

基于 VAR 和 GM 组合模型的电力消费预测

徐刚¹, 王维国²

(1. 中国科学院 大连化学物理研究所, 辽宁 大连 116023; 2. 东北财经大学 数学与数量经济学院, 辽宁 大连 116025)

摘要: 组合预测模型较单一模型具有更准确的预测精度, 利用中国电力消费 1978 年~2007 年的年度数据, 将 VAR 模型与灰色预测的 GM(1,1)、无偏 GM(1,1) 两种模型有机组合, 建立灰色 VAR 组合预测模型, 对 2008 年~2015 年的电力消费量进行预测。结果表明, 该模型平均相对误差为 4.88%, 拟合精度较高, 具有较好的应用价值。

关键词: 电力消费预测; VAR 模型; 灰色预测; 组合预测

中图分类号: F407.2 文献标识码: A 文章编号: 1008-2204(2010)06-0056-04

Electricity Consumption Forecast: Based on VAR and GM Combination Model

XU Gang¹, WANG Wei-guo²

(1. Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China;
2. College of Quantitative Economics, Dongbei University of Finance&Economics, Dalian 116025, China)

Abstract: A combined forecast model is more accurate than a single forecasting model. Based on the electricity consumption data from 1978 to 2007, by combing the VAR model, the GM(1,1) model and the WGM(1,1) model, we set up the grey VAR combined forecast model and forecast the electricity consumption from 2008 to 2015. The research results show that the forecast values are accurate: the average relative error is only 4.88%, so it can be used in the application as one of the efficient methods.

Key words: electricity consumption forecast; VAR model; grey forecast; combination forecast

电力是重要的能源之一, 也是国民经济的基础。准确地预测电力的未来消费量, 可以更合理地生产和使用电力。中国含电力等在内的能源消费量预测研究大致可以分为两类: 一是采用单一方法预测, 如多元统计^[1]、时间序列法^[2]和 ARMA 模型^[3]等; 二是采用组合模型预测, 如灰色 GM(1,1) 和神经网络组合预测模型^[4]、BP 和 GM 的组合模型^[5]等。由于单一模型的局限性, 组合预测模型被越来越多地运用于实际中, 从而使拟合和预测精度大为提高。文章根据组合预测理论, 将 VAR 模型与灰色预测的 GM(1,1)、无偏 GM(1,1) 两种模型有机组合, 对电力消费量进行拟合和预测。

一、组合模型建立

(一) GM(1,1) 模型

设有原始数据序列: $x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}$

(n), 满足 $x^{(0)}(k) \geq 0, k = 1, 2, \dots, n$ 。利用该数据序列建立 GM(1,1) 模型。

步骤 1. 做一阶累加形成生成数据序列

$$x^{(1)}(k) = \sum_{m=1}^k x^{(0)}(m) \quad (k = 1, 2, \dots, n) \quad (1)$$

步骤 2. 确定数据矩阵 B, Y_n

取背景值为 $z_{(t)} = \frac{1}{2}(x_{(t+1)}^{(1)} + x_{(t)}^{(1)})$, 则

$$B = \begin{bmatrix} -z_{(1)} & 1 \\ -z_{(2)} & 1 \\ \vdots & \vdots \\ -z_{(n)} & 1 \end{bmatrix} \quad Y_n = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}$$

步骤 3. 求参数列

$$[\hat{a}\hat{u}]^T = (B^T B)^{-1} B^T Y_n \quad (2)$$

步骤 4. 建立生成数据序列模型

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{\hat{u}}{\hat{a}} \right) e^{-\hat{a}k} + \frac{\hat{u}}{\hat{a}} \quad (3)$$

步骤 5. 建立原始数据序列模型

$$\begin{aligned} \hat{x}^{(0)}(k) &= \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1) = \\ &= (1-e^u) \left(x^{(0)}(1) - \frac{\hat{u}}{\hat{a}} \right) e^{-\hat{a}(k-1)} \quad (k=2,3,\dots,n) \end{aligned} \quad (4)$$

式(4)中,当 $k=2,3,\dots,n$ 时, $\hat{x}^{(0)}(k)$ 是原始数据序列 $x^{(0)}(k)$ 的拟合值;当 $k>n$ 时, $\hat{x}^{(0)}(k)$ 是原始数据序列的预测值。^[6]

