陕西省 GDP 的预测方法比较与选择

曹 飞

(西安电子科技大学 人文学院, 西安 710071)

摘要:以陕西省 1993 年至 2010 年不变价格的 GDP 为样本,分别应用了二次函数模型、指数函数模型和自回归整合移动平均模型对原数列进行了拟合、分析与预测。经过比较,三者的误差依次为 17.948 3%, 8.752 9%, 3.389 0e-06。自回归整合移动平均模型与原序列高度拟合,误差几乎为零,因此不需要组合预测。

关键词:陕西省 GDP;预测方法;选择与比较。

中图分类号:F201; F224 **文献标志码**:A **文章编号**:1671-1807(2012)04-0104-03

GDP 是反映一国或者一个地区所有常住单位在核算期内生产活动的最终成果及衡量国民经济发展规模、速度、结构、效益的代表性指标,也是制定经济发展战略目标的主要指标,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控制,是处于严重衰退还是处于通胀威胁之中[1],因此,科学准确地预测 GDP 的变化趋势,对政府及其有关部门把握未来区域经济运行状况、制定发展战略,有着重要的现实意义。

目前 GDP 的预测方法主要有灰色模型、二次函数模型、指数函数模型和线性预测方法和自回归整合移动平均模型等,也有学者通过组合预测的方法,将误差小的预测模型赋予较大的权重,进行组合预测。笔者认为,在各种预测模型中,如果有一个预测模型与原数据高度吻合的话,则不需要进行组合预测。否则,反而会降低预测的精度。

1 二次函数模型与指数函数模型

党的十四大以来,陕西省经济发展态势良好,成绩斐然。消除了价格因素后,陕西省 GDP 从 1993 年的 590.553 6 亿元,增加到 2011 年的 8 841.467 2 亿元,平均增速在 10%以上。因此,对陕西省 GDP 的准确拟合与预测有利于为各级政府科学决策提供一定的参考依据。

1.1 二次函数模型

年份用 t 来表示,将 1993 年定义为起始年,即 t=1。将 2010 年定义为末尾年,即 t=14,以下模型定义相同。GDP 的预测值为 F_1。

$$\hat{F}_1(t) = 1463.592475 + 39.26932623 \times t^2 - 331.7017767 \times t$$

(5.260932) (11.38961) (-4.920191)

R-squared=0.981750

二次函数预测模型(模型一)的 F 检验通过,系数的 T 检验也通过,拟合度也达到 0.98。但从预测效果来看很不理想,平均误差达到 17.948 3%。因此,重新构建指数函数模型。

1.2 指数函数模型

把指数函数(模型二)的 GDP 的预测值用 F_2 表示。

$$\hat{F}_2(t) = 515.813517517035 \times e^{0.1497258963t}$$

(119.3771) (30.97675)

R-squared=0.983599

指数函数预测模型的 F 检验通过,系数的 T 检验也通过,拟合度也达到 0.98。预测效果比二次函数模型提高了许多,但平均误差也达到 8.752 9%。因此,重新构建 ARIMA 模型,以提高预测精度。

从图 1 来看,二次函数模型和指数函数模型与原数据有比较大的偏离,拟合效果较差。

2 ARIMA 模型[2-3]

2.1 时间序列的平稳化处理

ARIMA 模型(模型三)的前提是序列的平稳性,经检验,序列 y(用 y 代表 GDP 的原始数列)的 ADF

收稿日期:2012-02-27

基金项目:西安电子科技大学中央高校专项基金(72115464)

