

基于时间序列模型的短期中国能源消费量预测研究

陈 琛

云南财经大学, 云南 昆明

收稿日期: 2023年1月21日; 录用日期: 2023年2月11日; 发布日期: 2023年2月23日

摘 要

能源的使用与社会生产生活有着密切的联系, 对此能源消费量的准确预测对于制定能源生产和消费规划以及保持中国经济健康、持续发展具有极其重要的理论与现实意义。本文采用ARIMA模型, 利用中国1990~2019年的能源消费量进行时间序列建模, 2020~2021年能源消费量数据进行测试。通过相关的检验和分析, 疏系数ARIMA((5),2,1)模型的拟合效果较好, 对比2020和2021年真实数据发现预测平均误差仅为0.835%, 因此利用该模型对中国未来三年的能源消费量进行预测。预测结果表明, 未来三年中国能源消费量仍然增加, 但是增长速度放缓, 符合中国当前低碳经济发展背景。

关键词

能源消费量, 时间序列模型, 疏系数模型, 预测

Short-Term Energy Consumption Forecasting in China Based on Time Series Model

Chen Chen

Yunnan University of Finance and Economics, Kunming Yunnan

Received: Jan. 21st, 2023; accepted: Feb. 11th, 2023; published: Feb. 23rd, 2023

Abstract

The use of energy is closely related to social production and social life, the accurate prediction of energy consumption for the development of energy production and consumption planning, keeping China's economy healthy and sustainable development has the extremely important theoretical and practical significance. In this paper, ARIMA model is adopted to conduct time series mod-

eling with China's energy consumption from 1990 to 2019, and the data of energy consumption from 2020 to 2021 are used for testing. Through relevant testing and analysis, ARIMA((5),2,1) model has a good fitting effect. Compared with the real data of 2020 and 2021, it is found that the average error of prediction is only 0.835%. Therefore, this model is used to forecast the energy consumption of China in the next three years. The prediction results show that in the next three years, China's energy consumption will still increase, but the growth rate will slow down, which is in line with China's current low-carbon economic development background.

Keywords

Energy Consumption, Time Series Model, Sparse Coefficient Model, Forecast

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

能源的使用是各行业必发展必不可少的要素投入，是国民经济和环境可持续发展的关键因素。随着经济的增长，能源消费总量的也在逐年攀升，从能源供需角度来看，由图 1 可知 1990~2021 年中国能源消费量一直在持续增长，在 1990~2000 年的增长速度一直较慢，从 2001 年开始突然增长速度加快，在 2008 年增长减缓，但是之后消费的增速又再次升高。其主要原因可能是受到 2008 年发生的世界金融危机影响，此时全国的经济增长速度放缓，由此能源市场业的需求降低能源消费量下降。在渡过世界金融危机后市场回归正常，能源消费量增长速度也恢复。进一步从能源使用排放角度来看，能源大量消耗的同时会产生许多污染物，例如二氧化碳等。当前中国为控制二氧化碳排放、应对全球气候变暖提出了“双碳”目标，但是中国的能源储备情况造就了中国能源消费结构将长期以煤炭为主，而煤炭的燃烧又会对生态环境、以及“双碳”目标的实现带来一系列的环境问题。

