

基于K-L信息量法的安徽省工业用电量预测

石雪梅¹,葛斐¹,肖夕林²

(1. 国网安徽省电力公司 经济技术研究院, 安徽 合肥 230022; 2. 合肥工业大学 管理学院, 安徽 合肥 230009)

Forecasting Industrial Electricity Consumption of Anhui Province Based on K-L Information Method

SHI Xuemei¹, GE Fei¹, XIAO Xilin²

(1. Institute of Economic and Technology, State Grid Anhui Electric Power Company, Hefei 230022, Anhui, China;
2. School of Management, Hefei University of Technology, Hefei 230009, Anhui, China)

ABSTRACT: In view of the correlation between the regional industrial electricity consumption and economic indicators, the K-L information method is used to select the leading economic indicators which can indicate the change trend of regional electricity consumption on the monthly scale and the leading time of each leading indicator is calculated. Taking the economic leading indicator as the independent variable, the regional industrial electricity consumption as the dependent variable the multiple regression model is established and the best fitting equation is selected according to the AIC criterion and BIC criterion and the industrial electricity consumption forecast model is obtained. The model is used to predict monthly electricity consumption in Anhui Province from May 2014 to December 2014, and the results show that the prediction accuracy is high, the forecasting method can be used for industrial electricity consumption forecast.

KEY WORDS: economic leading indicators; K-L information method; electricity consumption forecasting; Anhui province

摘要: 针对区域工业用电量与经济指标的相关性, 运用K-L信息量法, 在月度尺度上筛选出能够指示区域工业用电量变化趋势的经济先行指标, 并获得各先行指标的先行期数。以经济先行指标为自变量、区域工业用电量为被因变量建立多元回归模型, 根据AIC准则和BIC准则选取最佳拟合方程, 得到工业用电量预测模型。运用模型预测安徽省2014年5月-12月的月用电量, 结果显示预测精度较高, 预测方法可以用于工业用电量预测。

关键词: 经济先行指标; K-L信息量法; 用电量预测; 安徽省

对于省级电力公司来说, 提高区域用电量预测的科学性和精确度, 具有重要意义。电力消费对国民经济变动非常敏感, 宏观经济运行状况对于区域

用电量具有一定的指示作用^[1-5]。为了从经济指标中挖掘与用电量变化有关的信息, 基于统计学方法的数据挖掘技术受到广泛应用^[6-8]。从已有文献来看, 传统的预测方法已经很难做出改进, 只有改进月度用电量的预测模式, 尽可能多地引入最新的经济信息参与预测, 才能进一步提高预测精度。

但是, 对经济先行指标筛选, 以及建模过程中指标的筛选过于依靠主观判断等问题。本研究将延续以上思路, 基于经济先行指标和多元回归模型建立区域用电量预测模型, 但在指标选取、先行期数判断和多元回归模型构建方面提出改进方案, 以增强模型的全面性、科学性和模型预测结果的精确性。

1 经济先行指标的筛选

K-L信息量法是在20世纪50年代由Kull-back和Leibler提出^[9], 用以衡量2个概率分布的相似程度, 即度量他们之间的距离, 越小即代表两者越接近越相似。其原理是以基准序列为理论分布, 备选指标为样本分布, 不断变化备选指标与基准序列时差, 计算K-L信息量。K-L信息量最小时对应的时差数确定为备选指标的最终时差。

对于离散变量, 设基准指标为 $y=\{y_1, y_2, \dots, y_n\}$, 由于任意满足 $P_i>0$, $\sum_{i=1}^n p_i=1$ 的序列 p 均可视为某随机变量的概率分布序列, 因此, 基准指标序列记为 p , 即

