \varepsilon_{t-j}^2 \end{cases}$$

GARCH模型的约束条件

- 参数非负

$$\omega > 0, \eta_i \geq 0, \lambda_j \geq 0$$

- 参数有界

$$\sum_{i=1}^{p} \eta_i + \sum_{j=1}^{q} \lambda_j < 1$$

EGARCH模型

$$
\begin{cases}
x_t = f(t, x_{t-1}, x_{t-2}, \cdots) + \varepsilon_t \\
\varepsilon_t = \sqrt{h_t}\, e_t \\
\ln(h_t) = \omega + \sum_{i=1}^{p} \eta_i \ln(h_{t-i}) + \sum_{j=1}^{q} \lambda_j g(e_t) \\
g(e_t) = \theta e_t + \gamma[|e_t| - E|e_t|]
\end{cases}
$$

IGARCH模型

$$
\begin{cases}
x_t = f(t, x_{t-1}, x_{t-2}, \cdots) + \varepsilon_t \\
\varepsilon_t = \sqrt{h_t} e_t \\
h_t = \omega + \sum_{i=1}^{p} \eta_i h_{t-i} + \sum_{j=1}^{q} \lambda_j \varepsilon_{t-j}^2 \\
\sum_{i=1}^{p} \eta_i + \sum_{j=1}^{q} \lambda_j = 1
\end{cases}
$$

GARCH-M模型

$$
\begin{cases}
x_t = f(t, x_{t-1}, x_{t-2}, \cdots) + \delta\sqrt{h_t} + \varepsilon_t \\
\varepsilon_t = \sqrt{h_t}\, e_t \\
h_t = \omega + \sum_{i=1}^{p} \eta_i h_{t-i} + \sum_{j=1}^{q} \lambda_j \varepsilon_{t-j}^2
\end{cases}
$$

AR-GARCH模型

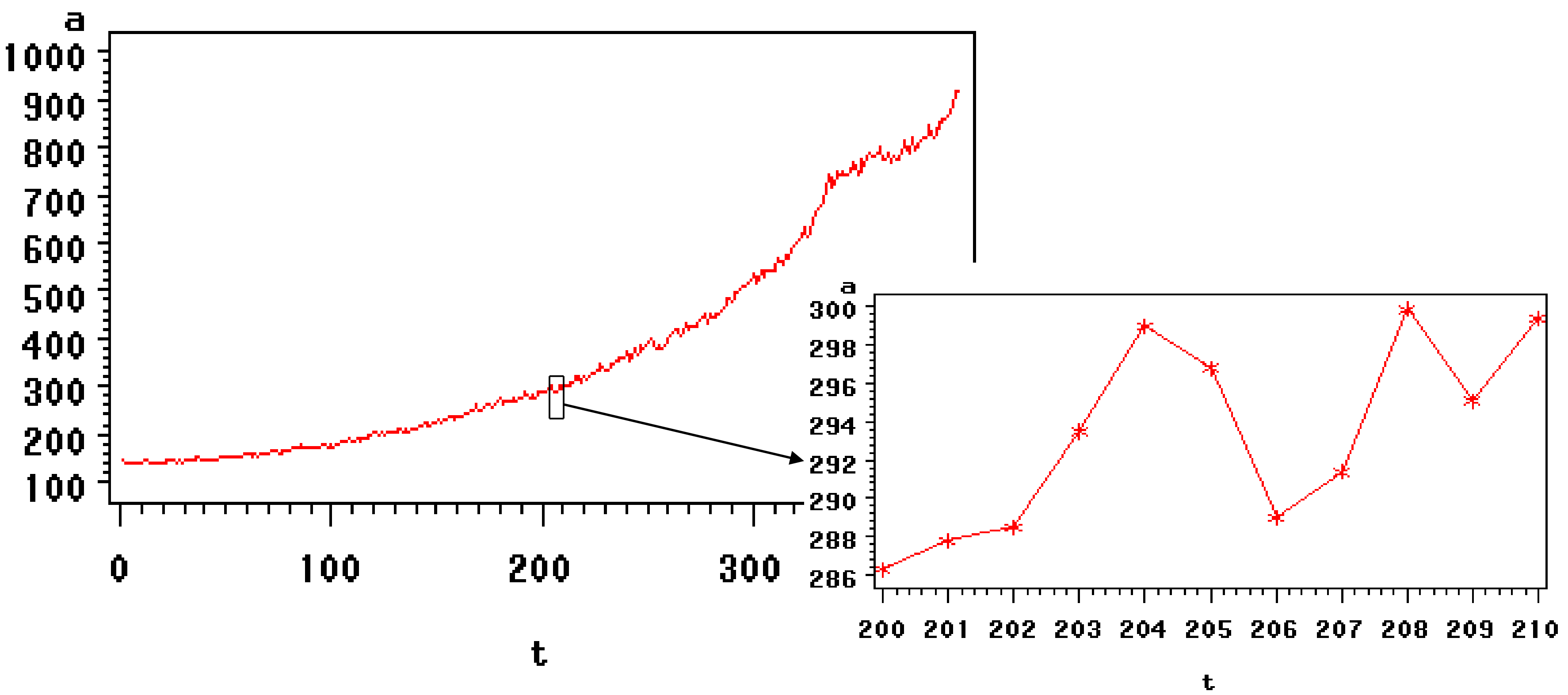
$$
\begin{cases}
x_t = f(t, x_{t-1}, x_{t-2}, \cdots) + \varepsilon_t \\
\varepsilon_t = \sum_{k=1}^{m} \beta_k \varepsilon_{t-k} + \upsilon_t \\
\upsilon_t = \sqrt{h_t} e_t \\
h_t = \omega + \sum_{i=1}^{p} \eta_i h_{t-i} + \sum_{j=1}^{q} \lambda_j \upsilon_{t-j}^2
\end{cases}
$$

