# 时间序列分析（TIME SERIES ANALYSIS）

主讲：吴尚
复旦大学管理学院统计与数据科学系

# 时间序列模型

- MA(q)， AR(p)，ARMA(p, q)
- 差分算子
- 非平稳时间序列ARIMA(p, d, q)模型

# 非平稳时间序列的例子

- $Y_t = M_t + X_t$，$M_t$均值不为常数， $X_t$平稳

- $Y_t = \phi Y_{t-1} + e_t$，$|\phi| \geq 1$

- $Y_t = M_t + e_t$，$M_t = M_{t-1} + \varepsilon_t$，$\{e_t\}$和$\{\varepsilon_t\}$为独立白噪声

- 如何将上述序列平稳化?

# 差分算子

- 一阶差分，$\nabla Y_t = Y_t - Y_{t-1}$
- $d$阶差分， $\nabla^d Y_t = \nabla^{d-1}(Y_t - Y_{t-1}) = \nabla^{d-1} Y_t - \nabla^{d-1} Y_{t-1}$
- $s$步差分， $\nabla_s Y_t = Y_t - Y_{t-s}$
- $\nabla Y_t = (1-B)Y_t$， $\nabla^d Y_t = (1-B)^d Y_t$
- $\nabla_s Y_t = (1-B^s)Y_t$

# 二阶差分例子

- $Y_t = M_t + e_t$，$M_t = M_{t-1} + W_t$，$W_t = W_{t-1} + \varepsilon_t$
- $\{e_t\}$和$\{\varepsilon_t\}$为独立白噪声
- 如何通过差分将$Y_t$变成平稳过程，是什么模型?

# ARIMA(p, d, q)模型

- 如果一个时间序列$\{Y_t\}$的$d$阶差分$W_t = \nabla^d Y_t$是一个平稳的$ARMA(p,q)$过程，则称$\{Y_t\}$为**自回归滑动平均求和**模型，记为$ARIMA(p,d,q)$模型，通常取$d = 1$ 或 2.
- 可简写为$\Phi(B)\nabla^d Y_t = \Theta(B)e_t$
- 其中 $\Phi(x) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p, \phi_p \neq 0$,

  $$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q, \theta_q \neq 0$$
- 且$\Phi(x) = 0$的根都在单位圆外。
- 学会识别给定模型的具体参数（多项式分解）

# ARIMA(p, d, q)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)=随机游动

# ARIMA(p, d, q)的求和形式

- 以随机游动$Y_t = Y_{t-1} + e_t$为例
- $Y_t = e_t + e_{t-1} + e_{t-2} + e_{t-3} + \cdots$
- 该方差趋于无穷，所以我们不妨假设当$t < -m$时，$Y_t = 0$
- 可得ARIMA(p, 1, q)模型$Y_t = Y_{t-1} + W_t$的求和形式$Y_t = \sum_{j=-m}^{t} W_j$
- 类似的，如果$X_t = X_{t-1} + Y_t$, 当$t < -m$时，$X_t = 0$
- 则 $X_t = \sum_{j=-m}^{t} Y_j = \sum_{j=-m}^{t} \sum_{i=-m}^{j} W_i = \sum_{j=0}^{t+m} (j + 1) W_{t-j}$ 为ARIMA(p, 2, q)模型的求和形式
- 该求和形式可用于研究ARIMA模型的自协方差和自相关系数特征

# IMA(1, 1)模型

- $Y_t = Y_{t-1} + e_t - \theta e_{t-1}$
- $Y_t = \sum_{j=-m}^{t}(e_j - \theta e_{j-1}) = e_t + (1-\theta)e_{t-1} + \cdots + (1-\theta)e_{-m} - \theta e_{-m-1}$
- $Var(Y_t) = [1 + \theta^2 + (1-\theta)^2(t+m)]\sigma_e^2$
- $\gamma_{t,t-k} = [1 - \theta + \theta^2 + (1-\theta)^2(t+m-k)]\sigma_e^2$
- $\rho_{t,t-k} = \frac{1-\theta+\theta^2+(1-\theta)^2(t+m-k)}{\sqrt{[1+\theta^2+(1-\theta)^2(t+m)][1+\theta^2+(1-\theta)^2(t+m-k)]}} \approx \sqrt{\frac{t+m-k}{t+m}}$
- 对于多个滞后期数 $k$，$Y_t$ 和 $Y_{t-k}$ 高度正相关

