

doi:10.16055/j.issn.1672-058X.2015.0008.012

ARIMA 模型在重庆市能源消耗量预测中的应用*

李岩岩

(重庆工商大学 数学与统计学院,重庆 400067)

摘要:能源是人类赖以生存和发展的重要物质基础,对经济的快速发展和人们生活水平的大幅度提高起着无可取代的作用,随着经济的发展人们对能源的需求日益增多,由于地球上的资源是有限的,对未来能源消耗量的准确预测显得尤为重要;运用 ARIMA 模型对《2013 重庆市统计年鉴》中重庆市 1981-2012 年能源消耗量数据进行分析,结果显示:ARIMA(2,3,2)模型预测未来的结果较为准确,为重庆市资源消耗量提供了可靠的依据。

关键词:ARIMA 模型;时间序列;能源消耗;预测

中图分类号:O224 **文献标志码:**A **文章编号:**1672-058X(2015)08-0054-07

重庆市是中国面积最大的城市,辖区总面积为 8.24 万 km²,为北京、天津、上海三市面积的 2.39 倍。改革开放以来,经济取得了突飞猛进的增长。2013 年,工业增加值 5 249.65 亿元,比 2012 年增长 13.1%,占重庆市地区生产总值的 41.5%。伴随着工业化的快速发展,能源处于高耗阶段,能源资源短缺已成为重庆市经济发展的一大瓶颈。国内外不同的机构和许多学者针对能源消耗问题也运用不同的方法做了大量的研究,国家统计局和能源办常会采用定性预测方法对能源问题做一些研究然后制定合理的政策;当然对此问题的研究比较深入的还是广大学者们,他们常用的预测的方法有:回归预测方法、自适应预测、投入产出法、灰色关联法、生长理论、指数平滑法及神经网络预测法等,这些方法都能从一定的方面反映问题,但是由于没有考虑数据的平稳性及数据间的关联性,预测结果往往准确度不高,ARIMA 模型克服这方面的局限性,它可以很好的拟合非平稳时间序列,能大大提高预测的精度。

国内学者运用 ARIMA 模型做了很多研究,其中有刘勇^[1]针对我国的能源消费预测和池启水^[2]针对我国煤炭消费的预测;还有一些地方性的预测,蒋燕^[2]对广西全社会固定资产投资的预测及龚国勇^[4]对深圳 GDP 的预测,李伟^[4]对重庆市 GDP 的预测也用到了 ARIMA 模型,但是针对重庆市能源消耗问题还没有学者专门运用 ARIMA 方法进行过预测,现今要求整个社会都要以降低能源消耗为目标,为了尽快全面的掌握重庆市资源消耗状况,为重庆市节能降耗提供科学、有效的统计数据支撑,现通过收集、分析时间序列数据,运用 ARIMA 数学模型,科学分析重庆市能源消费量,估计模型以预测未来能源消耗量,为相关部门制订节能降耗措施和对策提供科学合理的依据。

收稿日期:2014-10-01;修回日期:2014-12-09.

* 基金项目:重庆市教委科技项目(KJ1400613).

作者简介:李岩岩(1988-),女,河南开封人,硕士研究生,从事统计学经济社会统计研究.

1 基于 ARIMA 模型的重庆市能源消耗量预测方法

1.1 ARIMA 模型理论

ARIMA 模型全称为差分自回归移动平均模型 (Auto Regression Integrated Moving Average Model, 简记 ARIMA), 是一类常用的时间序列预测模型, 是由博克思 (Box) 和詹金斯 (Jenkins) 于 20 世纪 70 年代初提出一著名时间序列预测方法, 所以又称为 Box-Jenkins 模型、博克思-詹金斯法。其中 ARIMA(p, d, q) 称为差分自回归移动平均模型, AR 是自回归, p 为自回归项; MA 为移动平均, q 为移动平均项数, d 为时间序列成为平稳时所做的差分次数。ARIMA 模型通常借助时间序列的随机特性来描述事物的发展变化规律, 指将非平稳时间序列转化为平稳时间序列, 然后运用时间序列的过去值、当期值以及滞后随机扰动项的加权来建立模型, 从而解释并预测时间序列的变化发展规律。ARIMA 模型根据原序列是否平稳以及回归中所含部分的不同, 包括自回归移动平均过程 (ARMA(p, q): Auto Regression Moving Average) 及差分自回归移动平均 (ARIMA(p, d, q): Auto Regression Integrated Moving Average) 过程。

