

基于电力数据的能耗及碳排放测算模型构建与应用

王成围¹, 李沛¹, 杨至元¹, 陈政¹, 何耿生¹, 曾金灿¹, 李姚旺², 杜尔顺²

(1. 南方电网能源发展研究院, 广州 510663; 2. 清华四川能源互联网研究院, 成都 610213)

摘要: 针对当前相关政府部门、企业对能源消费和碳排放等数据时效性和高频次更新的现实需求, 以电力数据为核心, 多源数据相融合, 通过对比分析多种模型在多种拟合方法下的表现, 确定了岭回归赋权下的多元线性回归(ridge-multivariable linear regression, R-MLR)能耗测算模型, 并在此基础上建立了碳排放监测模型。应用线性模型的可叠加性原理实现了年度数据建模、月度数据预测的功能, 解决了统计数据频率不一造成的建模困难的关键问题, 同时避免了传统数据拆分方法在数据源头引入的误差。岭回归赋权方法的应用解决了模型输入数据的多重共线性问题。所提模型适用范围广, 在区域、行业等不同空间尺度和年度、季度、月度等不同时间尺度实现了能源和碳排放监测与预测的功能, 具有“以简应繁”的效果, 填补了当前能源和碳排放数据高频统计数据上的空白, 可为相关政府部门、行业协会等进行能源消费和碳排放的监测以及制定相关政策提供理论依据和数据支撑。

关键词: 能源消费; 碳排放监测; 能耗强度; 碳排放强度; 岭回归

Construction and Application of Energy Consumption and Carbon Emission Calculation Model Based on Power Data

WANG Chengwei¹, LI Pei¹, YANG Zhiyuan¹, CHEN Zhen¹, HE Gengsheng¹, ZENG Jincan¹,

LI Yaowang², DU Ershun²

(1. Energy Development Research Institute, CSG, Guangzhou 510663, China; 2. Sichuan Energy Internet Research Institute, Tsinghua University, Chengdu 610213, China)

Abstract: In response to the current practical demand of timely and high-frequency data updates of energy consumption and carbon emissions in relevant government departments and enterprises, based on power data as the core and integrating multiple sources of data, a ridge-multivariable linear regression model for energy consumption calculation is proposed, by comparing and analyzing the performance of various models under various fitting methodologies. Based on this, a model for carbon emission monitoring is proposed. The superposition principle of linear models is applied to achieve the functions of annual data modeling and monthly data prediction, solving the key difficulty of modelling with inconsistent frequency of statistical data and avoiding the errors introduced by traditional data splitting methods at the data source. The application of the ridge regression weighting method solves the multicollinearity problem of the input data. The proposed model has a wide range of applicability and realizes the functions of monitoring and predicting energy and carbon emissions at different spatial scales such as regions and industries and at different time scales such as annual, quarterly and monthly. It has the effect of "solving complexity with simplicity" and fills the gap of high-frequency statistics data in energy and carbon emissions and provides theoretical basis and data support for relevant government departments, industry associations, etc. to monitor energy consumption and carbon emissions and formulate relevant policies.

Key words: energy consumption; carbon emission monitoring; energy consumption intensity; carbon emission intensity; ridge regression

基金项目: 国家自然科学基金资助项目(52477103); 中国南方电网有限责任公司科技项目(ZBKJXM20232244, ZBKJXM20232245, ZBKJXM20232246)。

Foundation item: Supported by the National Natural Science Foundation of China (52477103); the Science and Technology Project of China Southern Power Grid Co., Ltd. (ZBKJXM20232244, ZBKJXM20232245, ZBKJXM20232246).

0 引言

随着全球气候变化加剧,减少碳排放、实现碳达峰和碳中和已成为国际社会的广泛共识。我国作为世界上最大的碳排放国,在碳排放方面承担着重大责任。能源消费与碳排放之间存在着紧密联系,能源消耗过程中,大量化石燃料被燃烧,释放出二氧化碳等温室气体^[1-3]。随着全球能源消费的持续增长和节能减排压力的持续增大,相关政府部门、企业对能源消费和碳排放等数据更新频率的要求越来越高^[4-9]。准确预测能源消费总量和及时的碳排放监测有助于相关部门制定合理的能源和减排政策,评估节能降碳效果,应对气候变化挑战。

国内外能源消费预测研究方法涉及多个学科领域,包括能源经济学、统计学、机器学习等。同时,越来越多的学者开始关注可再生能源的消费预测^[10-12]以及能源消费与经济、人口、技术等因素之间的动态关系^[13-15]。在碳排放监测领域,全球范围内已经建立了一些碳排放监测与研究平台,如欧盟的碳排放交易系统(EU emissions trading system, EU-ETS)、联合国政府间气候变化专门委员会(intergovernmental panel on climate change, IPCC)等。这些平台为碳排放监测和研究提供了重要的数据支持和技术手段。同时,卫星遥感监测^[16-17]、在线监测^[18-19]等直接测量方法和排放因子模型^[20-21]、统计预测模型^[22-25]等间接测量方法不断涌现。

尽管能源消费预测和碳排放监测研究方面均取得了一系列的成果^[26-31],二者仍面临诸多共同的问题和挑战。当前能源统计数据在时效性上仍存在一定的滞后,且统计频率为年度,月度、季度的能源消费数据的统计和预测研究尚不充分^[32-33]。为了解决能源数据统计时效性的问题,填补能源数据在月度、季度等统计频次上的空白,当前研究人员一般做法是首先选取一个月度、季度标的变量作为拆分依据,然后利用Denton等模型^[34]将年度数据进行月度、季度拆分,最后进行模型拟合。从准确性的角度来看,该做法在数据拆分过程中不可避免地引入误差,还将出现模型输入数据既有统计数据又有手工转换数据的情况,无法保证输入数据源头的一致性。