(二) 无偏 GM(1,1) 模型

由于 GM(1,1) 模型对指数序列建模总存在偏差,故可对 GM(1,1) 模型的参数 a 和 u 加以修正来表示原始数据序列的参数 a 和 u ,修正后的模型为无偏差模型,记为 WPGM(1,1)。^[7]其建模步骤为:

步骤 1~3 同 GM(1,1) 模型建模步骤 1~3。

步骤 4. 求无偏 GM(1,1) 模型参数

$$\hat{a} = \ln \frac{2-\hat{a}}{2+\hat{a}} \quad \hat{A} = \frac{2\hat{u}}{2+\hat{a}}$$

步骤 5. 建立原始数据序列模型

$$\hat{x}^{(0)}(k) = \hat{A}e^{\hat{a}(k-1)} \quad (k=1,2,\dots,n)$$

与 GM(1,1) 模型相比,无偏 GM(1,1) 模型不存在 GM(1,1) 模型所固有的偏差,因而消除了 GM(1,1) 模型在原始数据增长率较大时失效的现象。

(三) 向量自回归(VAR)模型

向量自回归(VAR)是基于数据的统计性质建立模型,向量自回归模型是把系统中每一个内生变量作为系统中所有内生变量的滞后值的函数来构造模型。VAR 模型常用于预测相互联系的时间序列系统及分析随机扰动对变量系统的冲击。VAR(p) 模型的数学表达式为

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + B x_t + \varepsilon_t \quad t=1,2,\dots,T \quad (5)$$

其中: y_t 是 k 维内生变量; x_t 是 d 维外生变量; p 是滞后阶数; T 是样本个数。 $k \times k$ 维矩阵 $A_1 \dots A_p$ 和 $k \times d$ 维矩阵 B 是要被估计的系数矩阵。 ε_t 是 k 维

扰动向量,它们相互之间可以同期相关,但不与自己的滞后期相关及不与等式右边的变量相关,假设 Σ 是 ε_t 的协方差矩阵,是一个 $k \times k$ 的正定矩阵。^[8] 能源总消费量由煤炭、石油、天然气、电力^①等的消费量构成,其中每个因素的变化都会对其他因素产生影响,因此,可以采用 VAR 模型进行预测。

(四) 组合预测模型

组合预测模型如下:

$$Y = \omega_1 Y_{WG} + \omega_2 Y_G + \omega_3 Y_V \quad (6)$$

Y_{WG} , Y_G 和 Y_V 分别为电力消费的无偏 GM 预测模型、GM 预测模型和 VAR 预测模型, ω_1 , ω_2 和 ω_3 分别为其权重。

三、实证检验和分析

(一) 数据的选取

文章采用的一次能源总消费量(EN)、煤炭(COAL)、石油(OIL)、天然气(GAS)和电力(WATER)等的消费量数据均来自《中国统计年鉴》(1979~2008),单位是万吨标准煤。数据期间是1978年~2007年(1978年~2005年的数据用于建模,2006年和2007年实际数据不参加建模,留作检验)。为了克服数据中的异方差和数据的剧烈波动,取自然对数并用 LEN, LCOAL, LOIL, LGAS 和 LWATER 表示,运用 Eviews6.0 软件进行分析。

(二) VAR 模型

1. 平稳性检验

由于大多数经济变量的时间序列为非平稳序列,这些时间序列的数字特征随时间的变化而变化。因此,笔者利用 DF-GLS 方法分别对上述变量进行了平稳性检验,并运用 AIC 准则来判断 DF-GLS 检验中的滞后阶数,结果如表 1 所示。

表 1 各变量的平稳性检验

变量	检验类型(C, T, L)	DF-GLS 统计量	1% 临界值	5% 临界值	检验结论
LEN	(C,0,1)	0.144 695	-2.656 915	-1.954 414	不平稳
Δ LEN ^{**}	(C,0,1)	-2.184 126	-2.656 915	-1.954 414	平稳
LCOAL	(C,0,1)	0.124 042	-2.656 915	-1.954 414	不平稳
Δ LCOAL ^{**}	(C,0,1)	-8.688 151	-2.656 915	-1.954 414	平稳
LGAS	(C,T,1)	-0.397 973	-3.770 000	-3.190 000	不平稳
Δ LGAS [*]	(C,T,1)	-6.569 967	-3.770 000	-3.190 000	平稳
LOIL	(C,T,1)	-1.779 109	-3.770 000	-3.190 000	不平稳
Δ LOIL [*]	(C,T,1)	-4.581 702	-3.770 000	-3.190 000	平稳
LWATER	(C,T,1)	-2.792 951	-3.770 000	-3.190 000	不平稳
Δ LWATER [*]	(C,T,1)	-5.335 568	-3.770 000	-3.190 000	平稳

