

《计量经济学》建模案例

案例 1：用回归模型预测木材剩余物

伊春林区位于黑龙江省东北部。全区有森林面积 2189732 公顷，木材蓄积量为 23246.02 万 m^3 。森林覆盖率为 62.5%，是我国主要的木材工业基地之一。1999 年伊春林区木材采伐量为 532 万 m^3 。按此速度 44 年之后，1999 年的蓄积量将被采伐一空。所以目前亟待调整木材采伐规划与方式，保护森林生态环境。为缓解森林资源危机，并解决部分职工就业问题，除了做好木材的深加工外，还要充分利用木材剩余物生产林业产品，如纸浆、纸袋、纸板等。因此预测林区的年木材剩余物是安排木材剩余物加工生产的一个关键环节。下面，利用简单线性回归模型预测林区每年的木材剩余物。显然引起木材剩余物变化的关键因素是年木材采伐量。

伊春林区 16 个林业局 1999 年木材剩余物和年木材采伐量数据见附表。散点图见图 2.14。观测点近似服从线性关系。建立一元线性回归模型如下：

$$y_t = b_0 + b_1x_t + u_t$$

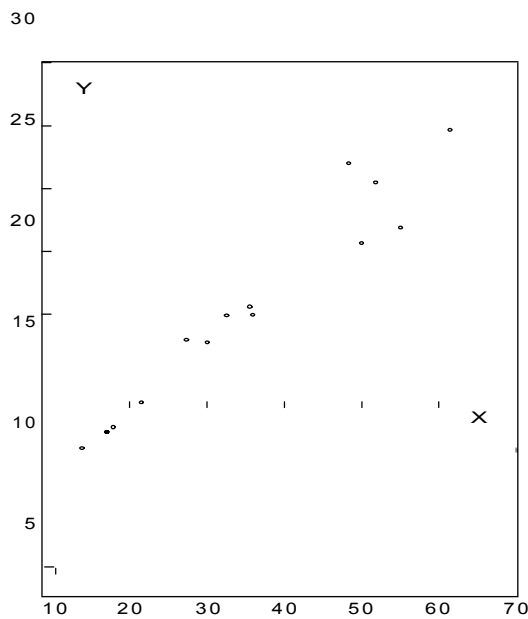


图 年剩余物 y_t 和年木材采伐量 x_t 散点图

Dependent Variable: Y
Method: Least Squares
Date: 10/09/03 Time: 15:38
Sample: 1 16
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.762928	1.220966	-0.624856	0.5421
X	0.404280	0.033377	12.11266	0.0000
R-squared	0.912890	Mean dependent var		12.67938
Adjusted R-squared	0.906668	S.D. dependent var		6.665466
S.E. of regression	2.036319	Akaike info criterion		4.376633
Sum squared resid	58.05231	Schwarz criterion		4.473207
Log likelihood	-33.01306	F-statistic		146.7166
Durbin-Watson stat	1.481946	Prob(F-statistic)		0.000000

图 1 Eviews 输出结果

Eviews 估计结果见图 1。

下面分析 Eviews 输出结果。先看图 1 的最上部分。**LS** 表示本次回归是最小二乘回归。被解释变量是 y_t 。本次估计用了 16 对样本观测值。输出格式的中间部分给出 5 列。第 1 列给出截距项 (C) 和解释变量 x_t 。第 2 列给出相应项的回归参数估计值 (\hat{b}_0 和 \hat{b}_1)。第

根据 Eviews 输出结果 (图 2.15), 写出 OLS 估计式如下:

$$\hat{y}_t = -0.7629 + 0.4043 x_t$$

$$(-0.6) \quad (12.1) \quad R^2 = 0.91, \quad s. e. = 2.04$$

其中括号内数字是相应 t 统计量的值。 $s.e.$ 是回归函数的标准误差, 即 $\hat{\sigma} = \sqrt{\sum \hat{u}_t^2 / (16-2)}$ 。 R^2 是可决系数。 $R^2 = 0.91$ 说明上式的拟合情况较好。 y_t 变差的 91%由变量 x_t 解释。检验回归系数显著性的原假设和备择假设是 (给定 $\alpha = 0.05$)

$$H_0: b_1 = 0; \quad H_1: b_1 \neq 0$$

obs	Actual	Fitted	Residual	Residual Plot
1	26.1300	24.0599	2.07015	
2	23.4900	18.7638	4.72621	
3	21.9700	20.1788	1.79123	
4	11.5300	13.7507	-2.22072	
5	7.18000	6.43325	0.74675	
6	6.80000	6.10983	0.69017	
7	18.4300	21.4725	-3.04246	
8	11.6900	12.4570	-0.76702	
9	6.80000	6.10983	0.69017	
10	9.69000	10.2739	-0.58391	
11	7.99000	7.92909	0.06091	
12	12.1500	13.5890	-1.43900	
13	6.80000	6.10983	0.69017	
14	17.2000	19.4511	-2.25106	
15	9.50000	11.3655	-1.86547	
16	5.52000	4.81613	0.70387	

图 2.16 残差图

因为 $t = 12.1 > t_{0.05(14)} = 2.15$, 所以检验结果是拒绝 $b_1 = 0$, 即认为年木材剩余物和年木材采伐量之间存在回归关系。上述模型的经济解释是, 对于伊春林区每采伐 1 m^3 木材, 将平均产生 0.4 m^3 的剩余物。

图 2.16 给出相应的残差图。Actual 表示 y_t 的实际观测值, Fitted 表示 y_t 的拟合值 \hat{y}_t , Residual 表示残差 \hat{u}_t 。残差图中的两条虚线与中心线的距离表示残差的一个标准差, 即 $s.e.$ 。通过残差图可以看到, 大部分残差值都落在了正、负一个标准差之内。

估计 b_1 的置信区间。由

$$t = P \left\{ \left| \frac{\hat{\beta}_1 - \beta_1}{s(\hat{\beta}_1)} \right| \leq t_{0.05(14)} \right\} = 0.95$$

得

$$\left| \hat{b}_1 - b_1 \right| \leq t_{0.05(14)} s(\hat{b}_1)$$

b_1 的置信区间是

$$\begin{aligned} & \left[\hat{b}_1 - t_{0.05(14)} s(\hat{b}_1), \hat{b}_1 + t_{0.05(14)} s(\hat{b}_1) \right] \\ & [0.4043 - 2.15 \times 0.0334, \quad 0.4043 + 2.15 \times 0.0334] \\ & [0.3325, 0.4761] \end{aligned}$$

以 95% 的置信度认为, b_1 的真值范围应在 $[0.3325, 0.4761]$ 范围中。

下面求 y_t 的点预测和置信区间预测。假设乌伊岭林业局 2000 年计划采伐木材 20 万 m^3 , 求木材剩余物的点预测值。

$$\begin{aligned} \hat{y}_{2000} &= -0.7629 + 0.4043 x_{2000} \\ &= -0.7629 + 0.4043 \times 20 = 7.3231 \text{ 万 m}^3 \end{aligned}$$

$$\begin{aligned} s^2(\hat{y}_{2000}) &= \hat{\sigma}^2 \left(\frac{1}{T} + \frac{(x_F - \bar{x})^2}{\sum (x - \bar{x})^2} \right) \\ &= 4.1453 \left(\frac{1}{16} + \frac{(20 - 33.25)^2}{3722.2606} \right) = 0.4546 \\ s(\hat{y}_{2000}) &= \sqrt{0.4546} = 0.6742 \end{aligned}$$

