§2.1 简单的非线性单方程计量经济学模型

一、非线性单方程计量经济学模型概述

二、非线性普通最小二乘估计

三、例题及讨论

四、非线性单方程模型的最大似然估计

说明

- 非线性计量经济学模型在计量经济学模型中占据重要的位置 ； 已经形成内容广泛的体系，包括变量非线性模型、参数非线性模型、随机误差项违背基本假设的非线性问题等;
- 非线性模型理论与方法已经形成了一个与线性模型相对应的体系，包括从最小二乘原理出发的一整套方法和从最大或然原理出发的一整套方法。
- 本节主要涉及最基础的、具有广泛应用价值的非线性单方程模型的最小二乘估计。

一、非线性单方程计量经济学模型概述

1. 解释变量非线性问题

- 现实经济现象中变量之间往往呈现非线性关系

  需求量与价格之间的关系

  成本与产量的关系

  税收与税率的关系

  基尼系数与经济发展水平的关系

- 通过变量置换就可以化为线性模型

2. 可以化为线性的包含参数非线性的问题

- 函数变换

$$\boxed{Q = AK^{\alpha}L^{\beta}}$$

$$\ln Q = \ln A + \alpha \ln K + \beta \ln L + \ln \mu$$

- 级数展开

$$\boxed{Q = A(\delta_1 K^{-\rho} + \delta_2 L^{-\rho})^{-\frac{1}{\rho}}}$$

$$\ln Q = \ln A - \tfrac{1}{\rho}\ln(\delta_1 K^{-\rho} + \delta_2 L^{-\rho}) + \ln \mu$$

$$\ln Q \approx \ln A + \delta_1 \ln K + \delta_2 \ln L - \frac{1}{2}\rho\delta_1\delta_2(\ln(\frac{K}{L}))^2 + \ln \mu$$

3.不可以化为线性的包含参数非线性的问题

$$Q = AK^{\alpha}L^{\beta} + \mu$$

$$Q = A(\delta_1 K^{-\rho} + \delta_2 L^{-\rho})^{-\frac{1}{\rho}} + \mu$$

- 与上页的方程比较，哪种形式更合理？
- 直接作为非线性模型更合理。

二、非线性普通最小二乘法

1. 普通最小二乘原理

$$\boxed{y_i = f(x_i, \beta) + \mu_i}$$

残差平方和

$$\boxed{S(\hat{\beta}) = \sum_{i=1}^{n}(y_i - f(x_i, \hat{\beta}))^2}$$

取极小值的一阶条件

$$\boxed{\frac{dS}{d\hat{\beta}} = -2\sum_{i=1}^{n}(y_i - f(x_i, \hat{\beta}))(\frac{-df(x_i, \hat{\beta})}{d\hat{\beta}}) = 0}$$

$$\boxed{\sum_{i=1}^{n}(y_i - f(x_i, \hat{\beta}))(\frac{df(x_i, \hat{\beta})}{d\hat{\beta}}) = 0}$$

如何求解非线性方程？

2. 高斯－牛顿(Gauss-Newton)迭代法

- 高斯—牛顿迭代法的原理

对原始模型展开台劳级数，取一阶近似值

$$\boxed{f(x_i,\hat{\beta}) \approx f(x_i,\hat{\beta}_{(0)}) + \left.\frac{df(x_i,\hat{\beta})}{d\hat{\beta}}\right|_{\hat{\beta}_{(0)}}(\hat{\beta}-\hat{\beta}_{(0)})}$$

$$z_i(\hat{\beta}) = \frac{df(x_i,\hat{\beta})}{d\hat{\beta}}$$

$$S(\hat{\beta}) = \sum_{i=1}^{n}(y_i - f(x_i,\hat{\beta}_{(0)}) - z_i(\hat{\beta}_{(0)})(\hat{\beta}-\hat{\beta}_{(0)}))^2$$

$$= \sum_{i=1}^{n}(y_i - f(x_i,\hat{\beta}_{(0)}) + z_i(\hat{\beta}_{(0)})\hat{\beta}_{(0)} - z_i(\hat{\beta}_{(0)})\hat{\beta})^2$$

$$= \sum_{i=1}^{n}(\tilde{y}_i(\hat{\beta}_{(0)}) - z_i(\hat{\beta}_{(0)})\hat{\beta})^2$$