1.2 ARIMA 模型结构

ARIMA 模型实际就是非平稳时间序列经过差分处理后的 ARMA 模型, ARMA 模型是目前最常用的拟合平稳序列的模型。ARMA 模型又可以细分为 AR 模型、MA 模型和 ARMA 模型三大类。

1.2.1 ARMA 模型

具有如下的模型称为 p 阶自回归 (auto regression) 模型, 简记为 AR(p):

$$x_t = \varphi_0 + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \dots + \varphi_p x_{t-p} + \varepsilon_t \quad (1)$$

其中, 模型的最高阶数为 p , 随机干扰项 $\{\varepsilon_t\}$ 为零均值白噪声序列即 $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ 。

具有如下结构的模型称为 q 阶移动平均 (moving average) 模型, 简记为 MA(q):

$$x_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (2)$$

其中, 模型的最高阶数为 q , 随机干扰项 $\{\varepsilon_t\}$ 为零均值白噪声序列即 $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ 。

把具有如下结构的模型称为自回归移动平均 (auto regression moving average) 模型, 简记为 ARMA(p, q):

$$x_t = \varphi_0 + \varphi_1 x_{t-1} + \dots + \varphi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (3)$$

限制条件与 AR(p)、MA(q) 相同, ARMA(p, q) 模型实际上就是 AR(p) 与 MA(q) 的结合。

1.2.2 ARIMA 模型具体结构

具有如下结构的模型称为求和自回归移动平均 (auto regression integrated moving average) 模型, 简记为 ARIMA(p, d, q) 模型^[6]:

$$\begin{cases} \Phi(B) \nabla^d x_t = \Theta(B) \varepsilon_t \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ Ex_s \varepsilon_t = 0, \forall s < t \end{cases} \quad (4)$$

式(4)中: $\nabla^d = (1-B)^d$, d 为差分次数; $\Phi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$, 为平稳可逆 ARMA(p, q) 模型的自回归系数多项式; $\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$, 为平稳可逆 ARMA(p, q) 模型的移动平滑系数多项式式(4)可以简记为

$$\nabla^d x_t = \frac{\Theta(B)}{\Phi(B)} \varepsilon_t \quad (5)$$

式(5)中 $\{\varepsilon_t\}$ 为零均值白噪声序列。

1.3 ARIMA 模型拟合基本步骤

(1) 检验时间序列的平稳性, 一般采用观察时序图和单位根检验的方法。能够用 ARMA 模型拟合的序列都是平稳时间序列, 对于非平稳时间序列, 需要先对原始序列进行若干次差分运算使其转化为平稳时间

序列,这个差分转换次数就是 d ,然后再用 ARMA 模型拟合。

(2) 根据时间序列模型的识别理论选取相应的模型。这需要借助与序列的相关图,AR 过程的 AC(自相关函数)和 PAC(偏自相关函数),与 MA 过程的 AC 和 PAC 相比较,有相反的变化理论模式。对于 AR 模型,其 AC 从图形上表现为“拖尾性”,而 PAC 表现为“截尾性”。MA 模型,其 AC 表现为“截尾性”,PAC 表现为“拖尾性”。若序列的 AC 和 PAC 均表现为“拖尾性”,则适合于 ARMA 模型。此时可以初步确定 p, q 的值。

(3) 模型参数估计。根据前面确定的 d 及初步确定的 p 和 q ,运用 AIC 和 SC 最小法则以及拟合优度最大等综合指标,选取最优的 p 和 q ,确定最终模型的参数。