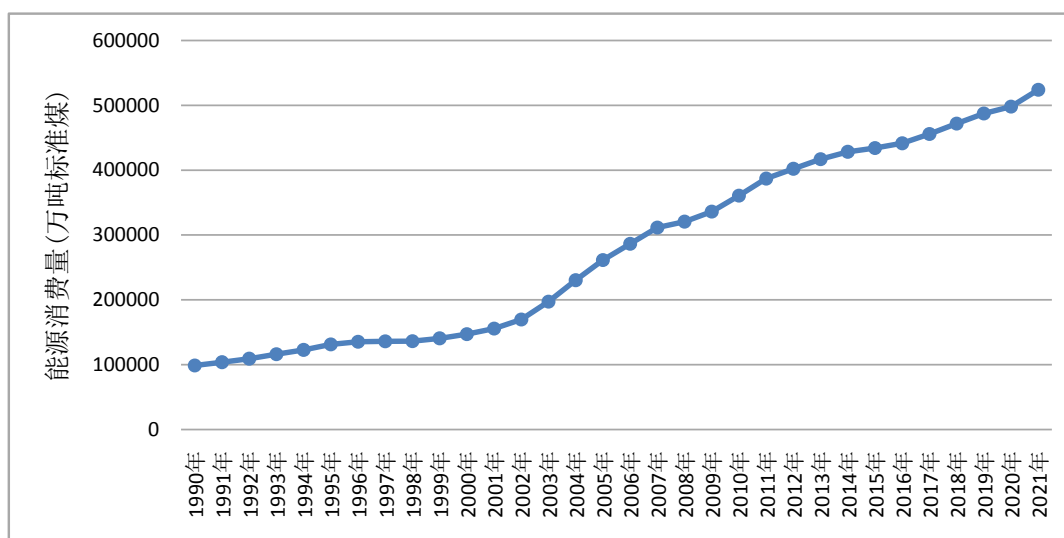


Figure 1. Time sequence diagram of Total energy consumption

图 1. 能源消费总量时序图

能源的使用与社会生产生活有着密切的联系,在当前的低碳经济环境下,中国能源大量消耗带来的资源、环境等问题已经不容忽视。本文利用最新的能源消费数据,通过构建时间序列模型对能源消费总量进行拟合及预测。有效预测能源消费量以及变化趋势,对制定能源生产和消费规划以及保持中国经济健康、持续发展具有极其重要的理论与现实意义。

2. 文献综述

通过对预测能源消费量的相关文献进行梳理,从研究方法角度将文章分为两类:第一类,依靠统计学方法的传统预测方法,包括时间序列法、回归分析法、灰色系统等方法。

其中时间序列法是运用变量的历史数据推测未来的发展趋势,主要为 ARIMA 模型,但是该模型只适用于短期预测,应用于中长期预测时误差较大。例如张峰和刘伟[1]运用 ARIMA 模型对北京市的能源消费量进行预测,得出 ARIMA 模型能够较为准确的预测能源消费量且未来北京市能源消费量的增长速度不断加快的结论。刘勇和汪旭晖[2]利用 ARIMA 模型在最小方差意义下达到了能源消费量的最优预测,得出中国能源消费量仍将会保持较高的增长水平结论。刘爱芹[3]将 Holt-Winters 无季节性指数平滑模型和 ARIMA 模型组合,利用组合模型对中国能源消费总量预测得出与刘勇等人同样的结论。

回归分析法是考虑到其他变量对于能源消费量的影响而建立的回归方程用于预测能源消费量。如张露[4]等人在选取影响能源消费的相关变量后,建立回归模型得到能源消费总量预测模型。林卫斌[5]等人在分析经济增长、效率改进和结构变化对能源消费量影响的基础上构建了能源需求回归模型,然后结合情景分析法对未来能源消费量进行预测。虽然回归分析模型考虑到其他变量的影响,但是需要知道未来时期中影响变量的值,且情景预测法的配合可能使得误差更大。

灰色预测法是通过灰色理论进行预测,适合中短期预测。其特点是可以处理数据量少、信息不充足的系统,适用于指数增长模型,通常用模型 GM(1,1)。如陈洪涛和周德利[6]用 GM(1,1)模型预测了中国能源消费量,得出中国能源消费量呈准指数增长的结论。还有改进的灰色预测法,如花玲和谢乃明[7]运用 GM(1,1)以及基于缓冲算子的 GM(1,1)同时对能源消费量进行预测,结果表明改进的 GM 模型可以更好的预测能源消费量。还有对 ARIMA 和 GM 模型相结合的改进方法,如宋宇辰[8]等人先利用 ARIMA 模型和灰色预测模型分别预测内蒙古制造业能源消费量,然后采用标准差法计算权重得到组合预测模型,得出组合预测模型的预测精度高于单一模型预测结果的结论。但是灰色预测模型仅适用于指数增长模型。

