

第4章 非线性回归模型的线性化

- (1) 多项式函数模型
- (2) 双曲线函数模型
- (3) 对数函数模型
- (4) 生长曲线 (**logistic**) 模型
(比教材中的模型复杂些)
- (5) 指数函数模型
- (6) 幂函数模型
- (7) 不可线性化的非线性回归模型估计方法
(不要求掌握)

第4章 非线性回归模型的线性化

有时候变量之间的关系是非线性的。虽然其形式是非线性的，但可以通过适当的变换，转化为线性模型，然后利用线性回归模型的估计与检验方法进行处理。称此类模型为可线性化的非线性模型。

以下非线性回归模型是无法用最小二乘法估计参数的。可采用非线性方法进行估计。估计过程非常复杂和困难，计算机的出现大大方便了非线性回归模型的估计。专用软件使这种计算变得非常容易。但本章不是介绍这类模型的估计。

$$y_t = \alpha_0 + \alpha_1 x_t^{\beta_1} + u_t$$

$$y_t = \alpha_0 e^{\alpha_1 x_t} + u_t$$

下面介绍几种典型的可以做线性化处理的非线性模型。

(1)多项式函数模型(1)

一种多项式方程的表达形式是

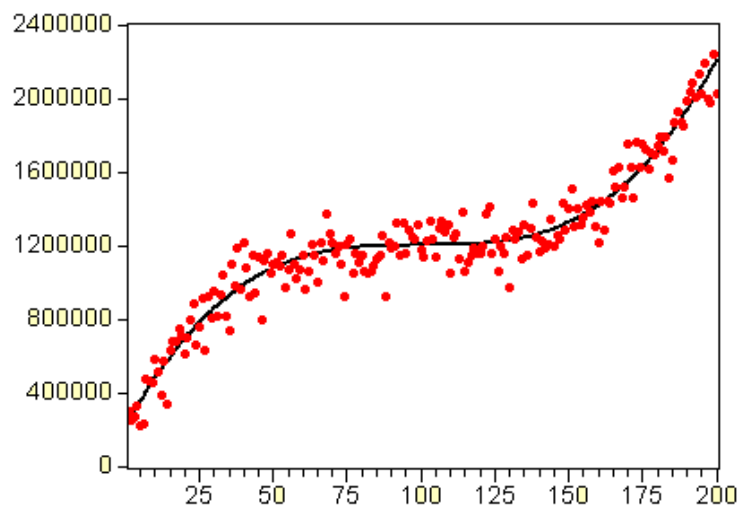
$$y_t = b_0 + b_1 x_t + b_2 x_t^2 + b_3 x_t^3 + u_t$$

令 $x_{t1} = x_t$, $x_{t2} = x_t^2$, $x_{t3} = x_t^3$, 上式变为

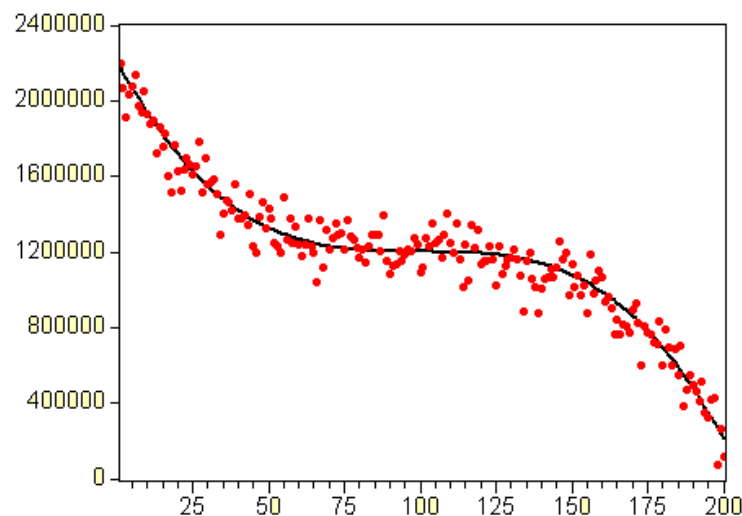
$$y_t = b_0 + b_1 x_{t1} + b_2 x_{t2} + b_3 x_{t3} + u_t$$

这是一个三元线性回归模型。如经济学中的

总成本与产品产量曲线与左图相似。



$(b_1 > 0, b_2 > 0, b_3 > 0)$



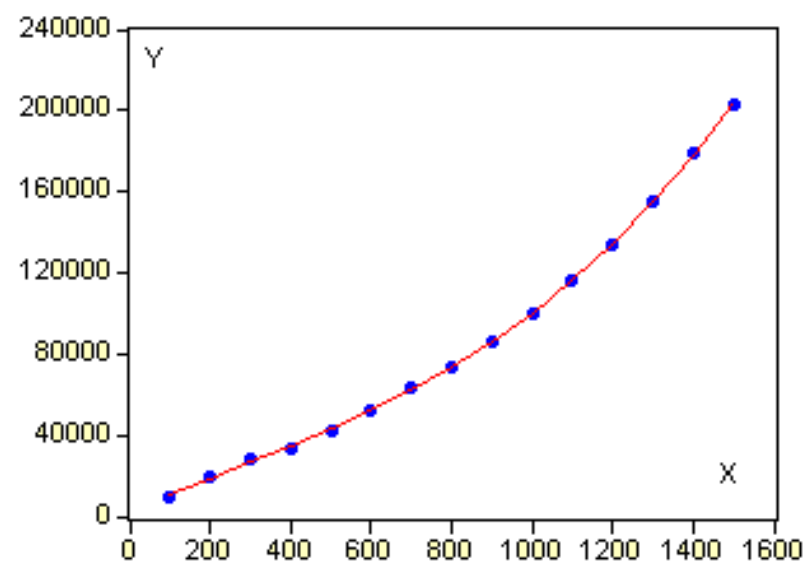
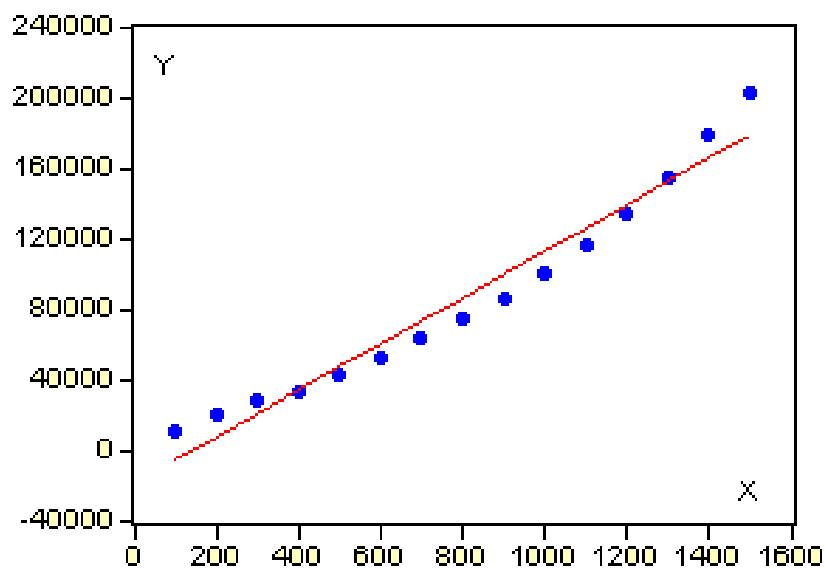
$(b_1 < 0, b_2 > 0, b_3 < 0)$

(第2版教材第111页)

(第3版教材第90页)

(1)多项式函数模型(1)

例4.1：总成本与产品产量的关系（课本91页）



$$y_t = b_0 + b_1 x_t + b_2 x_t^2 + b_3 x_t^3 + u_t$$

（第2版教材第112页）

（第3版教材第91页）

(1)多项式函数模型(1)

例4.1：总成本与产品产量的关系（课本91页）

Dependent Variable: Y
Method: Least Squares
Date: 09/20/07 Time: 17:51
Sample: 1 15
Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2434.652	1368.921	1.778519	0.1029
X	85.70278	7.170616	11.95194	0.0000
X^2	-0.028405	0.010242	-2.773303	0.0181
X^3	4.05E-05	4.22E-06	9.593420	0.0000
R-squared	0.999778	Mean dependent var	86353.33	
Adjusted R-squared	0.999717	S.D. dependent var	60016.44	
S.E. of regression	1009.303	Akaike info criterion	16.89509	
Sum squared resid	11205609	Schwarz criterion	17.08390	
Log likelihood	-122.7131	F-statistic	16497.11	
Durbin-Watson stat	2.275841	Prob(F-statistic)	0.000000	

$$\hat{C}_t = 2434.7 + 85.7 x_t - 0.028 x_t^2 + 0.00004 x_t^3$$

$$(1.8) \quad (12.0) \quad (-2.8) \quad (9.6)$$

$$R^2 = 0.9998, N = 15$$

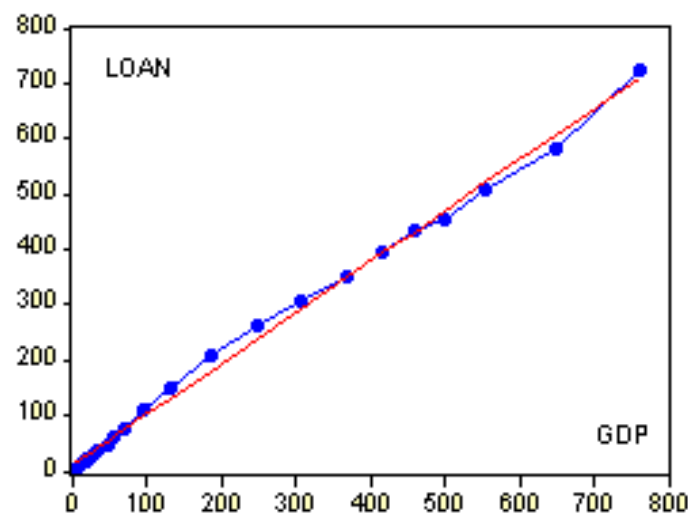
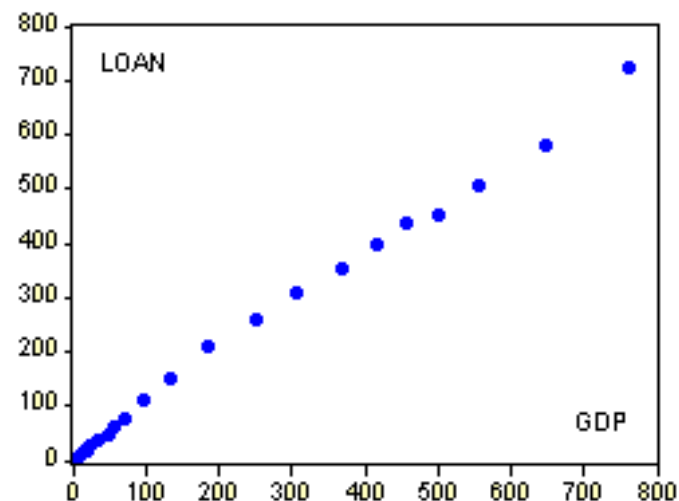
（第2版教材第114页）

