§2.2 非线性模型的几个专门问题
（教材§5.5）

一、一般非线性模型的最大似然估计
二、因变量的参数变换
三、异方差性的非线性方法
四、序列相关性的非线性方法
五、条件异方差性的非线性方法
附：第5章内容简介

一、一般非线性模型的最大似然估计

1. 一般非线性模型的描述

$$h(y_i,\theta) = g(x_i,\beta) + u_i \qquad i = 1,\cdots,n$$

$$(u_1,\cdots,u_n)' \sim N(0,\sigma^2 I)$$

随机项满足经典假设

$$x_i = \begin{pmatrix} x_{1i} & x_{2i} & \cdots & x_{ki} \end{pmatrix}$$

其中 $h(\cdot)$ 和 $g(\cdot)$ 是非线性函数，θ 和 β 是参数。

- 以上是一般非线性模型的完整描述。

- 模型参数的一种估计方法是最小二乘法，即最小化

$$\boxed{S(\theta, \beta) = \sum_i [h(y_i, \theta) - g(x_i, \beta)]^2}$$

- 模型参数的另一种估计方法是最大似然法。得到广泛应用。

2. 最大似然估计

$$y_i = g(x_i, \beta) + u_i$$

- y_i的密度函数

$$(2\pi\sigma^2)^{-1/2}\exp\left(-\frac{y_i - g(x_i,\beta)]^2}{2\sigma^2}\right)$$

$$\left|\frac{\partial u_i}{\partial y_i}\right|(2\pi\sigma^2)^{-1/2}\exp\left(-\frac{[h(y_i,\theta) - g(x_i,\beta)]^2}{2\sigma^2}\right)$$

$$J(y_i,\theta) = \left|\frac{\partial u_i}{\partial y_i}\right| = \left|\frac{\partial h(y_i,\theta)}{\partial y_i}\right| = J_i$$

雅可比行列式

雅可比行列式×正态分布密度函数

- 因变量样本的对数似然函数为：

$$\ln L = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 + \sum_i \ln J(y_i,\theta) - \frac{1}{2\sigma^2}\sum_i [h(y_i,\theta) - g(x_i,\beta)]^2$$

- 很明显若没有雅可比行列式项，参数的非线性最小二乘估计将是最大似然估计；然而，如果雅可比行列式包括θ，最小二乘法不是最大似然法。

- 最大化对数似然函数的一阶条件为：

$$\frac{\partial \ln L}{\partial \beta} = \frac{1}{\sigma^2} \sum_i u_i \frac{\partial g(x_i, \beta)}{\partial \beta} = 0$$

$$\frac{\partial \ln L}{\partial \theta} = \sum_i \frac{1}{J_i} \left(\frac{\partial J_i}{\partial \theta} \right) - \left(\frac{1}{\sigma^2} \right) \sum_i u_i \frac{\partial h(y_i, \theta)}{\partial \theta} = 0$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_i u_i^2 = 0$$

- 一般是得到中心化对数似然函数，然后最大化

$$\sigma^2 = \frac{1}{n}\sum_i u_i^2$$

$$\ln L_c = \sum_i \ln J(y_i, \theta) - \frac{n}{2}[1 + \ln(2\pi)] - \frac{n}{2}\ln\left[\frac{1}{n}\sum_i u_i^2\right]$$

- 如果变换的雅可比行列式是**1**，则不存在因变量的参数变换；如果变换的雅可比行列式包含θ，则称为因变量的参数变换模型。

二、因变量的参数变换

1. Box-Cox变换

- 一种将变量之间的非线性关系变换为线性关系的方法。
- **Box**和**Cox（1964）**提出的变换关系：

$$x^{(\lambda)}=\frac{x^{\lambda}-1}{\lambda}$$

要求变量**x**为正值。λ取值可以是整个实数域但多数应用有意义的取值范围为**[-2，2]**。

当λ**=2**，是二次变换；当λ**=0.5**，是平方根变换；当λ**=1**，是线性变换；当λ**=-1**，是倒数变换；当λ**=0**，是对数变换。

- 例如：

$$y^2 = \beta_0 + \beta_1 x_1 + \beta_2 x_2^{\ 2} \longrightarrow y^{(2)} = \alpha_0 + \alpha_1 x_1^{(1)} + \alpha_2 x_2^{(2)}$$

$$y^{(2)} = \frac{y^2 - 1}{2} \qquad x_1^{(1)} = \frac{x_1 - 1}{1} \qquad x_2^{(2)} = \frac{x_2^{\ 2} - 1}{2}$$

$$y = \beta_0 + \beta_1 \ln x \longrightarrow y^{(1)} = \alpha_0 + \alpha_1 x^{(0)}$$

$$\lim_{\lambda \to 0} x^{(\lambda)} = \lim_{\lambda \to 0} \frac{x^{\lambda} - 1}{\lambda} = \ln(x)$$

