

[文章编号] 1007-7405(2006)04-0375-06

前移回归分析新方法及其在福建 人均 GDP 预测中的应用

沈 军¹, 跃潮²

(1. 集美大学社会科学系, 福建 厦门 361021; 2. 集美大学计算机工程学院, 福建 厦门 361021)

[摘要] 介绍一种称为前移回归的时间序列预测新方法, 其建立依据是, 经济指标是多项相关因素的函数, 一个时间段内各项指标 (自变量) 的值会影响下一时间段待预测指标 (因变量) 的取值. 这种方法克服了以往时间序列预测只是自身拓展而不考虑多项因素 (变量) 的不足, 也弥补了回归分析预测法必须已知同时期各个自变量值才能预测的缺陷. 以福建省人均 GDP 的预测作为实例, 取得较好的效果.

[关键词] 经济预测; 时间序列; 前移回归分析; 人均 GDP

[中图分类号] O212.4

[文献标识码] A

0 引言

时间序列数据的形成往往是多项影响因素综合作用的结果, 时间序列的未来取值是可以预测的, 经济事物发展时产生的信息就是时间序列数据. 《福建省统计年鉴 (2005)》^[1]除了收录了全省 2004 年经济、社会各方面的统计数据外, 还汇总了历史重要年份和近二十多年的主要统计数据. 2005 年的经济普查, 使部分数据发生改动^[2]. 《2005 年福建省国民经济和社会发展统计公报》^[3]对 2005 年的经济情况作了全面阐述. 总结这一时期的经济发展数量规律, 对于今后建设海峡西岸经济带具有重要意义. 常用的定量预测方法主要有时间序列平滑预测法和回归分析预测法, 都存在一些不足, 本文采用新提出的前移回归分析预测法, 分析了资料较齐全的 1983 年至 2005 年的 11 项经济指标, 较好地进行了人均 GDP 的预测.

1 常用定量预测方法及其缺陷

预测, 是指以准确的调查统计资料和信息为依据, 从事物的历史、现状和规律性出发, 运用科学的方法, 对事物未来发展前景的测定. 经济预测是经济决策科学化的工具, 是政府编制计划、预见计划执行情况、加强计划指导的依据, 也是企业改善经营管理的有效手段之一.

预测有定性和定量两种, 定量预测的方法主要有时间序列平滑预测法和回归分析预测法两种, 都存在一些不足.

1.1 时间序列平滑预测法及其缺陷

时间序列数据是随时间变化的, 是时间的函数. 时间序列分析, 就是找出指标随时间的变化规律.

对于单一指标本身的预测, 一般是采用时间序列平滑预测法, 包括移动平均法、指数平滑法、差

[收稿日期] 2006-06-19

[基金项目] 福建省教育厅科技项目 (JA004235)

[作者简介] 沈军 (1959-) 女, 讲师, 从事政治经济学、经济分析与预测、军事理论研究.

分指数平滑法、自适应过滤法、直线模型预测法、多项式模型预测法、指数曲线模型预测法、修正指数曲线模型预测法、成长曲线预测模型和季节变动预测法等^[4~9]。这些方法是指标值的外推,并没有考虑影响因素,只是曲线的简单数学处理,虽然有参考意义,但不能作为决策的可靠依据。

1.2 回归分析预测法及其缺陷

时间序列数据又是随各项影响因素变化的,是多个影响因子(自变量)的函数,因此人们采用回归分析法^[4,6]。相关变量之间的关系可以是线性的,也可以是非线性的。这里只讨论线性回归分析。

当只有一个自变量时,就称为一元线性回归分析,其模型为:

$$Y_i = a + bx_i + \epsilon_i, i = 1, 2, \dots$$

其中, a 和 b 称为回归系数,代表各种随机因素对因变量的影响总和。

有多个自变量时,就是多元线性回归分析。设 x_1, x_2, \dots, x_p 是 p 个可以精确测量或可控制的变量。如果变量 y 与 x_1, x_2, \dots, x_p 之间的内在联系是线性的,那么进行 n 次试验,则可得 n 组数据: $(y_i, x_{i1}, x_{i2}, \dots, x_{ip}), i = 1, 2, \dots, n$ 。

