

Study on Energy Consumption Prediction in Yunnan Province Based on Combined Prediction Model

Xiaoxiao Yin

School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming Yunnan
Email: yinxiaoxiao2011@163.com

Received: Mar. 1st, 2015; accepted: Mar. 12th, 2015; published: Mar. 19th, 2015

Copyright © 2014 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

There is a strong relationship between energy and economic development of China, the protection of human health and the environment. It is essential to accurately predict energy consumption. This can provide a scientific basis for decision-making about energy work. Based on the historical data of energy consumption in Yunnan Province, on the basis of the three single prediction model, we establish combined prediction model to make energy predictions in Yunnan Province by non-optimal weighted linear combination model and minimum sum of square error of objective function for optimal weight linear combination model. The prediction results of these models are analyzed and compared. The results show that combined prediction model is performed better. This indicates that the combined prediction model is an useful theoretical tool for energy prediction in Yunnan Province.

Keywords

Single Prediction Model, Combined Prediction Model, Energy Consumption, Predictions

基于组合预测模型的云南省能源消费预测研究

尹潇潇

云南财经大学统计与数学学院, 云南 昆明
Email: yinxiaoxiao2011@163.com

收稿日期：2015年3月1日；录用日期：2015年3月12日；发布日期：2015年3月19日

摘要

能源问题关系着我国经济的发展、环境的保护以及人类的健康，准确预测能源消费总量至关重要，可以给能源工作决策提供科学依据。本文结合云南省能源消费总量的历史数据，在三个单一预测模型的基础上，基于均方误差的非最优权重线性组合法和最小预测误差平方和为目标函数的最优权重线性组合法构建云南省能源消费总量的组合预测模型。通过比较上述预测结果，发现组合预测模型具有更好的预测效果，研究表明该模型对云南省能源消费预测有重要的理论与现实意义。

关键词

单一预测模型，组合预测模型，能源消费，预测

1. 引言

近年来，随着我国经济的迅猛发展，人类对能源的需求量在不断的增加，能源消耗量也随之与日俱增，能源消费量与碳排放存在紧密的联系。据有关资料显示，世界能源需求的80%~85%来源于化石燃料，而每年使用化石燃料产生的碳排放量约占大气中碳排放总量的70%，化石燃料燃烧是目前大气中碳增加的首要原因[1]。大气中碳含量的增加，必然导致人们的健康受损，生态环境也会遭到破坏。为了降低碳排放，从能源本身来讲需要减少能源的消耗量。因此，做好能源消费总量预测研究工作，不仅为相关部门制定科学的能源战略规划和合理的能源消费政策提供依据，也有助于保护生态环境和人类的健康，促进我国经济健康、持续的发展，这具有非常重要的现实意义和战略意义。

一直以来，能耗问题备受大家的关注，国内很多学者对我国能源消费总量问题进行研究分析，他们采用的预测方法各式各样。目前组合预测模型备受欢迎，建立组合预测模型最关键的是确定各单一模型的权重系数。确定权重的方法有很多，如黄宜[2]等建立指数平滑模型和系统动力学模型分别对云南省能源消费进行预测，然后运用合作博弈中的Shapley值方法确定综合预测模型中各单一模型权重(通过分配总误差)，由于Shapley值法具有对合作的各方公平且易接受的特点，故本文采用该方法确定预测模型的权重能够提高预测的可靠性；邓鸿鹄[3]根据不同权重系数计算方法，构建了基于均方误差非最优权重组合法和最小误差平方和为目标函数最优权重线性组合法的组合模型，分别对北京市能源消费进行预测，结果表明组合模型提高了预测精度且前者组合模型预测效果没有后者好；常建娥和蒋太立[4]运用层次分析法确定组合预测模型中各单一模型的权重系数，该方法思想简洁，灵活性大，主要用于多目标决策中，特别适用于缺少定量数据进行定性判断的情况；熊浩云[5]等分别使用神经网络和时间序列法对我国能源消费总量进行预测，然后依据标准差法对各模型所占权重分配，进而建立我国能源消费的组合预测模型，组合预测模型不但克服了单一模型的缺陷，而且提高了预测精度；张志朝[6]综合各个模型的优缺点，使组合预测误差平方和达到最小，从而确定各单一预测方法的权重系数，得到最优线性组合预测模型，最终对安徽省人均GDP进行预测，研究结果表明该模型预测可靠性较高；江正华[7]在网络信息资源评价的应用研究中，基于最小方差原理，把运用熵权模糊综合评判法求得的客观权重与运用层次分析法求得的主观权重进行最优优化组合，计算出组合权重，从而修正了单一模型所确定的单纯主观或客观权重这样就能够根据专家调查的指标得分，建立基于方差的最优组合赋权网站评价模型，使评判结果更具科学、合理。

