法律声明

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

第六章 多元时间序列分析

主讲教师 周仕君

本章结构

- 平稳时间序列建模
- 虚假回归
- 单位根检验
- 协整
- 误差修正模型

6.1 平稳时间序列建模

- ARIMAX模型结构

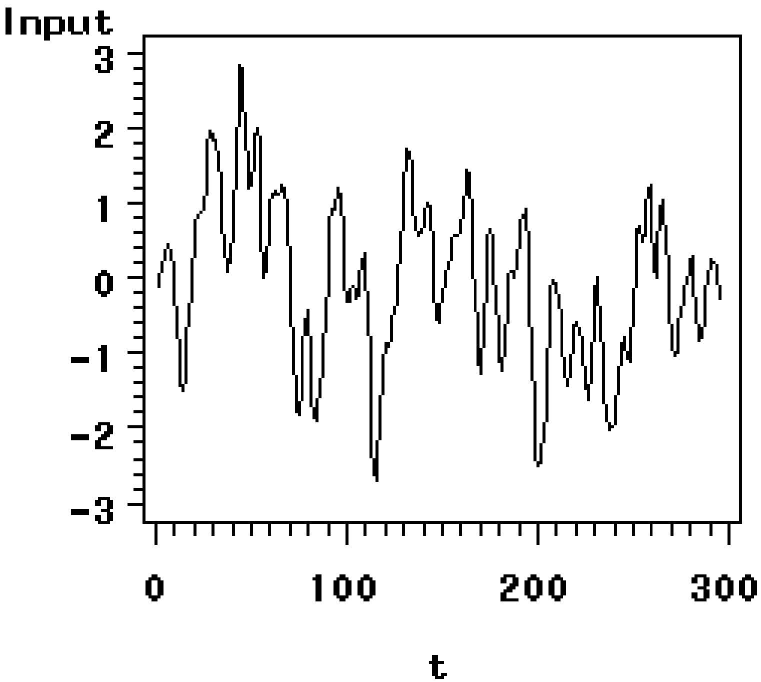
$$
\begin{cases}
y_t = \mu + \sum_{k=1}^{k} \dfrac{\Theta_i(B)}{\Phi_i(B)} B^{l_i} x_{it} + \varepsilon_t \\
\varepsilon_t = \dfrac{\Theta(B)}{\Phi(B)} a_t
\end{cases}
$$

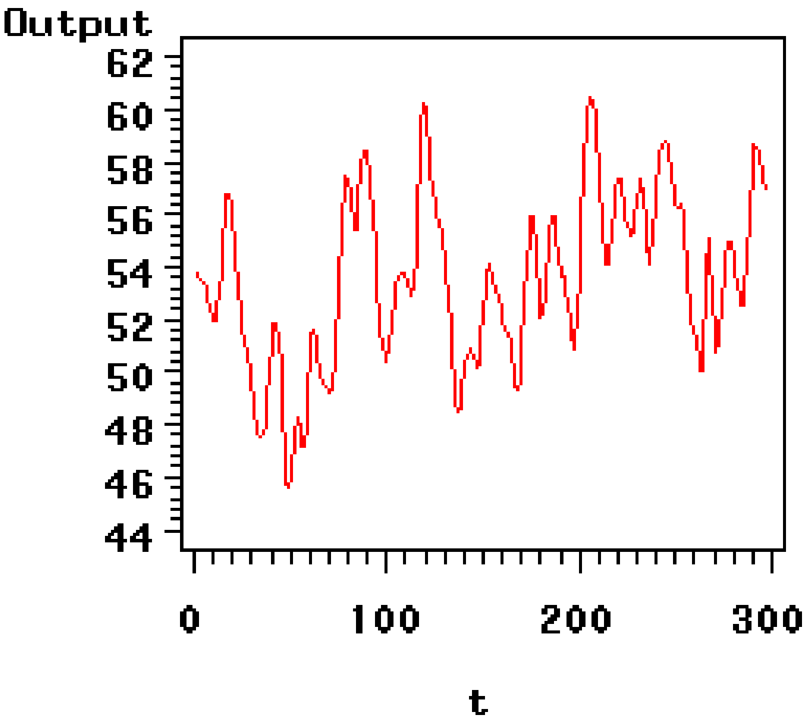
例6.1

- 在天然气炉中，输入的是天然气，输出的是CO_2，CO_2的输出浓度与天然气的输入速率有关。现在以中心化后的天然气输入速率为输入序列，建立CO_2的输出百分浓度模型。