（第3版教材第92页）

案例1：厦门市贷款总额与GDP的关系分析

(1990~2003, file:bank08)

obs	LOAN	GDP
1990	63.70000	57.10000
1991	78.00000	72.00000
1992	112.7000	97.70000
1993	151.8000	132.3000
1994	209.6000	187.0000
1995	260.8000	250.6000
1996	306.8000	306.4000
1997	352.3000	370.3000
1998	397.3000	418.1000
1999	435.3000	458.3000
2000	488.3000	501.2000
2001	552.0000	556.0000
2002	646.0000	648.0000
2003	898.0000	760.0000



案例1：厦门市贷款总额与GDP的关系分析

(1990~2003, file:bank08)

从散点图看，用多项式方程拟合比较合理。

$$Loan_t = \beta_0 + \beta_1 GDP_t + \beta_2 GDP_t^2 + \beta_3 GDP_t^3 + u_t$$

$$\hat{loan}_t = -24.5932 + 1.6354 GDP_t - 0.0026 GDP_t^2 + 0.0000027 GDP_t^3$$

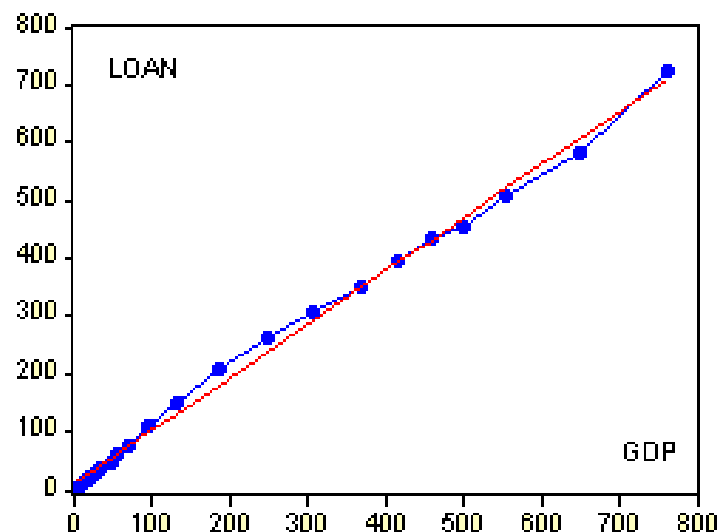
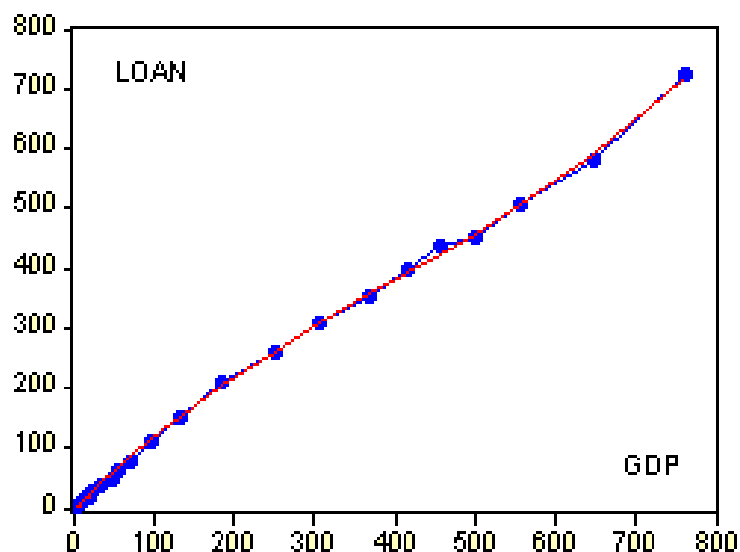
(-2.0)

(11.3)

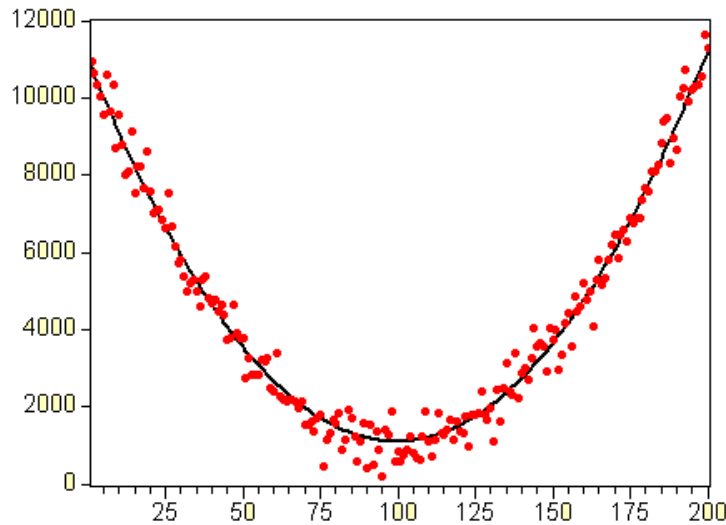
(-6.3)

(7.9)

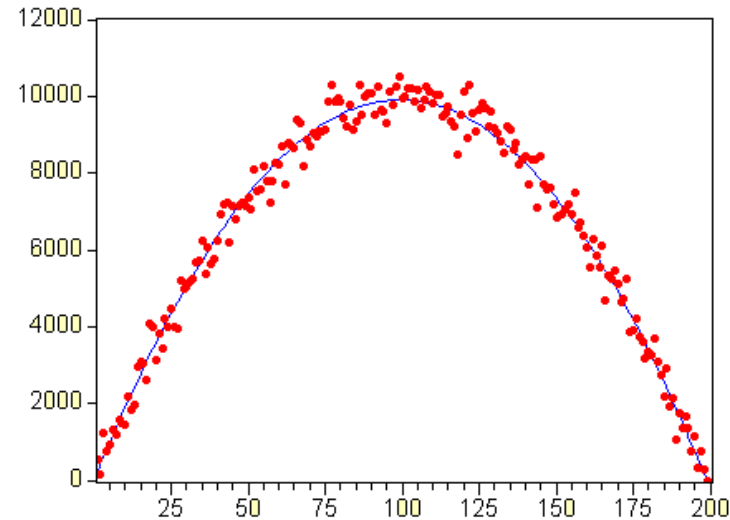
$R^2=0.9986, DW=2.6$



(1)多项式方程模型(2)



$$(b_1 > 0, b_2 > 0)$$



$$(b_1 < 0, b_2 < 0)$$

另一种多项式方程的表达形式是

$$y_t = b_0 + b_1 x_t + b_2 x_t^2 + u_t$$

令 $x_{t1} = x_t$, $x_{t2} = x_t^2$, 上式线性化为,

$$y_t = b_0 + b_1 x_{t1} + b_2 x_{t2} + u_t$$

(第2版教材第114页)

(第3版教材第93页)

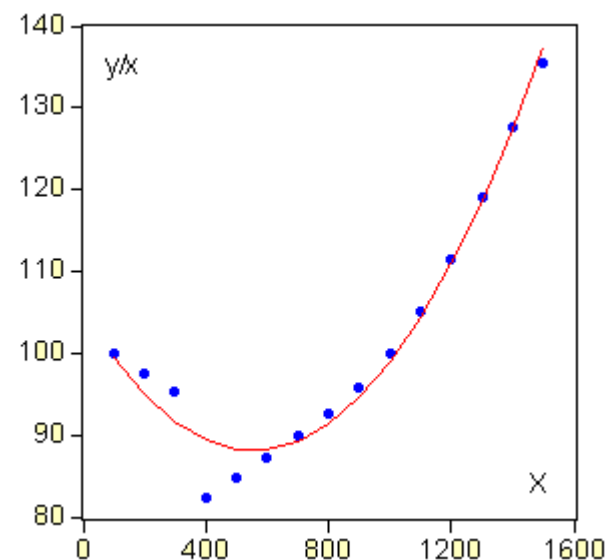
如经济学中的边际成本曲线、平均成本曲线与左图相似。

(1)多项式方程模型(2)

例4.1：平均成本与产品产量的关系（课本93页）

Dependent Variable: Y/X
Method: Least Squares
Date: 10/02/07 Time: 10:01
Sample: 1 15
Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	105.1552	2.475876	42.47191	0.0000
X	-0.061751	0.007121	-8.672084	0.0000
X^2	5.55E-05	4.33E-06	12.82378	0.0000
R-squared	0.972820	Mean dependent var	101.6315	
Adjusted R-squared	0.968290	S.D. dependent var	15.60920	
S.E. of regression	2.779592	Akaike info criterion	5.059342	
Sum squared resid	92.71358	Schwarz criterion	5.200952	
Log likelihood	-34.94506	F-statistic	214.7481	
Durbin-Watson stat	1.862198	Prob(F-statistic)	0.000000	