因为

$$E(\hat{y}_{2000}) = E(\hat{b}_0 + \hat{b}_1 x_{2000}) = b_0 + b_1 x_{2000} = E(y_{2000})$$

$$t = \frac{\hat{y}_{2000} - E(y_{2000})}{s(\hat{y}_{2000})} \sim t_{(T-2)}$$

则置信度为 0.95 的 2000 年平均木材剩余物 $E(\mathbf{y}_{2000})$ 的置信区间是

$$\begin{aligned} \hat{\mathbf{y}}_{2000} \pm t_{0.05(14)}s(\mathbf{y}_{2000}) &= 7.3231 \pm 2.15 \times 0.6742 \\ &= 5.8736, \quad 8.7726 \end{aligned}$$

从而得出预测结果，2000 年若采伐木材 20 万 m^3 ，产生木材剩余物的点估计值是 7.3231 万 m^3 。平均木材剩余物产出量的置信区间估计是在 [5.8736, 8.7726] 万 m^3 之间。从而为恰当安排 2000 年木材剩余物的加工生产提供依据。

附数据如下：

林业局名	年木材剩余物 \mathbf{y}_t (万 m^3)	年木材采伐量 \mathbf{x}_t (万 m^3)
乌伊岭	26.13	61.4
东风	23.49	48.3
新青	21.97	51.8
红星	11.53	35.9
五营	7.18	17.8
上甘岭	6.80	17.0
友好	18.43	55.0
翠峦	11.69	32.7
乌马河	6.80	17.0
美溪	9.69	27.3
大丰	7.99	21.5
南岔	12.15	35.5
带岭	6.80	17.0
朗乡	17.20	50.0
桃山	9.50	30.0
双丰	5.52	13.8
合计	202.87	532.00

案例 2 中国宏观消费分析

(摘自经济蓝皮书《2004 年：中国经济形势分析与预测》)

按照我国现行国民经济核算体系，国内生产总值（按支出法计算）是由最终消费、资本形成总额和货物与服务的净出口之和三部分组成。前两部分占绝大多数。其中最终消费又分为居民消费和政府消费两类。而居民消费又可分为农村居民消费和城镇居民消费。

在这种核算体系下，居民消费包括居民个人日常生活中衣、食、住、用等物质消费以及在文化生活服务性支出中属于物质产品的消费。

政府消费包括国家机关、国防、治安、文教、卫生、科研事业单位，经济建设部门的事业单位，人民团体等非生产机构使用的燃料、电力、办公用品、图书、设备等物质消费。

国内生产总值中最终消费与资本形成总额的比例关系，即旧核算体系下国民收入中消费与积累的比例关系是国民经济正常运行的最基本的比例关系。如果这一比例关系发生严重失调，最终会成为制约经济正常运行的严重障碍。

下面分析中国的消费问题。为消除物价变动因素以及异方差的影响，以下分析所用的数据均为不变价格数据（1952=1）以及分别取自然对数后的数据。数据见附表。

图 1 给出不变价格的国内生产总值与消费曲线，图 2 和图 3 分别给出国内生产总值与消费的年增长率曲线。

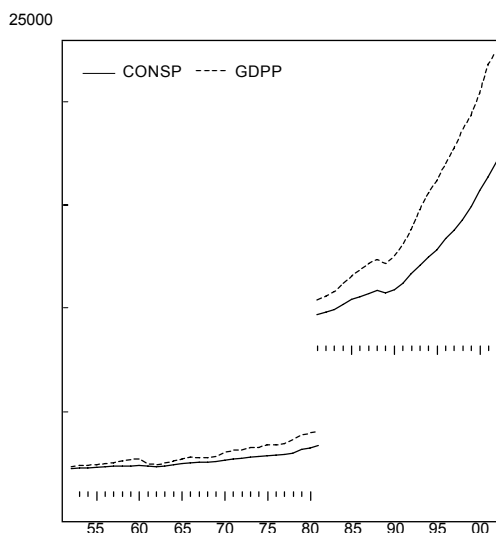


图 1 国内生产总值与消费（不变价格）曲线

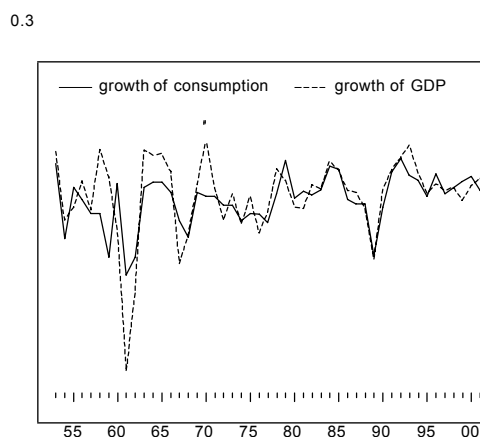


图 2 国内生产总值与消费年增长率曲线

由图 1、2 可以看出国内生产总值与消费的增长都很快。国内生产总值曲线的波动幅度相比较较大。消费曲线的波动幅度相对较小。这与宏观消费行为具有“惯性”有关。他既不可能随时间突然大幅增加，也不可能随时间突然大幅减少。

	1952-1978		1979-2002	
	平均增长率	年增长率的的标准差	平均增长率	年增长率的的标准差
GDP	5.76%	0.10	9.15%	0.044
消费	4.79%	0.05	9.18%	0.040

首先结合图 1.3 对国内生产总值序列的增长率变化做进一步分析。1952-1957 年国民收入呈较稳步发展。以不变价格计算，平均年增长率为 7.97%。1958 年开始的大跃进使经济发展速度突然加快。在计划经济体制下，这种人为的提高经济发展速度超出了国家物质基础所能承受的限度，所以在维持了短短两年超高速增长（1958 年的年增长率为 16.9%，1959 年的年增长率为 11.4%）之后，经济发展便出现了大倒退。1960 年几乎为零增长。1961 和 1962 年连续 2 年出现建国以来从未有过的负增长（分别为 -27.2% 和 -11.1%）。由于国家及时采取了一系列经济调整措施，1963-1966 年国民经济迅速得到恢复，并出现持续高速增长态势。上述 4 年的增长率分别为 17.8%，15.8%，16.1% 和 12.5%。1966 年开始的文化革命使中国经济进入一个很不稳定的发展阶段。1967 和 1968 年国民经济再度出现负增长，随后经济发展出现“振荡”现象。自 1978 年实行改革开放政策以来，在由计划经济向市场经济转变过程中，经济发展突飞猛进。1952-1978 年国民收入年平均增长率为 5.76%。1978-2002 年的年平均增长率为 9.15%。后一时期是前一时期的 1.6 倍（不变价格）。在后一时期里，经济增长速度如此之高，持续时间如此之长，发展趋势如此之稳定，在我国的发展史上是没有先例的。

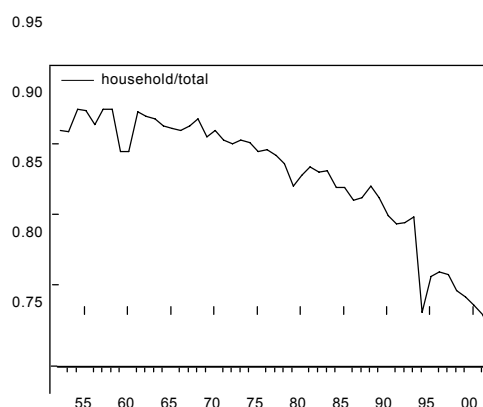
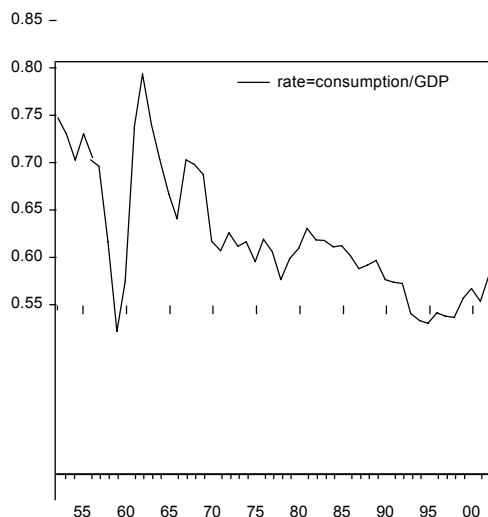