GARCH模型拟合步骤

- 回归拟合
- 残差自相关性检验
- 异方差自相关性检验
- ARCH模型定阶
- 参数估计
- 正态性检验

例5.12

- 使用条件异方差模型拟合某金融时间序列。



回归拟合

- 拟合模型

$$x_t = \alpha_1 x_{t-1} + \varepsilon_t$$

- 参数估计

$$\hat{\alpha}_1 = 1.0053$$

- 参数显著性检验
  - P值<0.0001，参数高度显著

残差自相关性检验

- 残差序列DW检验结果
  - Durbin h=-2.6011
  - $\Pr(Dh < -2.6011) < 0.0046$
- 拟合残差自回归模型
  - 方法：逐步回归
  - 模型口径

$$\varepsilon_t = -0.1559\varepsilon_{t-1} - 0.407\varepsilon_{t-2} + \upsilon_t$$

异方差自相关检验

- Portmantea Q检验
- 拉格朗日乘子（LM）检验

Portmantea Q检验

- 假设条件

$$H_0: \rho_1 = \rho_2 = \cdots = \rho_q = 0 \leftrightarrow H_1: \rho_1, \rho_2, \cdots, \rho_q \text{不全为零}$$

- 检验统计量

$$Q(q) = n(n+2)\sum_{i=1}^{q}\frac{\rho_i^2}{n-i} \sim \chi^2(q-1)$$

- 检验结果
  - 拒绝原假设 $Q(q) \geq \chi_{1-\alpha}^2(q-1)$
  - 接受原假设 $Q(q) < \chi_{1-\alpha}^2(q-1)$

LM检验

- 假设条件

$$H_0:\rho_1=\rho_2=\cdots=\rho_q=0 \leftrightarrow H_1:\rho_1,\rho_2,\cdots,\rho_q\text{不全为零}$$

- 检验统计量

$$LM(q)=W'W \quad , W=\left(\frac{\rho_1^2}{\hat{\sigma}^2},\frac{\rho_2^2}{\hat{\sigma}^2},\cdots,\frac{\rho_q^2}{\hat{\sigma}^2}\right)$$

- 检验结果
  - 拒绝原假设 $Q(q)\geq\chi^2_{1-\alpha}(q-1)$
  - 接受原假设 $Q(q)<\chi^2_{1-\alpha}(q-1)$

例5.12残差序列异方差检验

阶数	Portmantea Q 统计量	P 值	LM 统计量	P 值
1	21.0457	< 0.0001	20.9461	< 0.0001
2	69.5836	< 0.0001	57.6219	< 0.0001
3	91.2062	< 0.0001	62.8963	< 0.0001
4	103.7725	< 0.0001	63.0874	< 0.0001
5	105.1216	< 0.0001	65.9268	< 0.0001
6	105.1753	< 0.0001	68.5504	< 0.0001
7	105.4858	< 0.0001	68.6805	< 0.0001
8	116.6605	< 0.0001	86.6446	< 0.0001
9	131.2003	< 0.0001	102.3673	< 0.0001
10	185.9957	< 0.0001	135.5774	< 0.0001
11	205.3304	< 0.0001	135.6465	< 0.0001
12	417.1844	< 0.0001	230.3537	< 0.0001

ARCH模型拟合

- 定阶：GARCH(1,1)
- 参数估计：极大似然估计
- 拟合模型口径：AR(2)-GARCH(1,1)

$$
\begin{cases}
x_t = 1.0046 x_{t-1} + \varepsilon_t \\
\varepsilon_t = -0.1559 \varepsilon_{t-1} - 0.407 \varepsilon_{t-2} + \upsilon_t \\
\upsilon_t = \sqrt{h_t} e_t \\
h_t = 0.0951 + 0.8999 h_{t-1} + 0.1053 \upsilon_{t-1}^2
\end{cases}
$$