输入/输出序列时序图

- 输入序列
- 输出序列





一元分析

- 拟合输入序列

$$x_t = -0.1228 + \frac{a_t}{1 - 1.97607B + 1.37499B^2 - 0.34336B^3}$$

- 拟合输出序列

$$y_t = 53.90176 + \frac{a_t}{1 - 3.10703B + 1.34005B^2 - 0.21274B^4}$$

多元分析

协相关图

```
                           Crosscorrelations

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

-10     0.0015683         0.02301    |                 . | .                 |
 -9    0.00013502         0.00198    |                 . | .                 |
 -8    -0.0060480         -.08872    |                 **| .                 |
 -7    -0.0017624         -.02585    |                 .*| .                 |
 -6    -0.0080539         -.11814    |                 **| .                 |
 -5    -0.0000944         -.00138    |                 . | .                 |
 -4    -0.0012802         -.01878    |                 . | .                 |
 -3    -0.0031078         -.04559    |                 .*| .                 |
 -2    0.00065212         0.00957    |                 . | .                 |
 -1    -0.0019166         -.02811    |                 .*| .                 |
  0    -0.0003673         -.00539    |                 . | .                 |
  1     0.0038939         0.05712    |                 . |*.                 |
  2    -0.0016971         -.02489    |                 . | .                 |
  3     -0.019231         -.28210    |             ******| .                 |
  4     -0.022479         -.32974    |            *******| .                 |
  5     -0.030909         -.45341    |          *********| .                 |
  6     -0.018122         -.26583    |              *****| .                 |
  7     -0.011426         -.16761    |                ***| .                 |
  8    -0.0017355         -.02546    |                 .*| .                 |
  9     0.0022590         0.03314    |                 . |*.                 |
 10    -0.0035152         -.05156    |                 .*| .                 |

                     "." marks two standard errors
```

拟合回归模型

- 模型结构

$$y_t = \mu + \frac{\theta_0 - \theta_1 B - \theta_2 B^2}{1 - \phi_1 B} B^3 x_t + \varepsilon_t$$

- 模型口径

$$y_t = 53.32256 + \frac{-0.5648 - 0.42573B - 0.29964B^2}{1 - 0.60057B} B^3 x_t + \varepsilon_t$$