$$p_i=y_i/\left(\sum_{i=1}^n y_i\right), t=1, 2, \dots, n \quad (1)$$

设备选指标为 $x=\{x_1, x_2, \dots, x_n\}$, 序列记为 q , 即

$$q_t = x_t / \left(\sum_{i=1}^n x_i \right), t=1, 2, \dots, n \quad (2)$$

K-L信息量的计算公式为

$$I(p, q) = \sum_{i=1}^n p_i \ln \frac{p_i}{q_i} \quad (3)$$

当备选指标序列 x 与基准指标序列 y 完全一致时, K-L信息量等于0; 指标 x 与基准指标 y 越接近, K-L信息量绝对值越小, 越接近于0。

本研究以安徽省工业用电量为基准指标, 在全球、全国和区域尺度大范围搜集月度经济指标, 并运用K-L信息量法筛选安徽省工业用电量的经济先行指标, 并建立预测模型。

2 模型框架与实现

2.1 模型框架

图1所示为预测模型的概念框架, 根据框架, 模型建立的步骤如下: 1) 选取区域、全国及世界经济的指标, 作为备选指标, 获取经济指标的月度数据, 将数据分为两部分, 一部分数据用于建模, 另一小部分进行预测检验。2) 以区域用电量指标为基准指标, 扣除季节性因素的影响, 对所有备选指标进行-12阶~+12阶的滞后处理, 计算基准指标与所有备选指标不同滞后项的K-L信息量, 对每个备选指标选择最小的K-L信息量所对应的滞后阶数, 得到该备选指标与基准指标对应的最佳滞后阶数。如果某指标最佳滞后阶数小于0, 从概率上说, 该指标更有可能超前于基准指标发生变化, 可认为是基准指标的先行指标。3) 对先行指标进行滞后处理, 然后计算其与基准指标的相关性系数, 保留显著相关的先行指标。4) 以区域用电量指标为因变量, 以先行指标的最佳滞后项为解释变量, 建立时间序列多元回归模型, 根据AIC与BIC等准则, 以及对解释变量的方差膨胀因子的控制, 选取最优模型。5) 根据算例分析, 对模型进行检验与评价, 根据所选取的最优模型, 进行拟合与预测检验, 判断模型预测结果的精确度。

2.2 模型优选法

准则函数方法是最常用的模型优选方法, 例如AIC (Akaike information criterion) 准则^[10]和BIQ (Bayesian information criterion) 准则^[11]。通常在应用准则函数进行模型选择时, 将准则值最小的候选模型作为最

优模型。在一般的情况下, AIC值可以由式4) 计算得到:

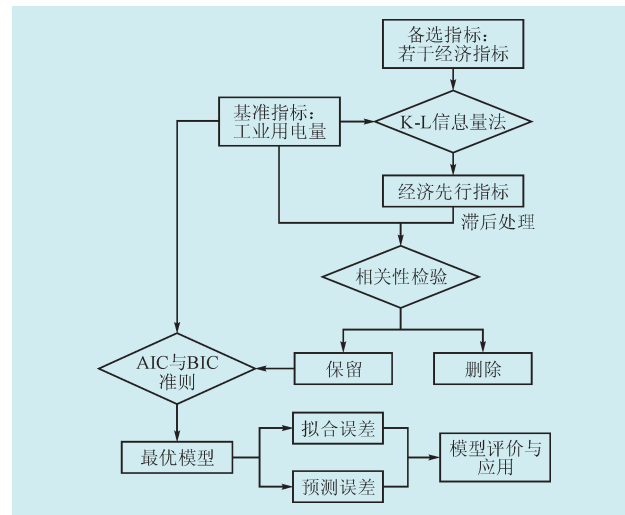


图1 预测模型框架

Fig. 1 Framework of forecasting model

$$AIC = 2k - 2\ln(L) \quad (4)$$

式中: k 为参数的数量; L 为似然函数。假设条件是模型的误差服从独立正态分布。AIC值越小, 模型越好, 越值得考虑, 这就是所谓的AIC准则。一般情况下, 当参数数量 k 增加时, 对数似然函数也将增加, 从而使AIC值变小。但当 k 过大时, 对数似然函数增速减缓, 导致AIC值反而增加, 使得模型变坏, 可见AIC准则有效且合理地控制了参数的数量 k 。

