

Application of the Combination Time Series Model in Forecasting the Gross Agriculture Output of Jiangxi Province

Xiaofa Tian

School of Mathematics and Statistics, Yunnan University of Finance and Economics, Kunming Yunnan
Email: xiaofa560@163.com

Received: Nov. 24th, 2015; accepted: Dec. 13th, 2015; published: Dec. 16th, 2015

Copyright © 2015 by author and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This article selects Jiangxi province's 1979-2014 agricultural output [1], using the Jiangxi agricultural gross output value of 1979-2009 data, respectively to set up the GM(1,1) model, the exponential curve model and mixed time series model. The data from 2010-2012 were used for the model test, and then on the basis of the three models, the combination of time series model was established to predict agricultural output of the Jiangxi province in recent years. Combined model prediction with a single model predicted results comparison shows that the results error of combination forecasting model is superior to the three model predictions respectively, and has more advantages in prediction of time series data. The prediction of 2015 Jiangxi province's agricultural output was 3385.07 billion Yuan, but it totally reached 3800 billion Yuan in 2015. For the wide gap with the expectation, we should increase investment in agriculture.

Keywords

GM(1,1) Model, Exponent Model, Mixed-Time Series Model, Combination Prediction Model, Gross Agriculture Output

组合时间序列模型在江西省农业总产值预测中的应用

田小发

云南财经大学统计与数学学院, 云南 昆明
Email: xiaofa560@163.com

收稿日期: 2015年11月24日; 录用日期: 2015年12月13日; 发布日期: 2015年12月16日

摘要

本文选取1979~2014年江西省农业总产值[1], 利用江西省农业总产值1979~2009年数据分别建立GM(1,1)模型、指数曲线模型和混合时间序列模型, 并用2010~2012年的数据进行模型检验, 然后在三个模型的基础上, 建立组合时间序列模型对江西省最近几年的农业总产值进行预测。通过组合模型预测与单个模型预测结果比较表明, 组合预测模型所得结果误差优于三个模型的分别预测, 在时间序列数据的预测中更有优势。预测2015年江西省农业总产值为3385.07亿元。但江西省2015年的目标是农业总产值达到3800亿元, 与目标差距较大, 因此应加大对农业的投入。

关键词

GM(1,1)模型, 指数模型, 混合时间序列模型, 组合预测模型, 农业总产值

1. 引言

目前, 国内外对经济数据的预测方法很多, 如移动平均、指数平滑、趋势分析、时间序列、灰色预测及神经网络等, 但每一种预测方法都有其局限性。近年来, 国内外学者提出许多对经济数据的预测方法, 根据方法可归纳为相关关系预测法和时间序列预测法两大类。相关关系预测法是用统计分析方法寻找现象与经济数据之间因果关系或结构比例关系, 并根据这些关系预测经济数据。由于江西省农业总产值受诸多因素的影响, 并且这些因素之间又保持着错综复杂的关系, 因此运用相关关系预测法预测农业总产值比较困难。如果选择预测模型的标准时追求预测精度的极大化, 最好选择时间序列模型[2]。

本文选用目前预测精度较高的 GM(1,1)模型和指数曲线模型以及混合时间序列三种模型, 建立构成组合模型的单项预测模型, 再使用最小误差法对各模型进行权重分配, 建立了江西省农业总产值的组合预测模型。

2. 组合时间序列模型

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幅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。

“组合预测”思想于1969年首次提出, 近年来引起我国学者的重视, 其应用范围也逐渐扩大。组合预测理论的基本原理是: 通过个体预测值的加权算术平均而得到其组合预测值, 在确定加权权重时, 以组合预测误差方差最小为原则。

GM(1,1)灰色预测模型要求负荷数据少、不考虑分布规律、不考虑变化趋势、运算方便、短期预测精度高。指数曲线趋势外推法简单方便, 适用于趋势明显的时间序列的预测。ARIMA模型较为复杂, 适用于数据随机性较强不易提取确定性因素的时间序列。结合这三种时间序列模型的优势, 构建组合模型,