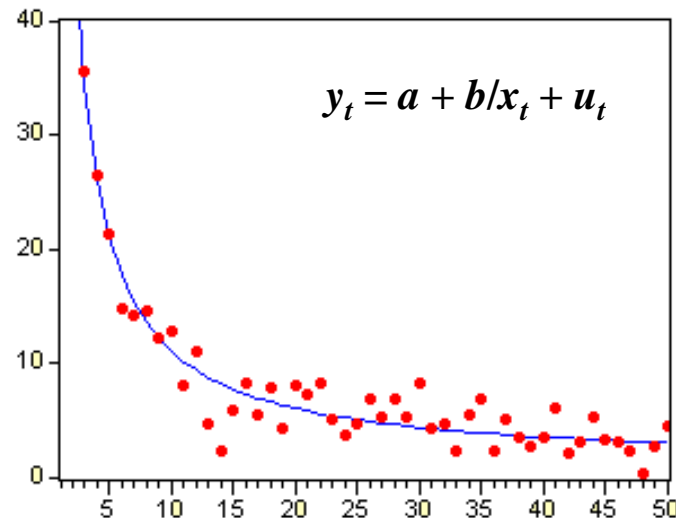
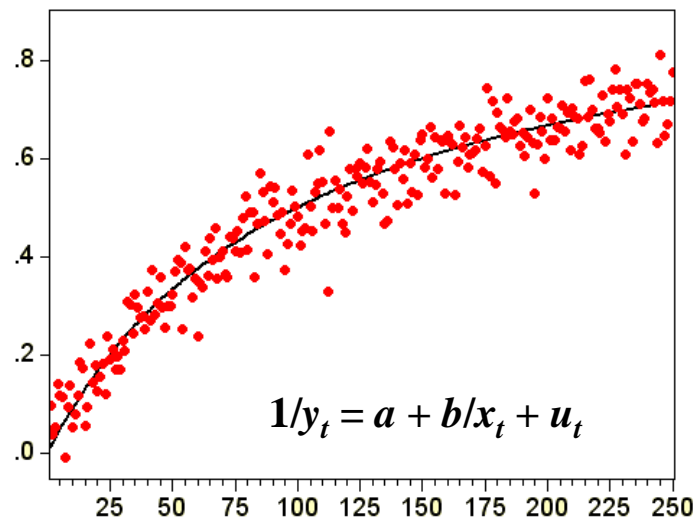
$$\frac{\hat{C}_t}{x_t} = 105.1 - 0.06 x_t + 0.00006 x_t^2$$

(42.5) (-8.7) (12.8) $R^2 = 0.97, N = 15$

（第2版教材第114页）
（第3版教材第93页）

(2) 双曲线函数模型

(第2版教材第114页)
(第3版教材第93页)



$$1/y_t = a + b/x_t + u_t \quad \text{或} \quad y_t = 1/(a + b/x_t + u_t)$$

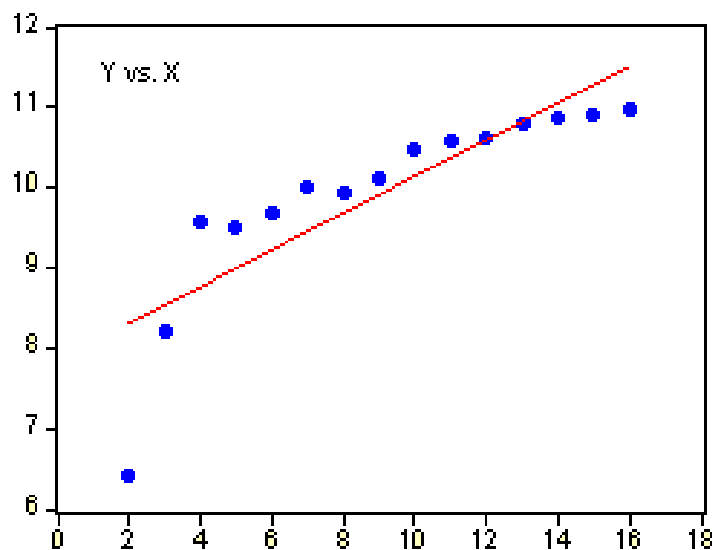
令 $y_t^* = 1/y_t$, $x_t^* = 1/x_t$, 得 $y_t^* = a + b x_t^* + u_t$

已变换为线性回归模型。双曲线函数还有另一种表达方式,

$$y_t = a + b/x_t + u_t$$

令 $x_t^* = 1/x_t$, 得 $y_t = a + b x_t^* + u_t$ 上式已变换成线性回归模型。

案例2：炼钢厂钢包容积Y与钢包使用次数X的关系（file:5nonli7）

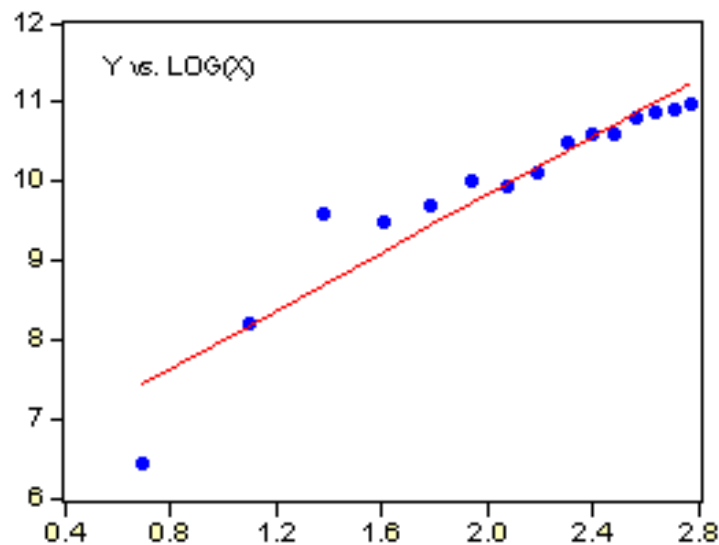


建立线性模型并估计

$$y = 7.85 + 0.27 x$$

(19.6) (5.7)

$$R^2 = 0.71, N = 15$$



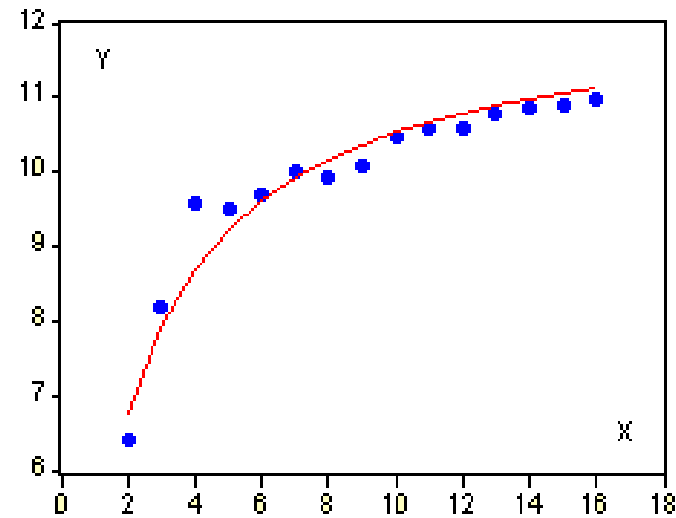
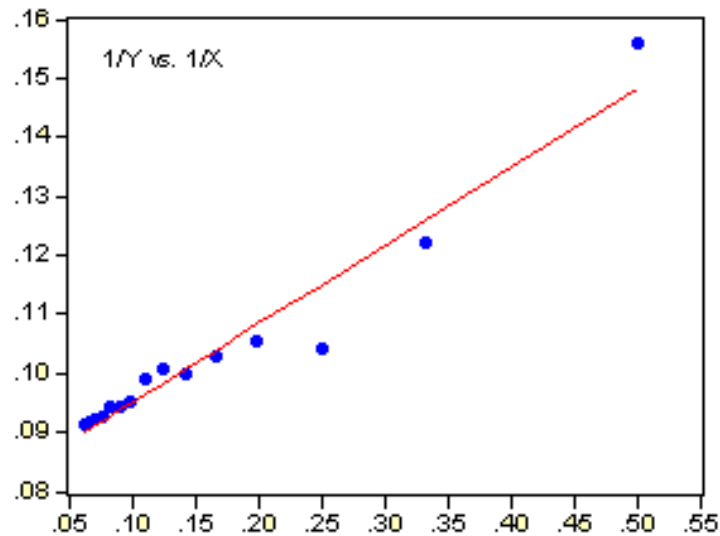
建立对数模型并估计

$$y = 6.16 + 1.83 \ln x$$

(16.0) (10.1)

$$R^2 = 0.89, N = 15$$

案例2：炼钢厂钢包容积Y与钢包使用次数X的关系（file:5nonli7）



双倒数线性化

双倒数模型

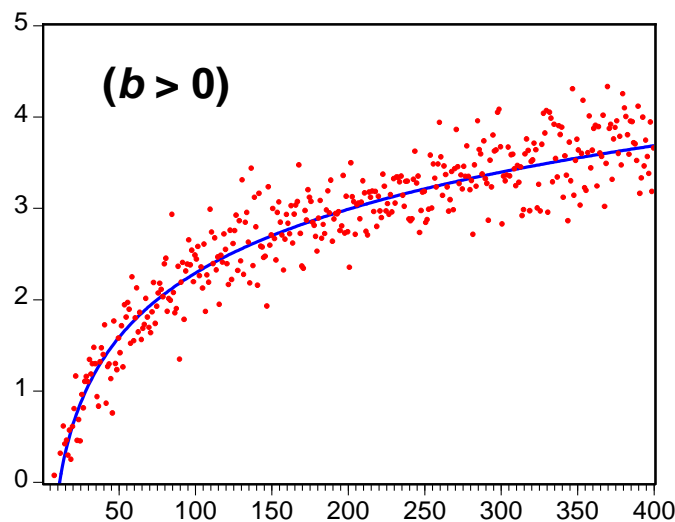
建立倒数模型并用1~14组数据估计，

$$1/y = 0.081 + 0.1339 (1/x)$$

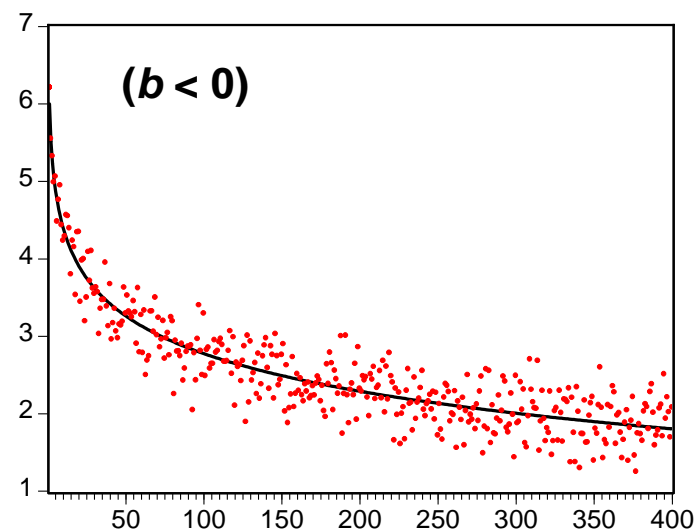
$$(42.1) \quad (14.1) \quad R^2 = 0.94, N = 15$$