(4) 对拟合后的模型进行白噪声检验。检验残差是否为纯随机序列,如果为白噪声序列证明模型还不是最优的需要进一步改进。

(5) 利用已经通过检验的模型进行预测分析。

2 ARIMA 模型在重庆市能源消耗量预测中的应用

2.1 数据描述

从《2013 年重庆市统计年鉴》获得 1981-2012 年重庆市能源消耗量的数据。参照《EViews 统计分析与应用》^[7]完成下面的操作,首先由原始数据可以绘制出 1981-2012 年重庆市能源消耗量的折线图,图 1 所示。其中 X 表示能源消耗量,单位是万吨标准煤。由图 1 可知,重庆市能源消耗量呈现指数上升的趋势,增长势头显著,可知时间序列 $\{X_t\}$ 为非平稳时间序列。

2.2 对数据进行平稳化处理

一般来说对于呈现指数趋势的时间序列,需要通过取对数的形式将其转化为线性趋势,然后再进行差分处理。通常来讲线性趋势可以通过一阶差分消除,二次曲线趋势可以通过二阶差分消除。针对本次时间序列数据,经过取对数 $Y=\ln X$ 以后,经过三次差分运算,序列达到平稳。三次差分后的序列用 Z 表示,其折线图用图 2 表示。对 Z 进行 ADF 单位根检验结果如表 1 所示。

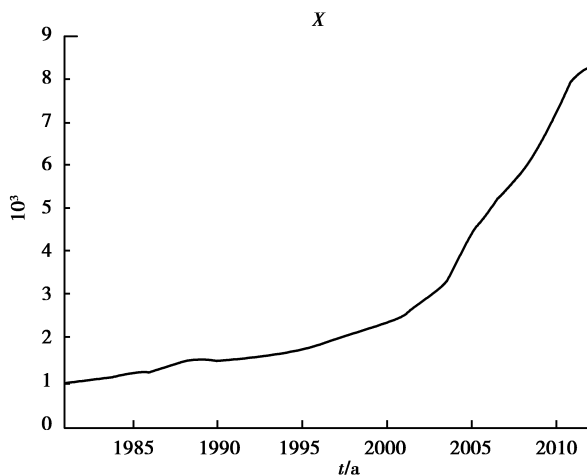


图 1 X 折线图

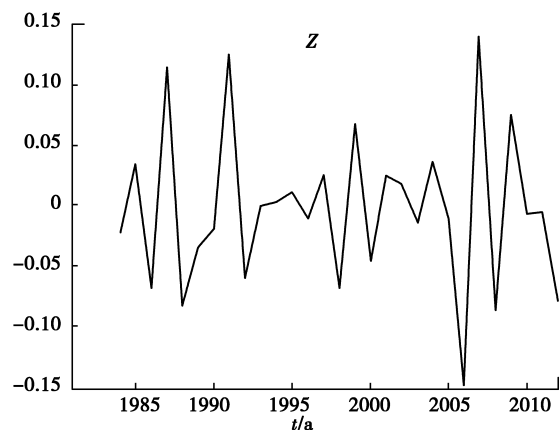


图 2 三次差分折线图

表 1 ADF 单位根检验

变量	ADF 值	Prob. *	1% level	5% level	10% level
Z	-4.555 388	0.001 5	-3.737 853	-2.991 878	-2.635 542

在表 1 中, ADF 检验 t 统计量为 $-4.555\ 388$, 其相应的概率值为 $0.001\ 5$, 在 1% 、 5% 、 10% 的检验水平下 t 统计量的临界值分别为 $-3.737\ 853$ 、 $-2.991\ 878$ 、 $-2.635\ 542$ 。序列 Y 三阶差分的 ADF 检验的 t 统计量都比 1% 、 5% 、 10% 检验水平下的临界值小, 因此可以拒绝原假设, 即可认为序列 Y 的三阶差分 (即 Z) 没有单位根, 也即序列 X 取对数后的三阶差分平稳 $I(3)$ 的。