与AIC准则相同, BIC准则是衡量统计模型拟合优良性的另一种标准。BIC值的计算式为:

$$BIC = -2\ln(L) + k \times \ln(n) \quad (5)$$

式中: L 为在该模型下的最大似然; n 为数据数量; k 为模型的变量个数。同AIC值相似, 模型的BIC值越小, 越值得考虑。

多重共线性是指线性回归模型中的解释变量之间由于存在精确相关关系或高度相关关系而使模型估计失真或难以估计准确。方差膨胀因子VIF (variance inflation factor) 是刻画多重共线性的方法之一, 如果 $VIF_i \geq 100$, 则剔除 X_i , 否则予以保留。最终得到的模型必然存在各个自变量 $VIF < 100$, 这是在自变量个数较多、且只做短期预测时所能接受的。

3 算例分析

3.1 指标初选

依据全面性、精炼性、可靠性、敏感性、稳定性、

时效性原则,选取经济领域内的备选指标。根据安徽省对外公布的经济统计资料,选取以下指标类型:1) 区域经济总体指标;2) 财政类;3) 投资类:固定资产投资总额、各行业固定资产投资完成额、房地产投资额;4) 分行业工业增加值;5) 分品种居民消费额;6) 进出口贸易;7) 工业品产量。此外,由于安徽省地处长三角地区,经济发展可能受到长三角经济波动的影响,因此,选取同处于长三角地区、经济发达、对外开放程度更高的上海市的部分经济指标,作为区域经济指标。

选取的国内指标类型有:1) 制造业指数;2) 非制造业指数;3) 货币供应量;4) 国际贸易与国际投资;5) 交通运输量;6) 行业景气指数。

国际指标类型有:1) 美日欧进出口贸易额;2) 美日欧失业率。

最终选取出国际国内140个经济指标,作为经济先行指标的备选指标。分别获取140个备选经济指标2011年1月—2014年6月的月度数据。同时,获取安徽省工业用电量2011年1月—2014年6月期间的月度数据。根据价格指数对价值量指标进行可比价格调整,转化为2014年6月可比价格。并运用季节性调整模型,对所有数据进行季节调整,扣除季节因素的影响。

3.2 指标筛选

对安徽省工业用电量与140项经济指标进行K-L信息量分析,筛选出工业用电量的经济先行指标,结果及其参数见表1。

经过K-L信息量法和考虑滞后阶数的相关性检验,发现有安徽省地方财政支出等21个指标,可作为安徽省工业用电量的先行指标,且时差相关性均在0.01的水平上显著,相关性系数均大于0.4。由此可见,安徽省工业用电量与这些经济先行指标存在显著的相关性,可用于构建预测模型。

3.3 预测模型

对筛选出的若干先行指标按照其最佳滞后阶数进行变换,作为解释变量,以安徽省工业用电量为因变量,建立回归模型。考虑到指标数据的公布有一定的滞后性,只考虑超前基准指标2期以上发生变化的经济先行指标,用于模型的解释变量。运用全指标模型、基于AIC准则选择出的最佳模型、基于BIC准则选出的最佳模型3种筛选原则分别选取最优回归模型,再根据3种模型的拟合情况,选取拟合

误差最小的模型,输出模型参数与拟合结果。经过比较,基于BIC准则选择出的最优模型成为安徽省工业用电量预测最优模型,模型变量及参数见表2。

表1 安徽省工业用电量的先行指标

Tab. 1 Leading indicators of industrial electricity consumption in Anhui

指标	最佳滞后阶数	相应K-L信息量值	相关性系数
地方财政支出(皖)	-9	0.007 2	0.736**
农林牧渔业固定资产投资(皖)	-7	0.026 2	0.657**
房地产竣工面积(皖)	-9	0.025 3	0.614**
集体企业工业增加值(皖)	-4	0.054 9	-0.620**
股份合作企业工业增加值(皖)	-3	0.062 3	-0.606**
硫酸产量(皖)	-4	0.003 1	0.589**
塑料制品产量(皖)	-9	0.007 4	0.537**
粗钢产量(皖)	-6	0.001 8	0.738**
电力电缆产量(皖)	-8	0.104 7	0.491**
重工业(沪)	-3	0.006 0	-0.692**
规模以上工业出口交货值(沪)	-7	0.010 0	-0.574**
电子信息产品制造业(沪)	-3	0.011 1	-0.569**
精品钢材制造业(沪)	-3	0.013 6	-0.724**
铁路货运量(全国)	-7	0.002 9	-0.453**
煤炭开采和洗选业景气指数(全国)	-5	0.005 6	-0.777**
黑色金属矿采选业景气指数(全国)	-3	0.005 4	-0.706**
石油加工、炼焦和核燃料加工业景气指数(全国)	-3	0.004 9	-0.728**
化学原料和化学制品制造业景气指数(全国)	-3	0.005 7	-0.687**
非金属矿物制品业景气指数(全国)	-5	0.005 9	-0.535**
欧元区进出口额	-6	0.001 9	-0.546**
美国失业率	-2	0.012 9	-0.752**