基于相关政府部门、企业对能源消费和碳排放等数据统计及时性的需求,针对当前测算能耗和碳

排放数据的方法所存在的问题和不足,本文将用电量等对能源消费具有显著影响的若干个自变量与历史能源消费数据建立关联关系,通过测试、比较多元线性回归模型(multivariable linear regression, MLR)、分布滞后模型(distributed lag, DL)、自回归分布滞后模型(auto regressive distributed lag, ARDL)和电能回归等多种模型在最小二乘法、岭回归、LASSO(least absolute shrinkage and selection operator)等多种数据拟合方法下的误差和模型拟合优度,从组合中选出最优模型,构建了基于岭回归赋权的多元线性回归(ridge-multivariable linear regression, R-MLR)模型,应用于区域及行业全口径能耗及碳排放测算。本研究在建模过程中根据线性可叠加性原理采用年度数据建模、月度数据测算的方式解决了历史统计数据 and 模型预测数据频率不一致所造成的建模困难的问题,不仅避免了传统数据拆分方法可能引入的误差,而且统一了不同时间尺度下模型的函数形式。研究成果可为相关政府部门、行业协会等进行能源消费和碳排放的监测以及制定相关政策提供理论依据和数据支撑。

1 模型构建原理

岭回归赋权的多元线性回归(R-MLR)能耗及碳排放监测模型原理如图1所示,模型可以根据数据的可获取情况自适应调整输入自变量的维度,针对区域模型引入区域灵敏变量、区域产业结构和宏观能源消费结构3个维度7个自变量,针对行业模型引入行业用电量、行业总产值、行业主要产品产量3个维度多个自变量。用电量数据获取方便、数据连续性好、统计准确率高,为模型输入的核心数据。同时,多维度自变量保证了信息输入丰度,降低了对单一数据源的依赖,减少了误判和错误测算的概率。在不同地区数据可获取情况不同的现实情况下,模型自变量自适应的特性保证了模型在多空间尺度应用的可推广性。本文所需数据来源于南方电网内部统计、地方政府统计年鉴、统计月报以及Wind数据库等。

本模型根据线性可叠加性原理,在自变量和因变量强线性相关性的事实基础上(表1统计显示自变量和因变量相关系数大于0.98)采用年度数据建模、月度数据测算的方式,在仅对常数项微调的前提下实现了在不同的时间尺度上函数形式的统一,

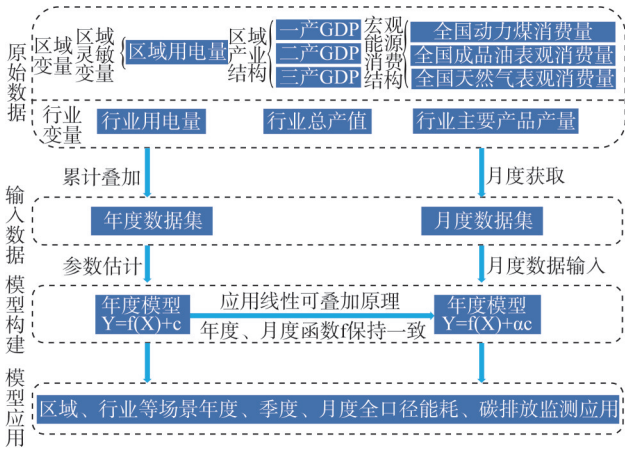


图 1 模型构建原理示意图

Fig. 1 Schematic diagram of the principle of model construction

解决了建模过程中历史统计数据和模型预测数据频率不一致所造成的建模困难的问题,保证了模型在多时间尺度应用的可复制性。模型数据来源为省区统计年鉴、电网统计数据、国家和行业相关部门统计数据等。

2 模型构建过程

2.1 能耗监测模型

2.1.1 能耗年度拟合模型

应用年度统计数据可拟合能耗年度模型,如式(1)所示。

$$y_t = f(x_t) + h(x_{t-1}, x_{t-2}, \dots, x_{t-k}) + g(y_{t-1}, y_{t-2}, \dots, y_{t-j}) + c + u_t \quad (1)$$

式中: y_t 为第 t 年年度能源消费总量; x_t 为第 t 年多维年度自变量组成的向量; $f(\cdot)$ 、 $h(\cdot)$ 和 $g(\cdot)$ 为线性函数,其中 $f(\cdot)$ 为 y_t 与当期自变量的拟合关系,

$h(\cdot)$ 和 $g(\cdot)$ 分别为 y_t 与自变量滞后项和因变量滞后项的拟合关系; k 和 j 分别为自变量和因变量的最大滞后量; c 为常数项; u_t 为残差。

式(1)可根据实际情况演化为 4 种回归模型。设定函数 $h(\cdot)$ 和 $g(\cdot)$ 所有参数为 0,即输入变量只考虑自变量当期值时,模型为多元线性回归模型(MLR);设定函数 $g(\cdot)$ 所有参数为 0,即输入变量同时考虑自变量当期值、滞后项,不考虑因变量滞后项时,模型为分布滞后回归模型(DL);当输入变量同时考虑自变量当期值、滞后项及因变量滞后项时,模型为自回归分布滞后模型(ARDL);特别的,当即输入变量只考虑自变量当期值且自变量向量只包含用电量一个元素时,模型简化为电能回归模型。

本文第 3 节将分别对 MLR、DL、ARDL 和电能回归 4 种模型进行对比测算,分析自变量维度及滞后项对模型拟合优度及测算精度的影响。

2.1.2 能耗监测月度模型

根据线性模型的可叠加性原理,在模型形式保持不变的前提下,年度数据拟合得到的模型参数在月度尺度依然适用,仅需在月度模型中将年度模型的常数项按比例分摊到月度即可。略去残差项,月度模型如式(2)所示,根据年度模型拟合得到的函数关系 $f(\cdot)$ 、 $h(\cdot)$ 和 $g(\cdot)$,输入月度自变量数据,即可测算月度能源消耗总量 $y_{t,m}$ 。

$$y_{t,m} = f(x_{t,m}) + h(x_{t-1,m}, x_{t-2,m}, \dots, x_{t-k,m}) + g(y_{t-1,m}, y_{t-2,m}, \dots, y_{t-j,m}) + \alpha c \quad (2)$$