拟合残差序列

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        0.89307    |                  . |******************  |
   2       -0.43103    |           *********| .                  |
   3       -0.12987    |                 ***| .                  |
   4        0.02319    |                  . | .                  |
   5        0.04054    |                  . |*.                  |
   6       -0.02657    |                  .*| .                  |
   7       -0.02496    |                  . | .                  |
   8        0.01993    |                  . | .                  |
   9        0.00001    |                  . | .                  |
  10        0.08721    |                  . |**                  |
```

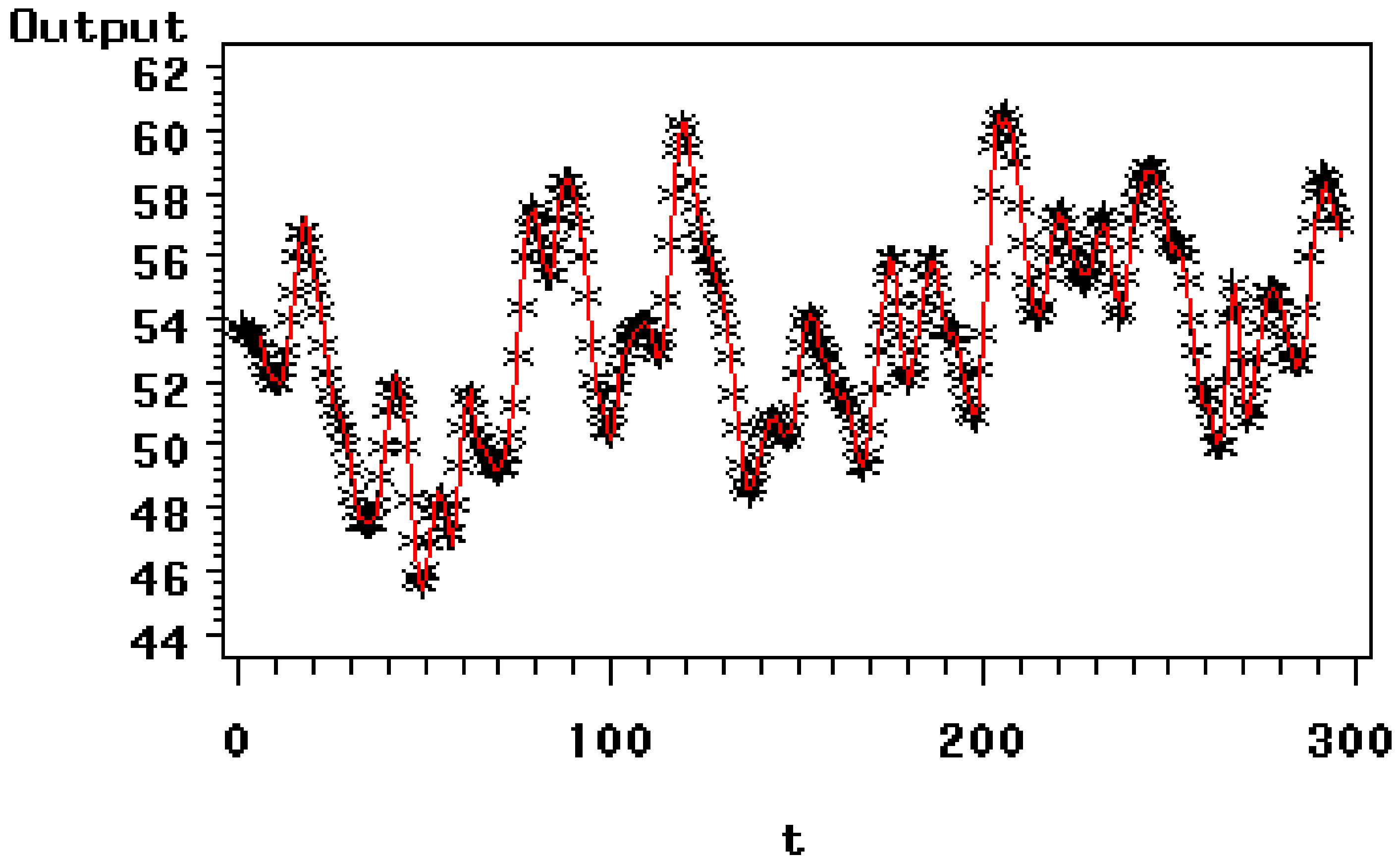
残差拟合模型

$$\varepsilon_t = \frac{1}{1-1.53B+0.64B^2}a_t$$

拟合模型

	模型结构	比较
一元模型	$y_t = 53.9 + \dfrac{a_t}{1 - 3.1B + 1.3B^2 - 0.2B^4}$	AIC=196.3 SBC=211.1
多元模型	$\begin{cases} y_t = 53.26 + \dfrac{-0.54 - 0.38B - 0.52B^2}{1 - 0.55B} B^3 x_t + \varepsilon_t \\ \varepsilon_t = \dfrac{1}{1 - 1.53B + 0.64B^2} a_t \end{cases}$	AIC=8.3 SBC=34.0

ARIMAX模型拟合效果图



6.2 虚假回归

- 假设条件 $H_0:\beta_1=0 \leftrightarrow H_1:\beta_1 \neq 0$

- 检验统计量 $t=\dfrac{\beta_1}{\sigma_\beta}$

- 虚假回归 $\Pr\{|t| \leq t_{\alpha/2}(n) | \text{非平稳序列}\} \geq \alpha$

6.3 单位根检验

- 定义
  - 通过检验特征根是在单位圆内还是单位圆上（外），来检验序列的平稳性
- 方法
  - DF检验
  - ADF检验
  - PP检验

DF检验

- 假设条件
  - 原假设：序列非平稳 $H_0: |\phi_1| \geq 1$
  - 备择假设：序列平稳 $H_0: |\phi_1| < 1$
- 检验统计量
  - $|\phi_1| < 1$时 $t(\phi_1) = \dfrac{\hat{\phi}_1 - \phi_1}{S(\hat{\phi}_1)} \xrightarrow{\text{渐近}} N(0,1)$
  - $|\phi_1| = 1$时

$$\tau = \frac{|\hat{\phi}_1| - 1}{S(\hat{\phi}_1)}$$

DF统计量

- $|\phi_1| < 1$时