倒数模型的估计结果最好。

(3) 对数函数模型



$$y_t = a + b \ln x_t + u_t, (b > 0)$$



$$y_t = a + b \ln x_t + u_t, (b < 0)$$

令 $x_t^* = \ln x_t$, 则

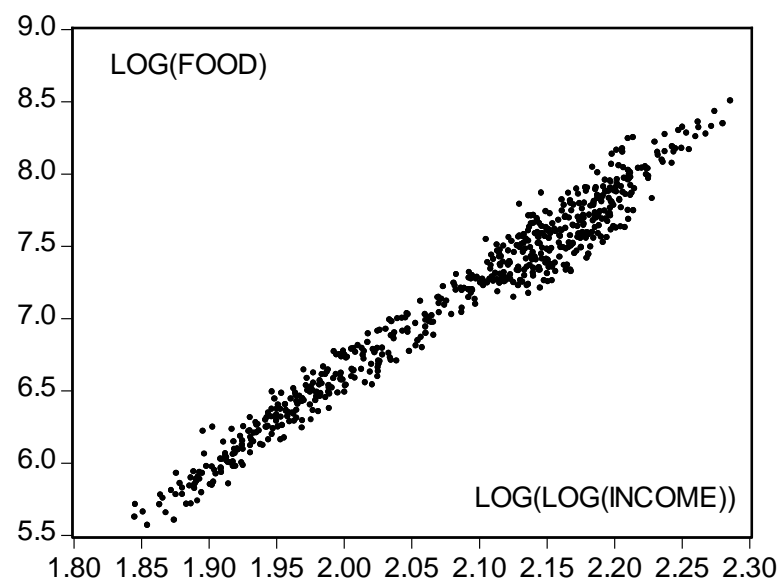
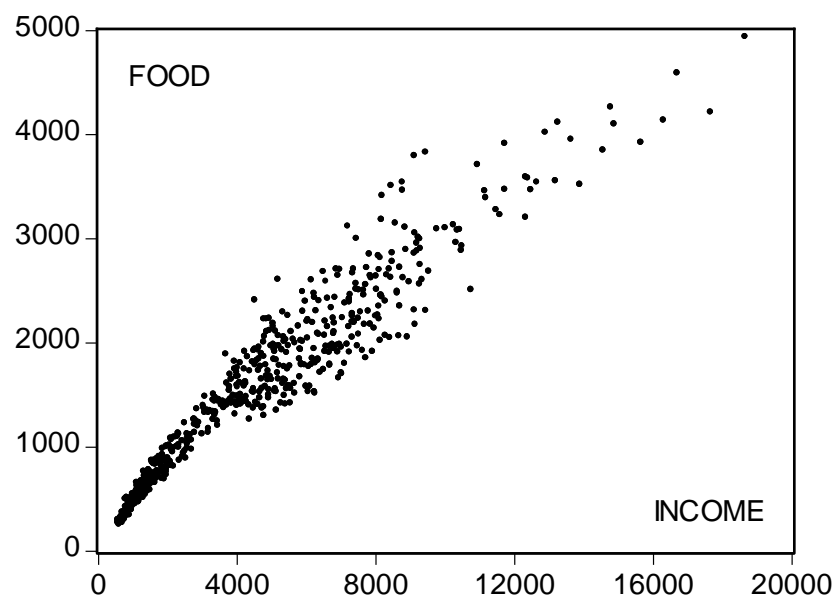
$$y_t = a + b x_t^* + u_t$$

变量 y_t 和 x_t^* 已变换成为线性关系。

(第2版教材第114页)

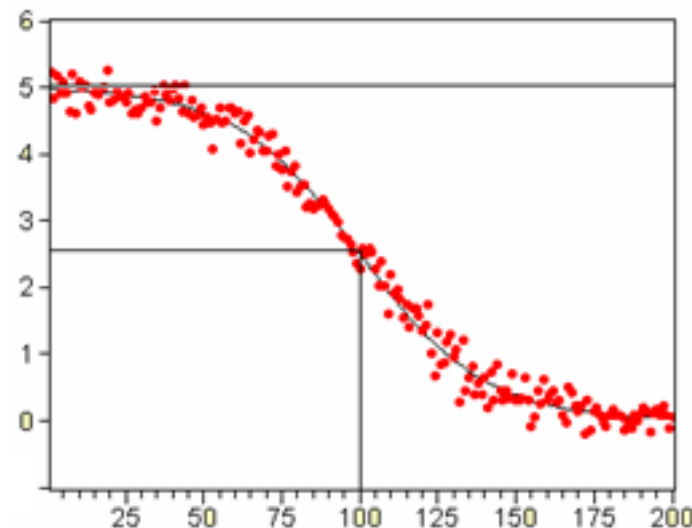
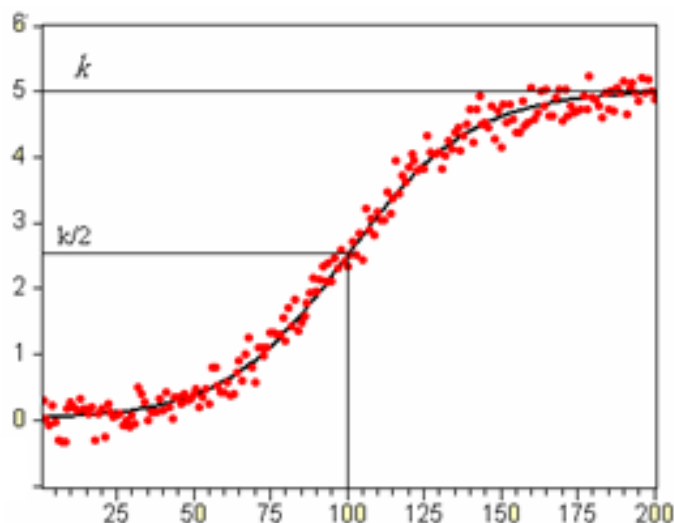
(第3版教材第93页)

案例：28个省市自治区1985~2005年城镇居民
人均食品支出（**food**）与人均收入（**income**）的关系



28个省市自治区1985~2005年城镇居民人均食品支出（**food**）与人均收入（**income**）的关系

(4) 生长曲线 (logistic) 模型（比教材中的模型复杂些）



$$y_t = \frac{k}{1 + e^{f(t)+u_t}} = \frac{k}{1 + e^{a-bt+u_t}}$$

（第2版教材第115页）

（第3版教材第94页）

美国人口统计学家**Pearl**和**Reed**广泛研究了有机体的生长，得到了上述数学模型。生长模型（或逻辑斯谛曲线，**Pearl-Reed**曲线）常用于描述有机体生长发育过程。其中 **k** 和 **0** 分别为 **y_t** 的上限和下限。

$$\lim_{t \rightarrow \infty} y_t = k, \quad \lim_{t \rightarrow -\infty} y_t = 0$$

a , **b** 为待估参数。曲线有拐点，曲线的上下两部分对称于拐点。

(4) 生长曲线 (logistic) 模型（与教材中的模型稍异）

为能运用最小二乘法估计参数 a, b ，必须事先估计出生曲线长上极限值 k 。
线性化过程如下。当上极限值 k 给定时，作如下变换，

$$\frac{k}{y_t} = 1 + e^{a-bt+u_t}$$

移项， $\frac{k}{y_t} - 1 = e^{a-bt+u_t}$ ，取自然对数， $Ln\left(\frac{k}{y_t} - 1\right) = a - bt + u_t$