多元线性回归模型可表示为:

$$\begin{aligned} Y_1 &= b_0 + b_1 x_{11} + b_2 x_{12} + \dots + b_p x_{1p} + \epsilon_1 \\ Y_2 &= b_0 + b_1 x_{21} + b_2 x_{22} + \dots + b_p x_{2p} + \epsilon_2 \\ &\dots \\ Y_n &= b_0 + b_1 x_{n1} + b_2 x_{n2} + \dots + b_p x_{np} + \epsilon_n \end{aligned}$$

其中, $b_0, b_1, b_2, \dots, b_p$ 是 $p+1$ 个待估参数,即回归系数, ϵ_i 表示第 i 次试验中的随机因素对 y 的影响,通常忽略。

通过解线性方程,求得参数 $b_0, b_1, b_2, \dots, b_p$ 就得到了 p 元线性回归方程:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p$$

建立回归方程的目的是要利用它来进行预报与控制。在实际问题中,事先并不能断定随机变量 Y 与 x_1, x_2, \dots, x_p 之间确有线性关系,在求解回归方程前,线性回归模型只是一种假设,所以在求出线性回归方程之后,还需对其进行统计检验,给以肯定或否定的结论。

回归分析预测法不考虑时间顺序,只是按照影响因子的线性表达式求得因变量的值。某个因变量(如某个经济指标)的值,只有在同时期的其他影响变量值确切可知的情况下才能求得。因此,从严格意义上讲,这不是预测,真正的预测是时间尚未到达时某项指标值的预估。

2 一种新型预测方法的原理

经济指标预测至今没有找到理想的定量方法,这就需要综合上述两种预测方法的优势,克服各自的不足,既考虑时间上的外推又考虑各项因素的影响。

作为讨论的实例,现将《福建省统计年鉴 2005》、2005 年经济普查公报和《2005 年福建省国民经济和社会发展统计公报》中综合出来的各项指标和人均 GDP 列于表 1。

一般来说,可以将人均国内生产总值(人均 GDP)当作因变量,其余指标当作自变量,求出回归方程,用来“预测”知道其它指标情况下的人均 GDP。这只是一般的思维。如果 2006 年尚未结束,2006 年各项指标未知,怎样预测 2006 年的人均 GDP 呢?

事物的发展是有一定的前兆或基础的。某年的国民经济各项指标值,反映了这一年为下一年各项事业打下的基础,是下一年某经济指标(如 GDP)的自变量。

基于“某一时间段的指标值是下一时间段某项待预测指标形成的基础”的观点,可以认为,后一时间段的因变量是前一时间段各自变量的函数,将回归模型写成:

$$Y_2 = b_0 + b_1 x_{11} + b_2 x_{12} + \dots + b_p x_{1p} + \epsilon_1,$$

$$Y_3 = b_0 + b_1 x_{21} + b_2 x_{22} + \dots + b_p x_{2p} + \epsilon_2,$$

...

$$Y_n = b_0 + b_1 x_{n1} + b_2 x_{n2} + \dots + b_p x_{np} + \epsilon_n.$$

其中, $m = n - 1$, 其余参数与前述回归模型相同. 整个模型比前述模型少一行, 实际数据处理中即是少一个样本, 这不影响方程的求解.