$$t(\phi_1) = \frac{\hat{\phi}_1 - \phi_1}{S(\hat{\phi}_1)} \xrightarrow{\text{渐近}} N(0,1)$$

- $|\phi_1| = 1$时

$$\tau = \frac{|\hat{\phi}_1| - 1}{S(\hat{\phi}_1)} \xrightarrow{\text{极限}} \frac{\int_0^1 W(r)dW(r)}{\sqrt{\int_0^1 [W(r)]^2 dr}}$$

DF检验的等价表达

- 等价假设

$$H_0: \ \rho = 0 \leftrightarrow H_1: \ \rho < 0$$

其中：$\rho = |\phi_1| - 1$

- 检验统计量

$$\tau = \frac{\hat{\rho}}{S(\hat{\rho})}$$

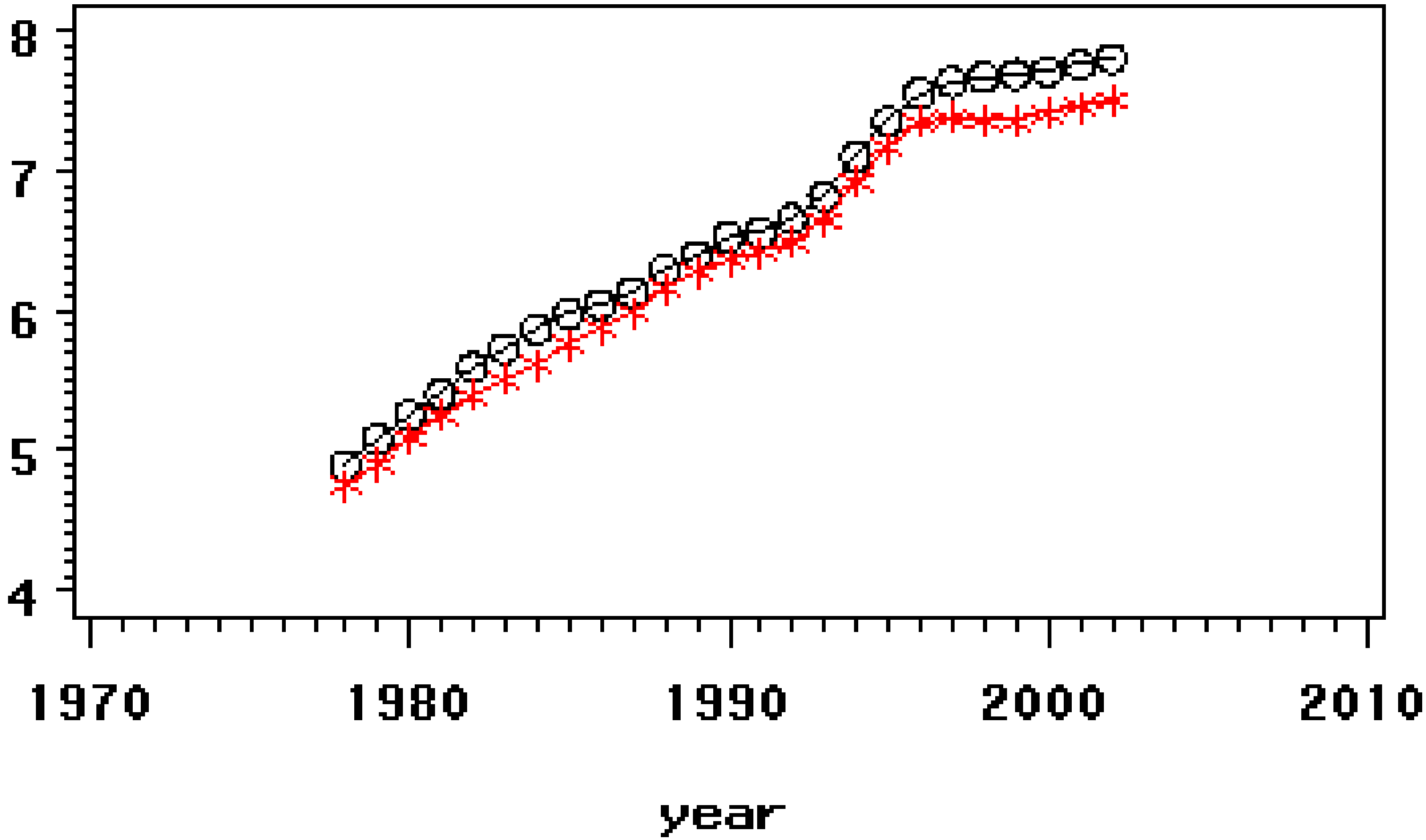
DF检验的三种类型

- 第一种类型 $x_t = \phi_1 x_{t-1} + \varepsilon_t$
- 第二种类型 $x_t = \mu + \phi_1 x_{t-1} + \varepsilon_t$
- 第三种类型 $x_t = \mu + \beta t + \phi_1 x_{t-1} + \varepsilon_t$

例6.2

- 对1978年－2002年中国农村居民家庭人均纯收入对数序列$\{\ln x_t\}$和生活消费支出对数序列$\{\ln y_t\}$进行检验

例6.2 时序图



例6.2 输入序列的DF检验

类型	延迟阶数	模型结构	τ检验统计量的值	$pr<\tau$
类型 1：无均值,无趋势模型	0	$\ln x_t=\varepsilon_t$	7.05	0.9999
	1	$\ln x_t=\dfrac{\varepsilon_t}{1-\phi_1 B}$	1.09	0.9224
类型 2：有均值,无趋势模型	0	$\ln x_t=\mu+\varepsilon_t$	-2.18	0.2190
	1	$\ln x_t=\mu+\dfrac{\varepsilon_t}{1-\phi_1 B}$	-1.16	0.6730
类型 3：有趋势模型	0	$\ln x_t=\mu+\beta t+\varepsilon_t$	-0.7	0.9617
	1	$\ln x_t=\mu+\beta t+\dfrac{\varepsilon_t}{1-\phi_1 B}$	-3.14	0.1215

例6.2 输出序列的DF检验

类型	延迟阶数	模型结构	τ 检验统计量的值	$pr<\tau$
类型 1：无均值,无趋势模型	0	$\ln y_t=\varepsilon_t$	6.54	0.9999
	1	$\ln y_t=\dfrac{\varepsilon_t}{1-\phi_1 B}$	1.08	0.9217
类型 2：有均值,无趋势模型	0	$\ln y_t=\mu+\varepsilon_t$	-2.27	0.1887
	1	$\ln y_t=\mu+\dfrac{\varepsilon_t}{1-\phi_1 B}$	-1.41	0.5598
类型 3：有趋势模型	0	$\ln y_t=\mu+\beta t+\varepsilon_t$	-0.46	0.9784
	1	$\ln y_t=\mu+\beta t+\dfrac{\varepsilon_t}{1-\phi_1 B}$	-3.31	0.0892

ADF检验

- DF检验只适用于AR(1)过程的平稳性检验。为了使检验能适用于AR(p)过程的平稳性检验，人们对检验进行了一定的修正，得到增广检验（Augmented Dickey – Fuller），简记为ADF检验

ADF检验的原理

- 若AR(p)序列有单位根存在，则自回归系数之和恰好等于1

$$\lambda^p - \phi_1 \lambda^{p-1} - \cdots - \phi_p = 0$$