注:(1) 符号 Δ 表示一阶差分,检验形式(C, T, L)中;(2) C 表示截距项, T 表示时间趋势项, L 表示滞后阶数;(3) * (* *) 表示以 1% (5%) 的显著水平拒绝原假设。

从平稳性检验结果可以看出,5 个变量 LEN, LCOAL, LOIL, LGAS 和 LWATER 的 DF-GLS 值均大于临界值,因此都是非平稳的,具有单位根。而经过

一阶差分后,5 个变量的 DF-GLS 值都小于临界值,因此它们都是一阶单整序列。

2. 滞后期的检验

建立 VAR 模型除了要满足平稳性条件外, 还要确定最大滞后期 P 。参考 LR(似然比)检验、赤池信息准则(AIC)和 HQ 信息准则这三种方法, 当三种方法确定的滞后期相等时, 该滞后期即为 VAR 模型的最大滞后期, 从而提高被估 VAR 模型的精确性, 从表 2 可以看出, 最大滞后期为 2。

表 2 三种方法滞后期的确定结果

滞后期	对数似然值	似然比检验	赤池信息准则	汉南-奎因信息准则
0	185.8215	NA	-13.90934	-13.83967
1	323.2733	211.4644	-22.55949	-22.14146
2	356.7176	38.58956*	-23.20905*	-22.44267*

3. VAR 模型的建立

依据表 2 的判断结果, 选定最大滞后期为 2。将 1978 年~2005 年的样本数据代入模型(1), 估计结果如下:

表 3 各方程统计检验结果

R^2 统计量	0.996685	0.992987	0.997876	0.989331	0.993507
调整 R^2 统计量	0.994476	0.988311	0.996460	0.982219	0.989179
残差平方和	0.011242	0.020415	0.010413	0.038950	0.046603
回归标准差	0.027377	0.036891	0.026348	0.050957	0.055739
F 统计量	451.0518	212.3804	704.7379	139.0987	229.5299
对数似然值	63.80786	56.05241	64.80399	47.65419	45.32193
赤池信息准则	-4.062143	-3.465570	-4.138768	-2.819553	-2.640149
施瓦茨信息准则	-3.529872	-2.933298	-3.606496	-2.287281	-2.107877
因变量均值	11.59574	11.27640	9.944475	7.797886	8.705899
因变量标准差	0.368337	0.341225	0.442841	0.382143	0.535830

表 4 总体评价指标

对数似然值	356.7176
赤池信息准则	-23.20905
施瓦茨信息准则	-20.54769

从模型的整体检验效果来看, 模型的 AIC 和 SC 值都很低, 该模型可用来进行预测。

4. VAR 模型平稳性检验

VAR 模型的单位圆曲线和全部特征根的位置图如图 1 所示。图 1 显示此 VAR 模型中不存在大于 1 的根, 是一个平稳系统。

5. 协整检验

协整(Cointegration)分析理论是近年来处理非平稳经济时间序列之间长期均衡关系的有力工具。两种最常用的检验方法是 EG 两步法和极大似然法。EG 两步法只能适合于单方程的协整检验, 而 JJ 法不仅能检验变量之间是否存在协整关系, 而且可准确确定出协整向量个数。文章采用 JJ 法检验多变量之间的协整关系, 结果如表 5 所示。

$$Y_t = \begin{bmatrix} -4.97 \\ -5.89 \\ -6.28 \\ 1.03 \\ -1.46 \end{bmatrix} +$$

$$\begin{bmatrix} 6.37 & -3.49 & -1.25 & -0.08 & -0.21 \\ 6.45 & -3.40 & -1.47 & -0.01 & -0.18 \\ 4.31 & -2.55 & -0.37 & -0.21 & -0.35 \\ 0.83 & -0.70 & 1.19 & 0.64 & 0.35 \\ 6.44 & -3.12 & -0.99 & -0.68 & 0.20 \end{bmatrix} Y_{t-1} +$$