第二类,非线性预测方法。传统模型大多基于线性假设,但是随着预测模型的发展,出现了利用计算机对数据训练以此预测的非线性模型,如神经网络、支持向量机(SVR)等方法,具有非线性的特点。如邢文婷[9]等人利用灰色变量 GM(1,N)模型、支持向量机(SVR)、卷积神经网络(CNN)三种模型对中国天然气进口量进行拟合,选取拟合精度最高的模型进行预测。其中 CNN 模型预测精度最高,得出中国天然气进口量增长速度逐渐下降但是仍呈上升趋势。姜洪殿[10]等人对中国新能源的开发应用现状进行梳理,然后通过对 GM(1,1)模型和 BP 神经网络模型两个模型进行加权组合构建了组合模型,利用该组合模型对中国新能源消费量进行预测。高迪等人[11]通过变权重组合合法组合的 ARIMA-BP 神经网络模型对上海市工业企业能源消费量进行预测。结果表明,与单一模型相比组合模型可以降低平均相对误差,从而提高预测结果的精确度。张磊[12]等人认为传统支持向量机(SVR)及其改进模型难以合理的对能源消费量进行预测,提出一个结合 AP 聚类算法构建最优化训练集,并用 SVR 模型进行预测的 AP-SVR 模型。结果表明,组合模型可以有效识别样本训练集中能源消费量数据累积规律的差异,拟合精度得到显著提高。

综上所述,各种预测方法均有自己的特点,需要根据数据再选择方法,考虑到能源消费量数据情况以及预测结果的精准度,本文采用 ARIMA 模型对能源消费量进行预测,该模型对于短期预测较为精准且可以对时间序列数据的结构和特征有本质上的认识[2]。

3. 模型介绍

3.1. ARIMA 模型

差分自回归移动平均模型(ARIMA 模型)是以随机理论为基础的时间序列预测方法。该模型的基本思想是在 t 时刻变量 x_t 可以由过去的自身数值和随即误差项解释, 即不需要建立因果关系, 只需要研究对象的过往数据就可以建模。由于现实中时间序列数据大多情况下是不平稳的, 可以通过对数据进行差分使得数据平稳, 模型的形式为[13]:

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

式中, ∇ 为差分算子, $\nabla x_t = x_t - x_{t-1}$; B 为延迟算子, $x_{t-p} = B^p x_t$; $\nabla^d = (1-B)^d$, $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ 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。

3.2. 参数显著性检验

利用数据拟合得到的模型应当检验参数是否显著为 0, 参数检验的假设条件为:

$$H_0: \beta_j = 0 \quad \forall 1 \leq j \leq m$$

$$H_1: \beta_j \neq 0 \quad \forall 1 \leq j \leq m$$

此处选择 t 统计量:

$$t = \sqrt{n-m} \frac{\hat{\beta}_j}{\sqrt{a_{jj} Q(\hat{\beta})}} \sim t(n-m)$$

若拒绝原假设, 认为该参数显著不为 0, 否则认为该参数不显著, 这时应当剔除不显著参数[13]。其中, n 为观测期数; m 为系数个数。

3.3. 白噪声检验

对于差分平稳后的序列如果是白噪声, 则没有回归的必要; 对于拟合模型若已经全部提取所有信息, 则残差序列应当为白噪声[13]。假设条件如下:

$$H_0: \rho_1 = \rho_2 = \dots = \rho_m = 0, \forall m \geq 1$$

$$H_1: \text{至少存在某个 } \rho_k \neq 0, \forall m \geq 1, k \leq m$$

此处选取 LB 统计量做白噪声检验:

$$LB = n(n+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{n-k} \sim \chi^2(m)$$

其中, n 为观测期数; m 为指定延迟期数。

4. 数据来源

本文目的是预测中国能源消费总量, 此处能源消费总量是指一定地域内, 国民经济各行业和居民家庭在一定时期消费的各种能源转化为标准煤的总和。包括: 原煤、原油、天然气、洗煤、焦炭、电力等