$$\overset{\lambda=1}{\Longrightarrow} 1 - \phi_1 - \cdots - \phi_p = 0$$

$$\Longrightarrow \phi_1 + \phi_2 + \cdots + \phi_p = 1$$

ADF检验

- 等价假设

$$H_0: \ \rho = 0 \leftrightarrow H_1: \ \rho < 0$$

其中：$\rho = \phi_1 + \phi_2 + \cdots + \phi_p - 1$

- 检验统计量

$$\tau = \frac{\hat{\rho}}{S(\hat{\rho})}$$

ADF检验的三种类型

- 第一种类型 $$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t$$
- 第二种类型 $$x_t = \mu + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t$$
- 第三种类型

$$x_t = \mu + \beta t + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t$$

例6.2续

- 对1978年－2002年中国农村居民家庭人均纯收入对数差分后序列$\{\nabla \ln x_t\}$和生活消费支出对数差分后序列$\{\nabla \ln y_t\}$进行检验

例6.2 $\{\nabla \ln x_t\}$ 序列的ADF检验

类型	延迟阶数	τ 检验统计量的值	$pr < \tau$
类型 1	0	-1.38	0.1498
	1	-1.37	0.1523
	2	-1.36	0.1555
	3	-1.34	0.1606
类型 2	0	-1.92	0.3196
	1	-2.23	0.2033
	2	-3.1	0.0421
	3	-1.9	0.3278
类型 3	0	-2.08	0.5311
	1	-2.41	0.3646
	2	-3.37	0.0820
	3	-1.99	0.5732

例6.2 $\{\nabla \ln y_t\}$ 序列的ADF检验

类型	延迟阶数	τ 检验统计量的值	$pr < \tau$
类型 1	0	-1.27	0.1813
	1	-1.8	0.0688
	2	-1.3	0.1699
	3	-1.06	0.2514
类型 2	0	-1.89	0.3297
	1	-3.35	0.0246
	2	-2.81	0.0737
	3	-1.72	0.4086
类型 3	0	-2.16	0.4897
	1	-3.62	0.0511
	2	-3.32	0.0902
	3	-2.21	0.4613

PP检验

- ADF检验主要适用于方差齐性场合，它对于异方差序列的平稳性检验效果不佳
- Phillips和 Perron于1988年对ADF检验进行了非参数修正，提出了PP检验统计量。
- PP检验统计量适用于异方差场合的平稳性检验，且服从相应的ADF检验统计量的极限分布

PP检验统计量

$$Z(\tau)=\tau(\hat{\sigma}^2/\hat{\sigma}_{Sl}^2)-(1/2)(\hat{\sigma}_{Sl}^2-\hat{\sigma}^2)T\sqrt{\hat{\sigma}_{Sl}^2\sum_{t=2}^{T}(x_{t-1}-\bar{x}_{T-1})^2}$$

- 其中：

(1) $\hat{\sigma}^2=T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^2$

(2) $\hat{\sigma}_{Sl}^2=T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^2+2T^{-1}\sum_{j=1}^{l}w_j(l)\sum_{t=j+1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_{t-j}$