$$\begin{bmatrix} 2.03 & -1.97 & -0.37 & -0.07 & -0.02 \\ 2.72 & -2.38 & -0.65 & -0.06 & -0.08 \\ 3.26 & -2.87 & -0.18 & -0.15 & 0.10 \\ -0.14 & -0.16 & -1.15 & 0.21 & -0.18 \\ -6.95 & 3.49 & 2.17 & 0.09 & 0.36 \end{bmatrix} Y_{t-2}$$

各方程拟合系数及总体评价指标如表 3 和表 4 所示。

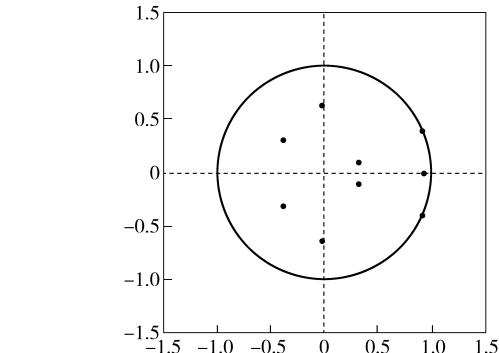


图 1 VAR 模型的特征根图

表 5 Johansen 最大特征值统计量检验结果

原假设 H_0	特征根	最大特征值统计量	5% 临界值	P 值
0 个协整向量*	0.832100	46.39411	37.16359	0.0034
至少 1 个协整向量	0.640820	26.62226	30.81507	0.1496
至少 2 个协整向量	0.501747	18.11282	24.25202	0.2628
至少 3 个协整向量	0.246365	7.354036	17.14769	0.6742
至少 4 个协整向量	0.125967	3.500552	3.841466	0.0613

注: * 表示在 5% 的显著性水平下拒绝原假设。

协整检验从检验不存在协整关系这一原假设开始逐步检验。从表 5 可以看出变量之间存在协整关系, 最大特征根检验认为存在一个协整向量, 其对应的协整方程形式为

$$\begin{aligned}\hat{L}ENERGY = & 0.7007 LCOAL + 0.1934 LOIL + \\& (0.00374) \quad (0.00825) \\& 0.1536 LGAS + 0.1889 LWATER \\& (0.00598) \quad (0.00514)\end{aligned}$$

圆括号内为标准差。

(三) 组合模型

在用 VAR 模型计算的基础上, 再用 1978 年~2005 年的电力消费量进行灰色拟合, 最后用组合模型进行预测。在组合模型中, 按照残差越小, 模型权重越大的原则, 令

$$\begin{aligned}\omega'_1 &= \frac{\sum_{i=1}^{27} e_{Vi}^2}{\sum_{i=1}^{27} e_{Vi}^2 + \sum_{i=1}^{27} e_{Gi}^2 + \sum_{i=1}^{27} e_{WGi}^2} = 0.462675 \\&\omega'_2 = \frac{\sum_{i=1}^{27} e_{Gi}^2}{\sum_{i=1}^{27} e_{Vi}^2 + \sum_{i=1}^{27} e_{Gi}^2 + \sum_{i=1}^{27} e_{WGi}^2} = 0.272425306 \\&\omega'_3 = \frac{\sum_{i=1}^{27} e_{WGi}^2}{\sum_{i=1}^{27} e_{Vi}^2 + \sum_{i=1}^{27} e_{Gi}^2 + \sum_{i=1}^{27} e_{WGi}^2} = 0.264900\end{aligned}$$

取 ω_1, ω_2 和 ω_3 分别为

$$\omega_1 = \max \{ \omega'_1, \omega'_2, \omega'_3 \} = 0.462675$$

$$\omega_2 = \max \{ \omega'_1, \omega'_2 \} = 0.272425306$$

$$\omega_3 = \max \{ \omega'_3 \} = 0.264900$$

最终组合模型为

$$\hat{Y} = 0.462675 Y_{WG} + 0.272425306 Y_G + 0.264900 Y_V$$

\hat{Y}, Y_{WG}, Y_G 和 Y_V 分别为水电消费的组合拟合量、灰色无偏 GM 拟合量、灰色 GM 拟合量和 VAR 拟合量。组合模型的拟合平均相对误差为 4.88%, 拟合精度达到了 95.12%, 拟合精度很高, 可以进行拟合, 拟合结果如图 2 所示。