式中: $x_{t,m}$ 为第 t 年第 m 月多维月度自变量组成的向量; m 为月份; $\alpha = 1/12$ 为年度模型常数项在月

表 1 模型变量之间的皮尔逊相关系数

Tab. 1 Pearson correlation coefficients between model variables

模型	第一产业 GDP	第二产业 GDP	第三产业 GDP	用电量	全国动力煤消费量	全国天然气表观消费量	全国成品油表观消费量
第一产业 GDP	1.00	0.99	0.99	0.98	0.92	0.99	0.98
第二产业 GDP	0.99	1.00	0.98	1.00	0.95	0.99	0.99
第三产业 GDP	0.99	0.98	1.00	0.97	0.87	0.99	0.95
用电量	0.98	1.00	0.97	1.00	0.96	0.99	0.99
全国动力煤消费量	0.92	0.95	0.87	0.96	1.00	0.91	0.97
全国天然气表观消费量	0.99	0.99	0.99	0.99	0.91	1.00	0.97
全国成品油表观消费量	0.98	0.99	0.95	0.99	0.97	0.97	1.00
南方区域能源消费总量	0.98	0.99	0.95	0.99	0.97	0.98	0.99

度模型中的分配系数。

季度能耗总量 $y_{t,s}$ 的测算可由式(3)计算得出。

$$y_{t,s} = \sum_{m \in S} y_{t,m} \quad (3)$$

式中： s 为季度； S 为属于季度 s 的月份所组成的集合。

2.1.3 近月数据预测

考虑到月度自变量公布具有延迟性，需要对近月或当月数据进行预测，以满足当月能耗测算的需求。假设预测月份的数据占其向前滚动一个完整年数据和的比例与去年同期月占其向前滚动一个完整年数据和的比例相等，近月或当月自变量数据可通过求解式(4)~(6)所列方程得到。当需要预测多个月份数据时，可按时间顺序从前向后依次分月预测，将已预测月份的数据作为已知量用于预测下一个月份。

当预测月份为3—11月时，求解式(4)得到第 t 年第 m 月的自变量数据。

$$\frac{x_{t,m}}{\sum_{i=1}^m x_{t,i} + \sum_{i=m+1}^{12} x_{t-1,i}} = \frac{x_{t-1,m}}{\sum_{i=1}^m x_{t-1,i} + \sum_{i=m+1}^{12} x_{t-2,i}} \quad (4)$$

式中： t 为年份； m 为预测月份， $3 \leq m \leq 11$ ； i 为求和索引。

由于春节影响，不同年份间1月和2月数据波动较大，因此1月和2月作为整体进行预测，不再拆分，公式如式(5)所示。

$$\frac{x_{t,1-2}}{x_{t,1-2} + \sum_{i=3}^{12} x_{t-1,i}} = \frac{x_{t-1,1-2}}{x_{t-1,1-2} + \sum_{i=3}^{12} x_{t-2,i}} \quad (5)$$

式中 $x_{t,1-2}$ 为预测自变量1月和2月份的和。

对于12月数据 $x_{t,12}$ 的预测，其计算公式如式(6)所示。

$$\frac{x_{t,12}}{\sum_{i=1}^{12} x_{t,i}} = \frac{x_{t-1,12}}{\sum_{i=1}^{12} x_{t-1,i}} \quad (6)$$

2.2 能耗监测模型赋权方法

针对式(1)所示的能耗监测模型，通过选取合适的赋权方法对历史统计数据进行拟合，即可求得模型中各项自变量的权重系数。本节介绍几种常用的赋权方法，尝试通过惩罚系数控制自变量的权重从而解决过拟合和多重共线性的问题。

2.2.1 多重共线性问题

在进行多元线性回归分析时，容易出现自变量

之间彼此强相关的现象，这种现象称为多重共线性。当出现严重共线性问题时会导致分析结果不稳定，甚至出现结果与实际完全相反的情况，使得本应该显著的自变量不显著，本不显著的自变量却呈现出显著性，这种情况下就需要消除多重共线性的影响。可采用皮尔逊相关性系数对数据相关性定量分析，皮尔逊相关系数由式(7)计算得出。

$$r = \frac{\sum (A - \bar{A})(B - \bar{B})}{\sqrt{\sum (A - \bar{A})^2} \sqrt{\sum (B - \bar{B})^2}} \quad (7)$$

式中： A 和 B 分别为两组数据的变量； r 为 A 和 B 的皮尔逊相关系数； $\bar{A}$ 和 $\bar{B}$ 分别为 A 和 B 的平均值。

皮尔逊相关系数是一个无量纲的统计指标，其取值范围为 $[-1, 1]$ ，相关系数小于0为负相关，大于0为正相关，等于0表示不存在相关性。相关系数的绝对值越大两变量间的相关程度越高。

表1列出了南方五省区(广东、广西、云南、贵州、海南)区域模型同期自变量之间、各自变量与因变量(南方五省区能源消费总量)之间的皮尔逊相关系数，所有自变量之间的相关系数都不小于0.87，相关系数平均值为0.97。同时，所有自变量均与能源消费总量存在较强的正相关性，相关系数不小于0.95。这表明自变量之间存在强线性相关，模型拟合过程中会出现严重的多重共线性问题。

本文将采用最小二乘法、岭回归法和LASSO回归法对所选取的模型进行参数拟合并对结果进行分析对比，解决拟合过程中的多重共线性问题。

2.2.2 最小二乘法赋权

最小二乘法拟合回归系数的过程为最小化残差项平方和，如式(8)所示。

$$\min \sum_i u_i^2 = \min \sum_i (y_i - \hat{y}_i)^2 \quad (8)$$

式中 $\hat{y}_i$ 为 y_i 的测算值。

对于多元线性回归，求解式(8)得到参数向量的矩阵估计形式如式(9)所示。

$$\hat{\mathbf{a}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \quad (9)$$

式中： $\hat{\mathbf{a}}$ 为参数估计列向量； $\mathbf{X}$ 为自变量矩阵； $\mathbf{X}'$ 为 $\mathbf{X}$ 的转置矩阵； $\mathbf{y}$ 为因变量列向量。使用最小二乘法进行参数估计的前提条件是矩阵 $\mathbf{X}$ 满列秩，否则 $\mathbf{X}'\mathbf{X}$ 无法求逆，严格的多重共线性会导致 $\mathbf{X}$ 不满列秩，近似的多重共线性造成矩阵 $\mathbf{X}'\mathbf{X}$ 几乎不可逆，其求逆结果使得参数估计的绝对值很大，对系