令 $y_t^* = Ln\left(\frac{k}{y_t} - 1\right)$ ，则

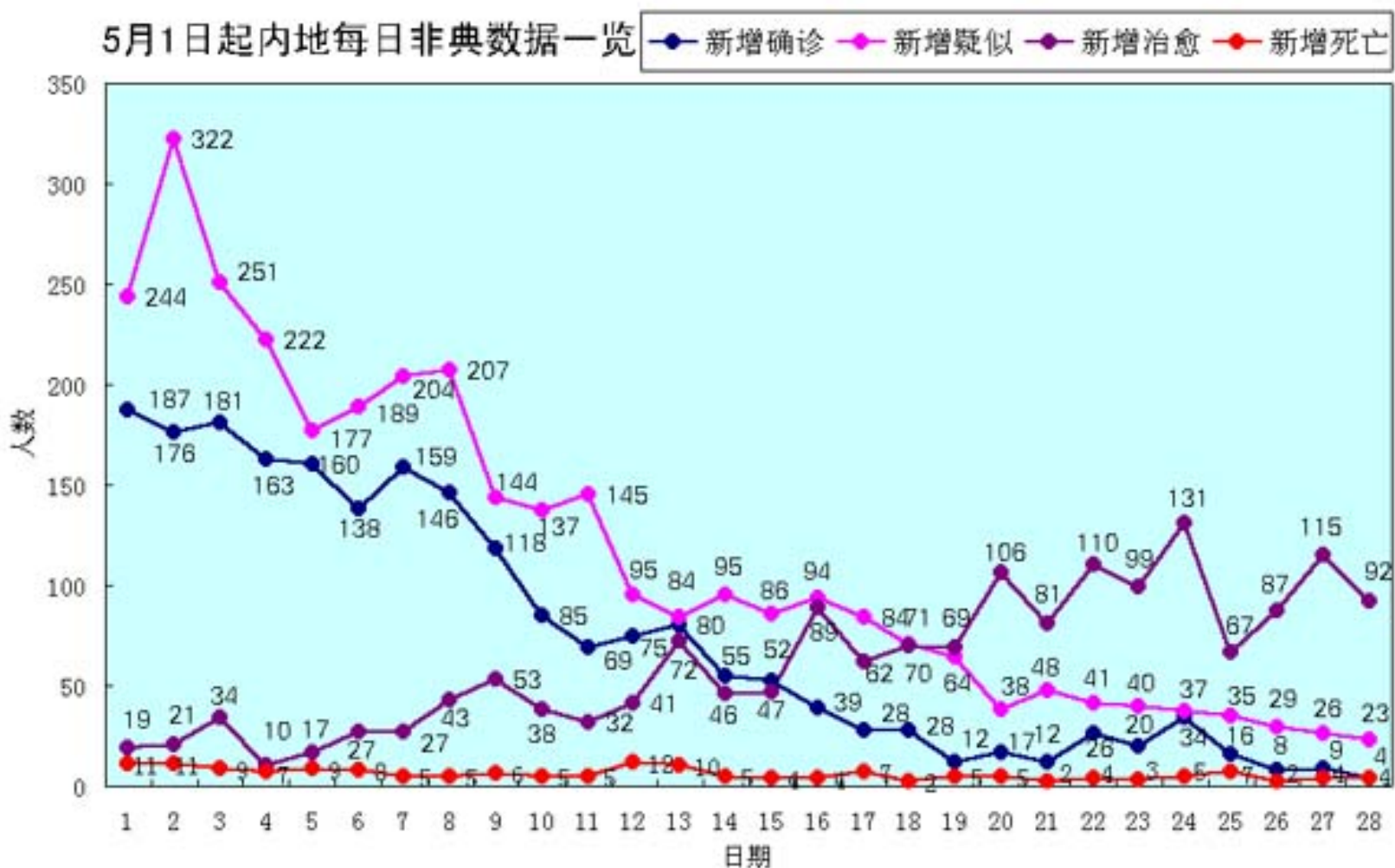
$$y_t^* = a - b t + u_t$$

（第2版教材第115页）
（第3版教材第94页）

此时可用最小二乘法估计 a 和 b 。

(4) 生长曲线 (logistic) 模型（与教材中的模型稍异）

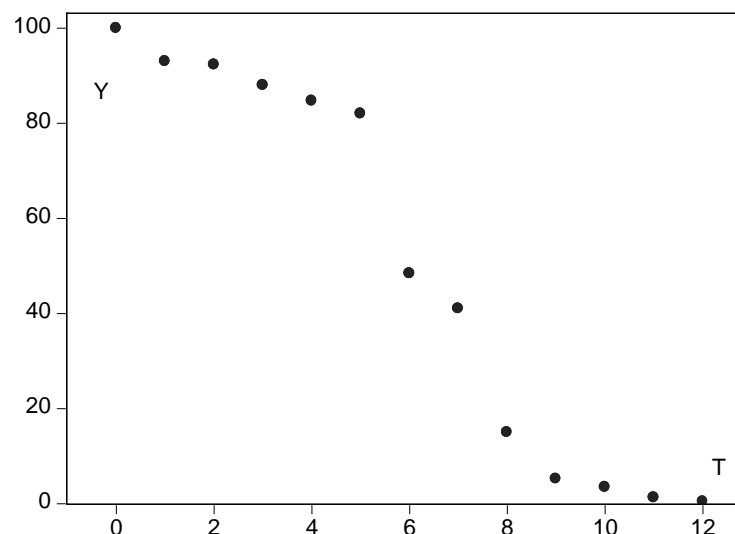
案例3：非典数据（2003-5-1~2003-5-28）



案例4：钉螺存活率曲线（file:nonli3）（生长曲线模型）

把一批钉螺埋入土中，以后每隔一个月取出部分钉螺，检测存活个数，计算存活率。数据见表。

y_t , 存活率 (%)	t , 土埋月数
100.0	0
93.0	1
92.3	2
88.0	3
84.	4
82.0	5
48.4	6
41.0	7
15.0	8
5.2	9
3.5	10
1.3	11
0.5	12



设定 y_t 的上渐近极限值 $k = 101$ （因为已有观测值 $y_t = 100$ ，所以令 $k = 101$ 更好些。），得估计结果如下：

案例4：钉螺存活率曲线（file:nonli3）（生长曲线模型）

Dependent Variable: LOG(101/Y-1)
 Method: Least Squares
 Date: 08/26/07 Time: 17:05
 Sample: 1 13
 Included observations: 13

	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.310784	0.292251	-14.75030	0.0000
T	0.765277	0.041330	18.51605	0.0000
R-squared	0.968913	Mean dependent var	0.280878	
Adjusted R-squared	0.966087	S.D. dependent var	3.027759	
S.E. of regression	0.557579	Akaike info criterion	1.810212	
Sum squared resid	3.419833	Schwarz criterion	1.897127	
Log likelihood	-9.766376	Hannan-Quinn criter.	1.792347	
F-statistic	342.8440	Durbin-Watson stat	1.356033	
Prob(F-statistic)	0.000000			

估计式是： $\log\left(\frac{101}{y_t} - 1\right) = -4.3108 + 0.7653 t$

(-14.8) (18.5) $R^2 = 0.97$

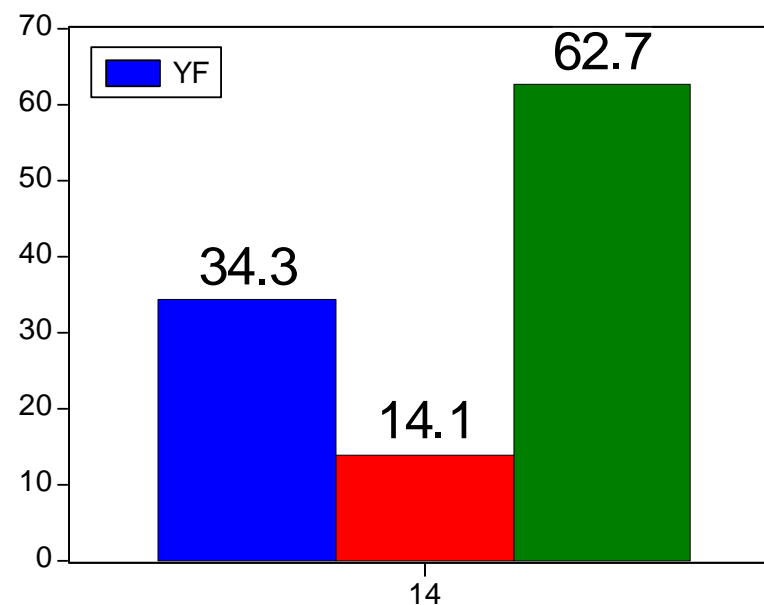
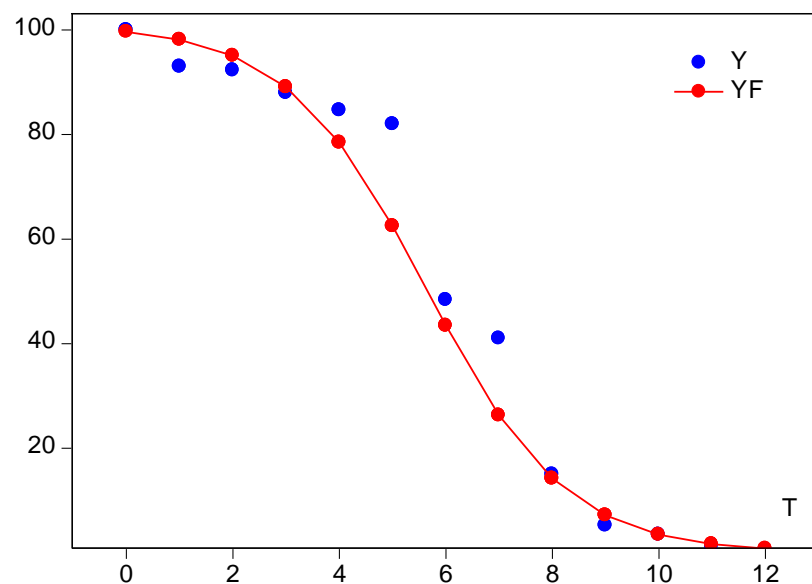
则逻辑函数的估计结果是 $\hat{y}_t = \frac{101}{1 + e^{-4.31 + 0.7653t}}$