各种能源。数据来源于《中国统计年鉴》和国家统计局。取 1990~2021 年中国能源消费总量进行分析，利用 1990~2019 年数据进行建模，2020、2021 年数据作为检验数据，并预测未来三年的能源消费量。

5. 模型建立

5.1. 平稳性检验

由图 1 能源消费总量时序图可知，能源消费总量数据呈逐步上升的趋势，故数据不平稳，需要对该序列进行处理。对我国能源消费量 x_t 数据利用 ADF 单位根检验做平稳性检验，原序列和一阶差分序列均不平稳，二阶差分后得到 $\Delta^2 x_t$ 序列平稳。但是观察 $\Delta^2 x_t$ 的时序图，如图 2 所示，除去 2008 年金融危机影响产生的异常，数据显示存在一定的异方差，因此考虑将数据做对数变换以消除异方差。

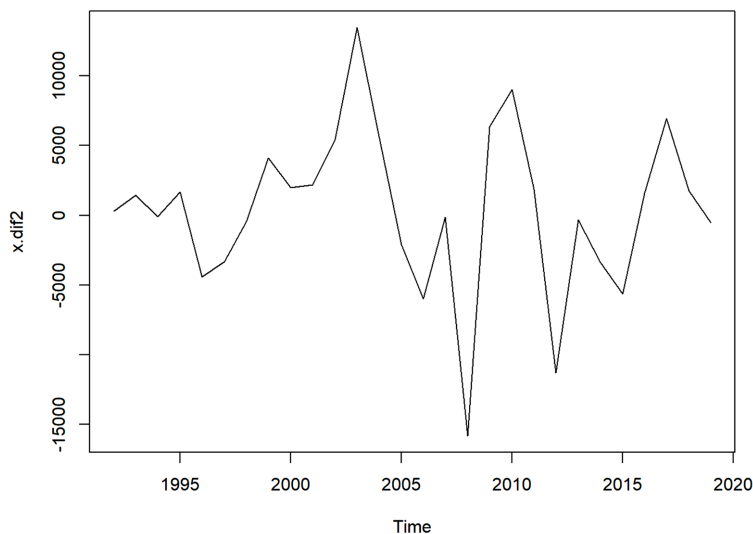


Figure 2. Sequence diagram of the original sequence in the second order difference case

图 2. 原序列二阶差分时序列图

对中国能源消费量 x_t 取对数以消除异方差性，用 $\ln x_t$ 表示。利用 ADF 单位根检验以确定 $\ln x_t$ 序列的平稳性，检验结果如表 1，根据结果可知 $\Delta^2 \ln x_t$ 平稳。

Table 1. ADF test results

表 1. ADF 检验结果

序列	检验形式	ADF 检验值	P 值	检验结果
$\ln x_t$	无常数均值无趋势项	7.57	0.990	不平稳
	有常数均值无趋势项	-0.844	0.738	
	有常数均值有趋势项	-0.49	0.976	
$\Delta \ln x_t$	无常数均值无趋势项	-1.02	0.307	不平稳
	有常数均值无趋势项	-1.61	0.471	
	有常数均值有趋势项	-1.65	0.698	
$\Delta^2 \ln x_t$	无常数均值无趋势项	-3.72	0.010	平稳
	有常数均值无趋势项	-3.65	0.013	
	有常数均值有趋势项	-3.59	0.050	

对 $\Delta^2 \ln x_t$ 平稳序列做白噪声检验, 以保证数据拟合模型具有意义。在延迟 5 阶时的 LB 统计量对应的 p 值 = 0.04788 < 0.05, 说明序列非白噪声序列, 可以进行拟合。

5.2. 自相关系数和偏自相关系数

由平稳性检验结果可知, 应该建立 ARIMA($p, 2, q$)模型。进一步还需要考察序列的自相关系数(ACF)和偏自相关系数(PACF), 以此确定 p, q 值。 $\Delta^2 \ln x_t$ 的 ACF 和 PACF 如图 3 所示, 除了延迟 5 阶的自相关系数外其余系数均在 2 倍标准差范围内, 可以判断该序列有短期相关性且在 5 阶截尾。同理可知, 偏自相关系数也是在 5 阶截尾。