数的估计不准确,甚至出现系数符号与理论预期相反,使得模型过拟合。

2.2.3 岭回归赋权

为了解决多重共线性问题,岭回归在残差平方和最小化的基础上加了一个参数L2范数平方的惩罚项,即:

$$\min \sum_i u_i^2 + \lambda \|\hat{\mathbf{a}}\|_2^2 \quad (10)$$

式中 λ 为惩罚系数。

参数向量的估计式变为:

$$\hat{\mathbf{a}} = (\mathbf{X}'\mathbf{X} + \lambda\mathbf{E})^{-1}\mathbf{X}'\mathbf{y} \quad (11)$$

式中 $\mathbf{E}$ 为单位矩阵。L2范数平方惩罚项的加入使得 $\mathbf{X}'\mathbf{X} + \lambda\mathbf{E}$ 满秩,保证了矩阵的可逆性,但是由于惩罚项的加入使得回归系数的估计不再是无偏估计,所以岭回归是以放弃无偏性、牺牲无偏性为代价解决病态矩阵问题的回归方法。

2.2.4 LASSO回归赋权

与岭回归不同,LASSO回归为了解决多重共线性问题在残差平方和最小化的基础上加了一个L1范数惩罚项,即:

$$\min \sum_i u_i^2 + \lambda \|\hat{\mathbf{a}}\|_1 \quad (12)$$

因为L1范数是绝对值的形式,在0点处不可导,因此它不再具有解析解,可以使用梯度下降法求解。LASSO回归能产生稀疏性,可以将一些不重要的回归系数缩减为0,达到剔除变量的目的。但由于剔除变量的影响,LASSO回归有可能会对变量惩罚过度而使得模型参估过程产生欠拟合。

3 模型优选

针对能耗测算式(1)演化出的电能回归模型、多元线性回归模型(MLR)、分布滞后模型(DL)、自回归分布滞后模型(ARDL)模型等4种模型分别在最小二乘、岭回归和LASSO赋权方法的应用下进行测算和分析,从而为能耗监测选取最优模型并解决拟合过程中的多重共线性问题。为综合评价各种模型在多种赋权方法下的表现,选取4种统计指标对模型测试结果进行比较。

3.1 模型评价指标

1)负参数数量/总参数数量:统计模型拟合过程中负参数的数量占总参数的数量的比例。由于模型自变量和因变量之间有很强的正相关性,大量负

参数的出现不符合预期,表示模型过拟合。

2)零参数数量/总参数数量:统计模型拟合过程中零参数的数量占总参数的数量的比例。零参数表示剔除相应的自变量,过多的零参数可能预示着模型的欠拟合。

3)决定系数 R^2 :定义式(13)该统计量表示自变量对因变量变化的解释程度,常用于评价模型拟合优度。 R^2 一般位于0到1之间, R^2 值越大表明模型拟合效果越好。

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (13)$$

式中 y_i 、 $\hat{y}_i$ 和 $\bar{y}$ 分别为第 i 年的真实值、预测值和所有年份真实值的平均值。

4)平均绝对百分比误差 M_{APE} :定义如式(14),此指标对相对误差敏感,不会因目标变量的量纲缩放而改变,便于测算结果横向对比。

$$M_{APE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (14)$$

式中 n 为统计数据的总年份。

3.2 模型测算结果

将南方五省区作为统一区域为研究对象,选取2008—2017年的年度历史数据为样本内数据进行模型拟合,选取2018—2022年的年度历史数据为样本外数据进行模型校验,测算4种模型在3种赋权方法下的拟合表现。DL和ARDL模型自变量的最大滞后量选取1、2、3分别进行测试,并从中选取结果最好的最大滞后量对应的结果进行统计。电能回归模型使用用电量单变量作为输入。

表2 模型拟合结果负参数、零参数占总参数个数的比例
Tab. 2 Proportion of negative parameters and zero parameters in the total number of parameters in the model fitting results

模型	负参数数量/总参数数量			零参数数量/总参数数量		
	最小二乘	岭回归	LASSO	最小二乘	岭回归	LASSO
电能回归	0/1	0/1	0/1	0/1	0/1	0/1
MLR	4/8	0/8	0/8	0/8	0/8	0/8
DL	7/16	0/16	0/24	0/16	0/16	22/24
ARDL	7/17	0/17	0/24	0/17	0/17	22/24

表2统计了4种模型在3种拟合方法下的“负参数数量/总参数数量”和“零参数数量/总参数数量”。电能回归模型由于只有一个参数,且用电量和能耗呈正线性相关,因此并未出现负参数或零参数。

MLR、DL 和 ARDL 3 种模型在最小二乘法赋权时均出现了大量的负参数,这与表 1 所示所有自变量均与因变量之间存在强正线性相关的事实相悖,预示着模型出现了过拟合。MLR、DL 和 ARDL 3 种模型在 LASSO 赋权时大部分参数为 0,表明 LASSO 在抑制多重共线性方面剔除了过多的变量,使模型欠拟合。岭回归赋权即无负参数也无零参数出现,表明其对所有的自变量进行了充分拟合。

表 3 统计了 4 种模型在 3 种拟合方法下样本外的决定性系数 R^2 和平均绝对百分误差 (M_{APE})。由于单自变量的限制,电能回归模型总体表现较差。岭回归赋权下的 MLR、DL、ARDL 3 种模型拟合优度均大于 0.9, M_{APE} 均小于 2%,表明模型拟合效果较好,其中 MLR 无论拟合优度还是误差均表现最好。对比最小二乘法、岭回归和 LASSO 3 种赋权方法在 3 种多元回归模型拟合中的表现,无论是从模型拟合优度的角度还是从平均绝对百分误差的角度,岭回归均表现出明显的优势,结合表 2 所示参数分析,可知原因是最小二乘法样本内过拟合、LASSO 样本内欠拟合,造成二者在样本外拟合优度下降,拟合误差上升。

表 3 样本外决定系数(R^2)和平均绝对百分误差(M_{APE})
Tab. 3 Out-of-sample goodness-of-fits (R^2) and mean absolute percentage errors (M_{APE})