案例4：钉螺存活率曲线（file:nonli3）（生长曲线模型）

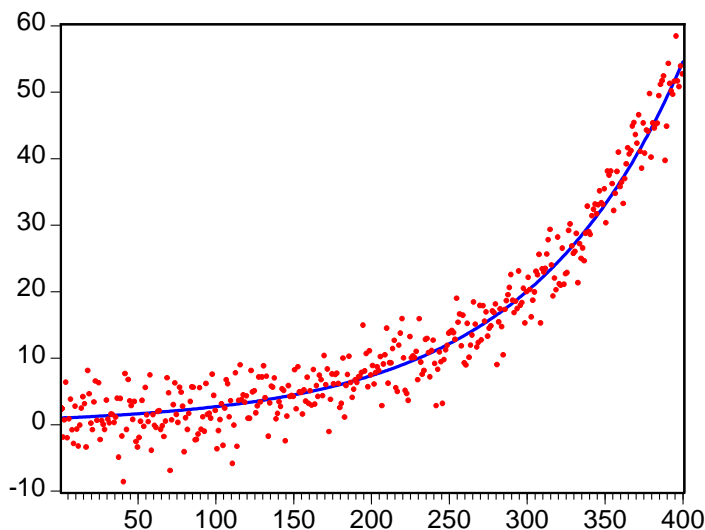
点预测：当 $t=6.5$ 月时，

$$\hat{y}_t = \frac{101}{1 + e^{-4.31 + 0.7653 \times 6.5}} = 34.3 \quad (\%)$$

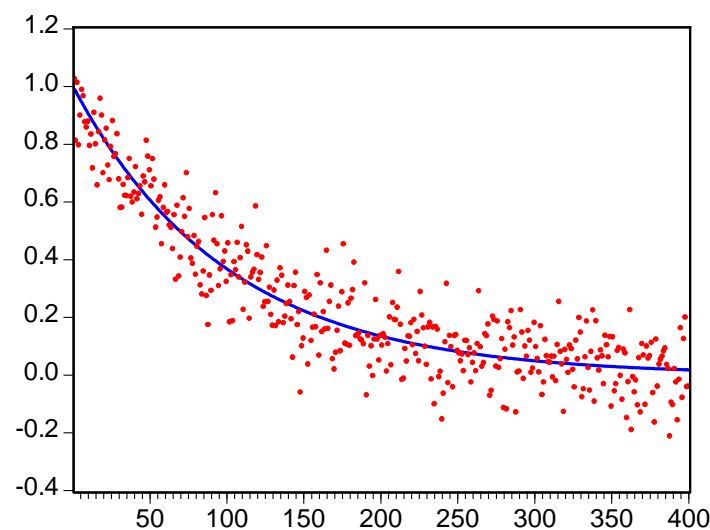
钉螺存活率样本值与拟合值。



(5) 指数函数模型



$$y_t = ae^{bx_t + u_t}, \quad (b < 0)$$



$$y_t = ae^{bx_t + u_t}, \quad (b > 0)$$

上式等号两侧同取自然对数，得

$$\ln y_t = \ln a + b x_t + u_t$$

令 $\ln y_t = y_t^*$, $\ln a = a^*$, 则

$$y_t^* = a^* + b x_t + u_t$$

变量 y_t^* 和 x_t 已变换成为线性关系。其中 u_t 表示随机误差项。

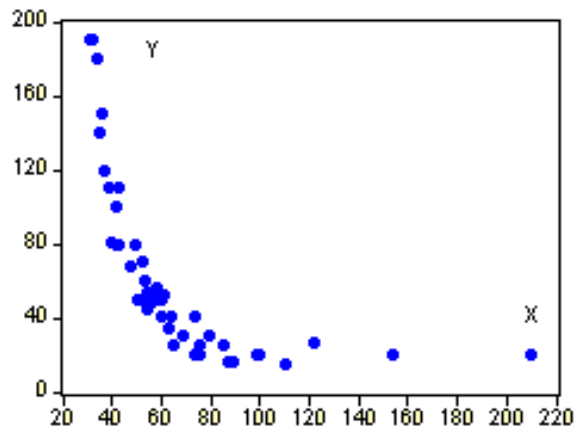
(第2版教材第115页)

(第3版教材第94页)

案例5：硫酸透明度与铁杂质含量的关系

(指数函数案例) (file:nonli01)

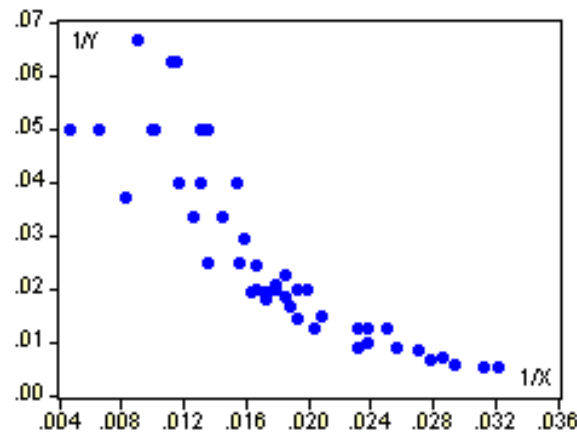
某硫酸厂生产的硫酸的透明度一直达不到优质指标。经分析透明度低与硫酸中金属杂质的含量太高有关。影响透明度的主要金属杂质是铁、钙、铅、镁等。通过正交试验的方法发现铁是影响硫酸透明度的最主要原因。测量了**47**个样本，得硫酸透明度（**y**）与铁杂质含量（**x**）的散点图如下：



$$(1) \quad y = 121.59 - 0.91 x$$

(10.1) (-5.7)

$R^2 = 0.42$, s.e. = 36.6, F= 32

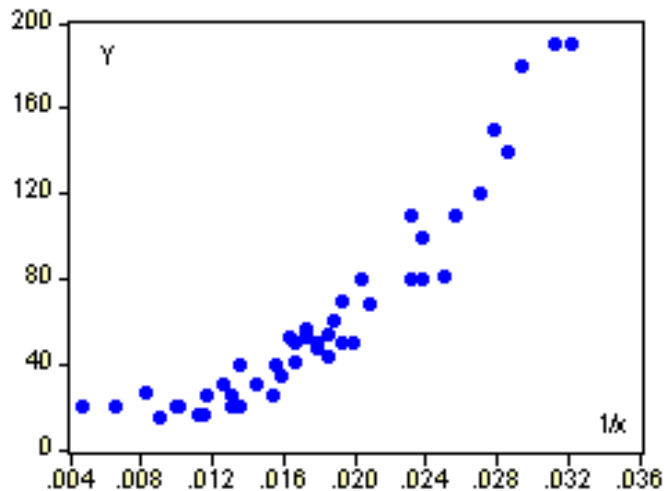


$$(2) \quad 1/y = 0.069 - 2.37 (1/x)$$

(18.6) (-11.9)

$R^2 = 0.76$, s.e. = 0.009, F= 142

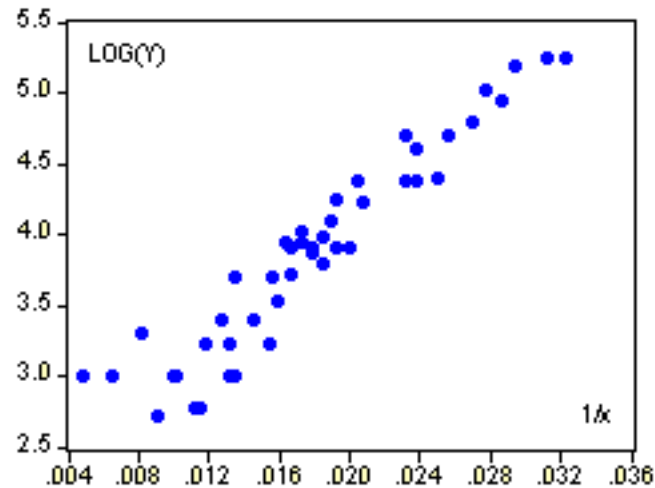
案例5：硫酸透明度与铁杂质含量的关系 (指数函数案例) (file:nonli01)



(3) $y = -54.40 + 6524.83 (1/x)$

(-7.2) (16.3)

$R^2 = 0.86$, s.e. = 18.2, F= 266



(4) $Ln y = 1.99 + 104.5 (1/x)$

(22.0) (21.6)

$R^2 = 0.91$, s.e. = 0.22, F= 468

还原, $Ln y = Ln(7.33) + 104.5 (1/x)$

$$y = 7.33e^{104.5(\frac{1}{x})}$$

案例5：硫酸透明度与铁杂质含量的关系 (指数函数案例) (file:nonli01)

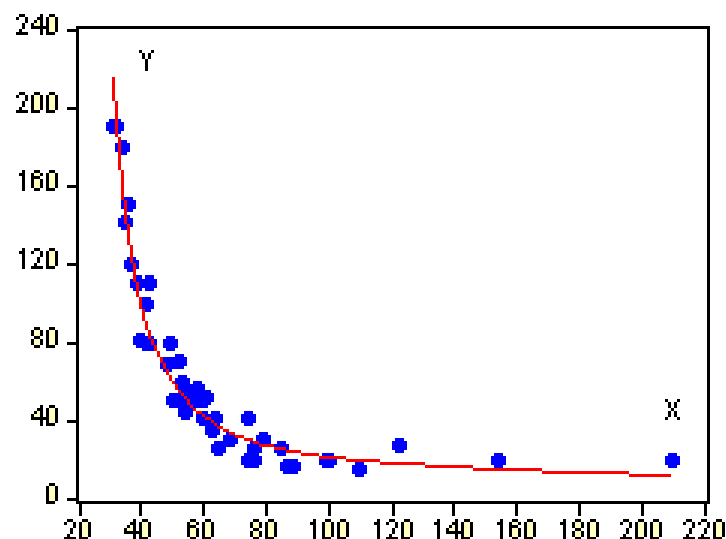
(5) 非线性估计结果是

$$y = 8.3e^{100.1\left(\frac{1}{x}\right)}$$

$$R^2 = 0.96$$

EViews命令: $Y=C(1)*EXP(C(2)*(1/X))$

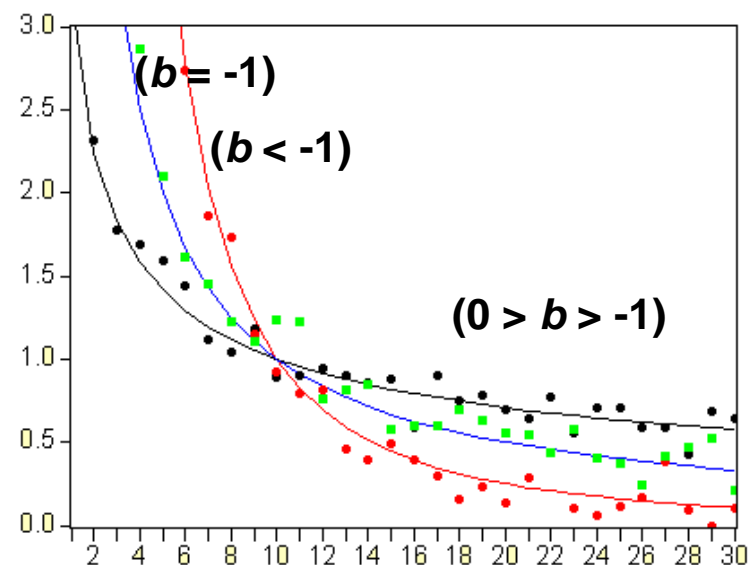
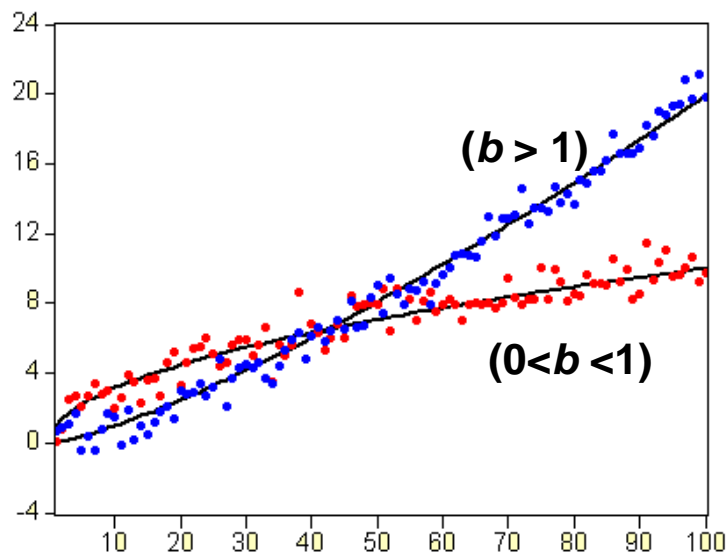
样本点与指数拟合曲线