# IMA(2, 2)模型

- $Y_t = 2Y_{t-1} - Y_{t-2} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$
- $Y_t = \sum_{j=0}^{t+m}(j+1)W_{t-j} = \sum_{j=0}^{t+m}(j+1)(e_{t-j} - \theta_1 e_{t-j-1} - \theta_2 e_{t-j-2}) = \sum_{j=0}^{t+m+2} \psi_j e_{t-j}$
- 其中$\psi_0 = 1$，$\psi_j = 1 + \theta_2 + (1 - \theta_1 - \theta_2)j$，$j = 1, \dots, t+m$
- $\psi_{t+m+1} = -(t+m+1)\theta_1 - (t+m)\theta_2$，$\psi_{t+m+2} = -(t+m+1)\theta_2$
- $Var(Y_t) = \sigma_e^2 \sum_{j=0}^{t+m+2} \psi_j^2 \approx \sigma_e^2 \sum_{j=1}^{t+m}(1-\theta_1-\theta_2)^2 j^2 \approx R(t+m)^3$
- 其中$R = \sigma_e^2(1-\theta_1-\theta_2)^2/3$，注意
- $\gamma_{t,t-k} = \sigma_e^2 \sum_{j=0}^{t+m+2-k} \psi_j \psi_{j+k}$
  $\approx \sigma_e^2 \sum_{j=1}^{t+m-k}(1-\theta_1-\theta_2)^2 j(j+k) \approx R(t+m-k)^2\left(t+m+\frac{k}{2}\right) >$
  $R(t+m-k)^2(t+m)$
- $\rho_{t,t-k} \approx \frac{R(t+m-k)^2\left(t+m+\frac{k}{2}\right)}{\sqrt{R(t+m)^3 R(t+m-k)^3}} > \sqrt{\frac{t+m-k}{t+m}}$
- $Y_t$ 和 $Y_{t-k}$ 高度正相关，且自相关系数比IMA(1, 1)同期更大

# ARI(1, 1)模型

- $Y_t = Y_{t-1} + W_t$，$W_t = \phi W_{t-1} + e_t$
- (1) $Y_t = (1+\phi)Y_{t-1} - \phi Y_{t-2} + e_t$
- $\psi_0 = 1$，$\psi_1 = 1+\phi$，$\psi_k = (1+\phi)\,\psi_{k-1} - \phi\psi_{k-2}$
- 通项为$\psi_k = 1 + \phi + \phi^2 + \cdots + \phi^k = \frac{1-\phi^{k+1}}{1-\phi}$，$\sum_{k=0}^{\infty} \psi_k^2 = \infty$
- (2) $Y_t = \sum_{j=-m}^{t}(e_j + \phi e_{j-1} + \phi^2 e_{j-2} + \cdots)$
- $Y_t = e_t + (1+\phi)e_{t-1} + (1+\phi+\phi^2)e_{t-2} + \cdots + \frac{1-\phi^{t+m-1}}{1-\phi}\sum_{k=0}^{\infty}\phi^k e_{-m-k}$
- 拓展：求解 $Var(Y_t)$，$\gamma_{t,t-k}$，$\rho_{t,t-k}$

# 差分消除多项式趋势

- $Y_t = \mu_t + X_t$，$\mu_t = \mu_0 + \mu_1 t + \mu_2 t^2 + \cdots + \mu_{d-1} t^{d-1}$
- $\nabla^d Y_t = \nabla^d X_t$
- 所以消去$\nabla^d$的同时要加上$d-1$次多项式
- 小心过度差分，例如真实模型为$Y_t = t + X_t$，$X_t$平稳，则两次差分后$\nabla^2 Y_t = \nabla^2 X_t$，然而$\nabla^2 X_t$并非可逆模型。这种情况不应该差分，而应该直接去除趋势项
- 思考：比较模型$Y_t = a + bt + e_t$和$\nabla^2 Y_t = e_t$