模型	样本外 R^2			样本外 $M_{APE}/\%$		
	最小二乘	岭回归	LASSO	最小二乘	岭回归	LASSO
电能回归	-0.02	-8.34	0.99	6.6	19.8	0.5
MLR	-6.59	0.94	-0.15	15.0	1.4	7.3
DL	0.74	0.92	0.87	2.8	1.6	2.1
ARDL	0.78	0.92	0.87	2.7	1.6	2.1

综合以上分析,岭回归赋权下的多元线性回归(R-MLR)模型在所有测算中表现最优。同时,电能回归模型与其他模型的对比表明多变量的引入明显优化了模型测算结果,提高了模型的鲁棒性。DL、ARDL 和 MLR 模型的对比表明,自变量或因变量滞后项作为输入变量引入模型都在不同程度上降低了模型拟合优度,增大了误差。

滞后项无法提高模型精度的原因在于自变量数据与能耗数据之间往往是同期涨落,存在同周期效应,并无明显的时滞效应,因此引入滞后项不是影响模型精准度的必要条件,反而因为滞后项的引

入增加了模型输入变量之间的多重共线性,使得模型的拟合变得不稳定。同时,针对月度和季度模型,由于月度和季度数据往往存在一定的季节效应,自变量和因变量的同周期效应在月度和季度数据上表现更为明显,使用滞后变量会因为错周期数据的引入而打乱自变量和因变量之间的同步季节效应,使模型拟合变得更加困难。因此,未引入滞后项的 MLR 模型拟合效果要优于引入滞后项的 DL、ARDL 模型。

4 碳排放监测模型构建

将能源消费总量折算为碳排放需要考虑能源消费结构。一般情况下,区域能源消费统计中煤炭、原油、天然气和一次电力为主要能源消费形式,其中一次电力一般指绿色电力,其碳排放因子近似为 0^[20]。因此,一次电力在消费过程中引起的碳排放量可以忽略不计,煤炭、原油和天然气消费承担区域主要的碳排放责任。在已知区域能耗总量的前提下,区域碳排放量 e 可由式(15)近似估算。

$$e \approx y(w_1 \cdot q_1 + w_2 \cdot q_2 + w_3 \cdot q_3) \quad (15)$$

式中: y 为区域年度、季度或月度能耗总量; e 为对应时间尺度下的碳排放总量; w_1 、 w_2 和 w_3 分别为区域统计年鉴公布的区域煤炭、原油、天然气的能源消费占比,由于能源消费占比不会突变,因此当期占比未公布时以最近一期公布数值代替; $q_1=2.66$ 、 $q_2=1.73$ 和 $q_3=1.56$ 分别为煤炭、原油、天然气折标煤后的碳排放系数,单位均为 tCO_2/t 标准煤。

对于行业测算,由于不同行业的能源消费结构差别较大,可根据行业能源消费特性依式(16)有针对性地建立行业能耗与碳排放之间的折算关系。

$$e \approx y \sum_{i \in I} w_i \cdot q_i \quad (16)$$

式中: I 为某行业主要能源消费品类的集合; w_i 为第 i 种能源折标煤后所占能源消费总量的比例; q_i 为第 i 种能源折标煤后的碳排放系数。

5 应用算例分析

5.1 区域应用场景

R-MLR 模型可对不同时间尺度能耗总量进行测算,进而计算得到相应的碳排放总量,从而解决了季度和月度能耗总量、碳排放总量数据缺失的问

题,为双碳目标下相关部门制定相关政策提供更为及时、细致的数据支撑。

应用R-MLR模型对某省年度能源消耗总量进行测算和校验,选取2008—2017年的数据拟合模型。如图2所示,以2018年能源消费统计值为基准并将其设为1,样本外2018—2022年最小百分误差为-0.72%,平均绝对误差为2.17%。测算结果表明模型误差较小,且测算值较好地拟合了5年内能源消耗总量的增长趋势。

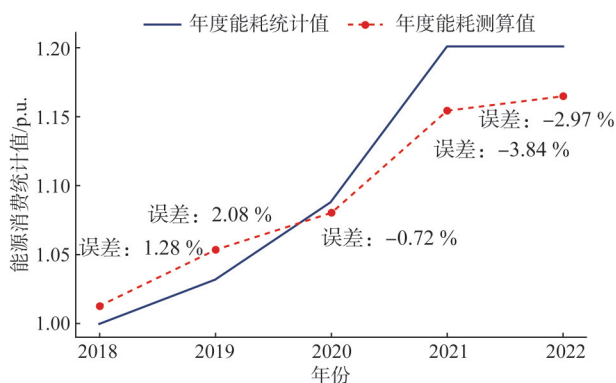


图2 某省年度能源消费测算校验(2018年能耗设为1)

Fig. 2 Calculation and verification of the annual energy consumption of a province (set the annual energy consumption of 2018 to be 1)

某省2018年一季度至2024年二季度能耗、碳排放总量及强度如图3所示。能耗、碳排放强度分别指单位GDP能耗、碳排放总量。能耗及碳排放总量呈现波动上升趋势,2024年二季度的能耗总量约为2018年一季度的1.16倍,同期碳排放量约为1.09倍。能耗总量最高点出现在2023年四季度为2018年一季度的1.17倍,同期碳排放为1.11倍。相对而言,能耗及碳排放强度则波动下降,2024年二季度能耗强度约为2018年一季度的78%,同期碳排放强度约为70%。能耗强度最低点出现在2023年四季度,约为2018年一季度的70%,同期碳排放强度为65%。

能耗和碳排放二者之间无论在总量还是强度方面,均呈现同涨同落的强正相关现象,且均呈现较强的季节特性。各年能耗、碳排放总量一季度较低,四季度最高,而强度则相反,呈现一季度较高,四季度最低的特性,能耗、碳排放总量与强度之间均呈现季节性负相关。其原因在于GDP存在强季节特性,往往一季度低、四季度高,且其变化