可以提高预测精度。

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} , ($e_{it} = y_t - f_{it}; i=1, 2, \dots, k; t=1, 2, \dots, n$), 第 i 种方法在组合预测中的权重为 w_i , 组合预测值和预测误差分别为 f_{ct} 和 e_{ct} ($t=1, 2, \dots, n$), 则

$$f_{ct} = \sum_{i=1}^k w_i f_{it}, \quad e_{ct} = y_t - f_{ct} = \sum_{i=1}^k w_i e_{it}$$

预测方法的预测误差平方和为 e_c^2 , 则有

$$e_c^2 = \sum_{t=1}^n e_{ct}^2 = \sum_{i=1}^k \sum_{j=1}^k \left[w_i w_j \left(\sum_{t=1}^n e_{it} e_{jt} \right) \right] = W^T E W$$

其中, $W = (w_1, w_2, \dots, w_k)^T$ 为组合权重向量, $E = (c_{ij})_{k \times k}$ 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 E 以及组合权重向量 w 有关, 预测误差信息矩阵 E 由参加组合预测的 k 种方法决定, 当预测误差信息矩阵 E 给定后, 通过线性规划的方法确定组合权重向量 w , 也即是求组合预测误差最小情况下的最优组合权重。

$$\begin{cases} \min e_c^2 = W^T E W \\ w > 0 \end{cases} \quad (1)$$

其中, $R_k = (1, 1, \dots, 1)_{1 \times k}^T$ 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 w , 再乘以单项预测值, 就是组合预测结果[3]。

3. 实证分析

3.1. GM(1,1)预测模型

$$\begin{aligned} \frac{dX^{(1)}}{dt} - 0.112839055X^{(1)} &= 953.0485 \\ X^{(0)}(1) &= 1054.9, \quad \frac{b}{a} = -8446.09, \quad X^{(0)}(1) - \frac{b}{a} = 9500.9871 \\ X^{(1)}(k+1) &= 9500.9871e^{0.112839055k} - 8446.09 \end{aligned}$$

预测公式:

$$X^{(0)}(k+1) = X^{(1)}(k+1) - X^{(1)}(k) \quad (2)$$

GM(1,1)模型检验

后验差比值: $c = s_2/s_1 = 2048/67237.5 = 0.03046 < 0.35$

小误差频率: $p = |B|/n = 6/6 = 1 > 0.95$

按照后验差比值 c 和小误差频率 p 可以得到预测精度等级为好[4]。

用上述的模型求得江西省 2010、2011、2012 年的农业总产值分别为: 1995.2 亿元, 2233.5 亿元, 2500.3 亿元。

3.2. 指数曲线趋势外推法

根据直线模型公式:

$$\begin{cases} b = \frac{\sum tY - n\bar{t}\bar{Y}}{\sum t^2 - n\bar{t}^2} = \frac{393.24 - 10 \times 5.5 \times 6.994}{385 - 11 \times (5.5)^2} = 0.104 \\ A = \bar{Y} - b\bar{t} = 6.994 - 0.104 \times 5.5 = 6.442 \end{cases}$$

$$a = e^A = e^{6.442} = 627.66$$

所以指数模型为:

$$\hat{y}_t = 627.66e^{0.104t}$$

3.3. 混合时间序列模型

3.3.1. 建立回归模型

从图 1 中回归估计的结果看, 模型拟合较好。可决系数 $R^2 = 0.967$, 从相对水平上看, 回归方程能减少因变量 y 的 96.7% 的方差波动。回归系数 t 检验的 p 值都小于 0.05, 回归系数都通过了显著性检验[5]。

所以回归方程为:

$$y = 0.115774t + 4.162545$$

3.3.2. 混合时间序列模型

根据 ARMA 模型定阶的基本原则表, 残差可以选择 AR(1)、MA(1)、ARMA(1,1)、ARMA(4,1) 等一系列时间序列模型。经验证, 最后选择如下混合时间序列模型:

从图 2 中可以看出, 模型的四个参数都通过了 t 检验, 所以这些变量的选用是恰当的; 且 F 统计量对应的 p 值为 0, 所以模型的整体拟合效果较好[6]。

模型残差序列的自相关函数和偏相关函数如下:

因为图3中 $Q(1) - Q(16)$ 对应的 p 值都比 0.05 大, 可以认为模型的残差序列为白噪声序列, 这也说明模型的拟合效果比较好。所以最终模型

$$y_t = \ln(\text{农业总产值}) = 4.24334 + 0.109552t + 0.809312y_{t-1} + 0.63010\varepsilon_{t-1}$$

根据建立的具体模型预测 2010~2012 年的数据, 预测结果和预测误差如表 1 所示。

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, 根据前述三种单项模型预测 2010~2012 年的数值和实际数值, 使组合预测误差最小情况下, 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。

3.4. 组合模型

计算残差 $e_{ct} = y_t - f_{ct} = \sum_{i=1}^k w_i e_{it}$, 得到预测误差信息矩阵:

$$E = \sum_{t=1}^n e_{it} e_{jt} = \begin{pmatrix} 18613.24 & 8567.54 & -8842.6 \\ 8567.54 & 6027.09 & -2205.16 \\ -8842.66 & -2205.16 & 14649.89 \end{pmatrix}$$

以最大误差平方达到最小的线性组合预测模型可表示为下列最优化问题

$$\begin{cases} \min e_c^2 = (w_1, w_2, w_3)^T E (w_1, w_2, w_3) \\ w_1 + w_2 + w_3 = 1 \\ w_1 > 0, w_2 > 0, w_3 > 0 \end{cases}$$

得到三个预测模型的权重分别为(0.113618,0.5166718,0.3697102)，所以组合模型为 $y = 0.113618y_1 + 0.5166718y_2 + 0.3697102y_3$ 。

其中 y_1 、 y_2 、 y_3 分别为GM(1,1)模型，指数曲线模型、混合时间序列模型预测的值。

通过计算可以得到组合预测的结果和预测误差如表2所示。

从表1和表2可以看出，虽然组合预测2011年的预测效果没有指数曲线预测效果好，预测2012年的预测效果没有混合时间序列预测效果好，可是发现组合预测误差比较稳定，并且预测误差相对较小，都小于3%。而且从权重系数可以看出，GM(1,1)模型权重较小，是因为它单个预测的误差相对较大。

$$\hat{y}_{2015} = w_1f_{13} + w_2f_{23} + w_3f_{33} = 395.53 + 1712.4 + 1274.14 = 3385.07$$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.162545	0.069208	60.14501	0.0000.
@TREND	0.115774	0.003963	29.21460	0.0000.
R-squared	0.967139	Mean dependent var		5.899148
Adjusted R-squared	0.966005	S.D. dependent var		1.070361
S.E. of regression	0.197349	Akaike info criterion		-0.345344
Sum squared resid	1.129454	Schwarz criterion		-0.252828
Log likelihood	7.352827	Hannan-Quinn criter.		-0.315186
F-statistic	853.4931	Durbin-Watson stat		0.202941
Prob (F-statistic)	0.000000			

Figure 1. After taking logarithm of agricultural output linear regression model test batches
图 1. 取对数后的农业总产值线性回归模型检验图

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.243340	0.331446	12.80250	0.0000
@TREND	0.109552	0.015450	7.090850	0.0000.
AR(1)	0.809312	0.128877	6.279730	0.0000
MA(1)	0.636010	0.168762	3.768682	0.0009
R-squared	0.994298	Mean dependent var		5.955870
Adjusted R-squared	0.993640	S.D. dependent var		1.040192
S.E. of regression	0.082954	Akaike info criterion		-2.017496
Sum squared resid	0.178915	Schwarz criterion		-1.830670
Log likelihood	34.26244	Hannan-Quinn criter.		-1.957729
F-statistic	1511.285	Durbin-Watson stat		2.104524
Prob (F-statistic)	0.000000			
Inverted AR Roots	0.81			
Inverted MA Roots	-0.64			

