

文章编号:1671-9352(2009)02-0056-04

# 组合预测模型在中国 GDP 预测中的应用

王莎莎<sup>1</sup>, 陈安<sup>2</sup>, 苏静<sup>1</sup>, 李硕<sup>1</sup>

(1. 山东大学数学学院, 山东 济南 250100; 2. 中国科学院科技政策与管理科学研究所, 北京 100080)

**摘要:**在 ARIMA、混合时间序列和 GM(1,1)模型基础上,利用中国经济发展数据建立一个组合预测模型,并把它应用于我国 GDP 的预测。所得结果误差优于三个模型的分别预测,表明组合预测模型在时间序列数据的预测中更有优势。

**关键词:** ARIMA 模型;组合预测模型;时间序列;GDP

**中图分类号:** F015      **文献标志码:** A

## Application of the combination prediction model in forecasting the GDP of China

WANG Sha-sha<sup>1</sup>, CHEN An<sup>2</sup>, SU Jing<sup>1</sup>, LI Shuo<sup>1</sup>

(1. School of Mathematics, Shandong University, Jinan 250100, Shandong, China;

2. Institute of Policy and Management, Chinese Academy of Sciences, Beijing 100080, China)

**Abstract:** On basis of the ARIMA model, mixed-time series model and GM(1,1) model, a combination forecast model was established by using the Chinese economic development data, and the forecasted GDP of China was applied. The resulted show that the error of this combination prediction model is smaller than the other three models, and denoted that the combination prediction model in forecasting the time-series data is more advantageous.

**Key words:** ARIMA model; combination prediction model; time series; GDP

国内生产总值(gross domestic product, GDP)是指在一定时期内(一个季度或一年),一个国家或地区的经济中所生产出的全部最终产品和劳务的价值,常被公认为衡量国家经济状况的最佳指标。它不但可反映一个国家的经济表现,更可以反映一国的国力与财富。国内生产总值能够提供经济状况的较完整的图像,通过它可以判断经济是在萎缩还是在膨胀,是需要刺激还是需要抑制,是处于严重衰退还是通胀威胁之中。自从 1985 年以来,国内生产总值的核算已经成为我国经济管理部门了解经济运行状况的重要手段和制定经济发展战略、规划、年度计划以及各种宏观经济政策的重要依据。因此,研究和建立国内生产总值模型具有重要的现实意义。时间序列 Box—Jenkins 法在预测中应用非常广泛,龚国勇<sup>[1]</sup>就是利用 ARIMA 模型对深圳 GDP 进行了预测,并且得到了较好的预测结果。丁文斌<sup>[2]</sup>对回归模型和时间序列模型进行了对比,回归模型预测在实际中应用较多,但建模过程中的一些要求和假设条件使人们在实际的操作中受到很大的限制。戴羽等<sup>[3]</sup>又利用另一种方法 GM(1,1)模型对安徽省 GDP 总量进行了预测,具有较高的精度。本文运用 EVIEWS 5.0<sup>[4]</sup>以及 MATLAB 7.0 软件,采用了中国 1980~2006 年的实际 GDP 数据,首先利用 1980~2002 年的数据建立时间序列 ARIMA( $p, d, q$ )模型、混合时间序列模型和灰色预测模型,并对 2003~2006 年的 GDP 进行预测,在此基础上进行组合预测<sup>[5]</sup>,这样可以进一步缩小预测的误差。通过比较,说明组合预测模型的应用方法及其优势,并且得到了更好的预测结果。

收稿日期:2008-07-18

基金项目:中科院知识创新工程重要方向资助项目(KZCXZ-YW-305-4);山东省自然科学基金资助项目(Y2007G08)

作者简介:王莎莎(1984-),女,硕士研究生,研究方向为运筹学与经济分析. Email:sha-19842002@163.com

## 1 预测方法的介绍

### 1.1 ARIMA 模型

ARIMA 模型的形式:考虑序列  $y_t$ ,若其能通过  $d$  次差分后变为平稳序列,即  $y_t \sim I(d)$ ,则

$$u_t = \Delta^d y_t = (1 - B)^d y_t, \quad (1.1)$$

$u_t$  为平稳序列,即  $u_t \sim I(0)$ ,于是可建立 ARMA( $p, q$ )模型:

$$u_t = c + \phi_1 u_{t-1} + \cdots + \phi_p u_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}. \quad (1.2)$$

