

文章编号:1005-2992(2018)06-0086-04

基于 CAR 模型的用电量需求预测研究

周小琳

(东北电力大学 经济管理学院, 吉林 吉林 132012)

摘 要: 以日均用电量需求为输出变量,国内生产总值、第二产业产值比重、第三产业产值比重和城镇化率为输入变量,建立了基于多变量时间序列 CAR 的用电量需求预测模型. 预测结果显示,所有年份用电量需求拟合值与观测值之间的相对误差均不超过 1%,说明该模型在用电量需求预测领域具有较高的适用性.

关 键 词: 用电量需求;多变量时间序列;CAR 模型

中图分类号: TM715

文献标识码: A

用电量需求预测是规划电力发展、指导电力建设的重要依据. 准确地预测全社会用电量需求是有效避免电力短缺和电力浪费、保障国家或地区社会经济系统平稳运行的关键. 国内学者使用多种方法进行用电量需求预测,这些方法各有利弊,但总体上来说,用电量需求预测的方法体系并不完善. 时间序列方法,如线性回归模型、指数平滑预测模型和趋势移动平均法、自回归模型^[1]等,其优点是参数少、模型简单易操作,但是模型的预测精度较低、无法体现天气等因素引起的非线性变化;支持向量机算法^[2~3]能够最小化结构风险,小样本情况下泛化能力较强、适用于非线性和高维度问题的短期预测;灰色关联法^[4~5]的优点是模型算法简单、参数少、容易实现,但该方法对模型参数具有较大的敏感性、模型预测精度不高;随机森林算法^[6]和分位数法^[7~8]不仅能够给予预测对象的确定性预测,还能够在出现天气变化等极端事件时保持较高的预测精度,但其缺点是算法复杂、输入变量多、难以实现;神经网络法^[9~10]学习能力强、在极端天气情况下仍然保持较高的预测精度,但模型算法结构复杂、运算量大、容易发生拟合,导致泛化能力不足.

假设某一地区用电量的时间序列为 $\{Y_t\}$,如果服从 $ARMA(n, n)$ 模型,用算子形式可以写成 $\varphi(B)Y_t = \theta(B)\varepsilon_t$. 在实际应用中,序列 $\{Y_t\}$ 除了受自身和白噪声 (ε_t) 滞后的影响外,有时还要受到控制变量 $\{X_t\}$ 的滞后影响,从而使模型变成: $\varphi(B)Y_t = \rho(B)X_t + \theta(B)\varepsilon_t$. 则此模型称为受控的自回归滑动平均模型,简记为 $CARMA(n)$ 模型. $CARMA(n)$ 模型不仅能够反映事物发展的自身运用规律,还能够体现复杂系统的动态特征,具有结构风险最小、适用于小样本、非线性、有效克服过拟合、泛化推广能力强、预测精度高的优点,但由于辨识 $CARMA(n)$ 模型比较复杂,实践中通常假设 $\theta(B) = 1$,即用一个充分高阶的多变量自回归模型(CAR模型)来逼近并替代 $CARMA(n)$ 模型,而 $CAR(n)$ 模型计算简单、且可以自动辨识. 国内文献关于 $CAR(n)$ 模型的研究较少,本文试图把 $CAR(n)$ 模型应用到用电量需求预测领域,以期找到更加精准的用电量需求预测方法.

收稿日期: 2018-06-05

第一作者: 周小琳(1978-),女,硕士,讲师,主要研究方向:电力市场

电子邮箱: 26870287@qq.com(周小琳)

1 多变量时间序列 CAR 模型度量方法

1.1 CAR 模型的参数估计

CAR(n) 模型参数估计的方法主要包括最小二乘法、递推最小二乘法、广义最小二乘法和递推增广最小二乘法等. 在总干扰的均值为零且相互独立的条件下, 采用最小二乘法或者递推最小二乘法进行参数估计. 当系统的总干扰是相关序列时, 最小二乘估计和递推最小二乘估计是非一致估计量, 广义最小二乘法、递推增广最小二乘法通过引进一个白色滤波器, 把具有相关性的总干扰转化为白噪声, 从而使得参数估计保持一致性. P 维输入向量、1 维输出向量的 CAR 模型可以表述为

$$Y_t = \sum_{i=1}^n \varphi_i Y_{t-i} + \sum_{i=0}^n \Gamma_i X_{t-i} + \varepsilon_t, \quad (1)$$