Figure 2. Hybrid model parametric test
图 2. 混合模型参数检验图

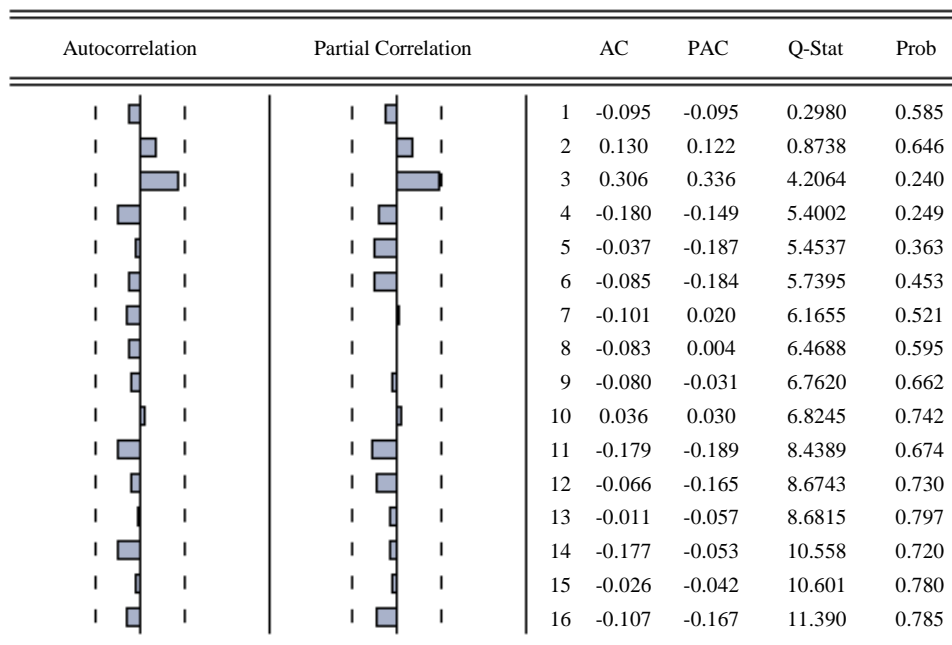


Figure 3. Hybrid model residual error sequence of autocorrelation function and partial correlation function test

图 3. 混合模型残差序列的自相关函数和偏相关函数检验图

Table 1. Three kinds of forecasting methods of prediction results

表 1. 三种预测方法的预测结果

年份	实际江西省农业总产值(亿元)	GM(1,1)		指数曲线趋势外推		混合时间序列	
		预测值	相对误差	预测值	相对误差	预测值	相对误差
2010	1900.6	1992.8	-4.85%	1970.4	-3.67%	1841.1	3.13%
2011	2207.3	2230.7	-1.06%	2186.4	0.95%	2102.3	4.76%
2012	2399.2	2497	-4.08%	2426	-1.12%	2390	0.38%

Table 2. Jiangxi agricultural output combination forecast result table

表 2. 江西省农业总产值组合预测效果表

年份	实际农业总产值(亿元)	预测农业总产值(亿元)	绝对误差	相对误差
2010	1900.6	1925.1	24.5	1.30%
2011	2207.3	2160.3	47	2.13%
2012	2399.2	2420.8	21.6	0.90%

4. 结论

本文通过对江西农业总产值历史数据的分析,结合时间序列模型的特点,采用组合预测误差最小方法进行权重分配,建立江西省农业总产值的组合时间序列预测模型,并对江西省 2015 年的农业总产值进行预测。研究表明,组合时间序列模型能一定程度上克服单项模型的缺陷,使不同年份的误差率趋于平缓,并且总体上降低预测模型的平均误差率。因此,对我国农业总产值的预测,组合时间序列模型比单项模型更优。

参考文献 (References)

- [1] 中华人民共和国国家统计局. 中国统计年鉴[M]. 北京: 中国统计出版社, 2013.
- [2] 陈显周, 区晶莹, 俞守华, 杨春. 组合时间序列模型及其在我国农业总产值预测中的应用[J]. 广东农业科学, 2012(13): 194-196.
- [3] 陈华友. 组合预测方法有效性理论及其应用[M]. 北京: 科学出版社, 2008.
- [4] 徐国祥. 统计预测与决策[M]. 上海: 上海财经大学出版社, 2008.
- [5] 何晓群. 实用回归分析[M]. 北京: 高等教育出版社, 2008.
- [6] 王燕. 应用时间序列分析[M]. 北京: 中国人民大学出版社, 2008.