- 如果已知被解释变量和解释变量各自进行何种λ的**B-C**变换，可以先变换，然后估计线性模型。
- 一般情况下，何种λ未知，作为一组参数引入模型，对变换后的模型进行非线性模型估计，同时得到λ和β的估计量。
- 许多应用软件，例如**GAUSS**、**SAS**可以实现。
- 这就引出了**B-C**变换的更重要的价值：如果不知道被解释变量和解释变量之间存在何种形式的函数关系，可以通过“**B-C**变换非线性模型估计”确定函数关系。

2. Box-Cox非线性回归模型的参数估计

$$u_i = f(y_i, x_i, \lambda, \beta) = y_i^{(\lambda_0)} - (\beta_0 + \beta_1 x_{1i}^{(\lambda_1)} + \cdots + \beta_k x_{ki}^{(\lambda_k)})$$

$$u = (u_1, \cdots, u_n)' \sim N(0, \sigma^2 I)$$

- 模型中被解释变量样本的对数似然函数为：

$$\boxed{\ln L(\lambda, \beta, \sigma^2 \mid y, x) = -\frac{1}{2} n[\ln(2\pi) + \ln(\sigma^2)] - \frac{1}{2}(u'u / \sigma^2) + (\lambda_0 - 1)\sum_{i=1}^{n} \ln y_i}$$

- 中心化对数似然函数：

$$\boxed{\ln L_c(\lambda,\beta\,|\,y,x) = -\frac{1}{2}n[1+\ln(2\pi)-\ln n] - \frac{1}{2}n\ln(u'u) + (\lambda_0-1)\sum_{i=1}^{n}\ln y_i}$$

$$\boxed{= -\frac{1}{2}n[1+\ln(2\pi)-\ln n] - \sum_{i=1}^{n}\ln y_i - \frac{1}{2}n\ln(u^{*\prime}u^{*})}$$

$$\boxed{u^{*} = u/[(y_1\times\cdots\times y_n)^{\lambda_0/n}]}$$

最大化中心化对数似然函数，就得到参数 λ 和 β 的最大似然估计 $\hat{\lambda}$ 和 $\hat{\beta}$ 。

σ^2的最大似然估计为：$\hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n} f^2(y_i, x_i, \hat{\lambda}, \hat{\beta})$

响应系数和弹性系数为：

$$\frac{\partial y_i}{\partial x_{ji}} = -\frac{\partial f}{\partial x_{ji}} \Bigg/ \frac{\partial f}{\partial y_i} = \beta_j x_{ji}^{\lambda_j - 1} / y_i^{\lambda_0 - 1}$$

$$\varepsilon_{yj} = \frac{\partial y_i}{\partial x_{ji}} \cdot \frac{x_{ji}}{y_i} = -\left(\frac{\partial f}{\partial x_{ji}} \Bigg/ \frac{\partial f}{\partial y_i}\right) \cdot \left(\frac{x_{ji}}{y_i}\right) = \beta_j x_{ji}^{\lambda_j} / y_i^{\lambda_0}$$

- 示例：

假定被解释变量**y**与解释变量**x**和**z**之间的关系为：

$$y^{(\lambda_1)} = \beta_1 + \beta_2 x^{(\lambda_2)} + \beta_3 z^{(\lambda_3)} + v$$

用模型 $y = \beta_1^* + \beta_2 x + \beta_3 z + u$

在 $\beta_1^* = 2, \beta_2 = \beta_3 = 1, u$ 是 i.i.d. $N(0, \sigma^2), \sigma^2 = 2.25$

