

# 运用 SAS 软件系统对山西省工业总产值的预测分析

贾亚军 (山西省统计局,山西 太原)

毕 华 (山西大学数学系,山西 太原)

刘维奇 (山西大学,山西 太原)

## 一、前言

经过建国后 50 年的努力,特别是经过改革开放以来 20 年的发展,山西工业经济发生了巨大的变化,取得了很大成就。山西已由一个落后的农业省发展成为全国重要的能源重化工基地,基本形成了以重工业为主,煤炭、电力、冶金、机械、化工、电子、建材、纺织等门类比较齐全的具有地方特色的工业生产体系。但在我国迅猛崛起的行列中,山西仍处于中下游位置。在经济大发展的社会背景下,在我国加入 WTO 后,在国家政策对中西部的倾斜的同时,山西省工业将在更大的范围、更深的层次上融入世界经济。考察工业发展必须从全局出发去分析它的原因,研究它的规律和发展趋势。工业总产值是衡量工业经济发展的重要指标。如何做好工业总产值的预测是制定工业发展决策的重要环节。本文运用国际著名的大型集成软件系统中的时间序列分析方法对山西省历年的工业总产值作了预测分析,为政府选择预测工业发展的统计模型提供参考。作者用求和自回归滑动平均(ARIMA)方法,作了各种参数搭配的试算比较,得到三个较好的模型 ARMA(0,1,1)、ARMA(0,1,3)和 ARMA(2,1,0),并分别计算出对最近十年(1989~1999)工业总产值的滚动预测误差为 6.78%、6.14%和 6.48%。

## 二、用 ARIMA 过程对工业总产值的建模

SAS/ETS 软件中的 ARIMA(求和自回归滑动平均)过程提供了一个综合工具包来进行一元时间序列的模型识别、参数估计及预测分析。数学上纯 ARIMA 模型记作:

$$W_t = \mu + \frac{1 - \theta_1 B - \dots - \theta_p B^p}{1 - \phi_1 B - \dots - \phi_p B^p} \varepsilon_t$$

其中  $t$  代表年份; $W_t$  表示响应序列  $Y_t$  或  $Y_t$  的差分; $\mu$  为均值项; $B$  是后移算子; $\varepsilon_t$  表示独立扰动或称作随机误差。

本文基于山西省 1949—1999 年工业总产值数据,采用 Box - Jenkins 方法建立 ARIMA 模型,建模步骤如下:

### 1. 模型识别(Identification)

在“识别”阶段,从原始数据的散点图可看出明显的指数趋势,对  $\{X_t\}$  作对数变换得到  $\{Y_t\}$  可将指数趋势转化为线性趋势,再进行差分以消除线性趋势,再结合自相关系数(ACF)图可以认为,通过一阶差分后的序列是近似平稳的(其自相关系数随时间间隔的增加而迅速下降)。

### 2. 模型的估计和诊断(Extimtion and Diagonosis)

作者针对  $\{Y_t\}$  的 ARIMA 模型  $(p,1,q)$  各种可能排列(明显不适合或模拟后提出警告的排列形式予以删除),采用 AIC 准则(Akaike Information Creterino)判定模型阶数,同时结合残差的自相关检验,确定了三个较优的模型,分别为:

ARIMA(0,1,1):  $(1 - B)y_t = (1 + 0.35472B)\varepsilon_t$

ARIMA(0,1,3):  $(1 - B)y_t = (1 + 0.37462B - 0.2535B^3)\varepsilon_t$

ARIMA(2,1,0):  $(1 - B)y_t = \frac{1}{1 - 0.39075B + 0.25726B^2}\varepsilon_t$

3. 模型的比较:作者用上述模型相应的程序对 1990~1999 年作了滚动预测,即用 1989 年以前的数据预测 1990 年的值;用

1990 年以前的数据预测 1991 年的值;...用 1998 年以前的数据预测 1999 年的值。从平均绝对滚动误差最小这个意义上来说,模型确定为 ARIMA(0,1,3)

在上述建模过程中采用的是对数变换数据,为了得到实际的预测值要进行反对数操作,但是会降低预测原始序列的均值,采用如下变换:

$$\text{forecast} = \exp(\text{forecast} + \frac{1}{2} \text{std} * \text{std})$$

其中 std 为预测的标准差;forecast 为预测值。用模型 ARIMA(0,1,3)作近十年(2000~2009)的预测,预测图如下:

## 三、讨论

本文基于 SAS/ETS 软件所作的时间序列分析方便而且灵活,并达到了较高的预测精度,所有的计算在数分钟内即可完成,从而便于推广应用及进一步改进模型。每天新的数据出来后,只要将其

用 ARIMA 方法滚动预测结果

年份	实际值	ARIM(0,1,1) 预测值	ARIM(0,1,3) 预测值	ARIM(2,1,0) 预测值
1990	6164830	6618340	6537401	6604529
1991	6621142	6986008	6927918	6992794
1992	7731540	7533819	7661307	7581724
1993	9996496	9040471	9064166	9125358
1994	12162056	12023589	12100523	12044632
1995	14090071	14200394	14084950	13975475
1996	17532423	16344586	15955828	16134377
1997	20209852	20927109	21023727	20916300
1998	20176209	23252905	23113781	23006522
1999	19343725	22299672	21738012	22084144
平均绝对滚动 预测误差(%)		6.78	6.14	6.48

加入本文所编写的 SAS 程序中(不必修改参数或其他 SAS 语句),运行程序即可自动算出由 ARIMA 方法给出的预测值,再结合专家的总体评价即可对下一年度进行预报。本模型具有较强的现实指导意义,由于它能对未来一年内山西省的工业产值作出一个宏观上的总量预期,把握在保持现有的政策环境、资本环境、劳动力数量、技术水平惯性发展条件下工业发展的趋势,这就为我们实行宏观控制,积极干预工业发展水平提供了量的指导。山西省正处于产业结构调整的关键时期,工业占全部行业的比重要下降,但是它在总量上还必须适应山西省国民经济的发展,避免成为经济全面发展的瓶颈。由于本模型能够对未来的工业产值做出较为精确的量的判断,这就为我们调整工业在全行业中的比重、合理安排山西省有限资源的配比以及相关政策的调整提供了一个客观、科学的依据。