构造并估计线性伪模型

$$\boxed{\tilde{y}_i(\hat{\beta}_{(0)}) = z_i(\hat{\beta}_{(0)})\beta + \varepsilon_i}$$

构造线性模型

$$\boxed{S(\hat{\beta}_{(1)}) = \sum_{i=1}^{n}(\tilde{y}_i(\hat{\beta}_{(0)}) - z_i(\hat{\beta}_{(0)})\hat{\beta}_{(1)})^2}$$

估计得到参数的第**1**次迭代值 $\hat{\beta}_{(1)}$

迭代

- # 高斯－牛顿迭代法的步骤

第一步：给出参数估计值$\hat{\beta}$的初值$\hat{\beta}_{(0)}$，将$f(x_i,\hat{\beta})$在$\hat{\beta}_{(0)}$处展开台劳级数，

取一阶近似值；

第二步：计算 $z_i = \left.\dfrac{df(x_i,\hat{\beta})}{d\hat{\beta}}\right|_{\hat{\beta}_{(0)}}$ 和 $\tilde{y}_i = y_i - f(x_i,\hat{\beta}_{(0)}) + z_i \cdot \hat{\beta}_{(0)}$的样本观测值；

第三步：采用普通最小二乘法估计模型 $\tilde{y}_i = z_i\beta + \varepsilon_i$，得到$\beta$的估计值$\hat{\beta}_{(1)}$；

第四步：用$\hat{\beta}_{(1)}$代替第一步中的$\hat{\beta}_{(0)}$，重复这一过程，直至收敛。

3. 牛顿－拉夫森(Newton-Raphson)迭代法

- 自学，掌握以下2个要点
- 牛顿－拉夫森迭代法的原理
  - 对残差平方和展开台劳级数，取二阶近似值；
  - 对残差平方和的近似值求极值；
  - 迭代。
- 与高斯－牛顿迭代法的区别
  - 直接对残差平方和展开台劳级数，而不是对其中的原模型展开；
  - 取二阶近似值，而不是取一阶近似值。

4. 应用中的一个困难

- 如何保证迭代所逼近的是总体极小值（即最小值）而不是局部极小值?
- 一是模拟试验：随机产生初始值→估计→改变初始值→再估计→反复试验，设定收敛标准（例如**100**次连续估计结果相同）→直到收敛。
- 一是利用检验统计量进行检验。

Gan Li, A Test for Global Maximum, Journal of American Statistical Association, 94(447),Sep 1999

5. 非线性普通最小二乘法在软件中的实现

- 给定初值
- 写出模型
- 估计模型
- 改变初值
- 反复估计

三、例题与讨论

例: 农民收入影响因素分析模型

- **分析与建模：**经过反复模拟，剔除从直观上看可能对农民收入产生影响但实际上并不显著的变量后，得到如下结论：改革开放以来，影响我国农民收入总量水平的主要因素是从事非农产业的农村劳动者人数、农副产品收购价格和农业生产的发展规模。用I表示农民纯收入总量水平、Q表示农业生产的发展规模、P表示农副产品收购价格、L表示从事非农产业的农村劳动者人数。收入采用当年价格；农业生产的发展规模以按可比价格计算的、包括种植业、林业、牧业、副业和渔业的农业总产值指数为样本数据；农副产品收购价格以价格指数为样本数据。