# ARIMA模型中的常数项

- 对于平稳模型，$\Phi(B)W_t = \theta_0 + \Theta(B)e_t$可以直接改写为$\Phi(B)(W_t - \frac{\theta_0}{\Phi(1)}) = \Theta(B)e_t$
- 减去$W_t$的均值 $\mu = \frac{\theta_0}{\Phi(1)} = \frac{\theta_0}{1-\phi_1-\phi_2-\cdots-\phi_p}$后变成零均值模型。
- $\Phi(B)\nabla^d Y_t = \theta_0 + \Theta(B)e_t$
- 由于$\Phi(B)\nabla^d \frac{t^d}{d!}\mu = \theta_0$，令$X_t = Y_t - \frac{t^d}{d!}\mu$，则$\Phi(B)\nabla^d X_t = \Theta(B)e_t$，注意这里$X_t$仍然可以有多项式趋势
- 例：$(1-0.5B)\nabla Y_t = 1 + e_t - 0.5e_{t-1}$
- 思考：$\nabla^2 Y_t = 6t + e_t$

# 其它消除不平稳性的变换

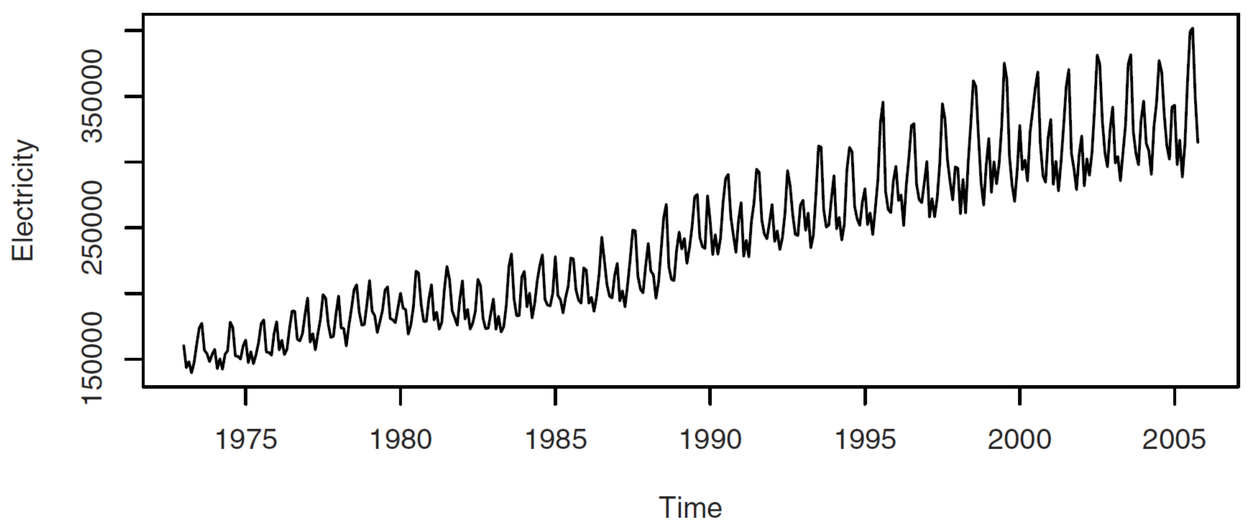
- (1) 对数变换：设$Y_t > 0$，$\mathrm{E}(Y_t) = \mu_t$，$\sqrt{\mathrm{Var}(Y_t)} = \mu_t \sigma$
- $\log Y_t = \log \mu_t + \log\left(1 + \frac{Y_t - \mu_t}{\mu_t}\right) \approx \log \mu_t + \frac{Y_t - \mu_t}{\mu_t}$
- 两边同时求期望和方差可得
- $\mathrm{E}[\log Y_t] \approx \log \mu_t$，$\mathrm{Var}[\log Y_t] \approx \sigma^2$
- (2) 对数差分变换：设$Y_t = (1 + X_t) Y_{t-1}$，则
- $\nabla[\log Y_t] = \log(1 + X_t) \approx X_t$
- 这里$\nabla[\log Y_t]$通常称为对数收益率，近似 $X_t = \frac{Y_t - Y_{t-1}}{Y_{t-1}}$