其中, Y_t 为一维输出向量, $X_t(X_{1t}, X_{2t}, \dots, X_{qt})^T$ 为 p 维输入向量(或者称为控制向量), $\varepsilon_t(\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{qt})^T$ 是零均值为 ε_{it} 、方差为 σ_i^2 的白噪声向量, $\Gamma_i = (\rho_{ts}^{(i)})$ 是元素为 $\rho_{ts}^{(i)}$ 的 $p \times p$ 阶参数阵.

令: $B = [\varphi_1, \dots, \varphi_n, \rho_{10}, \rho_{11}, \dots, \rho_{1n}, \dots, \rho_{p0}, \rho_{p1}, \dots, \rho_{pn}]^T = [\beta_1, \dots, \beta_{n+(n+1)n}]^T$

置 $t = n + 1, n + 2, \dots, N$, 并记

$$Y_N = (Y_{n+1}, Y_{n+2}, \dots, Y_N)^T,$$

$$E_N = (\varepsilon_{n+1}, \varepsilon_{n+2}, \dots, \varepsilon_N)^T,$$

$$W_N = \begin{bmatrix} Y_n & Y_{n-1} & \cdots & Y_1 & X_{1,n+1} & \cdots & -X_{1,1} & \cdots & X_{p,n+1} & -X_{p,n} & \cdots & -X_{p,1} \\ Y_{n+1} & Y_n & \cdots & Y_2 & X_{1,n+2} & \cdots & -X_{1,2} & \cdots & X_{p,n+2} & -X_{p,n+1} & \cdots & -X_{p,2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Y_{N-1} & Y_{N-2} & \cdots & Y_{N-n} & X_{1,N} & -X_{1,N-1} & \cdots & -X_{1,N-n} & \cdots & -X_{p,N} & -X_{p,N-1} & \cdots & -X_{p,N-n} \end{bmatrix}$$

则公式(1) 可以写成:

$$Y_N = W_N B + E_N, \quad (2)$$

因此, 可用最小二乘法求得 N 时刻参数 B 的最小二乘估计为

$$\hat{B}_N = (W_N^T W_N)^{-1} W_N^T Y_N. \quad (3)$$

若有新增观测数据 (Y_{N+1}, X_{N+1}) , 则从 $t = 1, 2, \dots, N$, 到时刻 $N + 1$ 时参数 B 的递推最小二乘估计为

$$\hat{B}_{N+1} = \hat{B}_N + K_{N+1} (Y_{N+1} - W_{N+1}^T \hat{B}_N), \quad (4)$$

其中, $K_{N+1} = (\lambda + W_{N+1}^T P_N W_{N+1})^{-1} (P_N W_{N+1})$, $P_{N+1} = [I - K_{N+1} W_{N+1}^T] P_N / \lambda$.

在递推时, 初值 \hat{B}_0 可从 0 开始, P_0 可取 λI , I 是单位阵, 而遗忘算子 λ 为 $0 < \lambda \leq 1$. 于是, 利用 N 组观察值估计 CAR 模型得到残差向量为 $\hat{\varepsilon}_t = Y_t - W_t^T \hat{B}_N$, 残差平方和为 $S_n = \sum_{t=n+1}^N \hat{\varepsilon}_t^2$, ε_t 的方差的估计量为 $\hat{\sigma}^2 =$

$$\sum_{t=n+1}^N \hat{\varepsilon}_t^2 / (N - n).$$

1.2 CAR 模型阶数的确定

CAR 模型可以用 F 检验的方法来确定阶数. 在给定的观测数据 (X_t, Y_t) , $t = 1, 2, \dots, N$ 下, 可相继地建立模型 CAR(n), $n = 1, 2, \dots$, 并比较 CAR(n) 和 CAR($n + 1$), F 统计量为

$$F = \frac{S_n - S_{n+1}}{S_{n+1}} \frac{N - (n + 1) - (n - 2)}{1 + p} - F(1 + p, N - (n + 1) - (n + 2)p)$$

在给定的显著性水平 α 下,当 $F \geq F_\alpha$ 时,拒绝 $CAR(n)$, 否则接受 $CAR(n)$ 模型,从而可确定阶数 n .