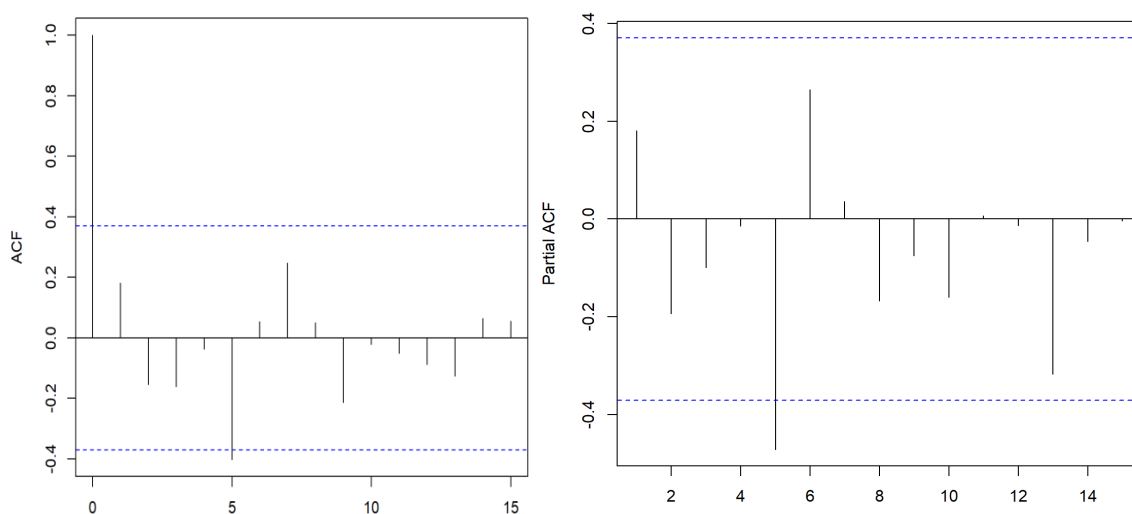


Figure 3. Autocorrelation and partial autocorrelation
图 3. 自相关和偏自相关图

5.3. 模型的构建与检验

同一个序列可以构造多个拟合模型, 因此应当充分考虑 p, q 的各种取值情况, 构建多个拟合模型。因为 $\Delta^2 \ln x_t$ 的 ACF 和 PACF 均在 5 阶截尾, 综合考虑认为可以尝试模型[13]: 1) ARIMA(5,2, i)、ARIMA($i, 2, 5$); 2) 疏系数模型: ARIMA((5),2, i)、ARIMA($i, 2, (5)$)、ARIMA((5),2,(5)), 其中 $i = 1, 2, \dots, 5$ 。

对上述模型通过 RStudio 软件利用最小二乘和极大似然估计混合方法进行拟合, 并对模型参数在 10% 置信水平下进行参数的显著性检验, 最终仅有 ARIMA(5,2,0)、ARIMA((5),2,0)、ARIMA((5),2,1)、ARIMA(0,2,(5))、ARIMA(1,2,(5))以上五个模型通过系数显著性检验, 回归结果见表 2。

再对通过参数检验的上述五个模型进行模型的显著性检验, 即对其残差序列进行白噪声检验。表 2 中 p 值为对拟合模型的残差利用 LB 统计量做白噪声检验滞后 6 阶和 12 阶的 p 值, 结果显示残差序列均为白噪声序列, 通过模型的显著性检验。

Table 2. Model regression results
表 2. 模型回归结果

	ARIMA(5,2,0)	ARIMA((5),2,0)	ARIMA((5),2,1)	ARIMA(0,2,(5))	ARIMA(1,2,(5))
x_{t-1}	0.3425	0	0	0	0.2995
x_{t-2}	-0.1686	0	0	0	0
x_{t-3}	-0.1941	0	0	0	0

