

# 上海市居民银行存款的数学模型及其分析

迟洪钦 胡荷芬

**提 要** 根据多元统计分析的方法, 利用 1985 年至 1996 年间的上海市国内生产总值、总收入、实际利率、股票、债券等数据, 建立若干预测上海市居民银行存款额的数学模型, 以 1997 年的数据作检验, 分析各个数学模型的预测信度.

**关键词** 统计分析; 非线性; 预测

**中图法分类号** O 212.4

## 0 引 言

近 20 年来我国由传统的指令性计划经济向市场经济体制逐步转型. 市场经济的大潮席卷中国大地. 如今金融改革正在成为改革深入发展的重点之一. 改革的目标是建立与社会主义市场经济相适应的现代金融体制和良好的金融秩序. 银行要适应金融竞争, 提高效率, 就必须引入和应用数学. 我们选择了国内生产总值、总收入、实际利率、股票的流通市值、债券发行额这 5 个因素作为基本变量, 考察这些因素对上海市居民存款的影响程度, 用多元统计分析的方法建立了若干数学模型. 对模型的各项指标进行了统计检验, 发现非线性数学模型对预测后期存款增长程度有较好的作用.

## 1 线性回归

回归分析的目的之一就是根据自变量的值得出因变量的值的估计值, 之二就是要获得一个以回归线作为估计根据所含误差的测度, 之三就是获得两个变量之间的联系程度或相关程度的测度.

回归分析是处理变量间相关关系的一种很有效的统计方法. 下面给出一个因变量与  $m$  个自变量的线性回归模型的计算方法.

设因变量  $Y$  与自变量  $X_1, X_2, \dots, X_m$  线性相关,  $n$  次观测数据  $(y_i, x_{i1}, x_{i2}, \dots, x_{im}) (i=1, 2, \dots, n)$  满足以下多元线性回归模型:

$$y_i = b_0 + b_1x_{i1} + \dots + b_mx_{im} + \epsilon_i,$$

收稿日期: 1999-01-15

第一作者迟洪钦, 男, 副教授, 上海师范大学理工信息学院, 上海, 200234

$$\dots\dots\dots$$

$$y_n = b_0 + b_1 x_{n1} + \dots + b_m x_{nm} + \epsilon_n,$$

其中  $\epsilon_i (i=1, \dots, n)$  是观测误差, 一般假定  $\epsilon$  服从分布  $N(0, \sigma^2)$  且相互独立. 记

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1m} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \dots & x_{nm} \end{bmatrix}$$

$$b = \begin{bmatrix} b_0 \\ \vdots \\ b_m \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix},$$

则可以写成矩阵形式:

$$Y = XB + \epsilon,$$

其中  $\epsilon$  服从分布  $N_n(0, \sigma^2 I_n)$ .

为了确定系数  $B$ , 只需要求解矩阵方程  $XB = Y$  所对应的法方程  $X^T X B = X^T Y$ . 选择国内生产总值(GDP)、总收入、实际利率(名义利率与通货膨胀率之差, 取一年定期存款年利率作为名义利率)、股票的流通市值(加权)、债券发行额这 5 个因素作为基本变量<sup>[1,2]</sup>, 建立数学模型 1 如下

$$Y = 9642022 - 0.53950X_1 + 2.66579X_2 - 1.29342X_3 + 0.02865X_4 + 2.24067X_5,$$

其中变量  $X_1$  表示国内生产总值,  $X_2$  表示总收入,  $X_3$  表示实际利率,  $X_4$  表示股票的流通市值,  $X_5$  表示债券发行额,  $Y$  表示上海市年底存款额(加权: 除以 1.08, 扣除公款私存等因素), 各变量的单位为亿元. 数据及回归分析的详细结果见表 1. 从模型中各变量系数可见, 总收入、股票的流通市值、债券发行额与银行存款增加为正相关, 国内生产总值、实际利率与银行存款增加为负相关. 股票的流通市值、债券发行额与银行存款同步增长, 反映了证券市场吸纳的资金, 从总体上来讲, 仅是对存款增长部分的分流, 而不是对存款原有存量的分流.

## 2 非线性回归

在某些应用问题中, 曲线回归函数可能比线性函数更适当. 下面给出最简单的非线性回归数学模型的计算方法.

设曲线回归函数为

$$y = a + bx + cx^2$$

这是一个二次抛物线方程.

为了求解系数  $a, b, c$  的值, 只需要求解下列方程组:

$$y = na + b \sum x + c \sum x^2$$

$$xy = a \sum x + b \sum x^2 + c \sum x^3$$

$$x^2 y = a \sum x^2 + b \sum x^3 + c \sum x^4$$

类似地可以给出多元非线性回归函数的计算方法.