基于能源对人类的重要性，做好云南省未来能源消费总量预测工作意义重大，可以为云南省政府制

定合理的能源战略及规划提供科学依据，保障云南省能源的合理利用以及健康稳定的发展。本文在一些学者研究的基础上，建立了基于几种单一预测模型的非最优权重线性组合模型，并通过均方误差确定各模型的权重。为了提高预测精度，综合考虑各单一模型的优缺点，采用最小误差平方和法确定这些单一模型的权重系数，最终得到最优权重线性组合预测模型，然后对云南省未来能源消费总量进行预测，其预测效果优于各单一模型和非最优权重线性组合预测模型。

2. 模型及方法

2.1. 组合预测模型的基本介绍

所谓组合预测模型，就是在几种单一预测模型的基础上，采用赋权方式构建模型。由于单一预测模型往往只能反映能源系统的局部信息，而将其进行有机组合，可以弥补单一模型的不足，提高预测精度，较全面的描述能源系统的趋势。本文就是采用组合预测模型对云南省能源消费总量进行预测分析的，这可使所预测的信息更加全面，预测的结果更加可靠。

然而建立组合预测模型的关键在于各单一模型在组合模型中所占权重的计算，目前计算权重的方法有很多，如层次分析法、Shapley 值法、标准差法和最小化方差法等等。根据不同权重系数的计算方法，组合预测模型可划分为两种，即非最优权重线性组合预测模型和最优权重线性组合预测模型。本文则分别采用基于均方误差的非最优权重线性组合法和基于最小预测误差平方和为目标函数的最优权重线性组合法来构建云南省能源消费总量的组合预测模型。

2.2. 非最优权重线性组合预测模型

参考文献[3]，知非最优权重线性组合模型是通过均方误差来计算各单一模型权重系数的一种组合预测模型，其基本原理是：对均方误差较小的单一预测模型可以赋予较大的权重，而对均方误差较大的单一预测模型则赋予较小的权重。该组合预测模型的构建过程如下：

- 1) 分别计算各单一预测模型的均方误差 MSE_i ：

$$MSE_i = \sqrt{\frac{1}{T} \sum_{t=1}^T e_{it}^2}$$

其中： e_{it} 为第 i 个模型在第 t 期的预测误差； T 为预测期数； MSE_i 为第 i 个模型的均方误差。

- 2) 分别计算各单一预测模型的权重系数 ω_i ：

$$\omega_i = \frac{MSE - MSE_i}{MSE \times (N - 1)}$$

其中： $MSE = \sum_{i=1}^N MSE_i$ ； ω_i 表示第 i 个模型在组合模型中所占的权重，且各个权重必须满足条件 $\begin{cases} 0 < \omega_i < 1 \\ \sum_{i=1}^N \omega_i = 1 \end{cases}$ ，

N 表示单一预测模型的个数。

- 3) 非最优权重线性组合预测模型为：

$$\widehat{y}_t = \sum_{i=1}^N \omega_i \widehat{y}_{it}$$

其中： \widehat{y}_t 为组合预测模型的拟合序列； \widehat{y}_{it} 为第 i 个单一预测模型的拟合序列。

2.3. 最优权重线性组合预测模型

根据文献[3]，获悉最优权重线性组合预测模型是通过使组合预测误差平方和最小，确定各单一预测

模型权重系数得到的组合模型。该模型构建的基本原理是：通过赋予各单一预测模型合适的权重后，使得组合模型得到的预测值与实际值之间的误差平方和达到最小。具体建模过程如下：