$$\boxed{I = AQ^{\alpha_1} P^{\alpha_2} L^{\alpha_3}}$$

- 农民收入及相关变量数据

年份	I（10亿元）	Q (1978=100)	P (1978=100)	L（100万人）
1978	62.45	100.0	100.0	31.52
1979	79.30	107.5	122.1	31.90
1980	96.50	109.0	130.8	35.02
1981	107.65	115.3	138.5	36.92
1982	120.80	128.4	141.5	38.05
1983	142.40	138.4	147.8	43.40
1984	185.85	155.4	153.7	58.88
1985	238.70	160.7	166.9	67.13
1986	285.52	166.1	177.6	75.22
1987	343.80	175.8	198.9	81.30
1988	442.60	182.6	244.6	86.11
1989	495.30	188.3	281.3	84.98
1990	524.66	202.6	274.0	86.74
1991	559.30	210.1	268.4	89.06
1992	613.66	223.5	277.5	97.65
1993	743.49	241.0	314.7	109.98
1994	979.39	261.7	440.3	119.64
1995	1271.16	290.2	527.9	127.07
1996	1567.33	317.5	550.1	130.28
1997	1721.71	333.7	525.3	135.27

讨论：NLS的初值及影响

- 由于农副产品收购价格和非农产业劳动者人数与农业生产规模指数严重共线性，以农民收入为被解释变量，农业生产规模指数为解释变量，**1978~1997**年数据为样本。

$$\boxed{i = Aq^{\alpha}}$$

线性化估计

收入年均增长19.1%，产值年均增长6.5%,该参数估计结果基本合理。

CPI

为什么如此之高？

人口

能否将它解释为“产值的收入弹性？”

Dependent Variable: LNI
Method: Least Squares
Date: 09/22/05 Time: 16:46
Sample(adjusted): 1980 1997
Included observations: 18 after adjusti[illegible]ndpoints
Convergence achieved after 6 iteratio[illegible]

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-8.057459	0.793770	-10.15087	0.0000
LNQ	**2.678082**	0.150524	17.79178	0.0000
AR(1)	1.013730	0.238470	[illegible]	[illegible]
AR(2)	-0.454939	0.239082	-1.902862	0.0778

R-squared	0.993104	Mean dependent var	5.997589
Adjusted R-squared	0.991627	S.D. dependent var	0.913254
S.E. of regression	0.083568	Akaike info criterion	-1.933171
Sum squared resid	0.097772	Schwarz criterion	-1.735311
Log likelihood	21.39854	F-statistic	672.0798
Durbin-Watson stat	1.878598	Prob(F-statistic)	0.000000

非线性估计（初值：1、5）

Dependent Variable: I
Method: Least Squares
Date: 09/22/05 Time: 16:55
Sample: 1978 1997
Included observations: 20
Convergence achieved after 5 iterations
I=C(1)*Q^C(2)

迭代收敛很快

与线性估计结果偏离大，经济意义不合理

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	7.26E-10	3.56E-09	0.204079	0.8406
C(2)	**4.945417**	0.000241	20518.14	0.0000

R-squared	0.735874	Mean dependent var	529.0785
Adjusted R-squared	0.721201	S.D. dependent var	[illegible]
S.E. of regression	262.9836	Akaike info criterion	14.07670
Sum squared resid	1244887.	Schwarz criterion	14.17627
Log likelihood	-138.7670	Durbin-Watson stat	0.193909

拟合效果较差

非线性估计（初值：0.001 、2）

Dependent Variable: I
Method: Least Squares
Date: 09/22/05   Time: 16:58
Sample: 1978 1997
Included observations: 20
Convergence achieved after 30 iterations
I=C(1)*Q^C(2)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.000691	0.000228	3.024843	0.0073
C(2)	**2.539090**	0.058477	43.42054	0.0000

R-squared	0.994203	Mean dependent var	529.0785
Adjusted R-squared	0.993880	S.D. dependent var	498.0612
S.E. of regression	38.96202	Akaike info criterion	10.25769
Sum squared resid	27324.70	Schwarz criterion	10.35726
Log likelihood	-100.5769	Durbin-Watson stat	0.689064