注:最佳滞后阶数小于0,表示变化超前于基准指标,是先行指标的可能性最大。相关性系数为考虑最佳滞后阶数之后,该指标与安徽省工业用电量的相关性系数,“**”表示在0.01的水平上显著(双尾),“*”表示在0.05的水平上显著(双尾)。

建模结果显示,股份合作企业工业增加值(皖)等11个指标成为多元回归模型的自变量,大部分自变量都通过了 t 检验,模型通过了方差膨胀因子小于100的检验。

对模型进行检验,判断模型的拟合残差是否符合白噪声特征。此外,为了检验模型的拟合与预测效果,进行拟合与预测的误差分析。模型拟合的残差分布见图2,模型的拟合结果见图3。结果显示,拟合残差符合白噪声特征,拟合结果与实际值变化趋势保持一致,拟合平均误差率为2.2%。拟合效果较好,因此,该模型可以用来预测安徽省的工业用电量。

表2 最优模型参数

Tab. 2 Parameters of the best model

对应回归系数	自变量名称	对应滞后阶数	t 值	$P_{ t }$
常数项 6 839 000	无	0	3.598	0.001 45**
X_1 46 320	股份合作企业工业增加值(皖)	-3	1.633	0.115 57
X_2 7 883	硫酸产量(皖)	-4	4.383	0.000 20***
X_3 -139	电力电缆产量(皖)	-8	-3.211	0.003 74**
X_4 149	重工业(沪)	-3	1.645	0.113 01
X_5 447.6	规模以上工业出口交货值(沪)	-7	2.721	0.011 92*
X_6 23.01	铁路货运量(全国)	-7	3.095	0.004 95**
X_7 -34 050	煤炭开采和洗选业景气指数(全国)	-5	-5.226	2.35e-05***
X_8 -45 340	黑色金属矿采选业景气指数(全国)	-3	-1.995	0.057 50·
X_9 -37 710	化学原料和化学制品制造业景气指数(全国)	-3	-3.334	0.002 77**
X_{10} 35 370	非金属矿物制品业景气指数(全国)	-5	2.653	0.013 91*
X_{11} 112 700	美国失业率	-5	2.428	0.023 06*