Continued

x_{t-4}	0.2205	0	0	0	0
x_{t-5}	-0.5144	-0.4179	-0.4654	0	0
ε_{t-1}	0	0	0.3867	0	0
ε_{t-2}	0	0	0	0	0
ε_{t-3}	0	0	0	0	0
ε_{t-4}	0	0	0	0	0
ε_{t-5}	0	0	0	-0.4494	-0.4456
AIC	-130.41	-131.52	-133.66	-131.46	-132.18
p 值	0.9918	0.4321	0.8156	0.3678	0.5479
	0.9743	0.7838	0.9868	0.6491	0.8109

对于通过模型检验的有效拟合模型本文通过 AIC 信息准则进一步对模型优化, 以期选择最优的模型对序列进行拟合和预测。AIC 信息准则是 AIC 的值越小越好, 由表 2 中可知, 认为 ARIMA((5),2,1)模型拟合中国能源消费总量的效果较好, 模型拟合结果为:

$$(1-B)^2 \ln x_t = \frac{1+0.3867B}{1+0.4654B^5} \varepsilon_t$$

进一步对该模型的残差做 LM 检验和 PortmanteauQ 检验, 以检验异方差, 检验结果如表 3 所示, 结果显示不拒绝原假设, 残差序列方差齐性, 说明没有异方差, 同时说明将数据取对数后可以消除异方差, 模型拟合较好。

Table 3. Results of heteroscedasticity test

表 3. 异方差检验结果

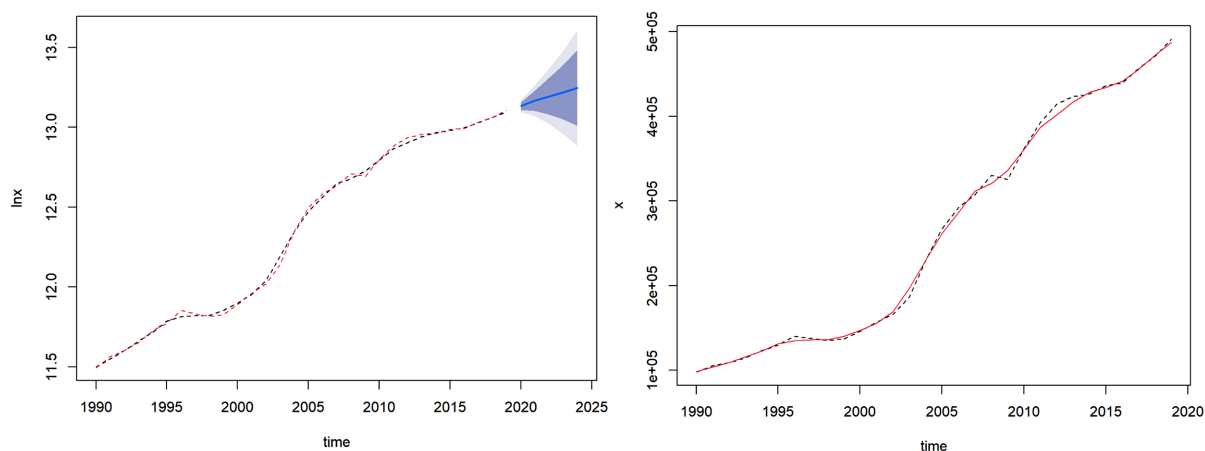
检验方法	Chi/X-squared	df	p 值
LM 检验	0.04346	1	0.8349
	0.39156	2	0.8222
	0.60786	3	0.8946
	0.63078	4	0.9596
	0.79565	5	0.9773
Portmanteau Q 检验	0.04332	1	0.8351
	0.37205	2	0.8303
	0.56682	3	0.9040
	0.5887	4	0.9643
	0.64414	5	0.9859