图3 年消费率曲线（1952-2002）

图4 居民消费与总消费比的变化曲线（1952-2002）

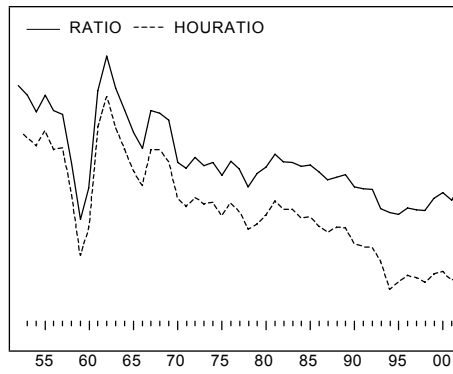


图 5 宏观消费比率与居民消费比率曲线（1952-2002）

下面分析消费率（消费额 / 国内生产总值，1952-2002）序列的变化。见图 3，总的来说变化幅度较大。

(1) 从趋势看，中国宏观消费比率值的变化是逐年下降。消费比率数据对时间 t (1952=1) 的回归结果如下：

$$\text{ratio} = 0.7581 - 0.0036t$$

(62.9) (-8.8) $R^2 = 0.61$ (1952-2002)

51 年间消费比率值平均每年减少 0.0036。

(2) 以 1978 年为界，改革开放之前 (1949~1978) 消费比率曲线波动大，改革开放之后 (1979~2002) 消费比率曲线波动小 (见图 5 和表 1)。1952~1978 年宏观消费比率值的均值是 0.7057，标准差是 0.0656。1979-2002 年宏观消费比率的均值是 0.6206。标准差是 0.0324。改革开放以后宏观消费比率值平均比改革开放前下降 0.085。随着时间的推移，消费比率的均值减小，标准差减小。改革开放之后标准差减小说明宏观消费比率值的波动在减小，中央政府调控宏观经济的能力逐步在提高。

(3) 宏观消费比率的最小值是 0.5660，最大值是 0.8379。都发生在上世纪 50 年代末和 60 年代初的经济困难时期。最小值 0.5660 发生在 1959 年是由于基本建设投资的极度扩张造成的 (1958 和 1959 年基本建设投资的年增长率分别是 87.7% 和 30.0%)。最大值是 0.8379 发生在 1962 年是由于执行经济调整政策，首先解决人民生活所致。

(4) 中国宏观消费比率值自 1993 年起跌破 0.60 大关。1995 年达到最低点 0.575。近 10 年来，宏观消费比率值基本上在 0.60 以下徘徊，平均值是 0.5876。在中央政府努力扩大消费的政策下虽然宏观消费比率值在 1999 和 2000 年回升至 0.60 以上，但 2001 和 2002 年又跌落到 0.60 以下。当然这并不意味着中国宏观消费绝对值的减少。相反，宏观消费总量一直在快速提高。因为固定资产投资以更快的速度增长，所以导致宏观消费比率值偏低。

(5) 图 4 给出居民消费占总消费的比率曲线。该比值从 0.91 直线下降至 0.76。这一方面反映出政府消费越削越增的过程，同时也反映出居民消费占总消费的比率变得越小。

表 1 中国消费比率数据的特征数

特征数名称	消费比率的特征数	
	(1952~1978)	(1979~2002)
均值	0.7057	0.6206
标准差	0.0656	0.0324
极大值	0.8379	0.6751
极小值	0.5660	0.5749
变异系数	0.0930	0.0522
样本容量	27	24

注：(1) 消费比率 = 中国宏观消费 / GDP。

(2) 1952~1999 年消费和 GDP 数据摘自《新中国五十年统计资料汇编》，1999 中国统计出版社。2000~2002 年消费和 GDP 数据摘自《中国统计年鉴》，2003，中国统计出版社。

(3) 消费比率数据的特征数用消费比率数据计算。

中国宏观消费比率的国际比较。

共选择 6 个工业发达国家和 4 个发展中国家和地区的 GDP 和宏观消费数据经计算后，与中国进行宏观消费比率的对比。6 个工业发达国家是英国、美国、法国、意大利、加拿大和日本 (GDP 和消费均为年度数据，德国由于数据不全未选)。4 个发展中国家和地区是菲律宾、墨西哥、香港 (GDP 和消费均为季节数据) 和韩国 (GDP 和消费为年度数据)。上述 10 个国家和地区的宏观消费比率曲线与中国宏观消费比率曲线的对比分别见图 7 和图 8。11 个国家和地区宏观消费比数据的 5 个特征数见表 2。结合图 7 和图 1.8 以及表 2，分析如下：

PDF 文件使用 “pdfFactory” 试用版本创建 www.fineprint.cn

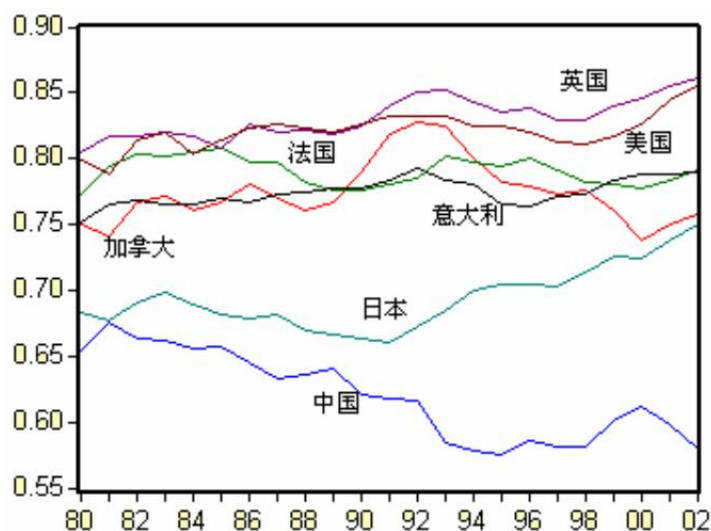


图 7 美国、英国、加拿大、法国、意大利、日本与中国的消费比率曲线比较

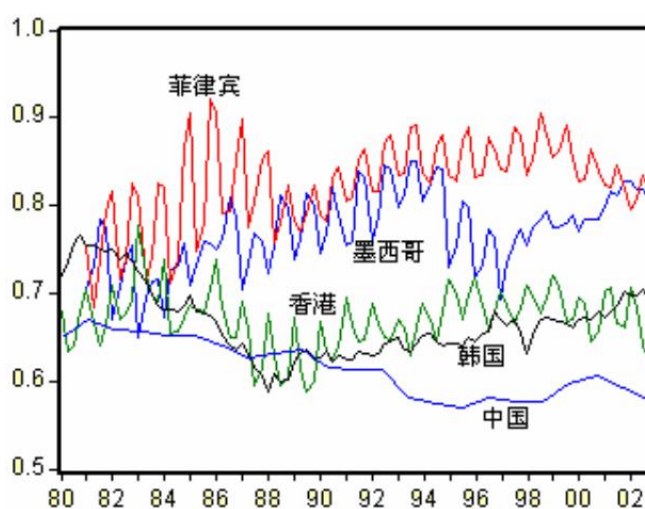


图 8 墨西哥、香港、菲律宾、韩国与中国大陆的消费比率曲线比较

(1) 在这 11 个国家和地区中，无论是和工业发达国家还是发展中国家和地区相比，中国的宏观消费比率是最低的。

(2) 年平均消费比率在 0.7 以上的国家按消费比率值大小顺序排列是英国、菲律宾、美国、法国、意大利、加拿大和墨西哥（见表 2 中第一栏）。年平均消费比率在 0.6~0.7 之间的国家是日本、香港、韩国和中国（见表 2 中第二栏）。显然，这种差别与文化传统有着密切的联系。前 7 个国家都是具有西方文化色彩的国家；而后 4 个国家都是具有东方文化色彩的国家。