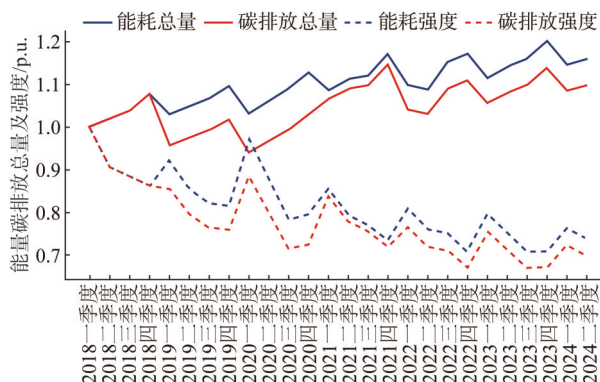


图3 某省季度能耗、碳排放总量及强度测算(2018年一季度测算值设为1)

Fig. 3 Calculation of the seasonal energy consumptions, carbon emissions and their intensities of a province (set the calculated values of the first season of 2018 to be 1)

幅度大于能耗、碳排放总量的变化幅度,使得能耗、碳排放强度在自然年内呈现前高后低的季节特性。

5.2 行业应用场景

将R-MLR模型的输入自变量调整为行业用电量、行业总产值、行业多种产品产量数据,同时计算各自变量与因变量的相关系数,挑选相关系数较高的自变量(如相关系数大于0.8)输入模型,即可对行业能源消费和碳排放情况进行测算。选取全国黑色金属冶炼及压延加工业(黑色金属),有色金属冶炼及压延加工业(有色金属),石油、煤炭及其他燃料加工业(燃料加工),化学原料及化学制品制造业(化工),电力、热力、燃气及水生产和供应业(电热气水),造纸和纸制品业(造纸),非金属矿物制品业(非金属),交通运输、仓储和邮政业(交通运输)等八大高耗能产业进行测算。

截至日前,八大行业能源消费总量数据统计值由国家统计局公布至2022年,有较大的延迟,已无法满足当前节能降碳工作形势下对能耗数据及时性的需求。模型所需自变量均为月度公布,因此,模型可对2022年至今的月度、季度的能源消费总量和碳排放进行测算,从而为行业协会以及相关部门提供更为及时的数据参考。图4显示了八大行业2018—2022年能耗测算的年均绝对误差。黑色金属冶炼及压延加工业、有色金属冶炼及压延加工业和石油、煤炭及其他燃料加工业五年平均绝对误差均小于2%,除交通运输、仓储和邮政业外其他行业误差均在5%以下,模型表现稳定,总体测算精度

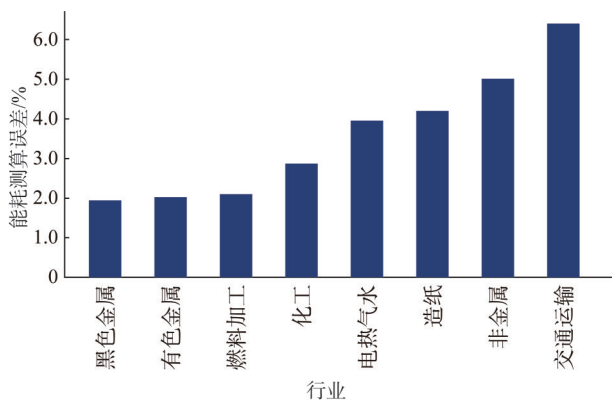


图4 全国八大高耗能行业能耗测算误差

Fig. 4 Energy consumption measurement errors in eight major energy-consuming industries across the country

较高。交通运输、仓储和邮政业近年受疫情影响较大，数据波动大，因此误差较大。

以能耗总量最高的黑色金属冶炼及压延加工业为例，图5显示其能耗总量以及由钢铁产量计算得到的能耗强度统计值(2007—2022年)及测算值(2007—2023年)。模型测算值与统计值趋势一致，误差较小。黑色金属冶炼及压延加工业能耗总量从2007年的4.78亿t上涨到2014年峰值的6.93亿t之后下降，2023年暂无统计值，其测算值为6.43亿t。2007—2023年能耗强度呈现快速下降趋势，由2007年846 kgce/t下降到2023年的测算值472 kgce/t，下降幅度达44%。

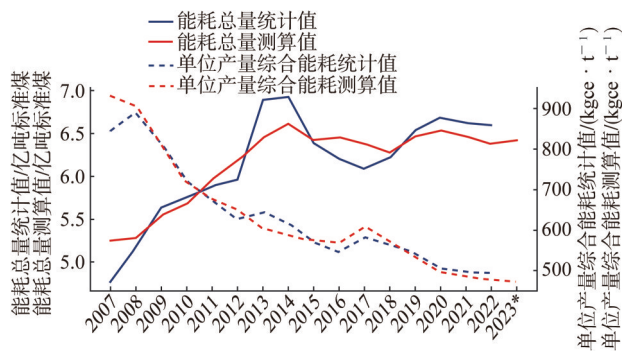


图5 黑色金属冶炼及压延加工业能耗总量及强度

Fig. 5 Energy consumptions and its intensities of the ferrous metal smelting and rolling processing industry

如图6所示，黑色金属冶炼及压延加工业碳排放总量与能耗总量呈现相同的波动趋势，其原因在于黑色金属冶炼及压延加工业行业为重点高耗能行业之一，碳排放主要来源于能源消耗，其能耗和碳排放数据之间具有极强的正相关性。从2007年的

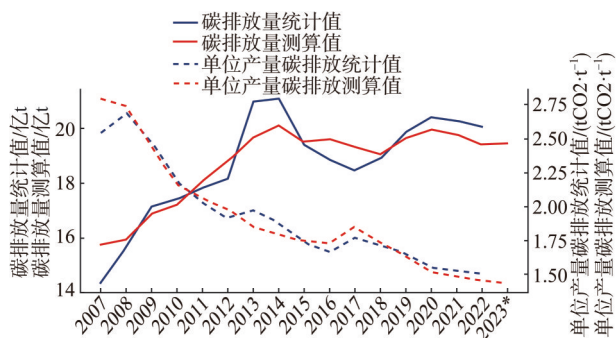


图6 黑色金属冶炼及压延加工业碳排放总量及强度

Fig. 6 Carbon emission and its intensity of the ferrous metal smelting and rolling processing industry

14.29亿t上涨到2014年峰值的21.11亿t之后下降，2023年的测算值分别为19.46亿t。2007—2023年碳排放强度呈现快速下降趋势，由2007年2.53 tCO₂/t下降到2023年的测算值1.43 tCO₂/t，下降幅度约为43%。