限制下生成一组样本。然后估计 B-C 非线性回归模型。

x	z	y	x	z	y
5. 381719	3. 336064	11. 00728	5. 331691	5. 716792	10. 80478
3. 009112	5. 232138	9. 48510	4. 800605	5. 066528	11. 58843
4. 375199	3. 546423	6. 444172	3. 941124	3. 141868	9. 877543
5. 472562	5. 590764	13. 44912	5. 720727	2. 194553	8. 04193
4. 701641	2. 12414	11. 50133	4. 536456	2. 789461	13. 27586
4. 047996	3. 869655	12. 35815	3. 084652	4. 212297	9. 501099
5. 855000	4. 078571	12. 99527	4. 508164	5. 634752	10. 30391
4. 820906	3. 933097	14. 13140	5. 632088	4. 536204	14. 21714
2. 453591	5. 121045	8. 881935	4. 889667	4. 324151	9. 392368
2. 453289	3. 521771	5. 953811	4. 028880	3. 201167	11. 79774
2. 512200	2. 927539	7. 790439	3. 825508	5. 426684	13. 31844
3. 462947	4. 706172	10. 50841	5. 364851	4. 563735	10. 90918
3. 816963	3. 639600	10. 39946	4. 445447	3. 734707	13. 38010
2. 723284	4. 589414	9. 679147	3. 971793	4. 092285	8. 883878
4. 203508	2. 221079	7. 571837	5. 899375	3. 708031	7. 875557
3. 493063	5. 657767	12. 43048	5. 514231	3. 824409	11. 76839
4. 294265	3. 375917	13. 01008	4. 328222	2. 222048	9. 82212
3. 345298	5. 094239	10. 55083	2. 191853	5. 540067	13. 19992
5. 096056	3. 493393	9. 798799	5. 982851	4. 974421	14. 95182
3. 257751	2. 634803	9. 098704	3. 787909	4. 886206	10. 57826

施加λ相同约束的估计结果

参数	*估计值*	*标准误差*	*t 统计量*
β_1	3.380	1.128	2.99645
β_2	0.746	0.278	2.68345
β_3	0.574	0.278	2.06475
λ	0.779	0.088	8.85227
σ^2	1.240	0.392	3.16327

真值：β1=2，β2=1，
β3=1，λ=1

未施加λ相同约束的估计结果

参数	估计值	标准误差	t 统计量
β_1	-8.471	26.127	-0.3242
β_2	25.846	73.338	0.35245
β_3	1.610	5.264	0.30585
λ_1	0.969	0.089	10.8876
λ_2	-1.536	2.298	-0.6684
λ_3	0.391	2.613	0.14964
σ^2	2.876	0.910	3.16044

为什么结果很差？

三、异方差性的非线性方法

1. 思路

- 将异方差问题看成一类非线性问题，采用**NML**估计，比较简单，可以同时得到参数估计量和反映异方差特征的量。

$$f(y_i, x_i, \beta) = u_i \qquad i = 1, \cdots, n$$

$$U = (u_1, \cdots, u_n)' \sim N(0, \sigma^2 \Omega(\alpha))$$

其中 $\Omega(\alpha)$ 为对角元为正的对角方阵，且依赖于参数

$\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_m)$. 即 u_i 不存在序列相关但存在异方差现象。

- 被解释变量样本的对数似然函数为：

$$\ln L(\beta,\alpha,\sigma^2 \mid y,x) = -\frac{1}{2}n[\ln(2\pi)+\ln(\sigma^2)]$$

$$-\frac{1}{2}\ln(|\Omega|) - \frac{1}{2}U'\Omega^{-1}U/\sigma^2 + \sum_{i=1}^{n}\ln(|\frac{\partial u_i}{\partial y_i}|)$$

将 σ^2 用 $\sigma^2 = U'\Omega^{-1}U/n$ 代替得到中心化对数似然函数为：

$$\ln L_c(\beta,\alpha \mid y,x) = -\frac{1}{2}n[1+\ln(2\pi)-\ln(n)]$$

$$-\frac{1}{2}\ln(|\Omega|) - \frac{1}{2}n\ln(U'\Omega^{-1}U) + \sum_{i=1}^{n}\ln(|\frac{\partial u_i}{\partial y_i}|)$$