1.3 CAR 模型时滞的确定

根据以上方法,确定 CAR 模型的阶数以后,其中某些参数 $\hat{\varphi}_1, \cdots, \hat{\varphi}_n, \cdots, \hat{\rho}_{p0}, \hat{\rho}_{p1}, \cdots, \hat{\rho}_{pn}$ 的估计值可能非常小,且接近于 0. 即,当 $\hat{\rho}_{i,0} = \hat{\rho}_{i,1} = \cdots = \hat{\rho}_{i,k-1} = 0$ 时,统计的时滞为 k ,模型的阶数小于 n ,需要通过 F 检验确定模型的时滞和真实阶数,从而得到真实模型的参数估计值. 首先,在给定的观测数据 $(X_t, Y_t), t = 1, 2, \cdots, N$ 下,通过递推最小二乘法估计 β_i 的 95% 的置信区间,判断此置信区间是否包含零点. 接着,删除零点的参数后,重新运用递推最小二乘法建立 CAR 模型. 最后,利用 F 检验判断删除的参数是否需要保留在模型中,从而确定模型的时滞和真实阶数.

2 基于 CAR 模型的中国用电需求量预测

2.1 数据来源与变量选取

用电需求量与一国的经济规模、产业结构和城镇化率密切相关. 本文选取平均每天电力消费量(亿千瓦时)代表用电量需求,用 $elec$ 表示;以国内生产总值(亿元)代表经济规模,用 GDP 表示;以国内生产总值中第二产业产值所占比重和第三产业产值所占比重代表产业结构,分别用 $sindrt$ 和 $tindrt$ 表示;城镇化率为城镇人口数占总人口数的比重,用 $urbanrt$ 表示. 本文全部数据均来源于国研网数据库.

2.2 用电需求量模型的建立

为克服异方差性,本文中所有变量均采用对数形式,选用 $\ln elec$ 作为输出变量, $\ln GDP$ 、 $\ln sindrt$ 、 $\ln tindrt$ 和 $\ln urbanrt$ 作为输入变量,构建如式(1)所示的 1 维输出向量、4 维输入向量的 CAR 模型. 建模所用 F 检验的显著性水平为 0.05,递推最小二乘法的遗忘因子为 1.0,检验结果为 $F = 2.81$. 定阶检验中 $CAR(n)$ 模型和 $CAR(n + 1)$ 模型的残差平方和分别为 0.00016 和 0.00021,定阶检验的 F 值为 0.07. 选定阶次全参数时和剔除不显著因素后模型的残差平方和分别为 0.00021 和 0.00023,是否为不显著因子检验的 F 值为 0.45, $F(\alpha = 0.05) = 4.46$. 剔除不显著项后的 CAR 模型为

$$\ln elec_t = 0.67 \ln elec_{t-1} + 0.13 \ln GDP_t + 0.25 \ln sindrt_s - 0.38 \ln sindrt_{t-1} \\ 0.71 \ln tindrt_t + 0.51 \ln urbanrt_t + 0.32 \ln urbanrt_{t-1}, \tag{5}$$

表 1 列出了 2001 年~2016 年我国平均每天电力消费量对数值的实际值和根据式(5)计算的拟合值,2001 年~2016 年,所有年份电力消费量拟合值与观测值之间的相对误差均不超过 1%,平均误差仅为 0.35%,可以看出,式(5)的整体拟合效果很好,误差偏离程度非常小,可以用该模型预测我国平均每天电力消费情况.