表 1 福建省 1983~2005 年 11 项经济指标

Tab 1 Economic indexes of Fujian province during 1983 to 2005

年份	农林牧渔 总产值 亿元	工业 增加值 /亿元	消费品 零售总额 亿元	消费品 价格 指数 ¹⁾	货物 周转量 /亿吨·km	旅客 周转量 /亿人·km	港口 吞吐量 万吨	固定资 产投资 /亿元	农村年 人均收入 /元	人口自然 增长率 /‰	人均 GDP 元
1983	68 08	37 76	62 59	100 000 0	131.78	91 50	869 55	26 97	301 84	18 22	487
1984	80 66	47 47	74 50	102 147 8	151.61	109 50	943 46	34 61	344 94	19 43	591
1985	99 05	62 09	96 04	113 659 8	161.97	130 33	1 114 09	55 62	396 45	17 70	737
1986	107 07	67 06	109 07	121 048 1	195.48	137 09	1 159 90	64 46	418 51	18 17	809
1987	132 97	82 69	126 06	132 474 2	225.02	159 38	1 303 36	81 60	484 88	19 21	999
1988	182 00	120 45	173 74	167 525 8	242.02	175 91	1 396 79	100 29	613 41	18 53	1 349
1989	209 92	142 45	202 30	199 226 8	270.06	173 66	1 614 99	101 65	697 34	18 57	1 589
1990	227 12	150 55	207 74	197 852 2	272.71	175 40	1 496 50	115 41	764 41	17 73	1 763
1991	253 51	188 29	230 99	204 725 1	267.26	186 70	1 706 38	145 62	850 05	13 77	2 041
1992	295 24	241 78	289 38	216 838 5	347.28	205 17	1 862 12	227 55	984 11	12 16	2 557
1993	386 34	387 92	388 32	250 257 7	434.02	232 27	2 679 09	368 45	1 210 51	11 10	3 633
1994	574 05	629 88	523 97	313 488 0	577.73	240 56	3 002 33	538 86	1 577 74	10 29	5 355
1995	738 63	764 64	659 37	361 168 4	608.61	247 65	3 460 80	681 17	2 048 59	9 30	6 787
1996	850 67	895 90	832 48	382 474 2	590.58	267 20	3 959 00	790 00	2 492 49	7 28	7 990
1997	925 56	1 066 29	987 69	389 003 4	605.78	253 15	4 485 00	898 47	2 785 67	6 32	9 179
1998	973 37	1 165 13	1 132 09	387 800 7	661.61	279 76	4 518 00	1 048 52	2 946 37	5 33	10 066
1999	1 010 82	1 268 95	1 246 25	384 364 3	746.71	301 58	5 285 00	1 040 00	3 091 39	5 21	10 797
2000	1 037 27	1 470 07	1 372 79	392 439 9	687.65	333 97	6 944 17	1 082 47	3 230 49	5 75	11 601
2001	1 061 61	1 645 34	1 499 46	387 371 1	779.92	372 72	8 278 42	1 134 48	3 380 72	6 04	12 362
2002	1 089 79	1 882 55	1 663 31	385 395 2	827.44	392 00	10 200 62	1 230 76	3 538 74	5 78	13 497
2003	1 150 79	2 147 00	1 740 45	388 488 0	1 223.82	386 19	12 495 48	1 507 87	3 733 93	5 85	14 979
2004	1 317 28	2 465 27	1 995 82	404 037 8	1 377.70	440 93	15 800 00	1 899 10	4 089 00	5 96	17 218
2005	1 390 27	2 869 57	2 345 82	412 972 5	1 577.10	481 14	19 600 00	2 344 73	4 450 36	5 98	18 621

说明: 1) 以 1983 年为 100

实际工作中可将因变量的值上移一条记录, 最后一行记录就缺失因变量值, 需将最后一行记录不列入计算. 为便于计算, 对于表 1, 就是将人均 GDP 上移一格到上一行, 1984 年的人均 GDP 移到 1983 年, 1985 年的人均 GDP 移到 1984 年, 以此类推, 2005 年的移到 2004 年, 然后删除 2005 年的那一行. 栏目“人均 GDP”改为“来年年均 GDP”. 经过这样处理后, 对此数据表进行前述回归分析, 用普通回归分析软件, 求解出回归系数. 得到的方程将 2005 年的各项指标代入, 就可预测出 2006 年的人均 GDP.

如果不对原始数据表进行修改, 则需要在程序中进行此项引变量值的“滑移”工作, 编写新的回归分析程序, 效果是一样的.

由于新方法是在原有回归分析方法上向前一时期记录滑移因变量值进行分析预测的, 故这种方法

可称为前移回归分析预测法, 简称前移回归法. 与之对应, 向后滑移的方法可称为后移回归分析预测法, 简称后移回归法. 这两种方法统称滑移回归法. 这里的“前”是指先前的时间段, 所产生的样本靠前; 这里的“后”是指相对的下一时间段, 所产生的样本较新而后出现. 如果某时间之前的某项指标缺失或无法取得, 可用后移回归法. 在地质勘探等领域钻井时, 要预测待掘进的更老层位的某项岩石化学、生物或油气指标, 后移回归法当属最好的方法. 由于这种方法考虑到了多种因素的综合作用, 可冠之以“多因素”, 称为多因素滑移回归预测法.