数学模型 2 对数学模型 1 进行修正, 在原有的 5 个基本变量基础之上, 引入  $X_1^2$  和  $X_4^2$  作

为修正变量,建立数学模型2如下:

$$Y = -109.55465 + 0.32400X_1 + 0.17915X_2 + 0.00094X_2^2 + 1.37830X_3 - 0.54917X_4 + 0.00044X_4^2 + 0.35861X_5,$$

其中变量 $X_1, X_2, X_3, X_4, X_5, Y$ 的单位为亿元.数据及回归分析的详细结果见表2.从模型2中各变量系数可见,只有股票的流通市值与银行存款增加为负相关,其余变量(包括两个修正变量)均与银行存款增加为正相关.

数学模型3在数学模型2基础之上,再增加 $X_2^2$ 作为修正变量,建立数学模型3如下

$$Y = -43.32512 + 0.12078X_1 + 0.51538X_2 + 0.00100X_2^2 + 1.14935X_3 - 0.40267X_4 + 0.00050X_4^2 + 1.33322X_5 - 0.00212X_5^2,$$

其中变量 $X_1, X_2, X_3, X_4, X_5, Y$ 的单位为亿元.数据及回归分析的详细结果见表3.从模型3中各变量系数可见,只有股票的流通市值和债券发行额的修正变量与银行存款增加为负相关,其余变量(包括另外两个修正变量)均与银行存款增加为正相关.

### 3 预测

预测从本质上来说是主观的,但预测的形成建立在对过去、现在及将来的各种信息综合分析基础之上,因而预测是对客观世界的反映.这里的所谓预测,采用的方法是将某年的国内生产总值(GDP)、总收入、实际利率(名义利率与通货膨胀率之差,取一年定期存款年利率作为名义利率)、股票的流通市值(加权)、债券发行额这5个基本变量的预计值代入相应的数学模型,计算得到该年的居民银行存款额的估计值.

将1997年数据代入上述3个数学模型进行检验,实际银行存款额为2367.44亿元(加权).用数学模型1算得银行存款估计额为1836.05亿元,绝对误差为-531.39亿元,相对误差为-22%.用数学模型2算得银行存款估计额为2428.09亿元,绝对误差为60.65亿元,相对误差为2%.用数学模型3算得银行存款估计额为2479.59亿元,绝对误差为112.15亿元,相对误差为5%.由此可见,数学模型2(7个变量的非线性回归分析)对于预测银行存款估计额效果最佳.在计算中发现,股票的流通市值对银行存款额影响很大,确切地讲,影响很敏感.即如果股票的流通市值有一个微小的变化,就会引起银行存款额有较大的变化.从数学模型2与数学模型1预测的结果对比可以看出,在线性数学模型中增加非线性项能大大改善预测效果.数学模型3没能进一步改善预测的结果,我们以为是变量 $X_2^2$ 与变量 $X_2$ 和 $X_4^2$ 产生共线性作用的结果.显然,债券发行额与总收入和股票的流通市值是相关的.以上的工作仅仅是初步的,无论是建立数学模型,还是对预测的结果进行分析研究,都还有许多工作可以做.

### 参 考 文 献

- 1 肖云茹. 概率统计计算方法. 天津: 南开大学出版社, 1994
- 2 潘德惠. 数学模型的统计方法. 沈阳: 辽宁科学技术出版社, 1986

表 1

国内生产总值	总收入	实际利率	股票	债券	银行存款	银行存款/1.08
466.75	70.66796	- 8.36	0	4.1653	62.79	58.13888889
490.83	86.28444	0.9	0	4.3362	80.69	74.71296296
545.46	97.222419	- 0.9	0	19.1599	106.79	98.87962963
648.3	118.39624	- 11.46	0	18.9722	125.74	116.4259259
696.54	144.66894	- 4.56	0	23.886	172.93	160.1203704
756.45	172.2308414	3.78	1.722	27.4642	223.05	206.5277778
893.77	196.7682751	- 2.94	4.1202	56.35	290.41	268.8981481
1114.32	239.946399	- 2.44	78.176	86.363	368	340.7407407
1511.61	407.4016094	- 9.22	308.868	46.564	519.53	481.0462963
1971.92	561.3024384	- 12.92	337.379	113.051	888.7	822.8713704
2462.57	688.4508024	- 7.72	410.9	253.082	1278.44	1183.740741
2902.2	787.2137669	- 1.73	986.125	415.1128	1718.56	1591.259259

SUMMARY OUTPUT

回归统计		Coefficients	标准误差	t Stat	
Multiple R	0.999892708	Intercept	96.42022269	27.46212858	3.511025098
R Square	0.999785427	X Variable 1	- 0.539502178	0.101219901	- 5.330001044
Adjusted R Square	0.999606617	X Variable 2	2.66578586	1.288518501	9.239566446
标准误差	9.822448898	X Variable 3	- 1.293421511	0.746969605	- 1.731558423
观测值	12	X Variable 4	0.028646493	0.034203445	0.837532373
		X Variable 5	2.240673077	0.13393083	16.73007683

方差分析

	df	SS	MS	F	Significance F
回归分析	5	2697261.59	539452.3179	5591.309175	6.48169E-11
残差	6	578.8830141	96.48050235		
总计	11	2697840.473			