作者简介:曹飞(1974—),男,陕西榆林人,西安电子科技大学人文学院副教授,经济学博士,研究方向:劳动经济学、经济预测。

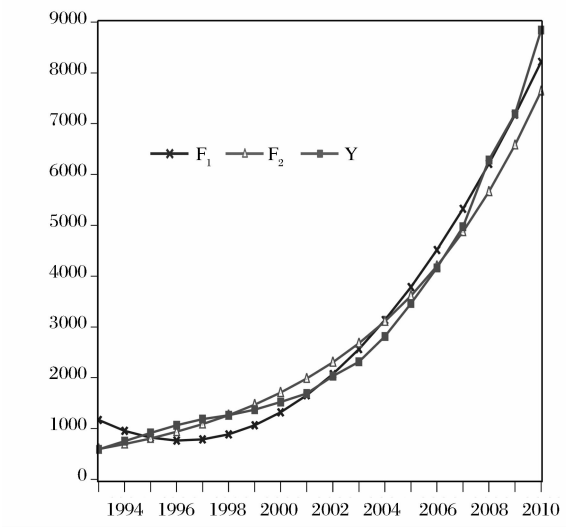

图 1　模型一、二和原数据(Y)的拟合效果

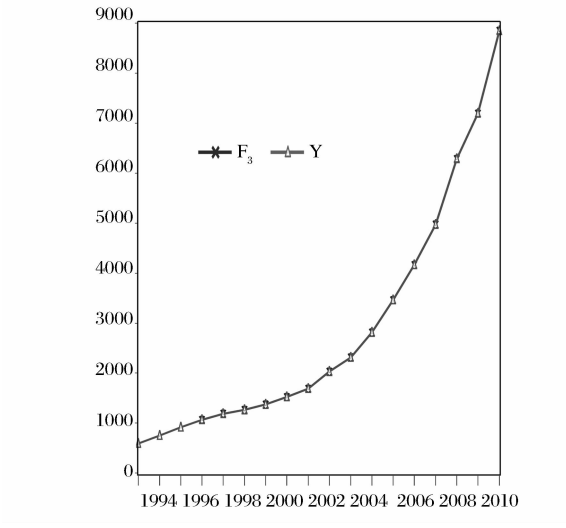

图 2　模型三和原数据(Y)的拟合效果

统计量大于显著性水平的临界值,表明该序列是非平稳的。为消除 y 序列的异方差,对其进行对数变换,令 y_1=LOG(y),对其进行平稳性检验,检验结果显示,该序列的 ADF 统计量通过了平稳性检验。

表 1　{y}、{y_1}序列的单位根检验结果

变量	检验形式(C,T,K)	ADF 值	5%临界值	结论
{y}	(C,T,3)	−0.504 013	−3.791 172	不平稳
{y1}	(C,T,3)	−4.332 888	−3.791 172	平稳

注:c,t 表示带有常数项和趋势项,k 表示滞后的阶数,按照 SIC 标准自动确定。

2.2　ARIMA 模型 p、q 的确定

从{y_1}的自相关图和偏相关图(图 3)中可以看出,在显著性水平在 5%的情况下,自相关系数 1 阶显著不为零,3 阶自相关系数虽然比较大,但其在回归方程中的 T 检验不能通过。偏相关图除 1 阶显著不为零,此后截尾。以 F_3 为被解释变量,分别对[C AR(1) AR(2) MA(1)]、[AR(2) MA(1)]、[C AR (1) MA(1)]、[AR(1) MA(1)]作回归分析,结果表明当以 AR(2)和 MA(1)为解释变量时,AIC 最小,R^2 最大,且 F 检验和 T 检验显著通过。因此,ARIMA(p,d,q)模型具体可确定为 ARIMA(2,0,1)。

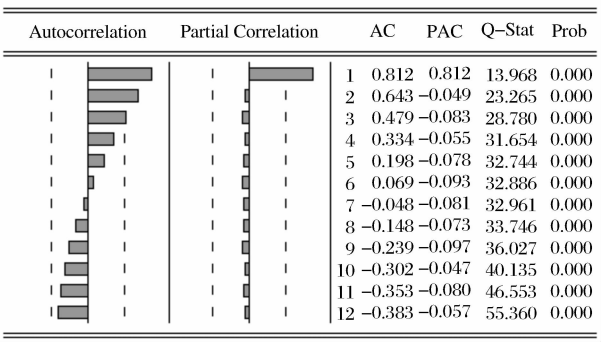

图 3　{ y1}序列的自相关图与偏相关图

2.3　ARIMA 模型参数估计与诊断检验

运用 ARIMA(2,0,1)模型进行预测之前,需要对模型进行参数估计与检验。经过多次估计,得到较为理想的模型结果。为了选择统计性质优良的模型,在同等或相似条件下,尤其是当滞后分布的长度确定时,通常选择信息准则量(AIC)较小的模型,本文 ARIMA(p,d,q)模型具体确定为 ARIMA(2,0,1)模型,F 检验和 T 检验显著通过检验,从预测效果来看,ARIMA 模型与原序列的平均误差为 3.389 0e−06,二者高度吻合,图 2 也明显反映出这一点。因此不需要与二次函数模型、指数函数模型进行组合预测,否则,会降低预测的精度。通过对该模型参差序列进行单位根检验,发现 ADF 值为−3.524 943,明显小于 1%(−2.771 926)、5%(−1.974 028)与 10%(−1.602 922)显著性水平的临界值,这表明该模型残差为白噪声序列。由此可确定 ARIMA(2,0,1)模型为平稳序列较理想的模型。