显著性:“***”表示在0.001的水平上显著,“**”表示在0.01的水平上显著,“*”表示在0.05的水平上显著,“·”表示在0.1的水平上显著。

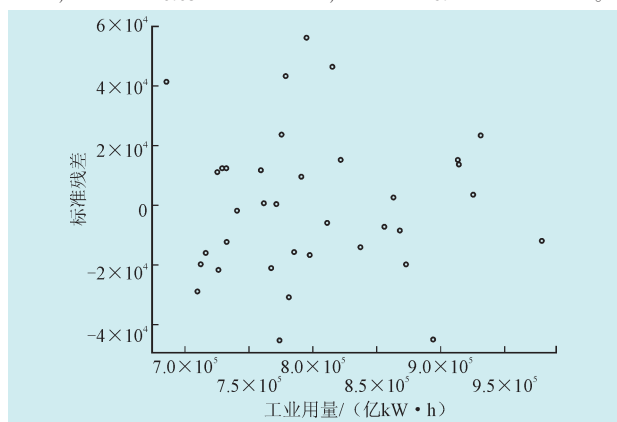


图2 模型拟合残差图

Fig. 2 Residual plot of the model

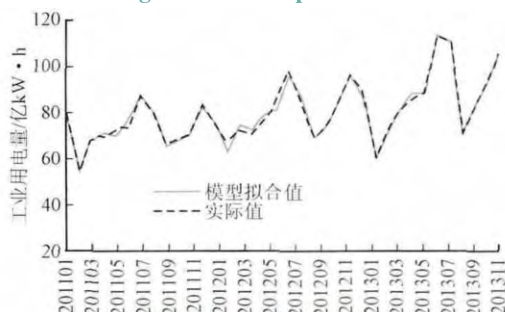


图3 模型拟合图

Fig. 3 Fitting chart of the model

3.4 用电量预测

运用筛选出的最优回归模型,预测2014年1月—6月的安徽省工业用电量。预测结果与实际值进行比较,得到预测误差(平均误差率为取绝对值后求平均),结果见图4。

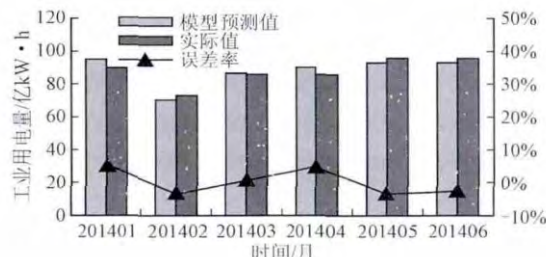


图4 安徽省工业用电量预测及其误差率

Fig. 4 Forecast and error of industrial electricity consumption in Anhui

模型预测结果显示,模型对安徽省工业用电量的预测精确度较高,预测误差率较稳定,2014年1—6月,预测误差率在-4%~5%之间波动,预测6个月的平均误差率为3.4%。

4 结语

区域用电量指标与经济指标的密切联系受到了研究者的广泛肯定。本研究基于K-L信息量法,从145个国内外经济指标中筛选出安徽省工业用电量的经济先行指标,发现指标间存在一定的时差相关关系,基于相关性较强的经济先行指标建立回归模型,用于预测安徽省工业用电量,取得了较好的预测效果。预测误差率较为稳定,均值为3.4%左右,因此可用作月度工业用电量预测。本研究建立模型所选取的指标,涵盖了区域工业产值、产品产量、全国行业景气指数、美国失业率等多个方面的指标,反映了区域经济与全国经济乃至世界经济的关联性,也进一步论证了经济发展与用电量的密切关系。

近年来,大数据研究技术的出现,为区域用电量预测提供了新的思路;自变量与因变量也许并不存在因果关系,而可能是同为结果的关系,但指标间只要存在一定的时差相关性,就可以用来建立预测模型。同时,在大数据时代,数据搜集、存储和更新的成本更低,为复杂动态模型研究提供了条件,而指标数量越多,包含的信息越丰富^[13-15]。本研究正是遵循了大数据研究的理念,略去了指标间的因果关系研究,同时,尽可能多地搜集备选指标,以保证

各方面信息的完整性。此外,运用K-L信息量法筛选的先行指标并不是固定不变的,能够不断更新、改进。随着经济社会不断发展,需要对数据和模型进行更新,以保证该预测方法长期有效。