(3) 从消费比率的标准差和变异系数来看，排除菲律宾、墨西哥和香港（这 3 个国家的数据为季节数据，他们的方差与其他国家无可比性），中国和韩国是消费比率值变化最大的国家。中国消费比率标准差是变化最小的法国和意大利的 3 倍多。在消费比率低于 0.7 的国家与地区中，日本和韩国的消费比率曲线是先降后升；香港呈震荡变化特征；而中国则是呈逐年下降趋势。

表 2 加拿大、中国等 11 个国家与地区宏观消费比数据的特征值比较

国别	均值	标准差	极大值	极小值	变异系数	样本容量
英国 (1980~2002, 年度数据)	0.8311	0.0154	0.8606	0.8051	0.0185	23
菲律宾 (1982~2002, 月度数据)	0.8259	(0.0499)	0.9203	0.6829	(0.0604)	88
美国 (1980~2002, 年度数据)	0.8213	0.0140	0.8544	0.7884	0.0170	23
法国 (1980~2002, 年度数据)	0.7905	0.0106	0.8074	0.7734	0.0134	23
意大利 (1980~2002, 年度数据)	0.7748	0.0103	0.7931	0.7512	0.0133	23
加拿大 (1980~2002, 年度数据)	0.7744	0.0243	0.8279	0.7384	0.0314	23

墨西哥	(1982~2002, 月度数据)	0.7709	(0.0446)	0.8516	0.6487	(0.0579)	88
日本	(1980~2002, 年度数据)	0.6940	0.0241	0.7501	0.6600	0.0347	23
香港	(1980~2002, 月度数据)	0.6708	(0.0339)	0.7780	0.5874	(0.0505)	92
韩国	(1980~2002, 年度数据)	0.6665	0.0420	0.7513	0.5976	0.0630	23

PDF 文件使用 “pdfFactory” 试用版本创建 www.fineprint.cn

注：(1) 香港、菲律宾和墨西哥宏观消费和 GDP 数据未经季节调整。

(2) 英国、美国、法国、意大利、加拿大和日本的 GDP 和消费数据摘自国际货币基金组织数据库 (网站: <http://www.imf.org>)。菲律宾、墨西哥、韩国和香港的 GDP 和消费数据摘自经合组织数据库 (网站: <http://www.oecd.org>)。消费比率数据是作者自己计算的。

(3) 括号中的特征数不参与比较 (这些特征数来自于月度数据, 无可比性)。

(4) 中国的消费比率值为什么呈一路下滑趋势? 主要原因是全国固定资产投资增长率 (2002 年是 13.1%) 多年来远远高于消费的增长率 (2002 年是 5.8%), 从而导致消费比率值连年下滑。

(5) 表 3 给出了上述 11 个国家与地区 GDP 增长率与宏观消费比率的相关系数。出了韩国例外, 其他国家与地区均显示 GDP 增长率与宏观消费比率呈显著的负相关关系。即若 GDP 增长率上升, 则宏观消费比率下降; 若 GDP 增长率下降, 则宏观消费比率上升。看来这是个一般规律。为什么会存在这样的规律? 因为消费的惯性要比投资大得多。GDP 的变化主要由投资控制。如果投资增加, 投资占 GDP 的比例增加, 必然导致消费比率下降; 反之消费比率上升。

表 3 加拿大、中国等 11 个国家与地区 GDP 增长率与宏观消费比率的相关系数

	相关系数	临界值 (5%水平)	临界值 (10%水平)
意大利	-0.64	0.41	0.34
英国	-0.84	0.41	0.34
香港	-0.62	0.21	0.17
日本	-0.85	0.41	0.34
墨西哥	-0.48	0.21	0.17
加拿大	-0.66	0.41	0.34
美国	-0.77	0.41	0.34
菲律宾	0.19	0.21	0.17
法国	0.28	0.41	0.34
韩国	-0.07	0.41	0.34
中国 (1952-2002)	-0.50	0.26	0.24
中国 (1980-2002)	-0.20	0.41	0.34

(6) 中国目前的宏观消费比率这样低好不好? 从长期看不好, 应该改变消费与 GDP 之间的这种低比例关系。原因有四。①宏观消费和固定资产投资是维持经济高增长的两个最重要因素。在经济高增长条件下, 消费比率偏低是靠连年的固定资产投资高增长率维持的。而连年的固定资产投资高增长率必然带来人力、物力和财力的瓶颈现象。中国近年来之所以没有出现像大跃进时期的物力和财力的瓶颈现象, 主要是依靠外国直接投资和借外债支撑的。但长期借外债后, 还款将成为一个沉重负担, 同时经济长期超高速发展, 高素质人才的缺乏将变得越来越突出。这些因素制约固定资产投资的超高速增长将随着时间的延长越来越突出。②若没有一个合理的消费比率做支撑, 高投资比率将得不到延续, 最终导致产品相对过剩和积压, 经济发展速度下降。③提高消费比率, 维持消费的高增长同样能带来经济的高增长。因为提高消费比率主要刺激的是第三产业的发展。第三产业的发展在促进经济增长的同时, 还可以扩大劳动力的就业。为人民政府解决待业问题减轻压力。目前在这方面还有很大的潜力。以 2002 年为例, 全国第三产业产值占 GDP 的比例只有 0.34。④以经济建设为中心, 不断提高中国人民的物质与精神生活水平是我们党和国家的工作重心, 宏观消费比率长期保持低位不是我们的目的。

基于我国 54 年经济发展经验以及目前的经济发 展规模, 把年消费率平均值控制在 0.65-0.70 是比较合理的模式。

居民消费是由农村居民消费和城镇居民消费两部分组成。由于我国农村人口占大多数, 建国初期农村居民消费额在居民消费额中占较大比例。1952 年为 68.7%。随后, 这个比值逐年下降。1988 年这个比值突破 60.0%, 2002 年已降至 43.6%。这与城镇居民的收入增长速度大于农村居民的收入增长速度, 非农业人口占全国总人口的比例逐年增大, 以及农村劳动力大量涌入城市有直接关系。

下面分析农村居民和城镇居民人均消费水平的变化。1952 年分别为 62 元和 148 元 (当年价格)。1997 年分别为 2071 元和 4914 元 (当年价格)。按不变价格计算, 分别提高了 7.2 倍。图 1.9 给出城镇居民与农

村居民人均消费比的变化曲线。51 年间出现三个半波动周期，波动幅度在 2.2 和 3.2 之间变化。第一个波峰发生在 1959 和 1960 年。由于农业的减产直接影响了农村居民的收入和消费，使这个比值由建国初期的 2.5 倍猛增至 3.2 倍。随着经济政策的调整和农业生产的恢复，上述比值降至 2.4 左右。从 1970-1977 年这个比值连续攀升至 2.9。这是因为文化革命期间农村的一系列政策极大地限制了农民的生产积极性，从而使农业居民的收入和消费增长速度变小。第二次波峰发生在 1976 和 1977 年。随着农村体制的改革（取消人民公社，实行联产承包责任制等），极大地调动了农民生产的积极性。1978-1984 年是我国农业大发展时期，农村居民收入和消费水平的迅速提高（农村超前城市）使上述人均消费比值迅速下降。1984 年以后随着城市经济体制改革的深入，城镇居民的消费水平增加速度超过了农业居民消费水平的增加速度，致使上