经  $d$  阶差分后的 ARMA( $p, q$ )模型称为 ARIMA( $p, d, q$ )模型(Autoregressive Integrated Moving Average Model),其中  $p$  为自回归模型的阶数, $q$  为移动平均的阶数, $\varepsilon_t$  为一个白噪声过程。有时,为了消除序列的异方差,需要对数据进行处理,一般情况下取其对数。

### 1.2 混合时间序列模型

设变量  $y_t$  表示时间序列, $\Delta y_t$  表示其一阶差分(增量)。考虑拟合模型如下:

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_k t^k + u_t. \quad (1.3)$$

传统时间序列分析利用模型(1.3)预测  $y_t$  的长期趋势,误差项  $u_t$  反映除了  $y_t$  的长期趋势以外的波动情况,是构成  $y_t$  的预测误差的一个主要来源,它的未来不可预测。

现代时间序列分析的一个有效应用是对该误差项  $u_t$  序列建立 ARMA 模型,将  $u_t$  用 ARMA 模型替代。ARMA 模型帮助解释模型(1.3)中时间变量  $t$  无法解释的那部分变差。这种混合时间序列模型为:

$$\begin{cases} y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_k t^k + u_t, \\ \phi(B)u_t = \theta(B)\varepsilon_t, \end{cases} \quad (1.4)$$

其中  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ ,  $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$  为滞后算子, $\varepsilon_t$  是服从正态分布的白噪声误差项,它的方差比  $u_t$  的方差要小。

### 1.3 灰色预测模型

GM(1,1)模型是最常用的一种灰色模型,它是由一个只包含单变量的一阶微分方程构成的模型,是作为 GDP 预测的一种有效模型。

设 GDP 数列

$$x^{(0)} = [x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)],$$

构成灰色模块微分方程

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u. \quad (1.5)$$

将灰参数代入时间函数,有

$$\hat{x}^{(1)}(k+1) = \left[ x^{(0)}(1) - \frac{u}{a} \right] e^{-ak} + \frac{u}{a}; (k=1, 2, \cdots, n), \quad (1.6)$$

根据累加序列的预测结果进行减法运算,推算出原始序列的预测值:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k). \quad (1.7)$$

### 1.4 组合预测模型

不同的预测方法根据相同的信息,往往能提供不同的结果,如果简单地将误差平方和较大的一些方法舍弃掉,将会丢弃一些有用的信息,难以有效利用,应予以避免。组合预测法是指通过建立一个组合预测模型,把多种预测方法所得到的预测结果进行综合。由于组合预测模型能够较大限度地利用各种预测样本信息,比单项预测模型考虑问题更系统、更全面,因而能够有效地减少单个预测模型受随机因素的影响,从而提高预测的精度和稳定性。

1969年, Bates 和 Granger 对组合预测方法进行了比较系统的研究,首次提出组合预测概念时建议的组合预测方法,通常称为最小方差法,即在均方预测误差指标下的加权平均组合预测(B-G模型)。

假设对于同一预测问题,有  $k(k \geq 2)$  种预测方法。记第  $t$  期实际观测值、第  $i$  种方法的预测值和预测误差分别为  $y_t$ 、 $f_{it}$  和  $e_{it}$ , ( $e_{it} = y_t - f_{it}$ ;  $i=1, 2, \cdots, k$ ;  $t=1, 2, \cdots, n$ ), 第  $i$  种方法在组合预测中的权重(或组合加

权系数)为  $\omega_i (i = 1, 2, \dots, k; \sum_{i=1}^k \omega_i = 1)$ , 第  $t$  期组合预测方法的预测值和预测误差分别为  $f_{ct}$  和  $e_{ct} (t = 1, 2, \dots, n)$ , 则  $f_{ct} = \sum_{i=1}^k \omega_i f_{it}$ ,  $e_{ct} = y_t - f_{ct} = \sum_{i=1}^k \omega_i e_{it}$ 。预测方法的预测误差平方和为  $e_c^2$ , 则有

$$e_c^2 = \sum_{t=1}^n e_{ct}^2 = \sum_{i=1}^k \sum_{j=1}^k [\omega_i \omega_j (\sum_{t=1}^n e_{it} e_{jt})] = \boldsymbol{\omega}^T \mathbf{E} \boldsymbol{\omega}, \quad (1.8)$$

其中,  $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_k)^T$  为组合权重向量,  $\mathbf{E} = (c_{ij})_{k \times k}$  是预测误差信息矩阵, 由  $k$  种单项预测的误差计算得到, 这里  $c_{ij} = \sum_{t=1}^n e_{it} e_{jt}$ 。