- 对异方差的结构给出假定，可以对模型的参数和异方差的结构参数进行最大似然估计。
- 针对不同的问题假定不同的异方差结构；针对同一个问题假定不同的异方差结构，进行估计和比较。
- 典型的异方差结构及其对应的对数似然函数，见教材。

$$u_i \sim N(0, \sigma_i^2)$$

$$\sigma_i^2 = \sigma^2 h_i(\alpha)$$

2. 例题

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + u$$

州	开支	收入	州	开支	收入	州	开支	收入
AL	275	6247	AK	821	10851	AZ	339	7374
AR	275	6183	CA	387	8850	CO	452	8001
CT	531	8914	DE	424	8604	DC	428	10022
FL	316	7505	GA	265	6700	HI	403	8380
ID	304	6813	IL	437	8745	IN	345	7696
IA	431	7873	KS	355	8001	KY	260	6615
LA	316	6640	ME	327	6333	MD	427	8306
MA	427	8063	MI	466	8442	MN	477	7847

MS	259	5736	MO	274	7342	MT	433	7051
NB	294	7391	NV	359	9032	NH	279	7277
NJ	423	8818	NM	388	6505	NY	447	8267
NC	335	6607	ND	311	7478	OH	322	7812
OK	320	6951	OR	397	7839	PA	412	7733
RI	342	7526	SC	315	6242	SD	321	6841
TN	268	6489	TX	315	7697	UT	417	6622
VT	353	6541	VA	356	7624	WA	415	8450
WV	320	6456	WI	NA	7597	WY	500	9096

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	832.9144	327.2925	2.544862	0.0143
X	-0.183420	0.082899	-2.212588	0.0318
X2	1.59E-05	5.19E-06	3.057433	0.0037

OLS

	Coefficient	Std. Error	z-Statistic	Prob.
C	832.9144	327.2925	2.544862	0.0109
X	-0.183420	0.082899	-2.212588	0.0269
X2	1.59E-05	5.19E-06	3.057433	0.0022

ML

参数	估计值	标准误差	t 统计量
β_0	831.02	354.36	2.3451
β_1	-1833.2	936.60	-1.9573
β_2	1588.7	611.87	2.5964
α	3.0100	1.3644	2.2061

－0.18332

0.000015887

参数	估计值	标准误差	t 统计量
β_0	831.21	362.81	2.2911
β_1	-1833.1	961.84	-1.9058
β_2	1588.5	630.65	2.5189
α	3.9094	1.7291	2.2610

- 线性模型，截面样本，一般存在异方差。
- 采用非线性最大似然法估计，可以得到关于异方差结构的估计结果。
- 在某些情况下，得到异方差结构的估计结果比模型参数估计量更重要。
- 这就是异方差性的非线性方法的意义所在。

四、序列相关性的非线性方法

- 见教材
- 首先假定模型随机误差项的序列相关结构。一般以**AR(1)**、**MA(1)**、**ARMA(1,1)**为常见。
- 求出随机误差项对被解释变量的偏导数表达式。
- 构造最大似然函数。
- 同时得到模型参数和随机误差项的序列相关结构的估计结果。