述比值又开始攀升，1993 年达到 3.2。1994 年以后由于城镇居民收入增长速度逐年下降，直接影响到消费，于是上述比值又开始回落。1997 年已回落至 2.4。1997 年以后粮食价格一路走低，农民收入和支出与城市居民相比增长越来越慢，致使城乡消费比率值 2002 年又攀升至 11。

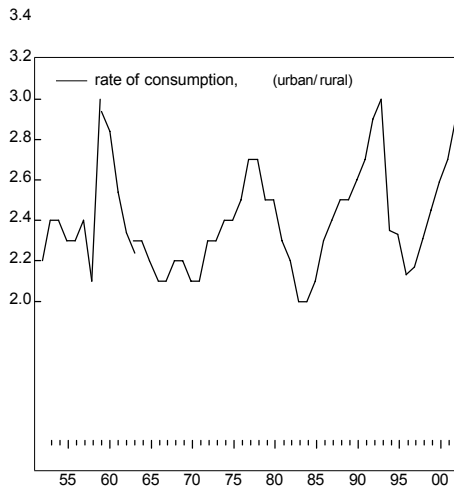


图 9 城镇与农村居民人均消费比的变化曲线

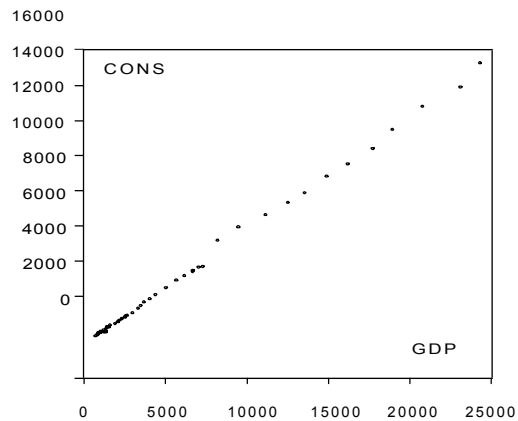


图 10 国内生产总值与消费额散点图

下面通过建立宏观消费计量经济模型进一步分析我国消费与国民收入的定量关系。（以下所用数据（1952-2002）均以不变价格（1952=1，单位：亿元人民币）计算。）

1952-2002 年国内生产总值与消费额散点图见图 1.10。说明消费与国内生产总值之间存在高度的线性关系。

用 CP_t 表示消费额（不变价格）， GDP_t 表示国内生产总值（不变价格），用 1952-2002 年数据得消费函数的 OLS 估计结果如下：

$$\hat{CP}_t = 164.0124 + 0.5919GDP_t \quad (1.1)$$

$$(5.2) \quad (159.9)$$

$$R^2 = 0.998, \quad DW = 0.67, \quad s.e. = 167.45$$

以上模型的 DW 值很小，严格地说模型存在自相关。为消除自相关 ($r=0.67$)，对变量进行广义差分。定义

$$GCP_t = CP_t - 0.665 CP_{t-1} \quad (1.2)$$

$$GGDP_t = GDP_t - 0.665 GDP_{t-1} \quad (1.3)$$

得估计的回归模型为，

$$GD\hat{CP}_t = 45.4845 + 0.5998 GGDP_t \quad (1.4)$$

$$(1.8) \quad (80.4)$$

$$R^2 = 0.9926, \quad DW = 1.63, \quad s.e. = 131.4$$

上模型中不存在自相关。消费函数的 GLS 估计结果是

$$\hat{CP}_t = 45.4845 + 0.5998GDP_t \quad (1.5)$$

消费函数的时间序列模型估计结果是

$$CP_t = 129.0977 + 0.6018 GDP_t + 0.7370\hat{u}_{t-1} + \hat{v}_t \quad (1.6)$$

$$(1.28) \quad (54.8) \quad (5.4)$$

$$R^2 = 0.9988, \quad DW = 1.7, \quad s.e. = 132.3$$

则长期关系是

$$CP_t = 129.0977 + 0.6018GDP_t \quad (1.7)$$

综上消费与国内生产总值的真实比值是 0.60。

下面研究消费（不变价格）对国内生产总值的弹性系数。对消费和国内生产总值取自然对数并进行 OLS 回归，得如下结果，

$$\widehat{\text{LnCP}}_t = 0.1932 + 0.9256 \text{LnGDP}_t \quad (1.8)$$

(3.0) (118.8)

$$R^2 = 0.9965, \quad DW = 0.77, \quad s.e. = 0.0584$$

对变量进行广义差分。定义

$$G\text{LnCP}_t = \text{LnCP}_t - 0.615 \text{LnCP}_{t-1} \quad (1.9)$$

$$G\text{LnGDP}_t = \text{LnGDP}_t - 0.615 \text{LnGDP}_{t-1} \quad (1.10)$$

得 GLS 估计结果如下：

$$G\widehat{\text{LnCP}}_t = 0.0814 + 0.9234 G\text{LnGDP}_t \quad (1.11)$$

(1.6) (57.6)

$$R^2 = 0.9857, \quad DW = 1.34, \quad s.e. = 0.047$$

对残差建立时间序列模型,

$$\text{LnCP}_t = 0.2103 + 0.9235 \text{LnGDP}_t + 0.6120 \hat{u}_{t-1} + \hat{v}_t \quad (1.12)$$

(1.6) (57.4) (5.2)

$$R^2 = 0.9977, \quad DW = 1.34, \quad s.e. = 0.0472$$

综上消费对国内生产总值的真实弹性是 0.92。

附表：1952-2003 年中国消费额、GDP 总值、消费增长率、GDP 增长率、消费比以及城农消费比数据。

obs	CONSP	GDPP	GROWTHCONSP	GROWTHGDP	RATIO	URBAN
1952	546.3000	692.2000	NA	NA	0.789223	2.4
1953	623.2173	806.8749	0.140797	0.165667	0.772384	2.6
1954	618.1604	830.0418	-0.008114	0.028712	0.744734	2.6
1955	675.7585	874.6596	0.093177	0.053754	0.772596	2.5
1956	722.8174	967.5612	0.069639	0.106215	0.747051	2.5
1957	752.4610	1015.601	0.041011	0.049651	0.740902	2.6
1958	783.8872	1187.139	0.041765	0.168903	0.660316	2.3
1959	748.5224	1322.374	-0.045115	0.113917	0.566044	3.2
1960	824.2267	1332.762	0.101138	0.007855	0.618435	3.1
1961	756.8176	969.8460	-0.081785	-0.272304	0.780348	2.8
1962	722.1577	861.8710	-0.045797	-0.111332	0.837895	2.6
1963	789.6848	1006.745	0.093507	0.168092	0.784394	2.5
1964	871.9268	1165.533	0.104145	0.157724	0.748093	2.5
1965	962.3438	1353.229	0.103698	0.161038	0.711147	2.4
1966	1042.439	1522.296	0.083229	0.124936	0.684781	2.3
1967	1070.745	1433.340	0.027154	-0.058436	0.747028	2.3
1968	1064.406	1433.103	-0.005920	-0.000165	0.742728	2.4
1969	1153.118	1575.803	0.083343	0.099574	0.731765	2.4
1970	1241.023	1876.370	0.076232	0.190739	0.661396	2.3
1971	1334.661	2049.667	0.075452	0.092358	0.651160	2.3
1972	1411.926	2107.026	0.057891	0.027985	0.670104	2.5
1973	1494.621	2278.330	0.058569	0.081301	0.656016	2.5
1974	1536.168	2324.880	0.027798	0.020431	0.660751	2.6
1975	1599.783	2500.794	0.041411	0.075666	0.639710	2.6
1976	1664.324	2508.359	0.040344	0.003025	0.663511	2.7
1977	1704.163	2621.917	0.023937	0.045272	0.649968	2.9
1978	1842.026	2966.196	0.080898	0.131308	0.621006	2.9
1979	2112.907	3286.162	0.147056	0.107871	0.642971	2.7
1980	2264.996	3463.821	0.071981	0.054063	0.653901	2.7
1981	2459.823	3643.461	0.086016	0.051862	0.675134	2.5
1982	2653.080	4003.213	0.078566	0.098739	0.662738	2.4
1983	2888.765	4365.876	0.088834	0.090593	0.661669	2.2