令实际的能源消费总量序列为 y_t ； \widehat{y}_t 表示组合预测模型的拟合序列；第 i 个单一预测模型的拟合序列记为 \widehat{y}_{it} ， ω_i 表示第 i 个模型在组合模型中所占的权重。实际值与组合预测模型拟合值之间的误差序列记为 $e_t = y_t - \widehat{y}_t$ ，权重系数矩阵 $W = (\omega_1, \omega_2, \omega_3, \dots, \omega_N)^T$ ，其中 N 表示单一预测模型的个数， $\omega_1, \omega_2, \omega_3, \dots, \omega_N$ 分别表示每个单一预测模型的权重系数。该组合预测模型记作：

$$\widehat{y}_t = \sum_{i=1}^N \omega_i \widehat{y}_{it} = (\widehat{y}_{1t}, \widehat{y}_{2t}, \dots, \widehat{y}_{Nt}) W$$

则实际值与模型拟合值的误差序列为：

$$e_t = y_t - \widehat{y}_t = y_t - \omega_1 \widehat{y}_{1t} - \omega_2 \widehat{y}_{2t} - \omega_3 \widehat{y}_{3t} - \dots - \omega_N \widehat{y}_{Nt} \quad (1)$$

对(1)整理得：

$$e_t = \omega_1 e_{1t} + \omega_2 e_{2t} + \omega_3 e_{3t} + \dots + \omega_N e_{Nt} = (e_{1t}, e_{2t}, e_{3t}, \dots, e_{Nt}) W$$

建立目标函数：误差平方和 $e_t^T e_t = [(e_{1t}, e_{2t}, e_{3t}, \dots, e_{Nt}) W]^T (e_{1t}, e_{2t}, e_{3t}, \dots, e_{Nt}) W$

现在最关键的是求解一个 $W = (\omega_1, \omega_2, \omega_3, \dots, \omega_N)^T$ ，使误差平方和 $e_t^T e_t$ 达到最小，且权重系数矩阵

$W = (\omega_1, \omega_2, \omega_3, \dots, \omega_N)^T$ 需满足： $\sum_{i=1}^N \omega_i = 1, 0 \leq \omega_i \leq 1$ 。

上述问题显然是一个最优化问题，因此，使用 matlab 软件可以求得单一预测模型的权重系数矩阵

$W = (\omega_1, \omega_2, \omega_3, \dots, \omega_N)^T$ ，从而得到最优权重线性组合预测模型 $\widehat{y}_t = \sum_{i=1}^N \omega_i \widehat{y}_{it}$ 。

3. 实证分析

3.1. 数据来源及说明

本文从 2011 和 2012 年《云南省统计年鉴》[8]获取了 1978 到 2011 年云南省的能源消费总量数据，我们用 1978~2007 年的数据建立预测模型，而 2008~2011 年的数据用于检验所建模型的拟合精度，从而确定预测效果最好的模型。

令云南省 1978~2007 年实际能源消费总量序列为 y_t ，其中 $t = 1, 2, \dots, 30$ (1978 年对应 $t = 1$)，时间 t 为自变量， y_t 为因变量。

3.2. 建立云南省能源消费总量预测模型

3.2.1. 建立单一预测模型

运用 SPSS 软件绘制 1978 年~2007 年云南省能源消费总量与时间 t 的散点图，见图 1。

观察散点图，根据云南省能源消费总量的变化趋势，尝试建立二次函数模型、三次曲线模型和指数模型对其预测分析。

对云南省能源消费总量做上述三种回归分析，我们可以得到以时间 t 为自变量，云南省能源消费总量 y_t 为因变量的预测模型表达式为：

二次函数模型：

$$y_{1t} = 8.672t^2 - 84.407t + 1293.945 \quad (2)$$

三次曲线模型：

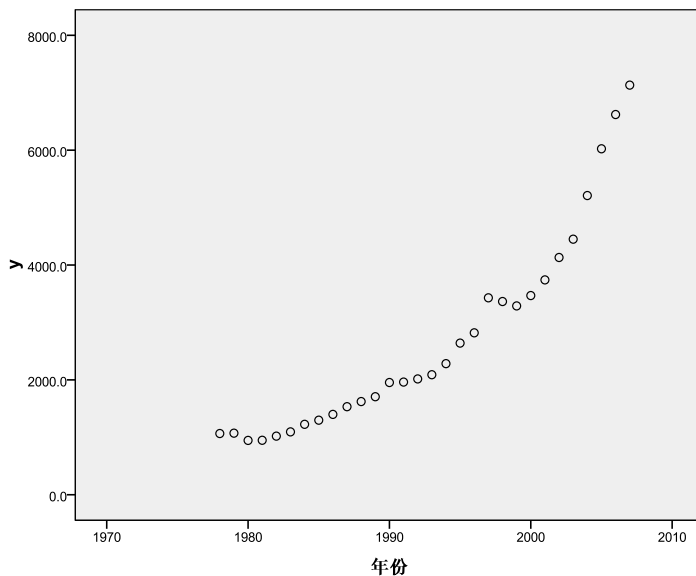


Figure 1. Scatter graph of total energy consumption
图 1. 能源消费总量的散点图

$$y_{2t} = 0.367t^3 - 8.371t^2 + 130.407t + 694.035 \quad (3)$$

指数模型:

$$y_{3t} = 770.441e^{0.069t} \quad (4)$$

其中上述三个式子中, $t = 1, 2, \dots, 30$ 。

根据(2)、(3)和(4)式, 可以分别求得二次函数模型、三次曲线模型和指数模型对云南省能源消费总量的拟合值 \hat{y}_{1t} 、 \hat{y}_{2t} 以及 \hat{y}_{3t} 。

根据云南省能源消费总量的真实值与预测值以及公式(5), 可以求得各个模型的平均相对误差绝对值, 即 $MAPE_1 = 8.6\%$, $MAPE_2 = 7.11\%$, $MAPE_3 = 6.55\%$, 且 $MAPE_3 < MAPE_2 < MAPE_1$ 。

$$\text{相对误差绝对值} = |\text{预测值} - \text{真实值}| \div \text{真实值} \times 100\% \quad (5)$$

3.2.2. 建立组合预测模型

虽然上面几种曲线模型中指数模型的拟合效果相对较好, 但无论我们选择哪个模型, 最终都会损失部分有用的信息。因此, 我们考虑给这些简单的模型赋予合适的权重使其有机结合, 构建组合预测模型。这就能够克服单一预测模型的缺陷, 提高预测精度, 充分利用简单模型提供的全部信息, 准确分析现象存在的规律, 使得预测效果优于单一模型。下面分别基于二次函数模型、三次曲线模型和指数模型这三个单一预测模型, 建立非最优和最优权重线性组合模型对云南省能源消费总量进行预测。

由于本文使用的数据为时间序列数据, 而且根据图 1, 我们发现云南省能源消费总量随时间基本趋于稳定的变化, 因此, 我们不妨做如下假设:

假设一: 云南省能源消费总量随时间变化相对稳定;

假设二: 云南省产业结构变化、技术进步等因素不会引起能源消费的显著变化。

1) 建立非最优权重线性组合预测模型

现在对二次函数模型、三次曲线模型以及指数模型赋予合适的权重, 建立组合预测模型。根据 2.2 对非最优权重线性组合预测模型建立方法的介绍, 则该组合预测模型的建模过程如下:

a) 根据公式 $MSE_i = \sqrt{\frac{1}{T} \sum_{t=1}^T e_{it}^2}$ ($i=1,2,3, T=30$), 分别计算二次函数模型、三次曲线模型以及指数模型的均方误差为: $MSE_1 = 274.79, MSE_2 = 202.67, MSE_3 = 323.93$ 。

b) 根据公式 $\omega_i = \frac{MSE - MSE_i}{MSE \times (N - 1)}$ ($i=1,2,3, MSE = \sum_{i=1}^N MSE_i = 801.39, N=3$), 分别计算二次函数模型、三次曲线模型以及指数模型在组合模型中所占的权重系数为: $\omega_1 = 0.3286, \omega_2 = 0.3736, \omega_3 = 0.2978$, 且各个权重满足条件 $0 < \omega_i < 1, \omega_1 + \omega_2 + \omega_3 = 1$ 。

c) 由 a)和 b), 得到非最优权重线性组合预测模型为:

$$\widehat{y}_t = 0.3286\widehat{y}_{1t} + 0.3736\widehat{y}_{2t} + 0.2978\widehat{y}_{3t} \quad (6)$$

其中: \widehat{y}_{1t} 为二次函数模型的拟合序列; \widehat{y}_{2t} 为三次曲线模型的拟合序列; \widehat{y}_{3t} 为指数模型的拟合序列。

根据(6)式, 可以得到云南省 1978~2007 年能源消费总量的拟合序列, 从而计算出非最优权重线性组合预测模型的平均相对误差绝对值为 $MAPE_4 = 6.07\%$ 。