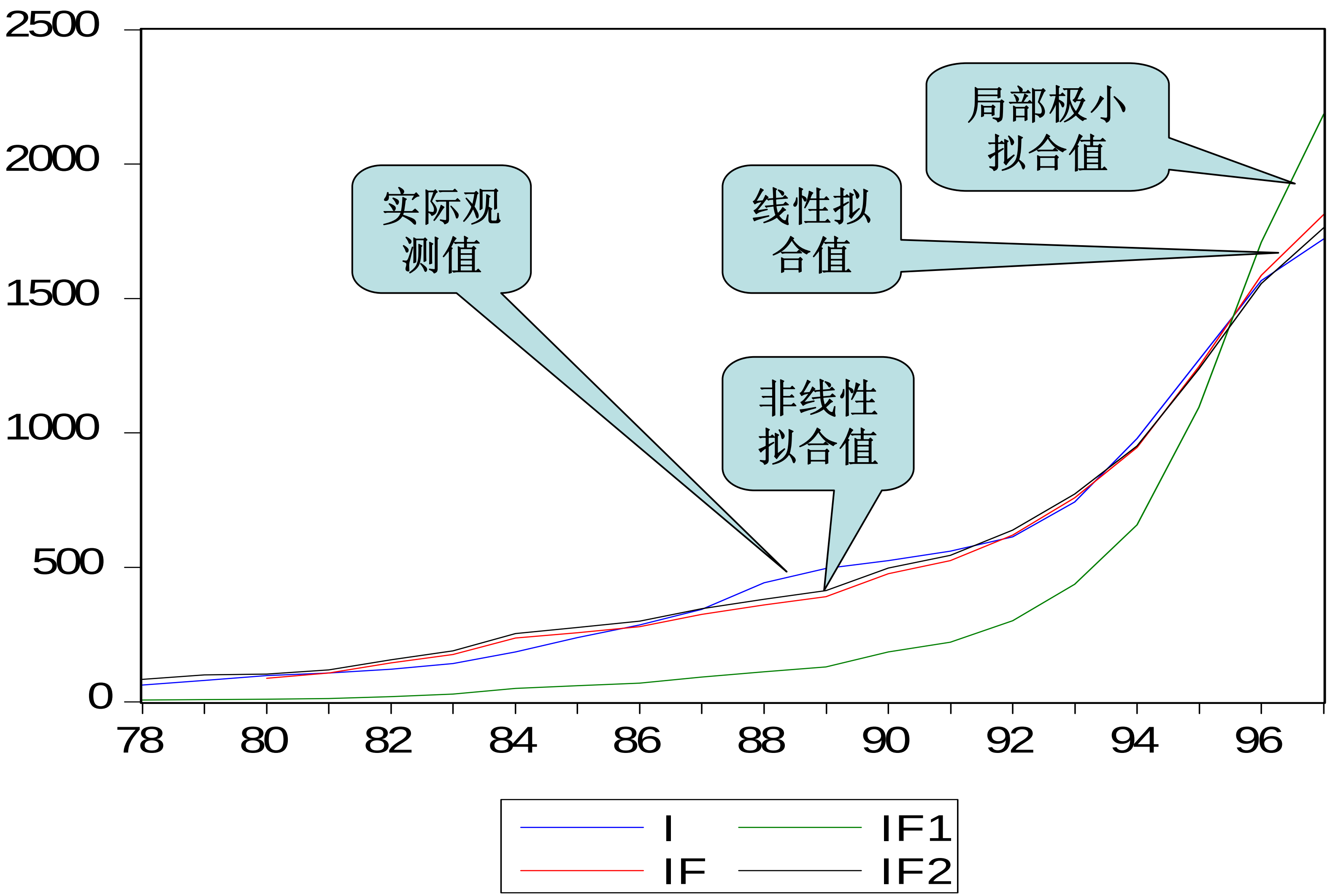
非线性估计（初值：0.1 、1）

Dependent Variable: I
Method: Least Squares
Date: 09/22/05 Time: 17:02
Sample: 1978 1997
Included observations: 20
Convergence achieved after 80 iterations
I=C(1)*Q^C(2)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.000691	0.000228	3.024999	0.0073
C(2)	**2.539090**	0.058478	43.41992	0.0000