1984	3280.282	5006.124	0.135531	0.146648	0.655254	2.2
1985	3707.188	5645.932	0.130143	0.127805	0.656612	2.3
1986	3964.204	6140.092	0.069329	0.087525	0.645626	2.5
1987	4207.294	6653.794	0.061321	0.083663	0.632315	2.6
1988	4460.610	7007.277	0.060209	0.053125	0.636568	2.7
1989	4269.959	6660.271	-0.042741	-0.049521	0.641109	2.7
1990	4502.585	7257.492	0.054480	0.089669	0.620405	2.8
1991	5060.990	8192.661	0.124019	0.128856	0.617747	2.9
1992	5826.347	9446.424	0.151227	0.153035	0.616778	3.1
1993	6511.858	11131.80	0.117657	0.178414	0.584978	3.2
1994	7215.488	12489.92	0.108054	0.122004	0.577705	2.55
1995	7767.802	13512.65	0.076546	0.081885	0.574854	2.53
1996	8708.014	14874.10	0.121040	0.100754	0.585448	2.33
1997	9411.197	16173.81	0.080751	0.087381	0.581879	2.37
1998	10287.88	17702.95	0.093154	0.094544	0.581139	2.51
1999	11368.09	18901.54	0.104998	0.067705	0.601437	2.65
2000	12678.88	20743.52	0.115304	0.097451	0.611221	2.79
2001	13794.27	23075.66	0.087972	0.112427	0.597784	2.9

建模案例 3. 全国味精需求量的计量经济模型

1. 依据经济理论选择影响味精需求量变化的因素

依据经济理论一种商品的需求量主要取决于四个因素，即①商品价格，②代用品价格，③消费者收入水平，④消费者偏好。模型为：

$$\text{商品需求量} = f(\text{商品价格}, \text{代用品价格}, \text{收入水平}, \text{消费者偏好})$$

对于特定商品味精，当建立模型时要对上述四个因素能否作为重要解释变量逐一鉴别。

商品价格：味精是一种生活常用品，当时又是一种价格较高的调味品。初步判断价格会对需求量产生影响。所以确定价格作为一个重要解释变量。

代用品价格：味精是一种独特的调味品，目前尚没有替代商品。所以不考虑代用品价格这一因素。

消费者收入：显然消费者收入应该是一个较重要的解释变量。

偏好：由于因偏好不食味精或大量食用味精的情形很少见，所以每人用量只会小范围内波动，所以不把偏好作为重要解释变量，而归并入随机误差项。

分析结果，针对味精需求量只考虑两个重要解释变量，商品价格和消费者收入水平。

$$\text{味精需求量} = f(\text{商品价格}, \text{收入水平})$$

2. 选择恰当的变量（既要考虑代表性，也要考虑可能性）

用销售量代替需求量。因需求量不易度量，味精是自由销售商品，不存在囤积现象，所以销售量可较好地代表需求量。味精商品价格即销售价格。

用人均消费水平代替收入水平。因为①消费水平与味精销售量关系更密切。②消费水平数据在统计年鉴上便于查找（收入水平的资料不全）。

$$\text{味精销售量} = f(\text{销售价格}, \text{人均消费水平})$$

用平均价格作为销售价格的代表变量。不同地区和不同品牌的味精价格是不一样的，应取平均价格（加权平均最好）。

取不变价格的人均消费水平：消费水平都是用当年价格计算的，应用物价指数进行修正。

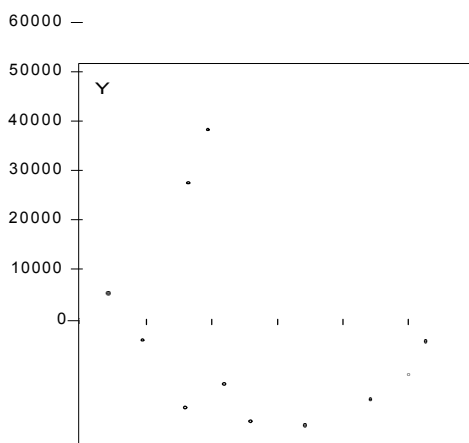
$$\text{味精销售量} = f(\text{平均销售价格}, \text{不变价格的消费水平})$$

3. 收集样本数据（抽样调查，引用数据）

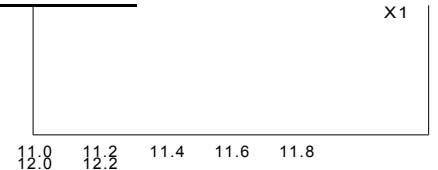
从中国统计年鉴和有关部门收集样本数据（1972-1982, $T=11$ ）。定义销售量为 y_t （吨），平均销售价格为 x_{1t} （元 / 公斤），不变价格的消费水平为 x_{2t} （元）。相关系数表如下：

	平均销售价格 (x_{1t})	不变价格的消费水平 (x_{2t})
味精销售量(y_t)	-0.3671	0.9771

注：临界值 $r_{0.05(9)} = 0.60$ 。



60000
50000
40000
30000
20000
10000
0

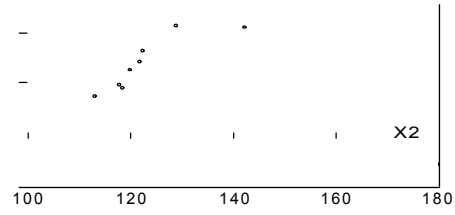


x_1

11.0 11.2 11.4 11.6 11.8

2002 15148.47 24275.01 0.098172 0.051975 0.624915 3.11

2002 15148.47 24275.01 0.098172 0.051975 0.624915 3.11



4. 确定模型形式并估计参数

$$\hat{y}_t = -144680.9 + 6313.4 x_{1t} + 690.4 x_{2t} \quad (1)$$

(-3.92) (2.17) (15.32) $R^2 = 0.97, DW = 1.8, t_{0.05(8)} = 2.3$

回归系数 6313.4 无显著性 (x_{1t} 与 x_{2t} 应该是负相关, 回归系数估计值却为正, 可见该估计值不可信)。剔除不显著变量 x_{1t} , 再次回归,

$$\hat{y}_t = -65373.6 + 642.4 x_{2t} \quad (2)$$

(-10.32) (13.8) $R^2 = 0.95, DW = 1.5, t_{0.05(9)} = 2.26$

问题: $b_1 = 6313.4$, 为什么检验结果是 $b_1 = 0$? 量纲的变化对回归结果会造成影响吗? 请同学们自己回答。

建模案例 4：用回归方法估计纯耕地面积

目前对土地的调查大多采用航空摄影，从照片上把各类资源图斑转绘到 1:10000 的地形图上，然后再从地形图上测绘图斑面积。

在处理如何获得实际耕地面积时，关键技术难题是如何将耕地图斑中包含的田埂、土坎、空隙地、宽度小于 2 米的路、沟、渠等面积从图斑中分离出来。因为它们在航空图片上的分辨率很低，无法直接勾绘，测算。

设一个毛耕地图斑面积用 S 表示，其中不能耕种的面积（扣除面积）用 ΔS 表示，则扣除系数， $y_i = \Delta S / S = (\text{扣除面积}) / (\text{毛耕地图斑面积})$ 。