6 结论

本文针对全口径能耗和碳排放监测在不同空间和时间尺度对测算精度和时效性的现实需求，通过对比分析多种模型在多种拟合方法下的表现确定了岭回归赋权下的多元线性回归(R-MLR)能耗总量测算模型，基于能耗总量和能源消费结构建立了碳排放监测模型。建模过程中保持不同时间尺度下模型的一致性，从而保证了年度、月度数据源头的一致性，避免了传统方法数据频率转换过程中引入误差的问题，增强了输入信息的可信度。通过多种拟合方法的对比寻优，有效解决拟合过程中的多重共线性问题。模型具有较好的通用性，适用范围广，在区域、行业等不同空间尺度和年度、季度、月度等不同时间尺度实现逻辑统一、形式统一和算法统一，具有“以简应繁”的效果。本文的研究成果可为相关政府部门、行业协会等进行能源消费和碳排放的监测以及制定相关政策提供理论依据和数据支撑。

通过模型测算，区域能耗和碳排放总量近年来呈现上升趋势，能耗和碳排放强度呈现下降趋势，并观察到能耗和碳排放总量呈现一季度低，四季度高的前低后高季节波动性，而能耗和碳排放强度则相反，呈现一季度高，四季度低的前高后低季节波动性。行业应用算例结果表明尽管黑色金属冶炼及

压延加工业能耗和碳排放总量近年来保持高位,但
是其能耗和碳排放强度均呈现快速下降的趋势。

参考文献

- [1] 杨华磊,杨敏.碳达峰碳中和:中国式现代化的能源转型之路[J].经济问题,2024(3):1-7.
YANG Hualei, YANG Min. Peak carbon emission and carbon neutrality: china's path to energy transition in modernization[J]. On Economic Problems, 2024(3): 1-7.
- [2] 赵焱,马欣欣,唐海峰.浅析电力企业“碳达峰,碳中和”能源数据体系构建[J].自动化博览,2023,40(9):30-33.
ZHAO Yan, MA Xinxin, TANG Haifeng. Analysis on the construction of an energy big data system for “carbon peaking and carbonneutrality” for power enterprises[J]. Automation Panorama, 2023, 40(9): 30-33.
- [3] 王坤,李永新,刘伟,等.我国碳达峰碳中和图景与典型国家能源转型过程的对比分析与经验启示[J].能源技术与管理,2023,48(4):1-6.
WANG Kun, LI Yongxin, LIU Wei, et al. Comparison and experience insights on China's carbon peak and carbon neutrality outlook and typical national energy transition processes[J]. Energy Technology and Management, 2023, 48(4): 1-6.
- [4] 李茂林,李宇宁,宋雯馨.基于人口与经济组合模型的区域能源消费预测[J].中国市场,2024(3):12-15.
- [5] 钟化鑫.基于清洁能源发展的中国一次能源消费预测研究[D].北京:中国石油大学,2022.
- [6] 王安,杨雨,杨锦伟.GPM(1,1,m)模型及其在中国能源消费预测中的应用[J].数学的实践与认识,2021,51(2):104-112.
WANG An, YANG Yu, YANG Jinwei. GPM(1,1,m) model and its application in forecasting energy consumption of China[J]. Mathematics in Practice and Theory, 2021, 51(2): 104-112.
- [7] 刘子华,曹瑞峰,赵志扬,等.基于碳排放流的综合能源系统碳排放监测方法[J].浙江电力,2023,42(10):65-72.
LIU Zihua, CAO Ruifeng, ZHAO Zhiyang, et al. A carbon emission monitoring method for integrated energy systems based on carbonemission flow[J]. Zhejiang Electric Power, 2023, 42(10): 65-72.
- [8] 叶鑫芳,钟志鹏,郑仁广,等.基于碳电强度的碳排放监测方法[J].能源与环境,2023(1):40-44.
- [9] 韩少秀,朱美峰,刘慧洁,等.碳排放监测预警体系研究——基于中国省级相关数据的分析[J].生态经济,2022,38(7):21-27.
HAN Shaoxiu, ZHU Meifeng, LIU Huijie, et al. Research on carbon emissions monitoring and early warning system: analysisbased on the provincial panel data of China[J]. Ecological Economy, 2022, 38(7): 21-27.
- [10] 纪艳菊,张玉,李翠.可再生能源预测技术在电力系统中的实时应用[J].现代工业经济和信息化,2023,13(12):143-145,149.
JI Yanju, ZHANG Yu, LI Cui. Real-time application of renewable energy forecasting in power systems[J]. Modern Industrial Economy and Informationization, 2023, 13(12): 143-145,149.
- [11] 刘显苗,邓韦斯,田伟达,等.基于AI算法的可再生能源预测系统研究[J].能源与环保,2023,45(8):190-194.
LIU Xianzhuo, DENG Weisi, TIAN Weida, et al. Research on renewable energy forecasting system based on AI algorithm[J]. China Energy and Environmental Protection, 2023, 45(8): 190-194.
- [12] 刘连义.基于新型累加灰色模型的可再生能源消耗量预测研究[D].邯郸:河北工程大学,2021.
- [13] 邵帅,崔兴华.能源供给侧与消费侧碳排放的责任核算与驱动因素——基于“收入者责任”视角的考察[J].上海交通大学学报(哲学社会科学版),2024,32(1):81-100.
SHAO Shuai, CUI Xinghua. Responsibility accounting and driving factors for carbon emissions on energy-supply and energy-consumption sides: an examination from the perspective of “income responsibility”[J]. Journal of Shanghai Jiaotong University (Philosophy and Social Sciences), 2024, 32(1): 81-100.
- [14] 王铁楠,卢静,陈星彤,等.考虑多重因素叠加情景的能源清洁低碳转型风险评估研究[J].中国电力,2024,57(3):183-189.
WANG Yinan, LU Jing, CHEN Xingtong, et al. Risk assessment of clean and low-carbon energy transformation considering superimposed scenario of multiple factors[J]. Electric Power, 2024, 57(3): 183-189.
- [15] 章政文,顾津锦,赵俊华,等.基于碳卫星与电力排放数据的碳计量[J].电力系统自动化,2024,48(1):2-9.
ZHANG Zhengwen, GU Jinjin, ZHAO Junhua, et al. Carbon measurement based on carbon satellite and electricity emission data[J]. Automation of Electric Power Systems, 2024, 48(1): 2-9.
- [16] 王震山.多源卫星遥感中国大气CO₂浓度及人为碳排放时空变化特征研究[D].南京:南京信息工程大学,2023.
- [17] 付金杯,李梦南,徐炜达,等.基于卫星遥感等多源数据分析发电厂CO₂排放的研究[J].中国环境科学,2024,44(4):1805-1815.
FU Jinbei, LI Mengnan, XU Weida, et al. CO₂ emissions from power plants based on multi-source data such as satellite remote sensing[J]. China Environmental, 2024, 44(4): 1805-1815.
- [18] 季彦黎.石油化工企业碳排放在线监测与管理技术研究[J].石油化工技术与经济,2023,39(4):22-25.
JI Yanjun. Research on online monitoring and management technology of carbon emission in petrochemical enterprises[J]. Techno-Economics in Petrochemicals, 2023, 39(4): 22-25.
- [19] 宋航.固定源碳排放数据实时在线监测与处理系统设计[D].南昌:南昌大学,2023.
- [20] 陈政,何耿生,尚楠.面向碳达峰碳中和的电网碳排放因子改进计算方法[J].南方电网技术,2024,18(1):153-162.
CHEN Zheng, HE Gengsheng, SHANG Nan. Improved calculation method of power grid carbon emission factor for carbon peak and carbon neutrality goals[J]. Southern Power System Technology, 2024, 18(1): 153-162.
- [21] 马大卫,苏阳,王正风,等.安徽省地级市发电碳排放因子特征与聚类研究[J].电力科技与环保,2023,39(6):524-532.
MA Dawei, SU Yang, WANG Zhengfeng, et al. Characteristics and clustering research of carbon emission factors of power generation for prefecture-level cities of Anhui Province[J]. Electric Power Environmental Protection, 2023, 39(6): 524-532.
- [22] 管梦姝.基于LCA-LEAP的钢铁行业能源—碳排放模型研究与应用[D].西安:西安电子科技大学,2021.
- [23] 曾金灿,王成围,杨晨,等.基于广义节点碳流理论的区域电