# 回顾 R 语言

- 阅读教材附录1：R入门
- 安装R，Rstudio
- 运行 install.packages(“TSA”)，每台电脑只需一次
- 每次写程序时，一开始要加上 library(TSA)
- 如果出错，可以卸载后尝试用其他版本的R，如3.6版本等

# 例：美国月度发电量

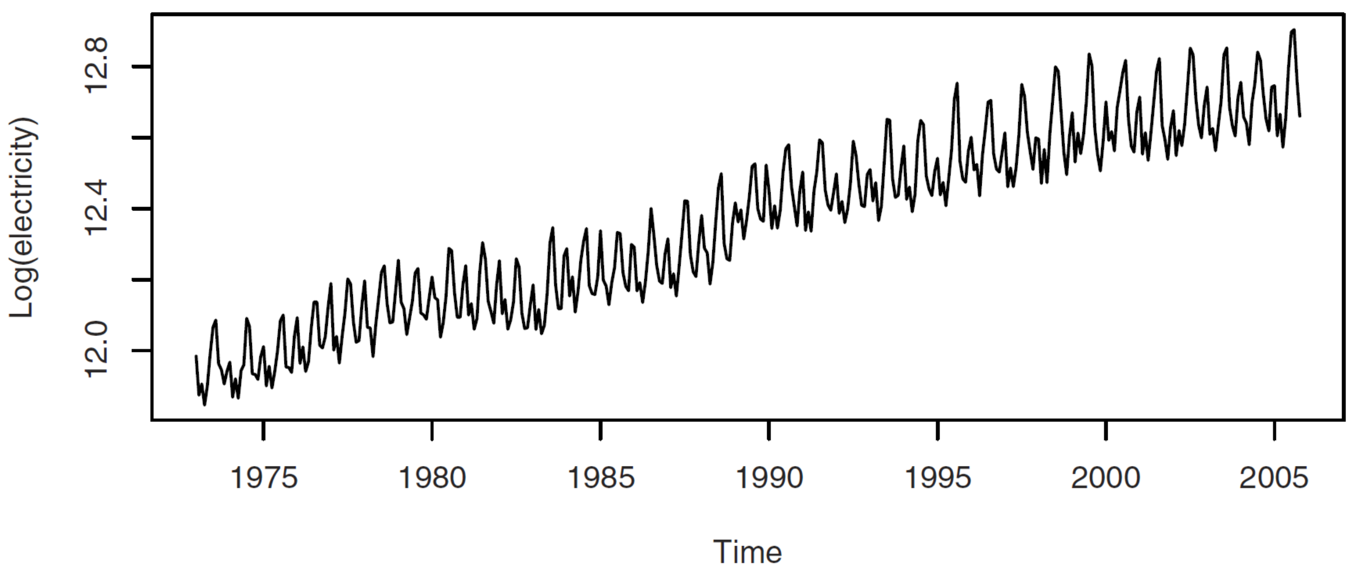
**Exhibit 5.8** **U.S. Electricity Generated by Month**




```
> data(electricity); plot(electricity)
```