样本内预测评价:

Forecast: YF	
Actual: Y	
Forecast sample: 1 47	
Included observations: 47	
Root Mean Squared Error	9.924945
Mean Absolute Error	7.722202
Mean Abs. Percent Error	17.16715
Theil Inequality Coefficient	0.065422
Bias Proportion	0.009562
Variance Proportion	0.001792
Covariance Proportion	0.988645

(6) 幂函数模型(全对数模型)



$$y_t = ax_t^b e^{u_t}$$

b 取不同值的图形分别见上图。对上式等号两侧同取对数，得

$$\textcolor{red}{Lny_t = Lna + b Lnx_t + u_t}$$

令 $y_t^* = Lny_t$, $a^* = Lna$, $x_t^* = Lnx_t$, 则上式表示为

$$\textcolor{red}{y_t^* = a^* + b x_t^* + u_t}$$

(第2版教材第116页)

(第3版教材第95页)

变量 y_t^* 和 x_t^* 之间已成线性关系。幂函数模型也称作**全对数模型**。

Cobb-Douglas生产函数（二元幂函数）

$Q = k L^\alpha C^{1-\alpha}$ 其中 Q 表示产量； L 表示劳动力投入量； C 表示资本投入量；

k 是常数； $0 < \alpha < 1$ 。更习惯的表达形式是

$y_t = \beta_0 x_{t1}^{\beta_1} x_{t2}^{\beta_2} e^{u_t}$ 上式两边同取对数，得：

（第2版教材第116页）

（第3版教材第95页）

$$\text{Lny}_t = \text{Ln}\beta_0 + \beta_1 \text{Lnx}_{t1} + \beta_2 \text{Lnx}_{t2} + u_t$$

取 $y_t^* = \text{Lny}_t, \beta_0^* = \text{Ln}\beta_0, x_{t1}^* = \text{Ln}x_{t1}, x_{t2}^* = \text{Ln}x_{t2}$ ，有

$$y_t^* = \beta_0^* + \beta_1 x_{t1}^* + \beta_2 x_{t2}^* + u_t$$

上式为线性模型。用 OLS 法估计后，再返回到原模型。

对于对数线性模型， $\text{Lny} = \text{Ln}\beta_0 + \beta_1 \text{Lnx}_{t1} + \beta_2 \text{Lnx}_{t2} + u_t$ ， β_1 和 β_2 称作弹性系数。

$$\beta_1 = \frac{\partial \text{Lny}_t}{\partial \text{Lnx}_{t1}} = \frac{y_t^{-1} \partial y_t}{x_{t1}^{-1} \partial x_{t1}} = \frac{\partial y_t / y_t}{\partial x_{t1} / x_{t1}} = \frac{x_{t1}}{y_t} \frac{\partial y_t}{\partial x_{t1}}$$

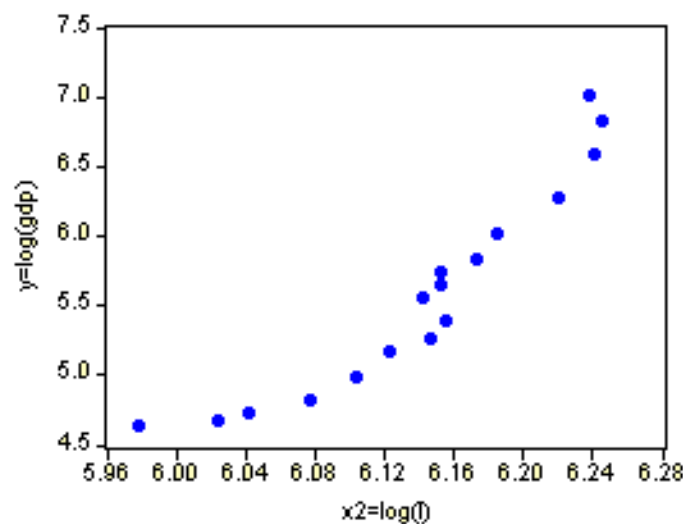
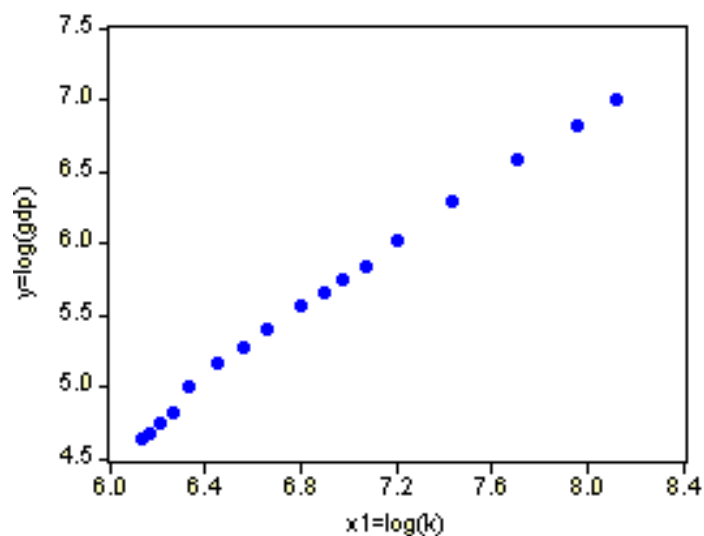
可见弹性系数是两个变量的变化率的比。注意，弹性系数是一个无量纲参数，

所以便于在不同变量之间比较相应弹性系数的大小。以 α_1 为例， $\alpha_1 = \frac{\partial y_t}{\partial x_{t1}}$

边际系数是弹性系数的一个分量。

例4.2：天津市GDP函数(教材第95页)

天津市 GDP_t （亿元），从业人员数（ L_t ，亿元），资金（ K_t ，亿元）数据(1980-1996)取对数后分别用Y、X1、X2表示。得估计模型如下：



(第2版教材第118页)

(第3版教材第95页)

例4.2：天津市GDP函数

(第2版教材第118页)
(第3版教材第95页)

Dependent Variable: Y
Method: Least Squares
Date: 02/02/07 Time: 19:33
Sample: 1980 1996
Included observations: 17

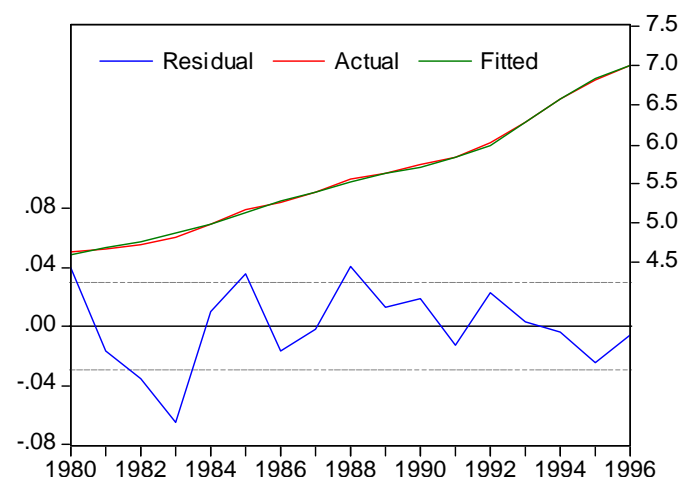
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-10.46386	1.287010	-8.130363	0.0000
X1	1.021124	0.029404	34.72712	0.0000
X2	1.471943	0.239290	6.151284	0.0000
R-squared	0.998608	Mean dependent var	5.600196	
Adjusted R-squared	0.998409	S.D. dependent var	0.749974	
S.E. of regression	0.029918	Akaike info criterion	-4.021929	
Sum squared resid	0.012531	Schwarz criterion	-3.874891	
Log likelihood	37.18639	F-statistic	5020.103	
Durbin-Watson stat	1.568312	Prob(F-statistic)	0.000000	

$$\hat{Y}_t = 10.46 + 1.02 X1_t + 1.47 X2_t$$

(-8.1) (34.7) (6.2)

$R^2 = 0.9986$, $DW = 1.7$, $N = 17$

因为 $1.02 + 1.47 = 2.47$ ，所以此生产函数属于规模报酬递增函数。



Forecast: YF
Actual: Y
Forecast sample: 1980 1996
Included observations: 17

Root Mean Squared Error	0.027150
Mean Absolute Error	0.021535
Mean Abs. Percent Error	0.410875
Theil Inequality Coefficient	0.002404
Bias Proportion	0.000000
Variance Proportion	0.000348
Covariance Proportion	0.999652