RESIDUAL OUTPUT

观测值	预测 Y	残差	标准残差	预测	实际
1	53.13930923	4.999579654	0.689182802	1836.049871	
2	70.1848362	4.528826761	0.624290387	2367.44	
3	105.4126662	- 6.533036585	- 0.900567002	绝对误差	- 531.3901286
4	119.6130917	- 3.187165789	- 0.439344905	相对误差	- 0.224457696
5	165.7105098	- 5.590139403	- 0.77059037		
6	204.1428317	2.384946066	0.328760401		
7	268.9540634	- 0.055915238	- 0.007707812		
8	333.7945395	6.946201253	0.957520986		
9	492.0568183	- 11.01052199	- 1.517780077		
10	808.5632552	14.30711512	1.97220934		
11	1191.950841	- 8.210100607	- 1.131747173		
12	1589.839049	1.420210751	0.195773423		

表 2

SUMMARY OUTPUT					
回归统计		Coefficients	标准误差	t Stat	
Multiple R	0.999938327	Intercept	-109.5546513	128.8093652	-0.850517749
R Square	0.999876659	X Variable 1	0.324000639	0.555615416	0.583138318
Adjusted R Square	0.999660811	X Variable 2	0.17915458	1.647129322	0.108767768
标准误差	9.120791804	X Variable 3	0.00094348	0.000621686	1.51761401
观测值	12	X Variable 4	1.37829692	1.928261059	0.714787509
		X Variable 5	-0.54917158	0.338964196	-1.620146279
		X Variable 6	0.000441761	0.000257154	1.717883755
		X Variable 7	0.358607111	1.176626548	0.304775641

方差分析

	df	SS	MS	F	Significance F
回归分析	7	2697507.717	385358.2453	4632.330861	1.19779E-7
残差	4	332.7553725	83.18884314		
总计	11	2697840.473			

RESIDUAL OUTPUT

观测值	预测 Y	残差	标准残差	预测	实际
1	49.01597961	9.122909283	1.658697388	2428.086403	2367.44
2	74.75250441	-0.039541446	-0.007189296	60.64640252	60.64640252
3	99.14094492	-0.261315287	-0.047511487	0.02561687	0.02561687
4	125.9398604	-9.513934426	-1.729792295		
5	164.0696938	-3.949323454	-0.718053013		
6	208.4929049	-1.965127114	-0.357292955		
7	265.7088697	3.189278416	0.579864122		
8	336.1685143	4.572226454	0.83130719		
9	486.3034025	-5.257106198	-0.955829774		
10	814.8993263	7.971044093	1.449269044		
11	1187.882982	-4.14224089	-0.753128627		
12	1590.986129	0.273130569	0.049659703		

表 3

SUMMARY OUTPUT					
回归统计		Coefficients	标准误差	t Stat	
Multiple R	0.999974012	Intercept	-43.32511616	101.918156	-0.425097136
R Square	0.999948024	X Variable 1	0.120780486	0.428345068	0.281970063
Adjusted R Square	0.999809421	X Variable 2	0.515375771	1.245717327	0.413718072
标准误差	6.836750197	X Variable 3	0.000979896	0.000466348	2.101211749
观测值	12	X Variable 4	1.149346818	1.449778664	0.792773991
		X Variable 5	-0.402665676	0.264135628	-1.524465589
		X Variable 6	0.000496371	0.000194626	2.550379573
		X Variable 7	1.333224025	1.004231572	1.327606164
		X Variable 8	-0.002124125	0.001046594	-2.02955886

方差分析

	df	SS	MS	F	Significance F
回归分析	8	2697700.249	337212.5312	7214.467501	2.45887E-06
残差	3	104.2234598	46.74116326		
总计	11	2697840.473			

RESIDUAL OUTPUT

观测值	预测 Y	残差	标准残差	预测	实际
1	50.27117634	7.868812546	2.203608747	2479.591668	
2	74.49740686	0.215556105	0.060373497	2367.44	
3	105.6543131	-6.774683436	-1.897470401	112.1516678	
4	121.0893927	-4.663466823	-1.306155534	0.047372549	
5	161.2629956	-1.142625213	-0.320029348		
6	204.5363404	1.991437418	0.557766808		
7	267.3264992	1.571648991	0.440191408		
8	339.3910073	1.349733442	0.378036742		
9	481.7127513	-0.666455013	-0.186662399		
10	822.2259983	0.644372012	0.180477361		
11	1184.194131	-0.453390478	-0.126986747		
12	1591.199099	0.060160359	0.016849865		

# Mathematical Models and Analysis of Resident Deposit in Shanghai

*Chi Hongqin Hu Hefen*

(College of Science, Engineering and Information)

**Abstract** Using multiple statistical analysis method this paper gives several mathematical models about Shanghai resident deposit based on the Shanghai statistical data from 1985 to 1996. These mathematical models are examined with Shanghai statistical data in 1997, and analyzed for predication credit.

**Key words** statistical analysis; nonlinear; predication