参考文献

- [1] THOMA M. Electrical energy usage over the business cycle[J]. Energy Economics, 2004, (26): 463-485.
- [2] 葛斐, 荣秀婷, 石雪梅, 等. 基于经济、气象因素的安徽省年最大负荷预测方法研究[J]. 中国电力, 2015, 48 (3): 84-87.
GE Fei, RONG Xiuting, SHI Xuemei, et al. The Anhui annual maximum load forecasting method research based on economic and meteorological factors[J]. Electric Power, 2015, 48 (3): 84-87(in Chinese).
- [3] 肖欣, 周渝慧, 张宁, 等. 城镇化进程与电力需求增长的关系研究[J]. 中国电力, 2015, 48 (2): 145-149.
XIAO Xin, ZHOU Yuhui, ZHANG Ning, et al. Study on the relationship between urbanization process and electricity demand growth[J]. Electric Power, 2015, 48 (2): 145-149(in Chinese).
- [4] 黄永高, 卢毅. 基于小波分析和灰色模型的用电量预测[J]. 电网与清洁能源, 2011, 27 (3): 34-37.
HUANG Yonggao, LU Yi. Electricity consumption forecasting based on wavelet analysis and gray model[J]. Power System and Clean Energy, 2011, 27 (3): 34-37(in Chinese).
- [5] 卓元志, 刘家军, 王明军, 等. 基于熵值法的组合模型用电量预测方法研究[J]. 电网与清洁能源, 2011, 27 (5): 47-50.
ZHUO Yuanzhi, LIU Jiajun, WANG Mingjun, et al. A combination model based on entropy method for annual electricity consumption forecasting[J]. Power System and Clean Energy, 2011, 27 (5): 47-50(in Chinese).
- [6] 刘畅, 高铁梅. 中国电力行业周期波动特征及电力需求影响因素分析——基于景气分析及误差修正模型的研究[J]. 资源科学, 2011, 33 (1): 169-177.
LIU Chang, GAO Tiemei. Characteristics of electricity industry cycle fluctuation and influential factors of electricity demand based on business analysis and the error correction model[J]. Resources Science, 2011, 33 (1): 169-177(in Chinese).
- [7] 崔巍, 刘洋, 张米尔, 等. 基于经济先行指标的省级电力市场需求分析[J]. 水电能源科学, 2012, 30 (5): 191-195.
CUI Wei, LIU Yang, ZHANG Mi'er, et al. Analysis of provincial power market demand based on economic leading indicators[J]. Water Resources and Power, 2012, 30 (5): 191-195(in Chinese).
- [8] 张维, 范玉宏, 周小兵. 全社会用电量预警指标研究[J]. 电力需求侧管理, 2012, 14 (2): 7-12.
ZHANG Wei, FAN Yuhong, ZHOU Xiaobing. Research on early warning index system of social electricity consumption[J]. Power Demand Side Management, 2012, 14 (2): 7-12(in Chinese).
- [9] KULLBACK S, LEIBLER R A. On information and sufficiency[J]. Annals of Mathematical Statistics, 1951, 22 (1): 79-86.
- [10] AKAIKE H. Information theory and an extension of the maximum likelihood principle[C]// Second International Symposium on Information Theory. Tsahkadsor, Armenia: Petrov B N, Csaki F, 1973: 267-281.
- [11] SCHWARZ G. Estimating the dimension of a model[J]. The Annals of Statistics, 1978, 6 (2): 461-464.
- [12] 谢家安. 基于体感温度的电力系统负荷分类及负荷预测[J]. 电网与清洁能源, 2012, 28 (8): 24-28.
XIE Jiaan. Load classification and load forecasting of power system based on apparent temperature[J]. Power System and Clean Energy, 2012, 28 (8): 24-28 (in Chinese).
- [13] 李国杰, 程学旗. 大数据研究: 未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考[J]. 中国科学院院刊, 2012 (6): 647-657.
LI Guojie, CHENG Xueqi. Research status and scientific thinking of big data[J]. Bulletin of the Chinese Academy of Sciences, 2012 (6): 647-657(in Chinese).
- [14] 俞立平. 大数据与大数据经济学[J]. 中国软科学, 2013 (7): 177-183.
YU Liping. Big data and big data economics[J]. China Soft Science, 2013 (7): 177-183(in Chinese).
- [15] 宋亚奇, 周国亮, 朱永利. 智能电网大数据处理技术现状与挑战[J]. 电网技术, 2013, 37 (4): 927-935.
SONG Yaqi, ZHOU Guoliang, ZHU Yongli. Present status and challenges of big data processing in smart grid[J]. Power System Technology, 2013, 37 (4): 927-935 (in Chinese).

收稿日期:2015-09-11。

作者简介:

石雪梅(1977—),女,硕士学历,高级工程师,主要从事电网规划及电力经济关系研究工作;

葛斐(1972—),男,硕士学历,高级工程师,长期从事能源电力经济及预测分析方面的技术研究与管理工作;

肖夕林(1992—),男,研究生,研究方向为管理科学。

(编辑 黄晶)