表 1 2001 年~2016 年中国日均电力消费量对数值的实际值与拟合值

年份	实际值	拟合值	误差	误差百分比/%	年份	实际值	拟合值	误差	误差百分比/%
2001	3.70	3.72	-0.02	-0.65	2009	4.62	4.63	-0.01	-0.19
2002	3.81	3.81	0.00	-0.10	2010	4.74	4.73	0.01	0.22
2003	3.95	3.94	0.01	0.23	2011	4.86	4.85	0.01	0.17
2004	4.10	4.09	0.01	0.20	2012	4.91	4.93	-0.02	-0.42
2005	4.22	4.23	0.00	-0.10	2013	5.00	4.98	0.02	0.45
2006	4.36	4.34	0.02	0.42	2014	5.04	5.05	-0.01	-0.21
2007	4.50	4.45	0.04	0.91	2015	5.05	5.06	-0.01	-0.24
2008	4.55	4.59	-0.04	-0.91	2016	5.10	5.09	0.01	0.20

数据来源:作者整理。

3 结 论

本文通过推导1维输出向量、 p 维输入向量的CAR模型,建立日均用电量需求的预测模型,并采用2001年~2016年我国国内生产总值、第二产业产值比重、第三产业产值比重和城镇化率的数据对模型进行检验。结果显示,本文构建的预测模型中,所有年份电力需求拟合值与观测值之间的相对误差均不超过1%,模型预测精度高、拟合效果好,可以用该模型对中国日均用电需求进行预测。

参 考 文 献

[1] 韩文花,汪素青,苏晓燕,等. 基于CS的SE-二次移动平均法的月度用电量预测[J]. 电气应用,2018,37(9):72-77.

[2] 曲朝阳,张率,刘洪涛. 基于用电影响因素回归的小区用电预测模型[J]. 东北电力大学学报,2015,35(1):73-77.

[3] 汲国强,李顺昕,赵伟博,等. 基于对数平均迪氏指数-布谷鸟搜索算法-最小二乘支持向量机的区域中长期电力需求预测[J]. 科学技术与工程,2018,18(10):213-218.

[4] 牛东晓,魏亚楠. 基于FHNN相似日聚类自适应权重的短期电力负荷组合预测[J]. 电力系统自动化,2013,37(3):54-57.

[5] 陈洁,吕豪杰. 基于离散型灰色模型的南京市工业用电量预测研究[J]. 南京工程学院学报:社科版,2018,18(1):50-54.

[6] 赵腾,王林童,张焰,等. 采用互信息与随机森林算法的用户用电关联因素辨识及用电量预测方法[J]. 中国电机工程学报,2016,36(3):604-614.

[7] 穆海振. 分位数回归方法在盛夏日电量预测中的应用[J]. 电力需求侧管理,2018,20(3):24-27.

[8] 薛承荣,顾洁,孙金华,等. 基于分位数回归的电力负荷预测影响因素研究及应用[J]. 电气应用,2015,34(12):134-138.

[9] 谷云东,张素杰,冯君淑. 大用户电力负荷的多模型模糊综合预测[J]. 电工技术学报,2015,30(23):110-115.

[10] 黄磊,陈浩,衣乔木,等. 基于k-means聚类 and BP神经网络组合模型的用电负荷预测[J]. 电力与能源,2016,37(1):56-60.

Research on Electricity Demand Forecast Based on CAR Model

Zhou Xiaolin

(School of Economics and Management, Northeast Electric Power University, Jilin Jilin 132012)

Abstract: Taking the average daily electricity demand as output variable and the GDP, the proportion of the secondary industry's output value, the proportion of the tertiary industry's output value and the urbanization rate as input variables, the electricity consumption demand prediction model based on the multi-variable time series was established. The predicted results show that the relative error between the fitting value and the observed value of electricity consumption demand is no more than 1% in all years, which indicates that this model has high applicability in the forecasting field of electricity demand.

Key words: Electricity consumption demand; Multi-variable time series; CAR model