(3) $\bar{x}_{T-1}=\dfrac{1}{T-1}\sum_{t=1}^{T-1}x_t$

例6.2续

- 对1978年－2002年中国农村居民家庭人均纯收入对数差分后序列$\{\nabla \ln x_t\}$和生活消费支出对数差分后序列$\{\nabla \ln y_t\}$进行PP检验

例6.2 $\{\nabla \ln x_t\}$序列的pp检验

类型	延迟阶数	τ检验统计量的值	$pr < \tau$
类型 1	0	-1.38	0.1498
	1	-1.4	0.1447
	2	-1.43	0.1381
	3	-1.39	0.1484
类型 2	0	-1.92	0.3196
	1	-2.05	0.2661
	2	-2.16	0.2240
	3	-2.06	0.2620
类型 3	0	-2.08	0.5311
	1	-2.22	0.4559
	2	-2.34	0.3995
	3	-2.22	0.4548

例6.2 $\{\nabla \ln y_t\}$序列的PP检验

类型	延迟阶数	τ检验统计量的值	$pr < \tau$
类型 1	0	-1.27	0.1813
	1	-1.37	0.1524
	2	-1.38	0.1513
	3	-1.3	0.1714
类型 2	0	-1.89	0.3297
	1	-2.17	0.2228
	2	-2.2	0.2110
	3	-2.06	0.2603
类型 3	0	-2.16	0.4897
	1	-2.43	0.3536
	2	-2.47	0.3397
	3	-2.31	0.4137

例6.2 二阶差分后序列的PP检验

类型	延迟阶数	$\{\nabla^2 \ln x_t\}$		$\{\nabla^2 \ln y_t\}$	
		τ 检验统计量的值	$pr < \tau$	τ 检验统计量的值	$pr < \tau$
类型 1	0	-4.29	0.0002	-3.42	0.0016
	1	-4.29	0.0002	-3.47	0.0014
	2	-4.31	0.0002	-3.44	0.0015
	3	-4.27	0.0002	-3.34	0.0019
类型 2	0	-4.23	0.0035	-3.37	0.0237
	1	-4.23	0.0035	-3.42	0.0210
	2	-4.25	0.0033	-3.39	0.0224
	3	-4.21	0.0037	-3.28	0.0283
类型 3	0	-4.12	0.0193	-3.27	0.0983
	1	-4.12	0.0194	-3.33	0.0881
	2	-4.15	0.0184	-3.29	0.0934
	3	-4.09	0.0205	-3.17	0.1158

6.4 协整

- 单整的概念
  - 如果序列平稳，说明序列不存在单位根，这时称序列为零阶单整序列，简记为 $x_t \sim I(0)$
  - 假如原序列一阶差分后平稳，说明序列存在一个单位根，这时称序列为一阶单整序列，简记为 $x_t \sim I(1)$
  - 假如原序列至少需要进行d阶差分才能实现平稳，说明原序列存在d个单位根，这时称原序列为阶单整序列，简记为 $x_t \sim I(d)$

单整的性质

- 若$x_t \sim I(0)$，对任意非零实数a，b，有

$$a + bx_t \sim I(0)$$

- 若$x_t \sim I(d)$，对任意非零实数a，b，有

$$a + bx_t \sim I(d)$$

- 若$x_t \sim I(0)$,$y_t \sim I(0)$ 对任意非零实数a，b，有

$$z_t = ax_t + by_t \sim I(0)$$

- 若$x_t \sim I(d)$，$y_t \sim I(c)$ 对任意非零实数a，b，有

$$z_t = ax_t + by_t \sim I(k) \quad k \le \max[d, c]$$

协整的概念