# 例：美国月度发电量

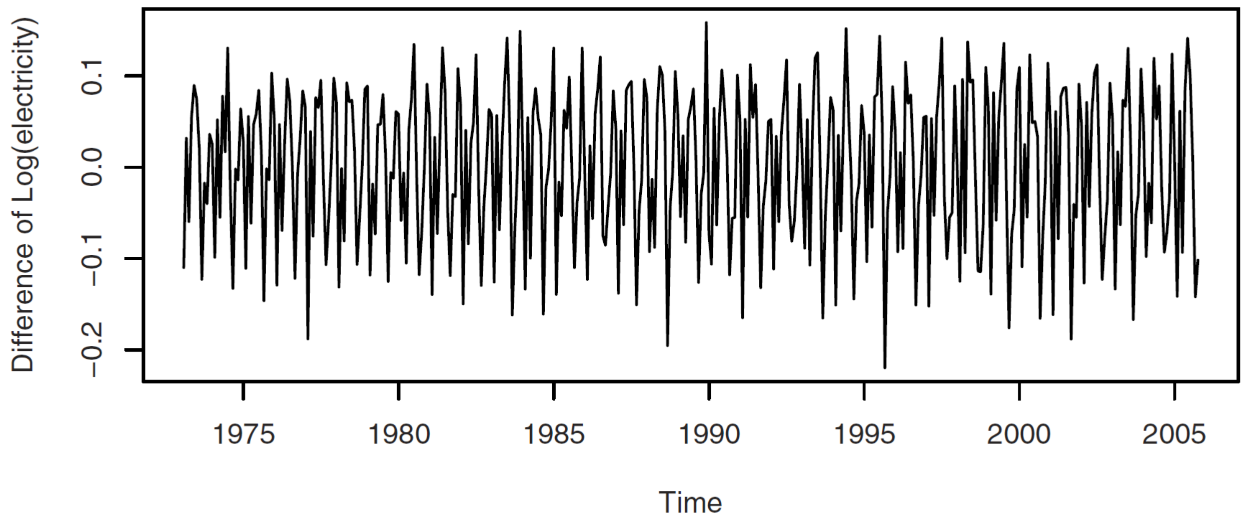
**Exhibit 5.9** **Time Series Plot of Logarithms of Electricity Values**




```
> plot(log(electricity),ylab='Log(electricity)')
```

# 例：美国月度发电量

**Exhibit 5.10 Difference of Logarithms for Electricity Time Series**



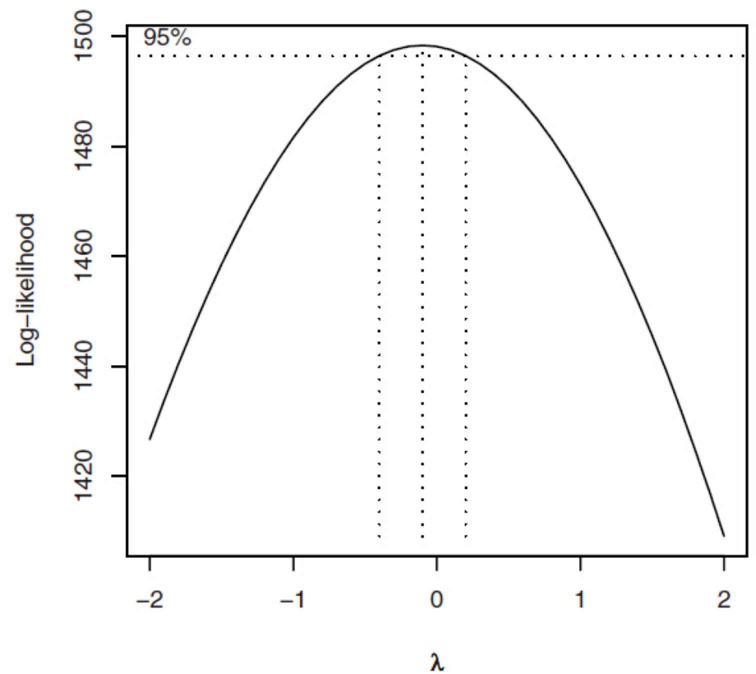

```
> plot(diff(log(electricity)),
   ylab='Difference of Log(electricity)')
```

# 幂变换

■ 给定参数$\lambda$的值，幂变换（Box-Cox变换）定义为

$$g(x) = \begin{cases} \dfrac{x^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log x & \lambda = 0 \end{cases}$$

**Exhibit 5.11 Log-likelihood versus Lambda**




```
> BoxCox.ar(electricity)
```