五、条件异方差性的非线性方法

ARCH

Auto Regressive Conditional Hoteroskedasticity

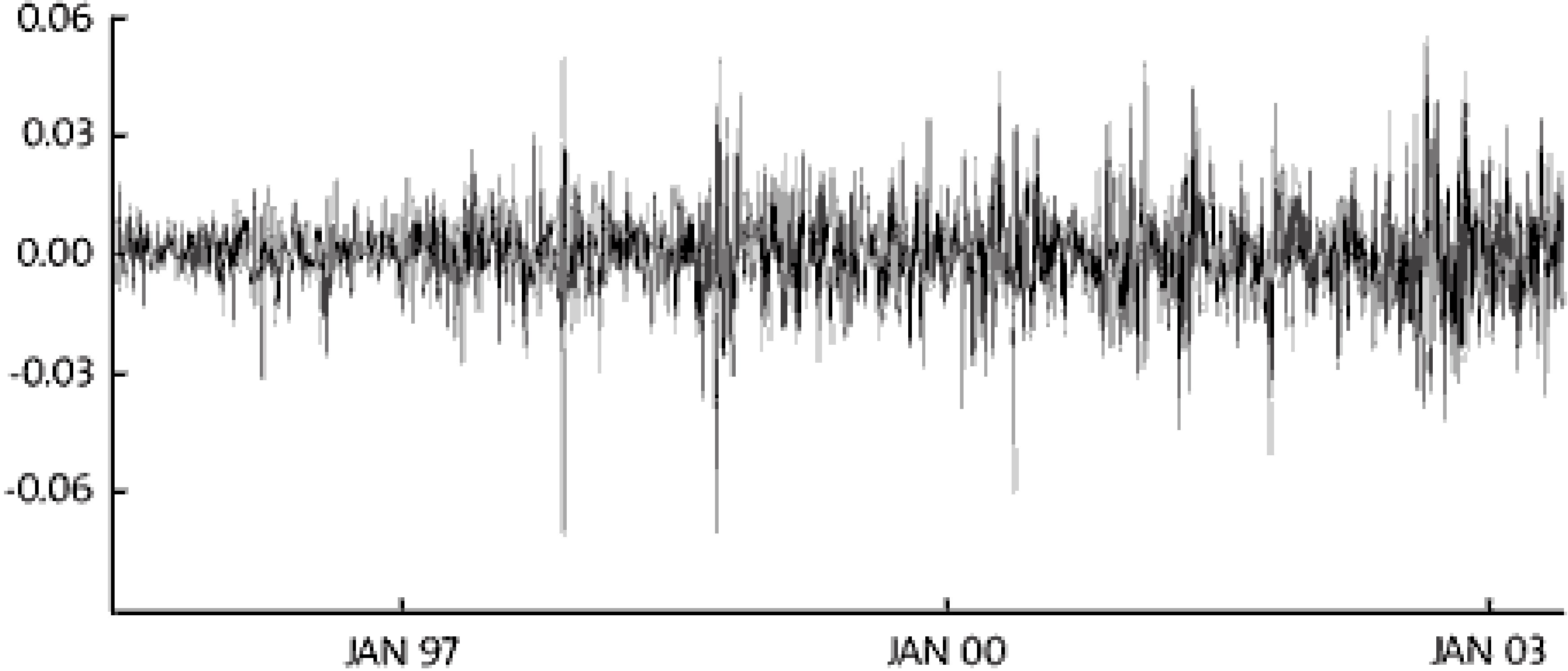
1. 条件异方差现象

- 通常横截面数据问题会产生异方差，而一般时间序列问题没有异方差现象。
- 如果时间序列数据问题出现异方差，经常以条件异方差形式。
- 所谓条件异方差，实际上是指“异方差”的“异”具有序列相关性。
- **Engle**于**1982**年分析英国通货膨胀率时首先发现条件异方差现象。
- 被广泛应用于金融市场时间序列分析。

- Engle, R.F.:1982, Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of U.K. Inflation," *Econometrica* 50: 987-1008.
- The application in Engle(1982) involved macroeconomic series such as the inflation rate, but Engle quickly realized that the ARCH model was useful in financial economics, as well.