2) 建立最优权重线性组合预测模型

根据 2.3 对最优权重线性组合预测模型的介绍, 我们知道该方法的基本原理是以误差平方和作为目标函数, 在一定的约束条件下, 求得该目标函数的最小值, 从而确定各单一模型的权重系数。该组合预测模型的建模过程如下:

令 1978 年~2007 年云南省能源实际消费总量序列为 y_t , 二次模型的拟合序列为 \widehat{y}_{1t} , 三次模型的拟合序列为 \widehat{y}_{2t} , 指数模型的拟合序列为 \widehat{y}_{3t} , 组合模型的拟合序列为 \widehat{y}_t , 则实际值与模型拟合值的误差序列为 $e_t = y_t - \widehat{y}_t$; $\omega_1, \omega_2, \omega_3$ 分别表示二次曲线模型、三次曲线模型和指数模型的权重系数, $0 \leq \omega_i \leq 1$ ($i=1,2,3$)且 $\sum_{i=1}^3 \omega_i = 1$, 权重系数矩阵记作 $W = (\omega_1, \omega_2, \omega_3)^T$ 。现需求解以下最优化问题:

$$\begin{cases} \min e_t^T e_t = [(e_{1t}, e_{2t}, e_{3t})W]^T (e_{1t}, e_{2t}, e_{3t})W \\ \sum_{i=1}^3 \omega_i = 1 \\ 0 \leq \omega_i \leq 1 \end{cases}$$

预测云南省能源消费总量的组合预测模型表达式为:

$$\widehat{y}_t = \omega_1 \widehat{y}_{1t} + \omega_2 \widehat{y}_{2t} + \omega_3 \widehat{y}_{3t} = (\widehat{y}_{1t}, \widehat{y}_{2t}, \widehat{y}_{3t})W$$

则实际值与模型拟合值的误差序列为:

$$e_t = y_t - \widehat{y}_t = y_t - \omega_1 \widehat{y}_{1t} - \omega_2 \widehat{y}_{2t} - \omega_3 \widehat{y}_{3t} \quad (7)$$

对(7)整理得:

$$e_t = \omega_1 e_{1t} + \omega_2 e_{2t} + \omega_3 e_{3t} = (e_{1t}, e_{2t}, e_{3t})W$$

使用 matlab 软件求得单一预测模型的权重系数矩阵 $W = (-0.304, 1.225, 0.079)^T$

从而得到最优权重线性组合预测模型为:

$$\widehat{y}_t = -0.304\widehat{y}_{1t} + 1.225\widehat{y}_{2t} + 0.079\widehat{y}_{3t} \quad (8)$$

通过模型(8), 对云南省 1978~2007 年的能源消费总量进行预测, 得到预测值, 然后根据输出结果, 计算出组合预测模型的平均相对误差绝对值为 $MAPE_5 = 8.1\%$ 。

3.2.3. 单一预测模型与组合预测模型拟合精度比较

上文通过建立三种单一预测模型即二次函数模型、三次曲线模型以及指数模型和两种组合预测模型(非最优权重和最优权重线性组合模型)分别对 1978 年至 2007 年云南省能源消费总量进行预测分析。各个预测模型的平均相对误差绝对值见表 1。

由表 1 可知, 非最优权重线性组合预测模型的拟合误差小于三种单一预测模型; 虽然最优权重线性组合预测模型的拟合误差与二次函数模型和三次曲线模型差不多, 且明显比指数模型的预测误差大, 但是通过各个模型的拟合结果, 可以得到它的误差平方和为 1299694.846, 小于其他三个单一预测模型相应的误差平方和。综上考虑, 组合预测模型的拟合精度较高, 其预测结果可靠。

现在通过单一预测模型和组合预测模型分别对 2008 年~2011 年云南省能源消费总量进行预测, 由此进一步比较简单模型和组合模型的拟合精度。预测结果见表 2。

由表 2 可以发现: 非最优和最优权重线性组合预测模型对 2008~2011 年云南省能源消费总量的拟合误差都比二次函数模型和指数模型的小。根据表 2 中的相对误差绝对值, 可以计算得到二次函数模型、三次曲线模型、指数模型、非最优权重线性组合模型以及最优权重线性组合模型拟合的平均相对误差绝对值分别为 8.3%、3.3%、13.7%、5.6%和 5.5%。从平均相对误差来看, 虽然三次曲线模型的相对小于组合预测模型, 但由于单一模型不能够全面反映所研究的问题, 因而综合理论与实际考虑, 组合预测模型的拟合效果优于简单模型。本文中两种组合预测模型相比较而言, 最优权重线性组合模型预测精度高于非最优权重线性组合模型。