由上面的公式知道, 组合预测方法的预测误差平方和  $e_c^2$  的大小与预测误差信息  $\mathbf{E}$  以及组合权重向量  $\boldsymbol{\omega}$  有关, 预测误差信息矩阵  $\mathbf{E}$  由参加组合的  $k$  中预测方法决定, 当预测误差信息矩阵  $\mathbf{E}$  给定后, 通过线性规划的方法确定组合权重向量  $\boldsymbol{\omega}$ , 也即是求组合预测误差最小情况下的最优组合权重。

$$\begin{cases} \min e_c^2 = \boldsymbol{\omega}^T \mathbf{E} \boldsymbol{\omega}, \\ \text{s. t. } \mathbf{R}_k^T \boldsymbol{\omega} = \mathbf{1}, \\ \boldsymbol{\omega} > \mathbf{0}, \end{cases} \quad (1.9)$$

其中,  $\mathbf{R}_k = (1, 1, \dots, 1)_{1 \times k}^T$  是元素全为 1 的  $k$  维向量。

只要知道各种单项预测方法的误差, 就可以计算出最有效的权重向量  $\boldsymbol{\omega}$ , 再乘以单项预测值, 就是组合预测结果了。

## 2 国民生产总值 GDP 的预测分析

本例由 ARIMA 预测、灰色预测和线性回归预测三种方法构成组合预测方法, 将其应用于中国 GDP 的预测中。由《中国统计年鉴 - 2007 年》给出的 1980 ~ 2006 年名义 GDP 的数据, 用 GDP 平减指数换算成 1952 年的不变价格计算的实际 GDP<sup>[6]</sup>, 并且作为组合预测模型的原始数据。根据实际 GDP 的 1980 ~ 2002 年的数据的变化趋势, 首先用 ARIMA 模型、混合时间序列模型和灰色预测模型分别进行预测。

由于中国经济一直保持明显的增长趋势, 在 ARIMA 模型中, GDP 序列非平稳。为了使该时间序列更容易变得平稳, 先对序列 GDP 取对数, 令  $X = \ln(\text{GDP})$ , 观察发现两阶差分后该模型变得平稳。然后用 EViews 5.0 软件进行拟合, 最终选择对  $X$  的 ARIMA(2, 2, 2) 模型。令  $Z$  为对  $X$  的二阶差分, 则拟合如下:

$$Z = 0.235Z_{t-1} + 0.107Z_{t-2} + 0.034\mu_{t-1} + 0.944\mu_{t-2},$$

其中  $Z = \Delta^2 X$ ,  $X = \ln(\text{GDP})$ 。

所以得到

$$\text{GDP}_t = e^{2X_{t-1} - X_{t-2} + 0.235Z_{t-1} + 0.107Z_{t-2} + 0.034\mu_{t-1} + 0.944\mu_{t-2}}。$$

混合时间序列模型中, 设 1980 年为第一年, 以此类推, 用  $t$  表示。先估计模型(1.4)的第一式如下:

$$\text{GDP}_t = 4\,009.48 + 46.41t^2 + \hat{\rho}_t,$$

利用自相关系数、偏自相关系数和 AIC 准则, 识别  $\hat{\rho}_t$  为 ARMA(1, 1), 估计的结果如下:

$$(1 - 0.67B)\hat{\rho}_t = (1 - 0.83B)\varepsilon_t。$$

GM(1, 1) 模型中, 关键是对参数  $a$ 、 $u$  的估计。利用 MATLAB 7.0 软件, 经过计算, 最终确定参数  $a = -0.09$ ,  $u = 3\,730.31$ 。

根据建立的具体模型预测 2003 ~ 2006 年的数据, 预测结果和预测误差如表 1 所示。

表 1 3 种预测方法的预测结果

Table 1 Forecasting results of three methods

年份	实际 GDP (亿元)	ARIMA(2, 2, 2)		混合时间序列		GM(1, 1)	
		预测值(亿元)	相对误差/%	预测值(亿元)	相对误差/%	预测值(亿元)	相对误差/%
2003	3 1640.51	3 1841.53	0.64	3 1138.26	-1.59	3 2668.12	3.25
2004	3 4831.47	3 4943.17	0.32	3 3282.18	-4.45	3 5834.38	2.88
2005	3 8465.36	3 8205.50	-0.68	3 5561.39	-7.55	3 9307.50	2.19
2006	4 2730.15	4 1776.05	-2.23	3 7962.01	-11.16	4 3117.41	0.91