对于每一个图斑，知道精确的扣除系数 y_i ，就很容易根据毛耕地图斑面积 S 计算出纯耕地面积。现在用回归分析方法，寻找影响扣除系数变化的主要因素，从而建立关于“扣除系数”的回归模型。

该论文研究的是湖南地区的耕地面积调查。湖南省属丘陵山区，地形复杂，各种地类犬牙交错，影响扣除系数的因素很多。如田埂宽度、地块大小、地块坡度、空隙地、地貌类型等。通过实际调查和分析，初步确定三个主要因素，即

“坡度”、“地块面积”和“田埂宽度”

论文作者在五个县共调查了 867 个样本点，其中水田样本 522 个，旱田样本 345 个。具体做法是首先把 867 个样本数据按“坡度”分成 25 个等级，然后再把属于同一个等级的样本数据用加权平均的方法求出另两个因素的观测值，“平均地块面积”和“平均田埂宽度”。整理样本数据如下：

i (序号)	y_i (扣除系数)	x_{1i} (坡度)	x_{2i} (平均地块面积)	x_{3i} (平均田埂宽度)
1	4.2356	0	1.9300	0.6318
2	4.8838	1	1.4918	0.7312
3	7.8300	2	1.1253	0.9731
...
25	39.4151	24	1.0600	4.0721

拟建模型为，

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i} + u_i$$

利用样本得估计的回归方程

$$y_i = 1.672 + 1.145 x_{1i} + 0.608 x_{2i} + 2.081 x_{3i}$$

$$(7.3) \quad (0.4) \quad (1.85) \quad F = 221.62$$

$$(F_{.05(3,21)} = 3.07, F_{.01(3,21)} = 4.87, t_{.05(21)} = 2.08, t_{.01(21)} = 2.84)$$

统计检验结果表明 x_{2i} , x_{3i} 为非重要解释变量。剔除之，用 y_i 对 x_{1i} 再次回归得，

$$y_i = 3.34 + 1.35 x_{1i}$$

实际的验证结果表明，用只考虑“地块坡度”计算出来的扣除系数估计“纯耕地面积”完全能满足精度要求，从而为减少野外作业强度（不必再测量“地块面积”和“田埂宽度”），迅速完成测算，提供了科学依据。

建模案例 5. 中国铅笔需求预测模型

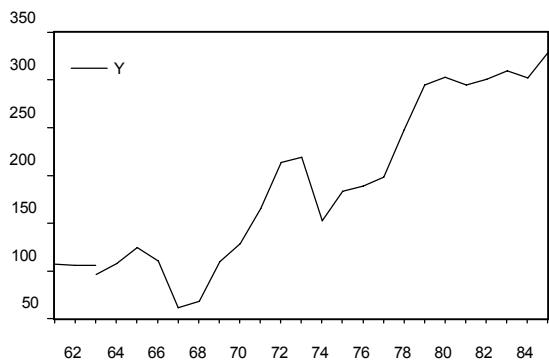
中国从上个世纪 30 年代开始生产铅笔。1985 年全国有 22 个厂家生产铅笔。产量居世界首位（33.9 亿支），占世界总产量的 1/3。改革开放以后，铅笔生产增长极为迅速。1979-1983 年平均年增长率为 8.5%。数据在 Eviews 文件 file:nonli6 中。铅笔销售量时间序列见图 4.21。1961-1964 年的销售量平稳状态是受到了经济收缩的影响。文革期间销售量出现两次下降，是受到了当时政治因素的影响。1969-1972 年的增长

是由于一度中断了的中小学教育逐步恢复的结果。1977-1978 年的增长是由于高考正式恢复的结果。1981 年中国开始生产自动铅笔，对传统铅笔市场冲击很大。1979-1985 年的缓慢增长是受到了自动铅笔上市的影响。

初始确定的影响铅笔销量的因素有全国人口、各类在校人数、设计人员数、居民消费水平、社会总产值、自动铅笔产量、价格因素、原材料供给量、政策因素等。经过多次筛选、组合和逐步回归分析，最后确定的被解释变量是 y_t （铅笔年销售量，千万支）；解释变量分别是 x_{t1} （自动铅笔年产量，百万支）； x_{t2} （全国人口数，百万人）； x_{t3} （居民年均消费水平，元）； x_{t4} （政策变量）。因政策因素影响铅笔销量出现大幅

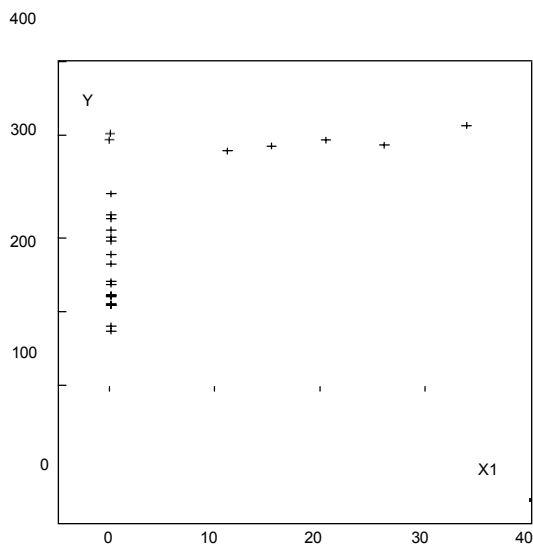
下降时，政策变量取负值。例如 1967、1968 年的 x_{t4} 取值-2，1966、1969-1971、1974-1977 年的 x_{t4} 取值-1）。

由图 4.22 知中国自生产自动铅笔起，自动铅笔产量与铅笔销量存在线性关系。由图 4.23 知全国人口与铅笔销量存在线性关系。说明人口越多，对铅笔的需求就越大。由图 4.24 知居民年均消费水平与铅笔销量存在近似对数的关系。散点图说明居民年均消费水平越高，则铅笔销量就越大。但这种增加随着居民消费水平的提高变得越来越缓慢。图 4.25 显示政策变量与铅笔销量也呈线性关系。