3 新型预测方法应用实例和效果

对表 1 的数据进行多元逐步回归分析, 来年年人均 GDP 作为因变量, 其余指标按顺序作为自变量 X_1, X_2, \dots, X_n . 进行各个变量的显著性检验, 引入变量的 F 值取 4.747, 剔除变量的 F 值取 4.667. 经 F 检验只有工业增加值、货物运输周转量和人口自然增长率 4 个指标效果显著, 其余指标效果不显著. 获得回归系数如表 2.

表 2 回归方程系数
Tab 2 Coefficient of regression equation

系数	b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
值	2 830 038	0	5 053	0	4 399	2 086	0	0	0	0	-168 274

由此得到回归方程:

$$\text{下年人均 GDP} = 2\,830.038 + 5.053 \times \text{上年工业增加值} + 4.399 \times \text{上年消费品价格指数} + 2.086 \times \text{上年货物周转量} - 168.274 \times \text{上年人口自然增长率}.$$

由这个方程计算 1984 年至 2005 年的人均 GDP 列于表 3 比较实际值与预测值, 发现有 3 年的误差超过 300 元, 其余误差都小于 300 元, 误差最大的也只有 426 元, 最小的为 2 元多, 在 100 元以内的有 11 年, 预测效果较好.

2006 年尚未结束, 各项指标无法得到, 但是根据方程预测的人均 GDP 为 21 432 元, 这有待于 2007 年初统计数据出炉后进行检验.

表 3 各年人均 GDP 预测值与实际值的比较
Tab 3 Comparison of predictive value and practical value

年份	实际值	预测值	误差值	年份	实际值	预测值	误差值
1984	591	669.729	78.729	1995	6 787	6 865.946	78.946
1985	737	566.007	-170.993	1996	7 990	7 987.715	-2.285
1986	809	1 003.259	194.259	1997	9 179	9 047.056	-131.944
1987	999	1 051.700	52.700	1998	10 066	10 130.100	64.099
1988	1 349	1 067.574	-281.426	1999	10 797	10 907.370	110.369
1989	1 589	1 562.477	-26.523	2000	11 601	11 614.650	13.649
1990	1 763	1 864.874	101.874	2001	12 362	12 452.440	90.444
1991	2 041	2 046.640	5.640	2002	13 497	13 459.580	-37.418
1992	2 557	2 922.588	365.588	2003	14 979	14 792.530	-186.474
1993	3 633	3 684.058	51.058	2004	17 218	16 957.740	-260.259
1994	5 355	4 928.925	-426.075	2005	18 621	18 937.060	316.058

元

事物的发展是多项因素综合作用的结果, 上述自变量和因变量共 11 项指标, 实际上可以将其中任何一项作为因变量, 其余作为自变量, 预测出这项指标 2006 年的取值. 若全部都轮流做一遍, 则可以求得全部指标的预测值, 见表 4. 不过, 如果以 2006 年的预测值来预测 2007 年的指标值, 效果会大打折扣. 预测 2007 年的指标, 应当用 2006 年的实际值.

表 4 的预测数据, 将经受 2007 年初发表的福建省经济和社会发展统计公报验证. 前些年采用上述方法, 预测效果较好. 但是, 2005 年经济普查, 修正了 2004 年数据, 而此前一些年的数据没有修正, 造成 2004 年在一些指标上出现突变, 可能影响回归方程的有效性和预测的准确度.

表 4 2006 年福建省部分经济指标预测

Tab 4 Prediction of Fujian economic indexes in 2006

指标	农林牧渔 总产值 /亿元	工业 增加值 /亿元	消费品 零售总额 /亿元	消费品 价格 指数 ¹⁾	货物 周转量 /亿吨·km	旅客 周转量 /亿人·km	港口 吞吐量 /万吨	固定资 产投资 /亿元	农村年 人均收入 /元	人口自然 增长率 /%	人均 GDP /元
预测值	1 535	3 471	2 597	397	1 927	540	22 770	2 645	4 869	4.7	21 432

说明: 1) 以 1983 年为 100 若以 2005 年为 100 则为 96.5

4 人均 GDP 预测实例给予的启示

数学是严密的科学, 可以用表 1 的数据和上述方程验证表 3 的结果. 仅用上年工业增加值、消费品价格指数、货物周转量和人口自然增长率 4 项参数就能计算出下年度的人均 GDP. 并且, 近 20 年都是这个规律, 这里面可能反映了福建省经济的内在规律, 给予如下启示:

- 1) 工业增加值、消费品价格指数、货物周转量和人口自然增长率 4 项指标是影响人均 GDP 直接的、重要的因素.
- 2) 工业增加值起重要作用, 表明福建省的发展仍以工业为主导.
- 3) 消费品价格指数影响着人均 GDP 的增长. 经济的发展, 在一定程度上容易出现物价上涨、货币贬值. 抑制物价上涨, 是一项基本政策, 要坚决贯彻执行.
- 4) 货物周转量反映了交通运输事业的发展, 交通运输是经济发展的动脉, 生产的发展必须有便利的运输条件. 福建省的交通在国内相对落后, 制约着经济的持续发展和经济腹地的形成, 政府应该加大发展交通事业的力度. 交通的发展将带动旅游、激发人气、加快物流.
- 5) 人口自然增长率在方程中得到体现, 影响人均 GDP 数值, 呈现反消长的关系. 人口自然增长率越高, 人均 GDP 就将越小. 这从道理上是显而易见的, 也说明搞好计划生育工作的重要意义.
- 6) 选入回归方程中的指标只有上述 4 项, 并不说明其它指标不影响人均 GDP. 上述 4 项指标不是孤立的. 工业增加值的取得, 间接反映了福建省在国民经济各方面的投入; 货物周转量的增加, 是交通相关行业发展的体现, 也由于工业发展拉动原材料和产品的运输; 人口自然增长率的控制也是全体人民共同努力的结果. 因此, 上述 4 项指标隐含了国民经济多个部门的综合作用.

5 结论

前移回归法是既反映时序又考虑各项影响因素的预测方法, 克服了单纯时间序列平滑分析和普通回归分析的缺陷. 新的预测方法可用于具有多元时间序列数据的各个领域, 如经济、农业、生物、地质、环境和气象等部门. 新方法在对福建省人均 GDP 预测的应用实例中得到了较满意的效果, 也给予人们若干启示. 在农业发展潜力受限的情况下, 以工业为主导, 强化交通运输, 抑制物价上涨, 控制人口增长, 是保持经济持续稳定协调发展的必要途径.

[参考文献]

- [1] 福建省统计局. 福建省统计年鉴 (2005) [M]. 北京: 中国统计出版社, 2005
- [2] 福建省统计局. 福建省第一次全国经济普查主要数据公报 [EB/OL]. [2005-12-22]. <http://www.stats.gov.cn/tjsj/0200601050044.htm>
- [3] 福建省统计局. 福建省 2005 年国民经济和社会发展统计公报 [EB/OL]. [2006-03-13]. <http://www.stats.gov.cn/xwfb/0200603130104.htm>
- [4] 彭念一, 胡宗义. 实用经济预测与决策 [M]. 长沙: 中南工业大学出版社, 1997: 11-76, 77-104
- [5] 钱钟威, 李生达. 经济预测的数量方法 [M]. 重庆: 重庆出版社, 1994: 135-192
- [6] James M Lattin, J Douglas Carroll, Paul E Green. Analyzing Multivariate Data [M]. 北京: 机械工业出版社, 2003: 38-78

A New Method of Forward Regression Analysis and Its Application in Prediction of Per Capita GDP in Fujian Province

SHEN Jun, DING Yuecha^①

(1. Department of Social Science, Jimei University, Xiamen 361021, China; 2. School of Computer Engineering, Jimei University, Xiamen 361021, China)

Abstract: This paper introduces the principle and application of a new forecasting method called forward regression analysis. Taking the prediction of per capita GDP in Fujian Province as an example, it found to be efficient. The method improved previous time series predicting method in which only self extension is done and multiple factors (variables) are not taken into consideration. Also, it got over the weakness of predicting by general regression analysis which relies on the simultaneous independent variables. An economy indicator is the function of the related economy factors. The indicators (independent variables) in a period may affect the indicator (dependent variable) to be predicted in the next period, which is the basis of the new method.

Key words: economy forecast; time series; forward regression analysis; per capita GDP

(责任编辑 朱雪莲)