4. 组合预测模型对未来云南省能源消费总量预测的应用

综上分析可知组合预测模型的预测效果整体上要优于单一预测模型, 现在选用最优权重线性组合模型对云南省未 2015 年~2019 年的能源消费总量进行预测分析。预测结果见表 3。

5. 结论

通过上面的实证分析, 可以得到以下结论:

- 1) 从理论与实际意义上考虑, 组合预测模型综合了各个简单模型的全部有用信息, 克服了单一预测模型的缺陷, 提高了预测的准确程度, 能够较全面的描述问题;
- 2) 未来几年内云南省能源消费总量在持续不断的增加, 到 2020 年将达到 16982.4 万吨标准煤。随着

Table 1. The absolute value of the average relative error

表 1. 平均相对误差绝对值

方程	二次	三次	指数	非最优	最优
平均相对误差绝对值(%)	8.6	7.11	6.55	6.07	8.1

Table 2. The prediction, actual, absolute value of the relative error from 2008 to 2011

表 2. 2008~2011 年预测值和真实值以及相对误差绝对值

年份	实际值	二次预测值	相对误差绝对值(%)	三次预测值	相对误差绝对值(%)	指数预测值	相对误差绝对值(%)	非最优组合模型预测值	相对误差绝对值(%)	最优组合模型预测值	相对误差绝对值(%)
2008	7510.82	7011.12	6.7	7625.42	1.5	6541.77	12.9	7100.85	5.5	7726.56	2.9
2009	8032.06	7473.05	7.0	8321.01	3.6	7009.09	12.7	7651.68	4.7	8475.15	5.5
2010	8674.17	7952.32	8.3	9070.33	4.6	7509.79	13.4	8238.22	5.0	9286.92	7.1
2011	9540.28	8448.94	11.4	9875.57	3.5	8046.26	15.7	8862.01	7.1	10164.7	6.5

Table 3. Prediction results of energy consumption in Yunnan Province from 2015 to 2019
表 3. 2015~2019 年云南省能源消费总量预测结果

年份	2015	2016	2017	2018	2019	2020
能源消费总量预测值(万吨标准煤)	11111.53	12130.19	13223.65	14394.86	15646.78	16982.4

经济的快速发展,能源需求也在不断的增加,而能源的过度消耗,将导致云南省未来能源供求紧张,生态环境遭到破坏,人们的健康也会受损。从长远来看,如果不适当控制能源的消耗,将会给国家的经济社会发展和人类的健康带来灾难。因此,准确的预测能源消费总量意义重大。

鉴于组合预测模型的预测准确度较高,故而可以将其用于其它领域的研究。为了实现国家经济健康、稳定、可持续发展,保护我们的生态环境和健康,我们应控制能源消费总量的过快增长。通过提高能源的利用率、开发可再生资源以及使用低碳能源等方法都可以减少能源的消耗。

致 谢

本文在云南财经大学统计与数学学院石磊教授的指导和帮助下,得以顺利完成,在此对石磊教授表示衷心的感谢!

参考文献 (References)

- [1] 沈满洪, 苏小龙 (2013) 能源低碳化研究文献评述. *低碳经济*, **1**, 49.
- [2] 黄宜 (2013) 基于 Shapley 值的云南省能源消费综合预测方法研究. *能源研究与信息*, **1**, 57-60.
- [3] 邓鸿鹄 (2013) 北京市能源消费预测方法比较研究. 硕士学位论文, 北京林业大学, 北京.
- [4] 常建娥, 蒋太立 (2007) 层次分析法确定权重的研究. *武汉理工大学学报·信息与管理工程版*, **1**, 153-156.
- [5] 熊浩云 (2010) 应用组合模型对我国能源消费的预测. *科学技术与工程*, **17**, 4268-4282.
- [6] 张志朝 (2009) 组合预测模型在安徽省人均 GDP 预测中的应用. *广西财经学院学报*, **5**, 32-35.
- [7] 江正华 (2014) 基于方差的最优组合赋权模型在网络信息资源评价中的应用. *计算机与应用*, **1**, 302-308.
- [8] 云南省统计局 (2011, 2012) 云南省统计年鉴. 中国统计出版社, 北京.