模型检验

- 检验方法：正态性检验
- 假设条件：$H_0: u_t \sim N(0,1) \leftrightarrow H_0: u_t \overset{not}{\sim} N(0,1)$
- 检验统计量

$$T_n = \frac{n}{6} b_1^2 + \frac{n}{24}(b_2 - 3)^2 \sim \chi^2(2)$$

- 检验结果
  - 拒绝原假设 $T_n \geq \chi^2_{1-\alpha}(2)$
  - 接受原假设 $T_n < \chi^2_{1-\alpha}(2)$

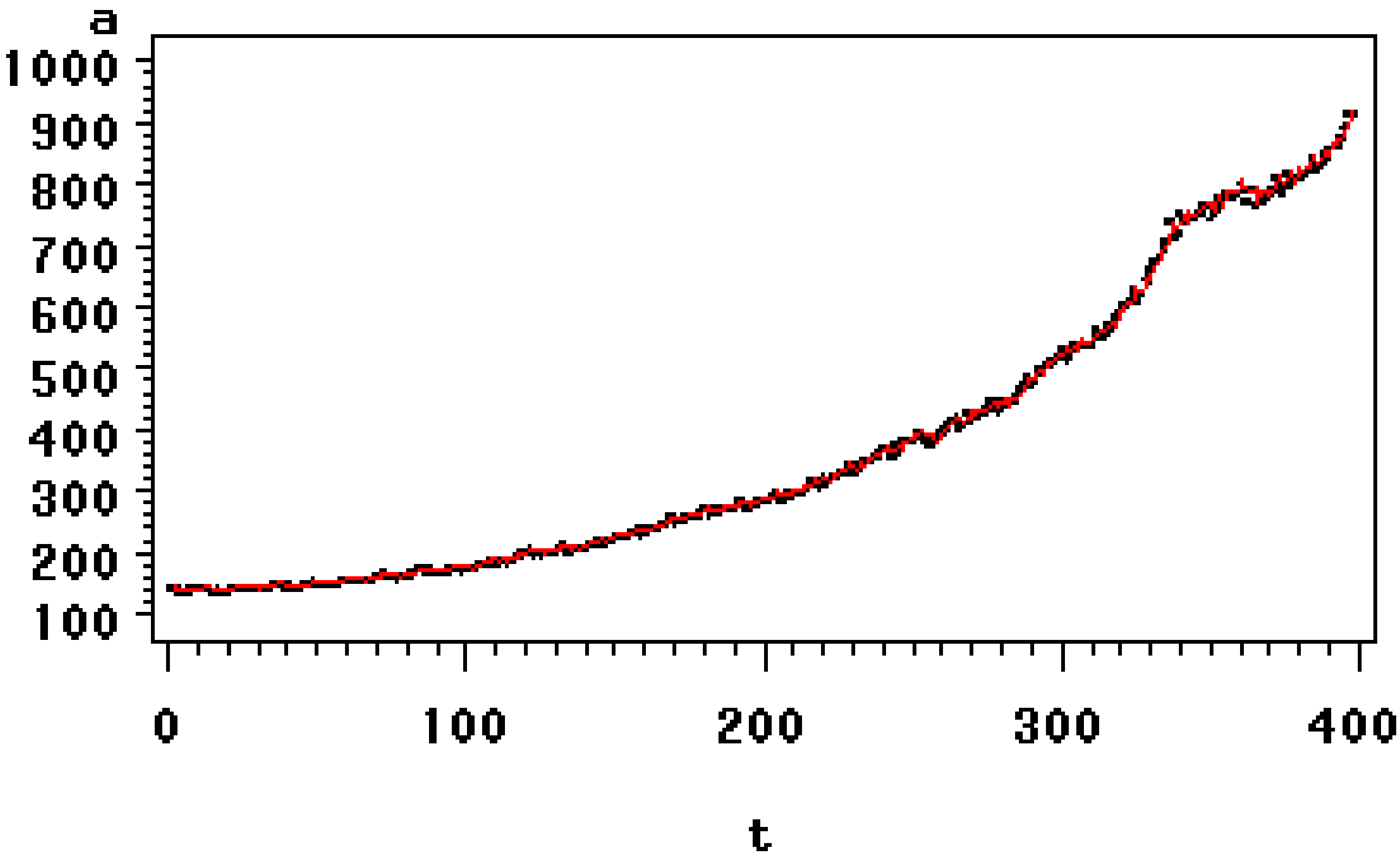
例5.13正态性检验结果

$$T_n = 1.1585$$

$$\text{P值} = 0.5603$$

- AR(2)-GARCH(1,1)模型显著成立

拟合效果图



疑问

- 问题答疑： http://www.xxwenda.com/
  - 可邀请老师或者其他人回答问题

联系我们

小象学院：互联网新技术在线教育领航者

– 微信公众号：小象
– 新浪微博：ChinaHadoop