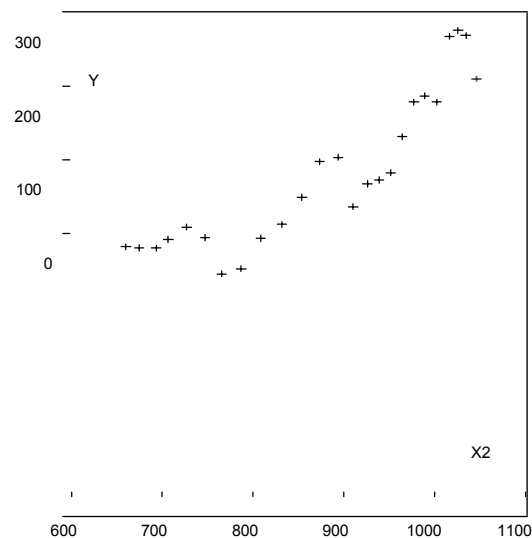


铅笔销售量时间序列（1961-1985）（文件名nonli6）

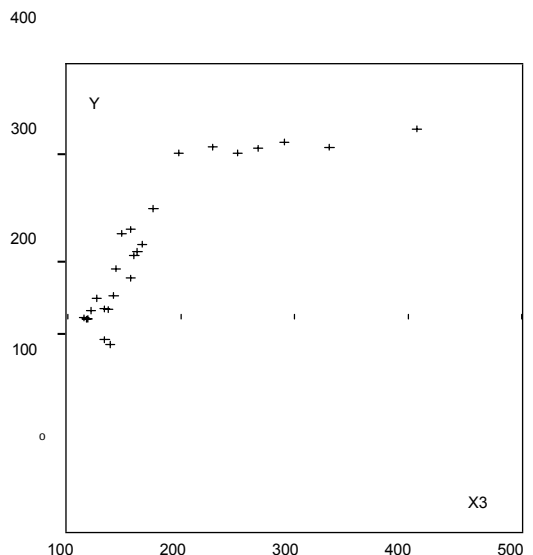
400



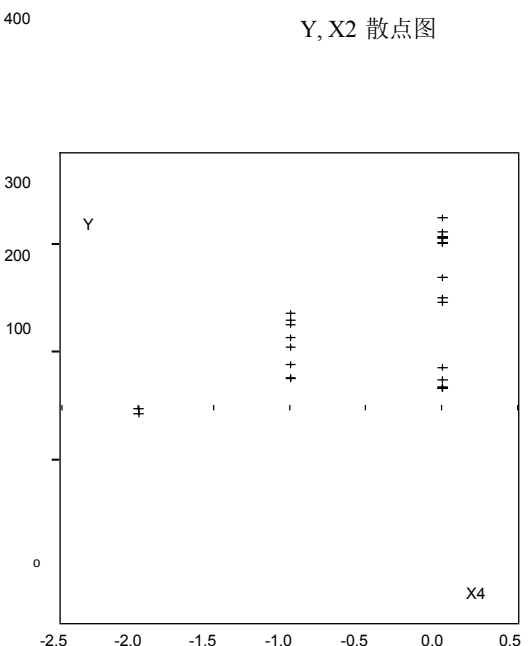
Y, X1 散点图



Y, X2 散点图



Y, X3 散点图



Y, X4 散点图

基于上述分析建立的模型形式是

$$y_t = b_0 + b_1 x_{t1} + b_2 x_{t2} + b_3 \ln(x_{t3}) + b_4 x_{t4} + u_t \quad (4.40)$$

y_t 与 x_{t3} 呈非线性关系。估计结果如下。

$$\hat{y}_t = -907.94 - 2.95 x_{t1} + 0.31 x_{t2} + 170.19 \ln x_{t3} + 45.51 x_{t4} \quad (4.41)$$

(-6.4) (-3.7) (4.8) (4.4) (12.6)

$$R^2 = 0.9885, DW = 2.09, F = 429, s.e. = 10.34$$

上式说明，在上述期间自动铅笔年产量每增加 1 百万支，平均使铅笔的年销售量减少 2950 万支。全国人口数每增加 1 百万人，平均使铅笔的年销售量增加 310 万支。对数的居民年均消费水平每增加 1 个单位，平均使铅笔的年销售量增加 17 亿支。一般性政策负面变动使铅笔的年销售量减少 4.551 亿支。当政策出现大的负面变动时，铅笔的年销量会减少 9.102 亿支。

当 y_t 对所有变量都进行线性回归时（见下式），显然估计结果不如(4.41)式好。

$$\hat{y}_t = -254.26 - 3.29 x_{t1} + 0.42 x_{t2} + 0.66 x_{t3} + 40.74 x_{t4} \quad (4.42)$$

(-12.0) (-3.0) (8.6) (3.5) (11.7)

$$R^2 = 0.9857, DW = 1.77, F = 346, s.e. = 11.5$$

建模案例 6 天津市粮食市场小麦批发价与面粉零售价的关系研究

首先看天津市粮食市场小麦批发价格的变化情况（图 1）。1995 年初，天津市粮食市场的小麦批发价格首先放开。在经历 5 个月的上扬之后，进入平稳波动期。从 1996 年 8 月份开始小麦批发价格一路走低。至 2002 年 12 月份，小麦批发价格降至是 1160 元/吨。

其次看面粉零售价的变化情况。因为面粉零售价格直接关系到居民的日常生活，所以开始时没有与小麦批发价格一起放开。当小麦批发价格一路看涨时，1995 年 1 月至 1996 年 6 月面粉零售价格一直处于 2.14 元/千克的水平上。1996 年 7 月起，面粉零售价格也开始在市场上放开。受小麦批发价格上涨的影响，一个月之内面粉零售价格从 2.14 元/千克涨到 2.74 元/千克。在这个价位上坚持了 11 个月之后，面粉零售价格开始下降。与小麦批发价格的下降相一致，在经历了 5 年零 7 个月的变化之后，面粉零售价格又恢复到接近开放前 2.14 元/千克的水平上（2.17 元）。

散点图如图 2。按时间分析这些观测点的变化情况（见图 3，逆时针方向运动）。见图 4，直接拟合这些数据效果将很差（ $R^2=0.027, r=0.17$ ）。

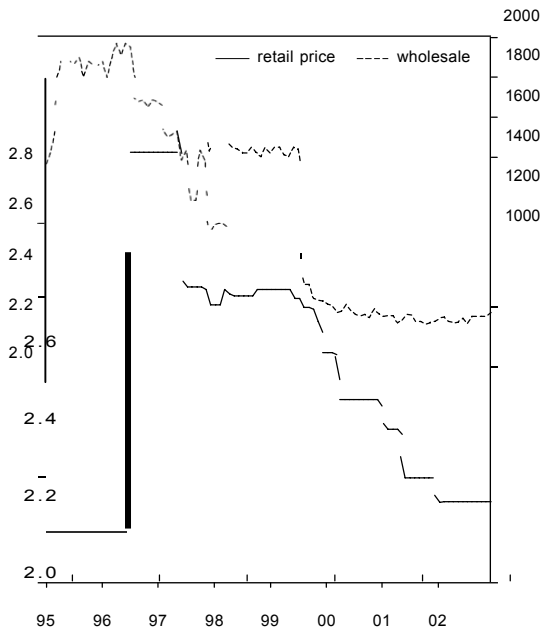


图 1

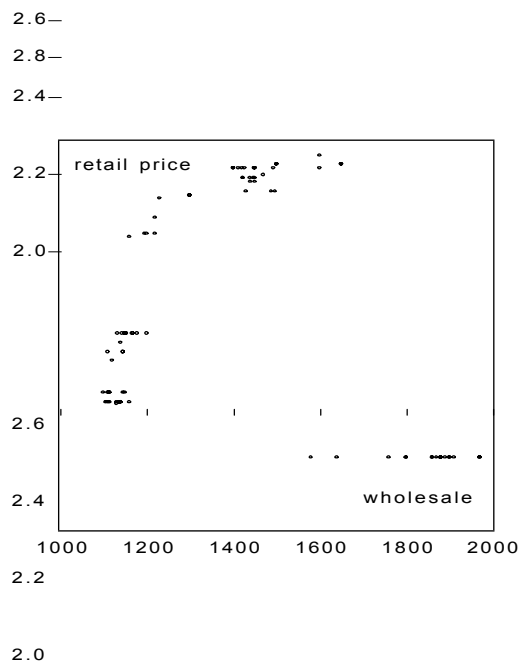


图 2

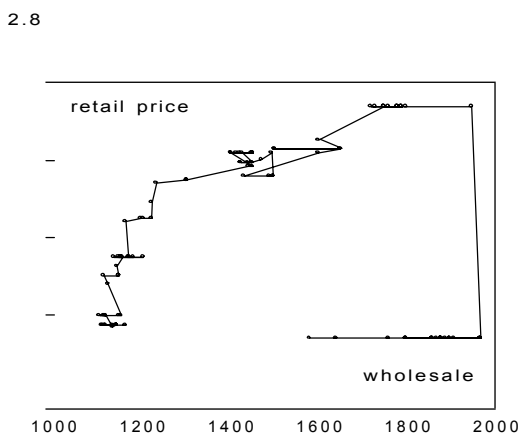


图 3



图 4

利用虚拟变量技术，在模型中加入虚拟变量。定义

$D = 0$, (1995: 1~1996:6, 面粉零售价格放开之前),

$D = 1$, (1996:7~2002:12, 面粉零售价格放开之后)。

取对数关系建立模型。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/658136142002007001>