从表1可以看出,这3种模型的预测结果各不相同。为了充分利用单项预测模型的信息,这里采用组合预测模型,纳入前述3种单项预测模型的数值,通过线性规划,求出各单项预测模型在组合预测模型中的权重为(0.710 346 3,0,0.289 653 7),组合预测的结果和预测误差如表2所示。

表2 我国 GDP 组合预测效果

Table 2 The result of combination prediction model in forecasting GDP of China

年份	实际 GDP(亿元)	预测 GDP(亿元)	误差(亿元)	相对误差/%
2003	3 1640.51	3 2080.92	440.41	1.39
2004	3 4831.47	3 5201.20	369.73	1.06
2005	3 8465.36	3 8524.55	59.19	0.15
2006	4 2730.15	4 2164.46	-565.69	-1.32

从表2可以看出,虽然组合预测2003~2004年的预测效果没有ARIMA(2,2,2)预测效果好,可是发现越往后组合预测的效果越好。而且从权重系数可以看出,混合时间序列模型的权重为0,表示这个模型不适合对GDP这个指标的预测,所以组合预测模型最终选择了ARIMA(2,2,2)模型和GM(1,1)模型。

用1980~2006年的数据再次建立组合模型,得到的数据如表3。

表3 我国 GDP 2007~2020 年预测结果

Table 3 Forecasting results of China's GDP from 2007 to 2020

年份	ARIMA(2,2,2)		混合时间序列		GM(1,1)		组合预测	
	预测值(亿元)	增长率/%	预测值(亿元)	增长率/%	预测值(亿元)	增长率/%	预测值(亿元)	增长率/%
2007	47 715.30	11.67	43 473.90	1.74	46 943.12	9.86	47 491.60	11.14
2008	52 723.12	10.50	46 535.79	7.04	51 463.41	9.63	52 358.13	10.25
2009	58 124.27	10.24	49 713.83	6.83	56 418.76	9.63	57 630.03	10.07
2010	64 115.34	10.30	53 007.81	6.63	61 850.02	9.63	63 459.13	10.11
2011	70 745.31	10.34	56 418.42	6.43	67 806.54	9.63	69 893.92	10.14
2012	78 061.06	10.34	59 944.26	6.25	74 335.31	9.63	76 981.74	10.14
2015	104 858.01		71 219.37		97 941.65		102 855.04	
2020	171 482.42		92 332.75		155 090.23		166 734.11	

从表3可得到2007年的经济预测增长率为11.4%,而且将来的5年经济增长率在10%左右,基本符合中国的国情。

### 3 结束语

(1) 我国国内生产总值正处于快速增长阶段,需要保持和改进现有的政策和措施。通过组合预测模型的预测以及和其他3种模型比较可知,在今后的10年内,我国国民经济仍将保持快速的增长,并且在2015年将达到10万亿元(1952年价格为基准)。

(2) 我国GDP在未来的10年内增速不会发生较大的变化,仍以比较快的速度发展,符合我国的经济规律。

(3) 组合预测法由于考虑了更多的影响因素,然后建立预测模型,充分利用各单项预测的有用信息,因此在GDP预测中可得到可信度更高的预测值,可以为有关部门提供科学依据。在实际应用中应注意,单一模型是组合预测模型的基础,应尽力提高单一模型的预测精度,对于精度很差的模型不予采用;同时选择单一模型时,尽可能从不同影响因素出发,充分利用数据所包含的有用信息。组合预测模型具有比其他模型更好的优势,结果更接近于现实的数据,所以,可以把它应用在其他数据的预测上。

#### 参考文献:

- [1] 龚国勇. ARIMA模型在深圳GDP预测中的应用[J]. 数学的实践与认识, 2008, 38(4): 54-57.
- [2] 丁文斌. GDP总量预测方法探讨研究[J]. 统计与预测, 2003, 6: 57-59.
- [3] 戴羽, 王媛媛, 王伦夫. 基于灰色GM(1,1)模型的安徽省GDP总量预测[J]. 重庆工学院学报: 自然科学版, 2008, 22(2): 74-78.
- [4] 易丹辉. 数据分析与Eviews应用[M]. 北京: 中国统计出版社, 2002.
- [5] 李闽榕, 李建平, 黄茂兴. 中国省域经济综合竞争力评价与预测研究[M]. 北京: 社会科学文献出版社, 2007.
- [6] 杨冠琼. 中国经济增长数据可信度检验研究——理论、模型与实证检验[M]. 北京: 经济管理出版社, 2006.

(编辑: 李晓红)