R-squared	0.994203	Mean dependent var	529.0785
Adjusted R-squared	0.993880	S.D. dependent var	498.0612
S.E. of regression	38.96202	Akaike info criterion	10.25769
Sum squared resid	27324.70	Schwarz criterion	10.35726
Log likelihood	-100.5769	Durbin-Watson stat	0.689064

拟合结果


2500
2000
1500
1000
500
0
78
80
82
84
86
88
90
92
94
96
实际观测值
线性拟合值
局部极小拟合值
非线性拟合值
I
IF
IF1
IF2

讨论

- 一般情况下，线性化估计和非线性估计结果差异不大。如果差异较大，在确认非线性估计结果为总体最小时，应该怀疑和检验线性模型。
- 非线性估计确实存在局部极小问题。
- 根据参数的经济意义和数值范围选取迭代初值。
- NLS估计的异方差和序列相关问题。
  - NLS不能直接处理。
  - 应用最大似然估计。

一种不可取的方法：改善拟合和D.W.值

E!为NLS估计的残差。不是理论意义上的广义差分法。

Dependent Variable: I
Method: Least Squares
Date: 09/22/05   Time: 20:25
Sample(adjusted): 1979 1997
Included observations: 19 after adjusting endpoints
Convergence achieved after 6 iterations
I=C(1)*Q^C(2)+C(3)*E1(-1)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.000860	0.000228	3.776467	0.0017
C(2)	2.499262	0.046937	53.24660	0.0000
C(3)	0.715660	0.198603	3.603461	0.0024

R-squared	0.996690	Mean dependent var	553.6379
Adjusted R-squared	0.996276	S.D. dependent var	499.1121
S.E. of regression	30.45702	Akaike info criterion	9.814449
Sum squared resid	14842.08	Schwarz criterion	9.963571
Log likelihood	-90.23727	Durbin-Watson stat	1.342729

四、非线性单方程模型的最大似然估计

1. 经典线性单方程模型的最大似然估计

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \mu_i$$ **i=1,2,…,n**

$$Y_i \sim N(\mathbf{X}_i \widehat{\boldsymbol{\beta}}, \sigma^2) \leftarrow \mu_i \sim N(0, \sigma^2)$$

$$L(\hat{\boldsymbol{\beta}}, \sigma^2) = P(Y_1, Y_2, \cdots, Y_n)$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} e^{-\frac{1}{2\sigma^2}\Sigma(Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki}))^2}$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} e^{-\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}$$

$$
\begin{aligned}
Max\, L^* &= Ln(L) \\
&= -nLn(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})
\end{aligned}
$$

$$
Min\, (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})
$$

$$
\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}
$$

参数估计结果与参数的OLS估计相同

2. 简单非线性单方程模型的最大似然估计

$$y_i = f(x_i, \beta) + \mu_i$$

i=1,2,…,n

$$Y_i \sim N(f(\mathbf{X_i}, \boldsymbol{\beta}), \sigma^2)$$

$$\mu_i \sim N(0, \sigma^2)$$

$$L(\hat{\beta}, \sigma^2) = P(Y_1, Y_2, \cdots, Y_n)$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} e^{-\frac{1}{2\sigma^2}\Sigma(Y_i - f(X_i, \hat{\beta}))^2}$$

$$\boxed{\begin{aligned} MaxL^* &= Ln(L) \\ &= -nLn(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\sum(Y_i - f(X_i, \hat{\beta}))^2 \end{aligned}}$$

$$\boxed{Min\sum(Y_i - f(X_i, \hat{\beta}))^2}$$

- 面临NLS同样的过程，得到相同的估计结果。