2.3 ARMA 模型识别和阶数确定

由于检验出序列 $\{Z_t\}$ 是平稳的, 因此可以建立相应的模型, 在建模之前需要识别模型的类别, 即需要确定是 $AR(p)$ 模型或 $MA(q)$ 模型或 $ARMA(p, q)$ 模型。模型的确定通常借助序列的相关图, 即序列的自相关函数 (AC) 和偏自相关函数 (PAC)。

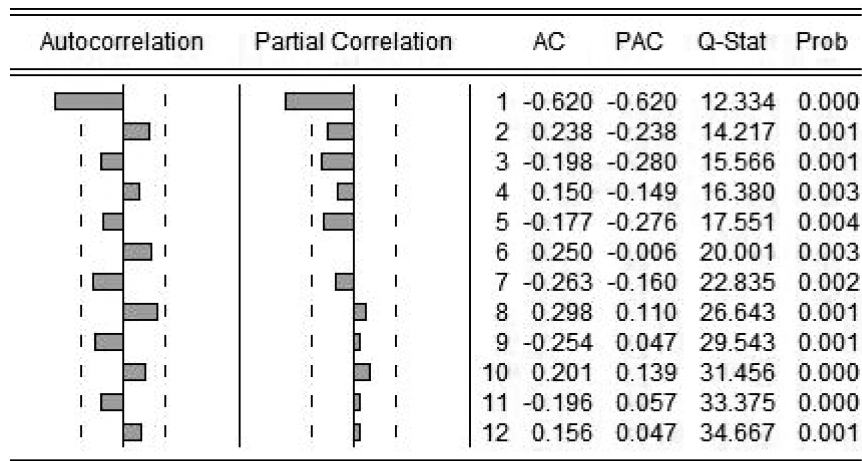


图 3 Z 的 AC 与 PAC 图

由图 3 可以看出自相关和偏自相关图均表现出“拖尾性”, 因此序列 $\{Z_t\}$ 适合用 $ARMA(p, q)$ 模型, 初步估计有 4 种模型形式: $ARMA(1, 1)$ 、 $ARMA(1, 2)$ 、 $ARMA(2, 1)$ 及 $ARMA(2, 2)$ 。

识别 ARMA 模型的形式后, 可以运用最佳准则定价函数法, 即 Akaike 提出的 AIC 准则, 准则是在模型参数极大似然估计的基础上, 对模型的阶数和相应的参数同时给出一组最佳估计。一般来讲, 在给出不同模型的 AIC 计算公式的基础上, 选取使 AIC 达到最小的那一组阶数为理想阶数。运用 Eviews 软件完成本过程, 需要注意的是, 对参数估计量的 t 检验显著性水平并不像简单回归方程那么严格, 更多的是考虑模型的整体拟合效果。AIC 准则及 SC 准则和调整后的可决系数都是选择模型的重要标准。

通过表 2 比较 4 个模型的 AIC 及其他各项指标, 虽说 $ARMA(1, 1)$ 和 $ARMA(1, 2)$ 的各项指标都比 $ARMA(2, 1)$ 和 $ARMA(2, 2)$ 的拟合效果好, 但是 $ARMA(1, 1)$ 和 $ARMA(1, 2)$ 的 $MA(q)$ 拟合过程不可逆, 所以只需比较 $ARMA(2, 1)$ 和 $ARMA(2, 2)$, 最终可以确定 $p=2, q=2$ 为最佳模型的阶数, 建立 $ARIMA(2, 3, 2)$ 模型:

$$Z_t = -0.000\ 208 - 1.080\ 346Z_{t-1} - 0.126\ 364Z_{t-2} + \varepsilon_t + 0.009\ 58\varepsilon_{t-1} + 0.922\ 219\varepsilon_{t-2}$$

其中 $Z = \Delta^3 Y, Y = \ln X$

表 2 模型指标结果比较

模型指标	ARMA(1,1)	ARMA(1,2)	ARMA(2,1)	ARMA(2,2)
R^2	0.819 088	0.845 491	0.641 344	0.702 277
调整后的 R^2	0.801 858	0.822 314	0.584 714	0.636 117
回归标准差	0.030 669	0.029 043	0.045 122	0.042 237
残差平方和	0.019 752	0.016 870	0.038 684	0.032 111