具体预测结果为

$$y_1t=1.041787y_1(t-2)+u+0.996673u(t-1)$$
$$(377.9161)\qquad(7.914155)$$
$$\text{R-squared}\qquad 0.996008$$

$\because y_1(t)=\text{LOG}(y)\qquad \therefore y(t)=\exp(y_1(t))$

用 F3 表示 y 序列的预测值,则有:

$$\hat{F}_3(t)=\exp\{1.041787y_1(t-2)+u+0.996673u(t-1)\}$$

3　三种模型预测结果与原序列的具体比较

表 2 三种预测方法的具体比较

年份	以不变价格计算的GDP的原值序列(亿元)	模型一预测结果 F1	模型二预测结果 F2	模型三预测结果 F3
1993	590.553 6	1 171.160 0	599.125 6	
1994	755.393 2	957.266 2	695.893 8	
1995	917.4587 1	821.911 1	808.291 7	
1996	1 066.588 0	765.094 6	938.843 6	1 066.590 6
1997	1 190.503 6	786.816 7	1 090.481 7	1 190.506 6
1998	1 266.297 0	887.077 6	1 266.611 7	1 266.300 1
1999	1 372.333 9	1 065.877 1	1 471.189 6	1 372.337 5
2000	1 523.394 1	1 323.215 1	1 708.809 9	1 523.398 0
2001	1 690.439 9	1 659.091 9	1 984.809 8	1 690.444 4
2002	2 028.973 4	2 073.507 3	2 305.388 0	2 028.978 7
2003	2 314.597 5	2 566.461 4	2 677.744 6	2 314.603 6
2004	2 812.736 95	3 137.954 1	3 110.242 7	2 812.744 6
2005	3 459.736 2	3 787.985 5	3 612.596 2	3 459.745 7
2006	4 164.714 7	4 516.555 5	4 196.087 6	4 164.726 5
2007	4 971.753 1	5 323.664 2	4 873.822 2	4 971.767 5
2008	6 284.003 5	6 209.311 6	5 661.021 5	6 284.022 2
2009	7 191.725 4	7 173.497 5	6 575.366 1	7 191.747 2
2010	8 841.467 2	8 216.222 2	7 637.391 8	8 841.494 9
	相对误差%	17.948 3	8.752 9	3.389 0e-06

通过利用 ARIMA(2,0,1)模型对陕西省 2011 年 GDP 的预测值为 10 432.155 6 亿元。据查陕西省 2011 年 GDP12 391.3 亿元,在消除价格因素后,与 ARIMA(2,0,1)模型非常接近。因此,ARIMA(2,0,1)模型高度拟合了陕西省 GDP 的实际发展情况,不需要与其他模型进行组合预测。否则,反而会降低预测的精度。

参考文献

[1] 许宪春. 国内生产总值核算的重要意义和作用[J]. 中国统计, 2003, 34(2) : 8— 9.

[2] 高铁梅. 计量经济分析方法与建模 EVIEWS 应用及实例[M]. 北京:清华大学出版社,2006:137—143.

[3] 张晓峒. EVIEWS 使用指南与案例[M]. 北京:机械工业出版社,2009:295—299.

Comparation and Selectiong of Pridiction Means of the Shaanxi Province's GDP

CAO Fei

(Institute of literature, Xi'an Electronics &Technology University, Xi'an 710071, China)

Abstract: Based on the data of the Shaanxi Province's constant price GDP from 1993 to 2010, the paper fitted, analyzed and pridicted the Shaanxi Province's constant price GDP by the quadratic function, exponential function and ARIMA model . by comparation, the errors of three models are in turn 17.9483%, 8.7529%, 3.3890e—06. The analyzed result indicated the ARIMA model is highly fitted the raw empirical data, error is almost zero, so combination forecasting model is not needed.

Key words: Shanxi Province's GDP; pridiction means; comparation and selectiong