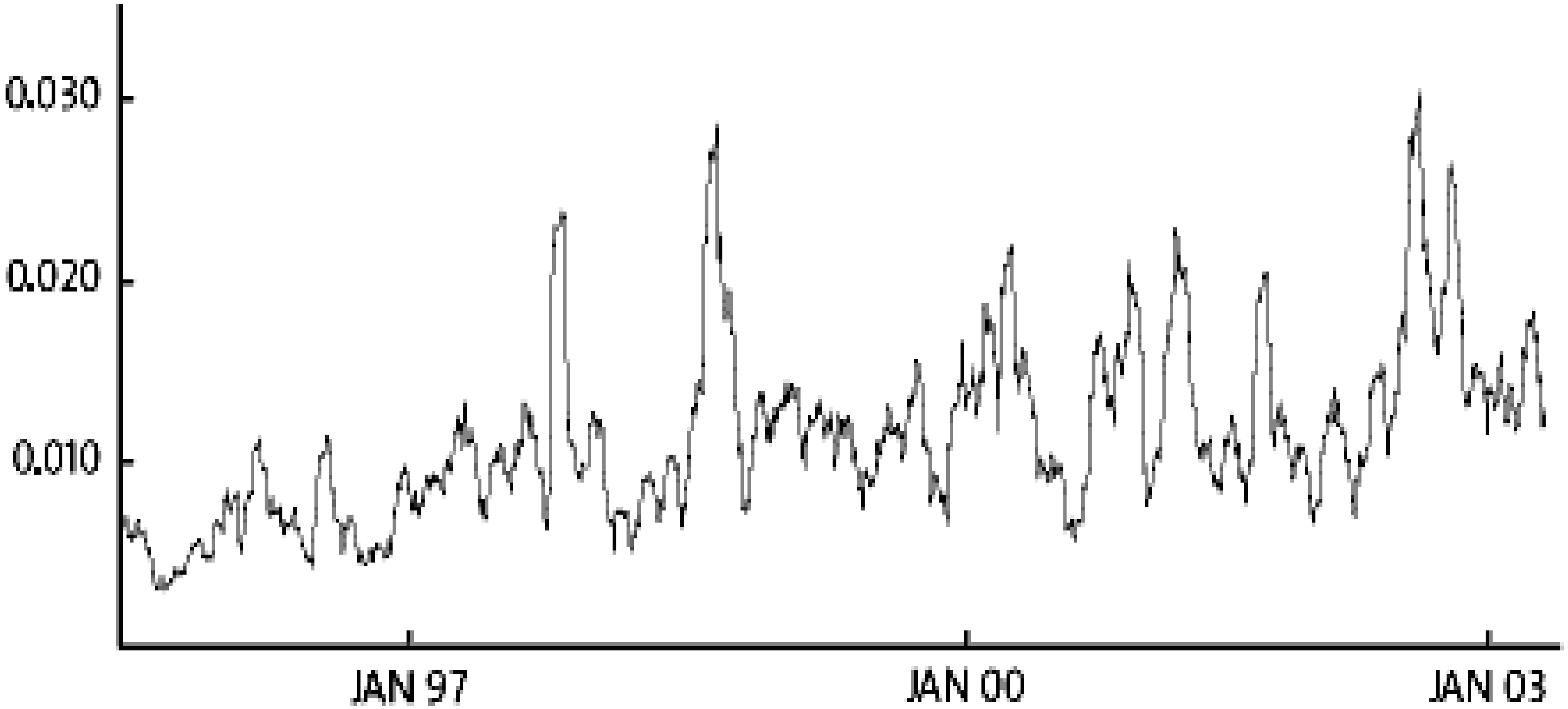
- **Risk evaluation is at the core of activities on financial markets.** Investors assess expected returns of an asset against its risk. Banks and other financial institutions would like to ensure that the value of their assets does not fall below some minimum level that would expose the bank to insolvency.
- **Such evaluations cannot be made without measuring the volatility of asset returns.** Robert Engle developed improved methods for carrying out these kinds of evaluations.

Percentage daily returns on an investment in the Standard & Poor 500 stock index May 16, 1995–April 29, 2003.



- The returns averaged 5.3 percent per year. At the same time there were days, when the fluctuations in prices were greater (plus or minus) than 5 percent. The standard deviation in daily returns measured over the entire period was 1.2 percent.
- Closer inspection reveals, however, that **the volatility varies over time: large changes (upwards or downwards) are often followed by further large fluctuations, and small changes tend to be followed by small fluctuations.**

Standard deviation for percentage daily returns on an investment in the Standard & Poor 500 stock index, May 16, 1995–April 29, 2003, computed from data for the four preceding weeks.



- **Many financial time series are characterized by similar time variation in volatility.**
- **Many financial economists are concerned with modeling volatility in asset returns.**

2. ARCH（q）模型

$$u_t = f(y_t, x_t, \beta)$$

具有异方差性

$$u_t \mid \psi_{t-1} \sim N(0, h_t)$$

异方差是有规律的：自回归

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 = \alpha_0 + \alpha(B) u_t^2$$

- 设样本有**n**个观察个数，则对数似然函数为：

$$\ln L(\beta,\alpha \mid y,x) = -\frac{1}{2}T\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln(h_t) - \frac{1}{2}\sum_{t=1}^{T}(u_t^2 / h_t) + \sum_{t=1}^{T}\ln\left|\frac{\partial u_t}{\partial y_t}\right|$$

- 模型估计的困难：长记忆下的高阶滞后。

3. GARCH（p, q）模型

$$u_t \mid \psi_{t-1} \sim N(0, h_t)$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{i=1}^{p} \theta_i h_{t-i}$$

$$\ln L(\beta, \alpha, \theta \mid y, x) = -\frac{1}{2} T \ln(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \ln(h_t) - \frac{1}{2} \sum_{t=1}^{T} (u_t^2 / h_t)$$

$$+ \sum_{t=1}^{T} \ln \left| \frac{\partial u_t}{\partial y_t} \right|$$

- **Bollerslev, T.:1986, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31, 307-327**
- The best-known extension is the generalized ARCH model (GARCH) developed by Tim Bollerslev in 1986. Here, the variance of the random error in a certain period depends not only on previous errors, but also on the variance itself in earlier periods.