续表2

模型指标	ARMA(1,1)	ARMA(1,2)	ARMA(2,1)	ARMA(2,2)
对数似然值	51.175 89	53.068 98	40.824 51	42.965 83
F 统计量	47.539 24	36.480 71	11.325 16	10.614 73
F 检验的显著水平	0.000 000	0.000 000	0.000 175	0.000 135
AIC 信息准则	-4.014 658	-4.089 082	-3.202 131	-3.301 377
SC 准则	-3.867 401	-3.892 740	-3.004 654	-3.054 530
DW 值	1.860 569	2.390 897	1.871 020	2.037 067

2.4 模型检验

模型检验也即对模型的残差序列进行白噪声检验,如果残差序列不是纯随机序列则意味着残差序列中还有信息未被提取,模型需要进一步改进。由图 4 知残差序列的拟合效果还是比较好的,但这还需要通过相关图进一步检验。对 ARIMA(2,3,2) 模型的残差进行白噪声检验(图 5),可以看出残差序列的自相关函数都在 95% 的置信区间内,由此可以得出残差序列相互独立即为白噪声的概率很大,故不能拒绝序列为白噪声序列的原假设,即可认为模型 ARIMA(2,3,2) 估计结果的残差序列不存在自相关。

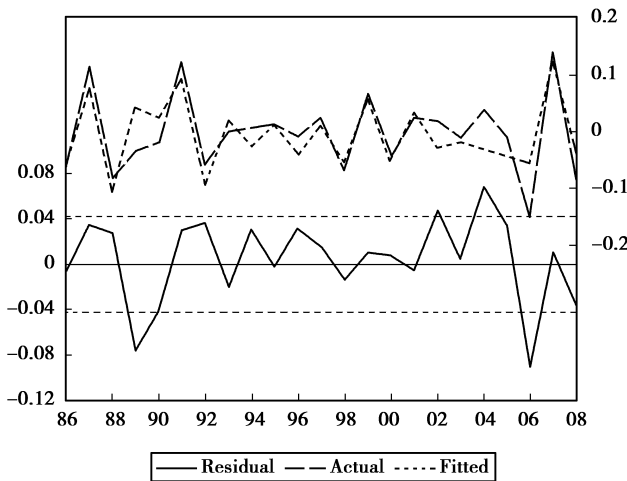


图 4 残差图

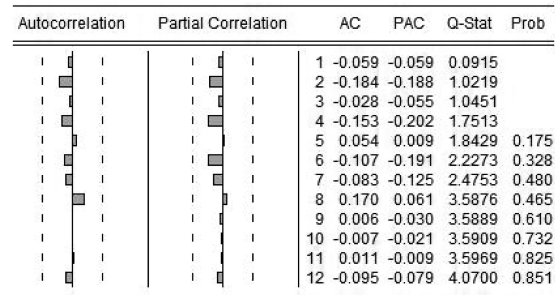


图 5 ARIMA(2,3,2) 模型的残差相关图及 Q 检验

3 预测与分析

根据时间序列 $\{Z_t\}$ 的 ARIMA(2,3,2) 模型:

$$Z_t = -0.000\ 208 - 1.080\ 346Z_{t-1} - 0.126\ 364Z_{t-2} + \varepsilon_t + 0.009\ 58\varepsilon_{t-1} + 0.922\ 219\varepsilon_{t-2}$$

可以推出时间序列 $\{Y_t\}$ 的 ARIMA(2,3,2) 模型的预测方程为

$$\begin{aligned} Z_t &= \Delta^3 Y_t = Y_t - 3\ln Y_{t-1} + 3\ln Y_{t-2} - \ln Y_{t-3} \\ Y_t &= 3\ln Y_{t-1} - 3\ln Y_{t-2} + \ln Y_{t-3} - 0.000\ 208 - 1.080\ 346Z_{t-1} - 0.126\ 364Z_{t-2} + \\ &\quad \varepsilon_t + 0.009\ 58\varepsilon_{t-1} + 0.922\ 219\varepsilon_{t-2} \end{aligned}$$