5.4. 预测

用 ARIMA((5),2,1)对中国 2020~2024 年的能源消费总量进行预测, 预测结果见表 4, 得到的预测图见图 4。由表 4 可知, 2020 年和 2021 年的预测值分别为 505306.27 万吨标准煤和 522938.95 万吨标准煤, 与实际值的相对误差分别为 1.47%和 0.20%, 说明预测结果与实际值相对接近, 并且对于 2021 年预测误差下降仅为 0.20%, 存在预测精度变高的情况, 进一步说明选择的模型可以较为准确的对未来能源消费量进行预测。同时由图 4 可知拟合结果较好, 其中图 4 中黑色虚线为预测值, 红色实线为预测值。

Table 4. Forecast results**表 4.** 预测结果

时间	2020	2021	2022	2023	2024
预测值(万吨标准煤)	505306.27	522938.95	537390.34	551540.20	566657.31
实际值(万吨标准煤)	498000	524000			
相对误差(%)	1.47	0.20			

**Figure 4.** Forecast of logarithm of energy consumption and energy consumption**图 4.** 能源消费量对数和能源消费量的预测图

6. 结论

ARIMA 模型可以作为预测中国能源消费总量的工具,并且预测精度较好。通过预测结果可以看出在未来 2022~2024 年中国能源消费总量仍然在持续增长。在 2021~2024 年能源消费总量将以 3.49%、2.76%、2.63%、2.74% 的增长率增长,可以看出能源增长速度有所放缓,符合中国当前低碳经济背景。目前中国很难摆脱以煤炭为主的能源消费情况,但是可以通过:

1) 加强科技创新。科技创新可以加快新能源与清洁能源的开发,同时还可以提高能源使用效率使得在同产出下能源消耗更少。

2) 有关部门制定相应促进节能政策,如在国外已经实行的碳税与碳交易政策,其中碳税政策是否可以借鉴国外经验并转化为适合中国国情的节能减排政策。其次中国的碳交易市场已经从部分试点到全国推行,应当在继续推行的同时做好其他辅助工作以使得达到节能减排作用。

3) 提高居民低碳意识,通过各种宣传途径大力推广低碳环保的好处,倡导居民绿色低碳生活。

参考文献

- [1] 张峰, 刘伟. 北京市能源消费预测与政策建议[J]. 中国人口·资源与环境, 2008(3): 99-102.
- [2] 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+32.
- [3] 刘爱芹. 基于组合模型的能源消费预测研究[J]. 中国人口·资源与环境, 2010, 20(11): 25-29.
- [4] 张露, 李永安, 王崇. 基于相关分析和回归分析的能源消费影响因素研究[J]. 中国能源, 2020, 42(6): 42-47.
- [5] 林卫斌, 苏剑, 周晔馨. 新常态下中国能源需求预测: 2015-2030 [J]. 学术研究, 2016(3): 106-112+178.
- [6] 陈洪涛, 周德群. 基于 GM(1,1)模型的中国能源消费预测研究[J]. 矿业研究与开发, 2007, 27(3): 77-79.
- [7] 花玲, 谢乃明. 政策冲击影响下中国能源消费预测分析及控制策略[J]. 中国管理科学, 2014, 22(7): 18-25.

- [8] 宋宇辰, 郭丽, 孟海东. 内蒙古制造业能源消费的组合预测模型[J]. 水电能源科学, 2012, 30(4): 205-209.
- [9] 邢文婷, 袁琳, 张巧. 基于机器学习方法的我国天然气进口预测[J]. 工业技术经济, 2022, 41(9): 136-144.
- [10] 姜洪殿, 董康银, 孙仁金, 宋显昱. 中国新能源消费预测及对策研究[J]. 可再生能源, 2016, 34(8): 1196-1202.
- [11] 高迪, 任庚坡, 李琦芬, 毛俊鹏, 桂雄威. 基于驱动因素分解的能源消费预测——以上海市为例[J]. 重庆理工大学学报(自然科学), 2021, 35(9): 269-277.
- [12] 张磊, 刘驰昉, 康淑瑰. 基于优化 AP-SVR 模型的能源消耗预测[J]. 山西大学学报(自然科学版), 2022, 45(4): 1004-1013.
- [13] 王燕. 时间序列分析: 基于 R [M]. 第二版. 北京: 中国人民大学出版社, 2015.