- 假定自变量序列为 $\{x_1\},\cdots,\{x_k\}$，响应变量序列为 $\{y_t\}$，构造回归模型

$$y_t = \beta_0 + \sum_{i=1}^{k} \beta_i x_{it} + \varepsilon_t$$

假定回归残差序列 $\{\varepsilon_t\}$ 平稳，我们称响应序列 $\{y_t\}$ 与自变量序列 $\{x_1\},\cdots,\{x_k\}$ 之间具有协整关系。

协整检验

- 假设条件
  - 原假设：多元非平稳序列之间不存在协整关系

$$H_0: \varepsilon_t \sim I(k), k \geq 1$$

  - 备择假设：多元非平稳序列之间存在协整关系

$$H_1: \varepsilon_t \sim I(0)$$

- 检验步骤
  - 建立响应序列与输入序列之间的回归模型
  - 对回归残差序列进行平稳性检验

例6.2续

- 对1978年－2002年中国农村居民家庭人均纯收入对数序列 $\ln\{x_t\}$ 和生活消费支出对数序列 $\ln\{y_t\}$ 进行EG检验。

构造回归模型

- 拟合模型
  - 一元线性模型
- 估计方法
  - 最小二乘估计
- 拟合模型口径

$$\ln y_t = 0.96832 \ln x_t + \varepsilon_t$$

残差序列单位根检验

类型	延迟阶数	τ 检验统计量的值	$pr < \tau$
类型 1	0	-1.33	0.1629
	1	-1.69	0.0845
	2	-1.93	0.0528
类型 2	0	-1.28	0.6222
	1	-1.64	0.4455
	2	-1.85	0.3484

我们可以以91.55%（1－0.0845）的把握断定残差序列平稳且具有一阶自相关性

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \upsilon_t$$

最终拟合模型

$$\ln y_t = 0.96821 \ln x_t + \frac{\upsilon_t}{1 - 0.83713B}$$

$$v_t \overset{i.i.d}{\sim} N(0, 0.000893)$$

误差修正模型

- 误差修正模型（Error Correction Model）简称为ECM，最初由Hendry和Anderson于1977年提出，它常常作为协整回归模型的补充模型出现
- 协整模型度量序列之间的长期均衡关系，而ECM模型则解释序列的短期波动关系

短期影响因素分析

- 响应序列的当期波动∇y_t主要会受到三方面短期波动的影响
  - 输入序列的当期波动 ∇x_t
  - 上一期的误差 ECM_{t-1}
  - 纯随机波动 ε_t

$$y_t - y_{t-1} = \beta x_t - \beta x_{t-1} - \varepsilon_{t-1} + \varepsilon_t$$

$$\nabla y_t = \beta \nabla x_t - ECM_{t-1} + \varepsilon_t$$

误差修正模型

$$\nabla y_t = \beta_0 \nabla x_t + \beta_1 ECM_{t-1} + \varepsilon_t$$

例6.2续

- 对1978年－2002年中国农村居民家庭人均纯收入对数序列 $\ln\{x_t\}$ 和生活消费支出对数序列 $\ln\{y_t\}$ 构造ECM模型

例6.2 构造ECM模型

- 拟合长期协整关系

$$\ln y_t = 0.96832 \ln x_t + \varepsilon_t$$

- 拟合短期波动（ECM模型）

$$\nabla \ln y_t = 0.9579 \nabla \ln x_t - 0.1537 ECM_{t-1} + \varepsilon_t$$

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