进而推出时间序列 $\{X_t\}$ 的预测方程为

$$X_t = e^{3\ln Y_{t-1} - 3\ln Y_{t-2} + \ln Y_{t-3} - 0.000\ 208 - 1.080\ 346Z_{t-1} - 0.126\ 364Z_{t-2} + \varepsilon_t + 0.009\ 58\varepsilon_{t-1} + 0.922\ 219\varepsilon_{t-2}}$$

对重庆市 2009-2012 年的能源消耗量做预测(单位:万吨标准煤),并与实际值做比较(表 3)。

表 3 预测结果

年份	实际值	预测值	相对误差/%
2009	6 431.63	6 383.10	0.755
2010	7 117.41	6 811.35	4.3
2011	7 951.12	7 309.77	8.066
2012	8 284.94	7 741.24	6.563

(1) 从 $\{X_t\}$ 的预测方程可以看出,重庆市资源消耗量与序列本身滞后一期值和滞后三期值以及残差项一期和二期有着密切的关系。由于参数估计的时候系数均为负值,这就意味着滞后一期值和滞后三期值以及残差项一期和二期对当期值成负相关关系。也就是说: $t-1$ 期资源消耗量增加则 t 期消耗量就会减少。因此,重庆市政府在安排每年资源消耗量时应注意这一点,不要寅吃卯粮,合理计划每年的资源消耗量,使社会经济更加稳定健康的发展。

(2) 图 6 中虚线是预测置信区间,可以看到随着向后预测期的增加,预测置信区间也变大,从而表明预测期越往后,模型预测的精度越差。表 3 所列的是 2009-2012 年序列 $\{X_t\}$ 的预测值、实际值及预测相对误差。从表 3 中可以看出,序列 $\{X_t\}$ 的预测值与实际值的相对误差比较小,预测的平均相对百分误差为 4.921%,一般认为平均相对百分误差在 10% 以内模型的预测效果比较好^[8](易丹辉,2005),可以用来预测。但是也要看到,随着预测期的增加,模型预测的相对误差也变大。这是 ARIMA 模型的一个缺陷,尽管如此它在短期预测方面相比其他的预测方法,预测的准确度还是比较高的,可以为重庆市能源消耗量短期预测提供借鉴。

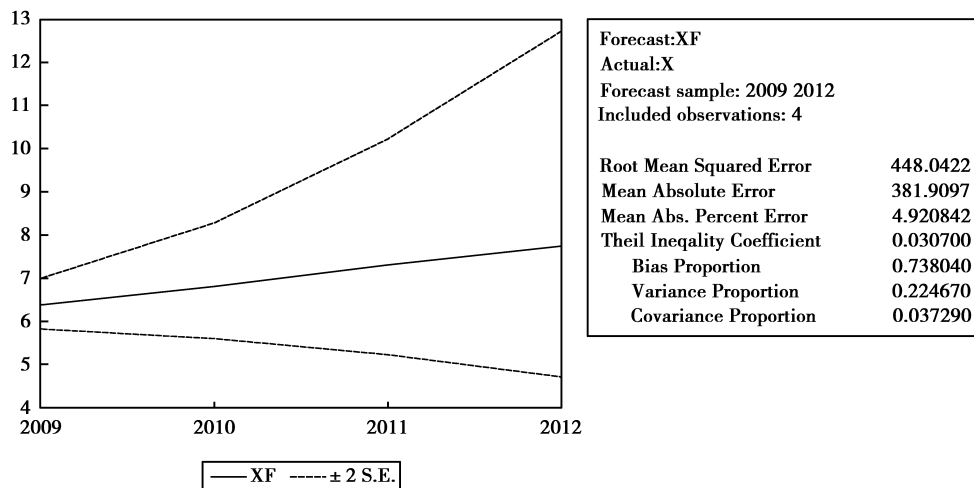


图 6 2009-2012 预测图

参考文献:

- [1] 刘勇,汪旭晖.ARIMA 模型在我国能源消费预测中的应用[J].经济经纬,2007(5):11-13
- [2] 池启水,刘晓雪.ARIMA 模型在煤炭消费预测中的应用分析[J].能源研究与信息,2007,23(2):117-123
- [3] 蒋燕.ARIMA 模型在广西全社会固定资产投资预测中的应用[J].数理统计与管理,2006,25(5):588-592
- [4] 龚国勇.ARIMA 模型在深圳 GDP 预测中的应用[J].数学的实践与认识,2008,38(4):53-57
- [5] 李伟,罗泽举.基于线性神经网络的重庆市 GDP 发展研究[J].重庆工商大学学报:自然科学版,2014,31(2):37-42
- [6] 王燕.应用时间序列分析[M].北京:中国人民大学出版社,2005
- [7] 樊欢欢,张凌云.EViews 统计分析与应用[M].北京:机械工业出版社,2009
- [8] 易丹辉.数据分析与 Eviews 的应用[M].北京:中国统计出版社,2005

Application of ARIMA Model in the Prediction of Energy Consumption of Chongqing

LI Yan-yan

(School of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing 400067, China)

Abstract: Resource is important basic material for human survival and development. It plays an irreplaceable role in the rapid development of economy and greatly improving people's living standards. With the development of economy, people's demand for energy is increasing, because the resources are limited on earth, therefore, it is particularly important to accurately predict future energy consumption. This paper uses ARIMA model to analyze 1981-2012 annual energy consumption data of Chongqing based on *Chongqing Statistical Yearbook 2013*. The result shows that the predict of future energy consumption by ARIMA (2,3,2) model is more accurate, which provides reliable basis for resource consumption of Chongqing.

Keywords: ARIMA model; time series; energy consumption; prediction

(上接第 42 页)

参考文献:

- [1] DANTZIT G B, RAMSER R H. The Truck Dispatching Problem[J]. *Management Science*, 1959(4):1128-1131
- [2] LAPORTE G, NOBERT Y, DESROEHERS M. Optimal Routing with Capacity and Distance Restrictions [J]. *Operations Research*, 1985(4):1050-1073
- [3] TAILLARD E. Parallel Interactive Search Method for Vehicle Routing Problem[J]. *Networks*, 1993,23:661-673
- [4] 郎茂祥. 配送车辆优化调度模型与算法[M]. 北京: 电子工业出版社, 2009
- [5] 李琳, 刘士新, 唐加福. 改进的蚁群算法求解带时间窗的车辆路径问题[J]. *控制与决策*, 2010,25(9):1379-1382
- [6] 李秀娟, 杨玥. 蚁群优化算法在物流车辆调度系统中的应用[J]. *计算机工程与应用*, 2013,33(10):2822-2826
- [7] 吴洁明. 物流配送车辆路径优化问题的仿真研究[J]. *计算机仿真*, 2011,28(7):357-360
- [8] 潘斌斌. 多目标路径规划问题的算法综述[J]. *重庆工商大学学报:自然科学版*, 2012,29(5):78-81

Research on the Vehicle Scheduling Method with Multiple Sequential Time and Batch

WU Chuan-liang¹, HE Ye-rong¹, WANG Jian-hua¹

(School of Economic and Management, Huainan Normal University, Huainan 232038, China)

Abstract: Through plan, scientific and accurate scheduling, scheduling vehicle with multiple sequential time and batch is to achieve the purpose of reducing logistics cost. Firstly, according to multiple sequential period change speed, this paper adopts the linear least squares method to change speed quantization. Secondly, adding time window constraints, capacity constraint, this paper construct good vehicle scheduling model of smultiple sequential time and batch. Finally, according to the basic ant algorithm, the model is improved and simulated.

Key words: multiple sequential time; vehicle scheduling; ant algorithm