案例6：中国私人轿车拥有量决定因素分析

（多元非线性回归）（file: nonli14）

1985-2002年中国私人轿车拥有量以年增长率23%，年均增长55万辆的速度飞速增长。

在建立中国私人轿车拥有量模型时，主要考虑如下因素：（1）城镇居民家庭人均可支配收入；（2）城镇总人口；（3）轿车产量；（4）公路交通完善程度；（5）轿车价格。

由于国产轿车价格与进口轿车价格差距较大，而且轿车种类很多，做分种类的轿车销售价格与销售量统计非常困难，所以因素“轿车价格”暂且略去不用。定义变量名如下：

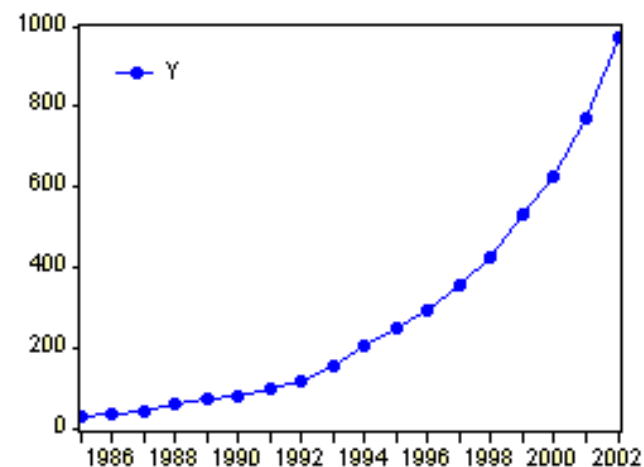
Y：中国私人轿车拥有量（万辆）

X1：城镇居民家庭人均可支配收入（元），

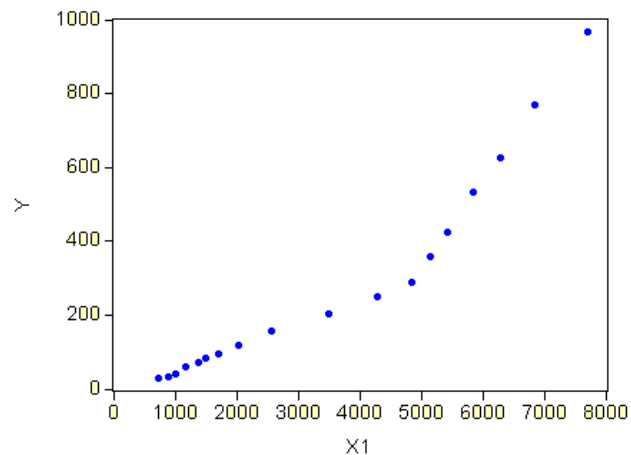
X2：全国城镇人口（亿人）

X3：全国汽车产量（万辆）

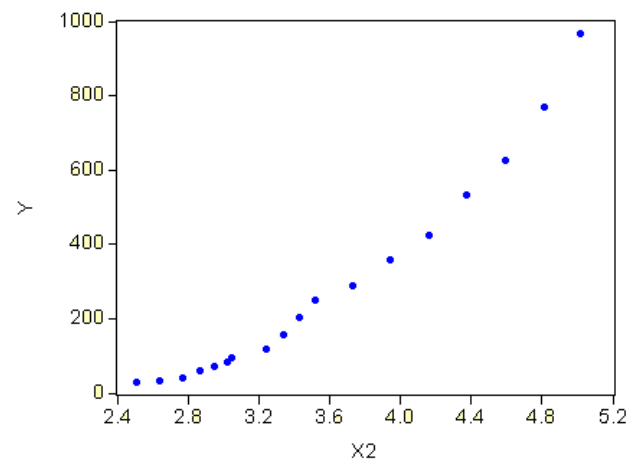
X4：全国公路长度（万公里）



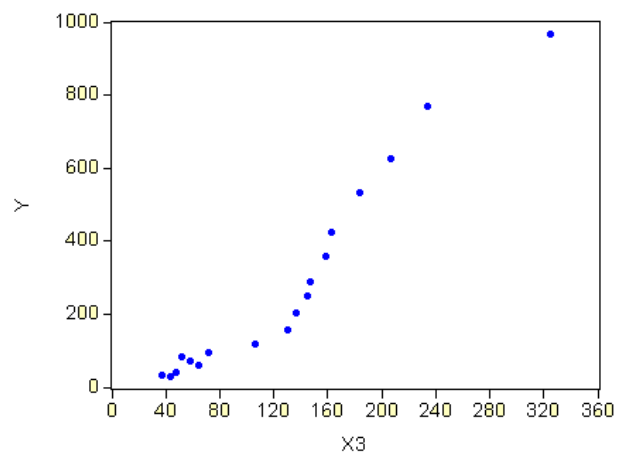
案例6：中国私人轿车拥有量决定因素分析 (多元非线性回归) (file: nonli14)



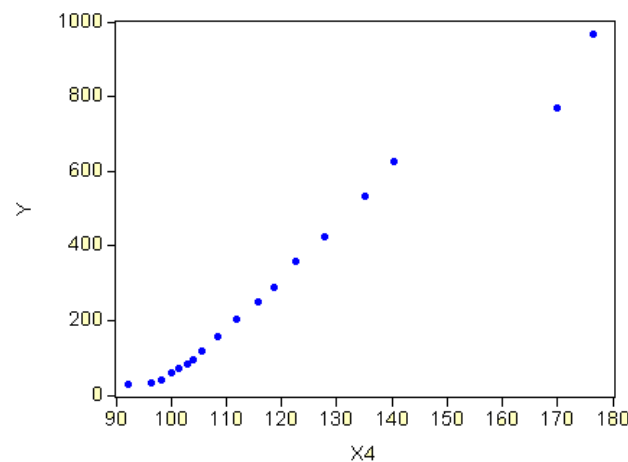
轿车拥有量与人均可支配收入



轿车拥有量与全国城镇人口



轿车拥有量与全国汽车产量



轿车拥有量与全国公路长度

案例6：中国私人轿车拥有量决定因素分析 (多元非线性回归) (file: nonli14)

Correlation Matrix					
	Y	X1	X2	X3	X4
Y	1.000000	0.950119	0.972308	0.966596	0.989264
X1	0.950119	1.000000	0.983022	0.958465	0.929555
X2	0.972308	0.983022	1.000000	0.962856	0.958785
X3	0.966596	0.958465	0.962856	1.000000	0.955281
X4	0.989264	0.929555	0.958785	0.955281	1.000000

$$Y = -925.66 + 0.0057X1 + 62.94 X2 + 0.41 X3 + 7.73X4$$

$$(-5.7) \quad (0.2) \quad (0.8) \quad (0.8) \quad (5.0)$$

$$R^2 = 0.99, DW=1.4, T=18, (1985\sim 2002)$$

看相关系数阵，Y与X1，X2，X3，X4的相关系数都在0.9以上，但输出结果中，解释变量X1，X2，X3的回归系数却通不过显著性检验。这预示解释变量之间一定存在多重共线性。

案例6：中国私人轿车拥有量决定因素分析 (多元非线性回归) (file: nonli14)

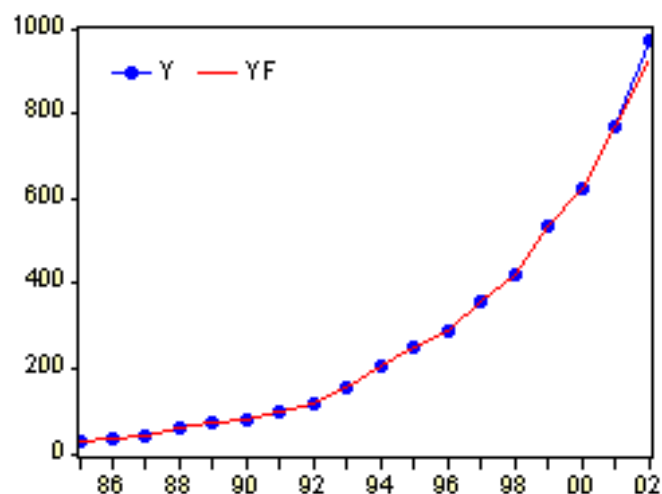
看散点图，把Y与X3,X4处理成线性关系，把Y与X1,X2处理成幂函数（抛物线）关系，得结果如下，

$$Y = 317.19 + 0.00000326X1^2 - 363 X2 + 74.41 X2^2 + 0.31X3 + 1.48X4$$

$$(2.9) \quad (4.4) \quad (-7.3) \quad (9.5) \quad (2.1) \quad (3.0)$$

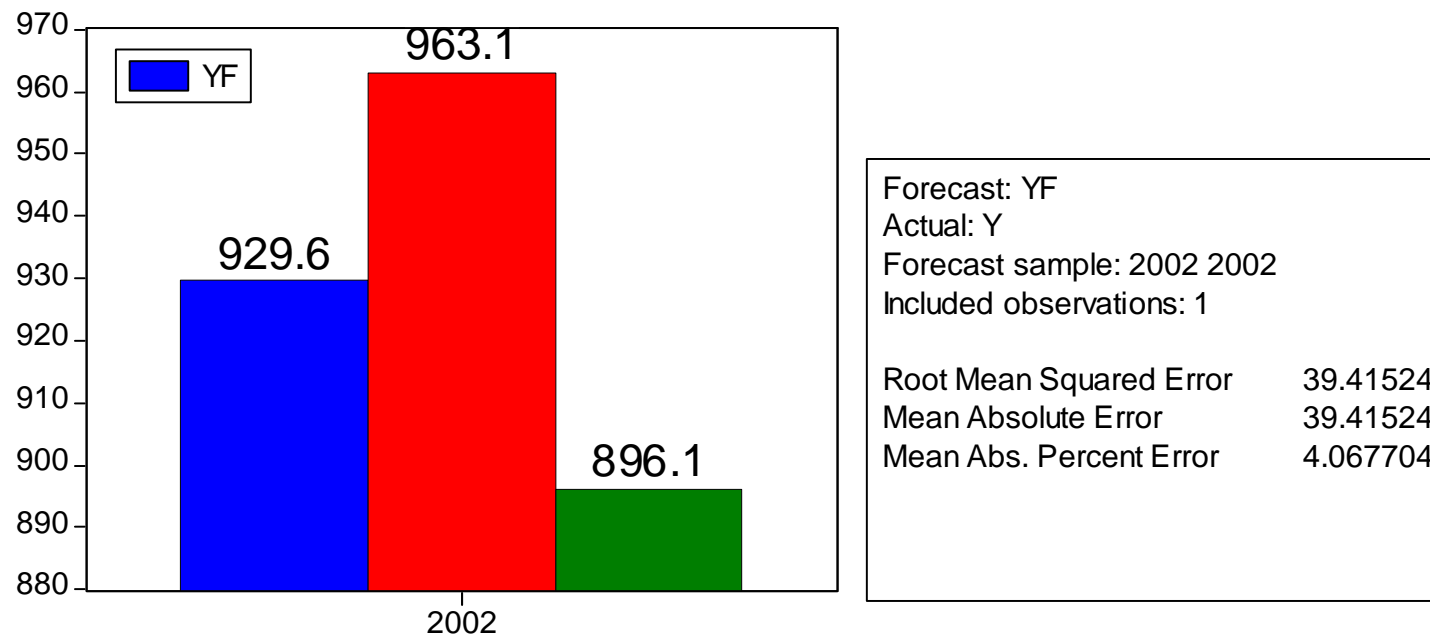
$$R^2 = 0.9991, DW=2.0, T=17, (1985\sim 2001)$$

每个变量都具有很高的显著性。拟合优度也提高了，没有异方差也没有自相关。样本内拟合如图。



案例6：中国私人轿车拥有量决定因素分析 (多元非线性回归) (file: nonli14)

作样本外1期预测。预测2002年 $Y = 929.5648$ 。预测误差0.04。EViews的计算结果。



(7) 不可线性化的非线性回归模型估计方法 (不要求掌握)

第4章结束.