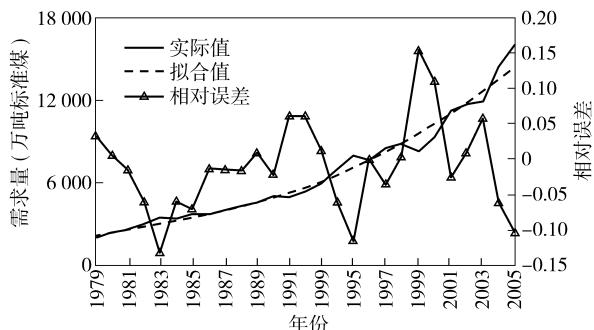


图 2 水电实际消费值与拟合值对比图

以 2006 年和 2007 年的电力消费数对拟合结果进一步检验, 2006 年和 2007 年两年的实际值为 17 731.4

和 19 387.6 万吨标准煤, 拟合值为 15 394.89 和 16 442.34 万吨标准煤, 由此计算的误差均方根为 2 658.36, 绝对误差平均为 2 640.88, 相对误差绝对值平均为 14.18, Theil 不等系数仅为 0.077042。

该模型较好地反映了数据的变化, 用 VAR 和 GM 组合模型预测了 2008 年~2015 年的电力消费量。由表 6 可以发现, 2008 年~2012 年中国电力消费量分别比 2005 年增长 9.29%, 16.89%, 25.09%, 33.91% 和 43.35%, 电力消费量不断攀升。

表 6 中国电力消费量预测 万吨标准煤

年份	2008	2009	2010	2011
预测值	17 572.57	18 793.96	20 112.38	21 530.83
年份	2012	2013	2014	2015
预测值	23 049.46	24 666.27	26 378.34	28 183.65

四、结论及建议

文章将 VAR 模型与灰色系统理论的两种模型相结合, 建立 VAR 和 GM 优化组合模型, 充分利用了灰色预测建模所需信息少、方法简单的优点以及 VAR 模型进行相关经济指标预测的使用方便、容易操作的特点, 对中国电力消费进行拟合, 结果表明组合模型具有较高的拟合精度, 可以作为未来电力消费预测的有效工具, 并对 2008 年~2015 年的中国电力消费进行了预测。预测结果表明, 中国电力未来消费量巨大, 这一方面说明随着中国工业化和居民生活水平的提高, 对电力的消费呈增长态势, 另一方面, 也涉及如何避免电力短缺的电力安全问题, 这也是一个挑战。因此, 在满足生产和生活的前提下, 合理生产和消费电力已日渐成为一个重要的议题。

注释:

① 据中国统计年鉴能源消费总量及构成表, 此处电力系指水电、风电与核电的消费量之和。

参考文献:

- [1] 徐明德, 李维杰. 线性回归分析与能源需求预测 [J]. 内蒙古师范大学学报: 自然科学版, 2003, 32(1): 17~20.
- [2] 林伯强. 结构变化、效率改进与能源需求预测——以中国电力行业为例 [J]. 经济研究, 2003(5): 57~63.
- [3] 黄俊, 周猛, 王俊海. ARMA 模型在我国能源消费预测中的应用 [J]. 统计与决策, 2004(12): 49~50.
- [4] 李亮, 孙廷容, 黄强, 等. 灰色 GM(1,1) 和神经网络组合的能源预测模型 [J]. 能源研究与利用, 2005(1): 10~13.
- [5] 付加锋, 蔡国田, 张雷. 基于 GM 和 BP 网络的我国能源消费量组合预测模型 [J]. 水电能源科学, 2006, 24(2): 1~3.
- [6] 刘思峰, 谢乃明. 灰色系统理论及其应用 [M]. 北京: 科学出版社, 2005: 130.
- [7] 吉培荣, 黄巍松, 胡翔勇. 无偏灰色预测模型 [J]. 系统工程与电子技术, 2000, 22(6): 6~8.
- [8] 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: 249.