- This development has turned out to be very useful; GARCH is the model most often applied today.
- Taylor(1986) suggested $p=q=1$, the most popular ARCH model in practice.

4. ARCH-M(q）模型

$$u_t = f(y_t, x_t, \beta, h_i) = y_i - x_i\beta - \gamma h_i$$

$$u_t \mid \psi_{t-1} \sim N(0, h_t)$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2$$

$$\ln L(\beta, \gamma, \alpha \mid y, x) = -\frac{1}{2} T \ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T} \ln(h_t) - \frac{1}{2}\sum_{t=1}^{T}(u_t^2 / h_t)$$

$$\frac{\partial u_t}{\partial y_t} = 1$$

为什么比ARCH少1项？

5. GARCH-M(p,q）模型

$$u_t = f(y_t, x_t, \beta, h_i) = y_i - x_i\beta - \gamma h_i$$

$$u_t \mid \psi_{t-1} \sim N(0, h_t)$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{i=1}^{p} \theta_i h_{t-i}$$

$$\ln L(\beta, \gamma, \alpha, \theta \mid y, x) = -\frac{1}{2} T \ln(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \ln(h_t) - \frac{1}{2} \sum_{t=1}^{T} (u_t^2 / h_t)$$

6. 检验

- **H_0：**不存在**ARCH**或**GARCH**
- 统计量**$LM=nR^2$**
- 用某种方法估计原模型，得到残差估计值
- 计算**LM**，进行**LM** 检验。

附：第5章内容简介

5.1 可线性化的非线性计量经济学模型

- 与§**2.1**相对应
- 变量非线性模型化为线性模型
  - 不涉及参数—函数变换：几种常见的变换函数
  - 涉及模型参数—仍然是一对一的变换
- 参数非线性模型化为线性模型
  - 估计线性模型参数，利用参数之间的关系计算原非线性模型参数

5.2 非线性模型估计中的优化计算方法

- 算法问题，已经进入应用软件程序中，一般不介绍。

5.3 非线性回归模型——一般估计方法

- 非线性回归模型的表达
  - 参数数量并不一定等于解释变量数量，这是非线性回归模型的重要特点。
  - 期望函数一般难以给定，但可以根据经济理论、行为规律、样本数据散点图给出，经估计后再检验。
- 非线性回归模型的**NLS**估计
  - **5**种选择初始值的思路
  - 如何通过参数变换实现参数约束：正数约束、区间约束、次序约束。

– 如何评估模型

A、如果参数估计量经济意义不合理时，从初始值方面检查；

B、如果参数估计量经济意义不合理时，从迭代过程是否平稳方面检查；

C、如果参数估计量经济意义不合理时，从期望函数是否合理方面检查；

D、显著性检验

E、共线性检验

F、异方差检验

G、残差的平稳性检验

H、残差的正态性检验

- 非线性回归模型的非线性**WLS**估计
  - 加权的目的：近期的观察比远期的观察更重要。
  - 权矩阵是一个对角元非负的对角**n**阶方阵。
- 非线性回归模型的非线性**ML**估计
  - 线性回归模型的**ML**估计中**J=1**，而在这里一般**J≠1**；
  - 线性回归模型的**OLS**估计与**ML**估计相等，而这里一般不相等，特殊情况下相等。
- 非线性回归模型的工具变量估计

5.4 非线性回归模型——统计推断

- 非线性回归模型的数学理论

5.5 非线性回归模型——专门问题

- 因变量的参数变换
  - 一般描述
  - **Box-Cox**变换
- 异方差性的非线性方法
- 序列相关性的非线性方法
- 条件异方差性的非线性方法

5.6 广义指数分布模型

- 广义指数分布模型
  - 广义指数分布模型的描述
  - 广义指数分布模型多峰的识别
  - 广义指数分布非线性模型的估计
  - 广义指数分布非线性模型的应用

5.7 非线性联立方程模型

- 特殊非线性联立方程模型——每个方程中都没有内生解释变量
  - GLS估计
  - ML估计
  - GMM估计
- 一般非线性联立方程模型
  - 非线性2SLS
  - 非线性3SLS
  - 非线性FIML

5.8 非均衡计量经济学模型

- 非均衡是一种经济现实。
- 描述非均衡的方程是非线性的。
- 更多是与其它方程共同构成联立方程模型。