- 网用电碳排放计算方法[J]. 广东电力, 2023, 36(11): 20 – 28.
- ZENG Jincan, WANG Chengwei, YANG Chen, et al. Measurement method of area electricity carbon emission based on generalized nodal carbon emission flow [J]. Guangdong Electric Power, 2023, 36(11): 20 – 28.
- [24] KANG C, ZHOU T, CHEN Q, et al. Carbon emission flow from generation to demand: a network-based model[J]. IEEE Transactions on Smart Grid, 2015, 6(5): 2386 – 2394.
- [25] BAI L, JIANG L, YANG D, et al. Quantifying the spatial heterogeneity influences of natural and socioeconomic factors and their interactions on air pollution using the geographical detector method: A case study of the Yangtze River Economic Belt, China[J]. Journal of Cleaner Production, 2019(232): 692 – 704.
- [26] 胡婉玲, 张金鑫, 王红玲. 我国碳排放数据质量管理现状及提升路径研究[J]. 社会科学动态, 2024(1): 18 – 23.
- HU Wanling, ZHANG Jinxin, WANG Hongling. Study on the present situation of carbon emission data quality management in China and its improving path[J]. Dynamics of Social Sciences, 2024(1): 18 – 23.
- [27] 邢宾. 碳市场数据质量控制方法概述[J]. 化工设计通讯, 2023, 49(9): 179 – 180, 193.
- XING Bin. Overview of carbon market data quality control methods [J]. Chemical Engineering Design Communications, 2023, 49(9): 179 – 180, 193.
- [28] 鞠月闯. 基于时空 LightGBM 模型的中国碳排放预测[D]. 南京: 南京邮电大学, 2023.
- [29] 丁长胜. 基于 Modbus 的能源监测系统设计与实现[D]. 天津: 天津科技大学, 2020.
- [30] 肖英, 何杰兰, 汪权方, 等. 中国能源消费差异化碳减排潜能分析[J]. 湖北大学学报(自然科学版), 2024, 46(5): 701 – 708.
- XIAO Ying, HE Jielan, WANG Quanfang, et al. Analysis of differentiated carbon emission reduction potential of energy consumption in China[J]. Journal of Hubei University (Natural Science), 2024, 46(5): 701 – 708.
- [31] 崔佳璇, 延肖何, 刘念. 计及电-碳耦合与协同自治的配电网储能优化配置策略[J]. 南方电网技术, 2025, 19(11): 110 – 121.
- CUI Jiaxuan, YAN Xiaohe, LIU Nian. Optimal allocation strategy of energy storage in distribution networks taking into account electricity-carbon coupling and cooperative autonomy [J]. Southern Power System Technology, 2025, 19(11): 110 – 121.
- [32] 谢平平, 陆秋瑜, 王雪林, 等. 适应新型电力系统发展风光储联合发电站碳排放核算方法研究[J]. 广东电力, 2025, 38(6): 1 – 10.
- XIE Pingping, LU Qiuyu, WANG Xuelin, et al. Research on carbon emission accounting method for wind-solar-storage combined power plant adapting to the development of new power system [J]. Guangdong Electric Power, 2025, 38(6): 1 – 10.
- [33] 崔勇, 韩一春, 郑谦, 等. 多能联盟低碳运营决策方法研究框架与展望[J]. 电测与仪表, 2024, 61(3): 10 – 19.
- CUI Yong, HAN Yichun, ZHENG Qian, et al. Research framework and prospect of multi-energy alliance low-carbon operation decision-making method [J]. Electrical Measurement & Instrumentation, 2024, 61(3): 10 – 19.
- [34] 纪晓斐. 季度 GDP 核算及其与年度 GDP 衔接方法对比研究[D]. 郑州: 河南大学, 2019.

收稿日期: 2024-09-02; 网络首发日期: 2025-01-29

作者简介:

王成国(1987), 男, 工程师, 博士, 研究方向为电碳耦合技术、电力系统稳定性分析, wang_cheng_wei_@126.com;

李沛(1993), 男, 工程师, 硕士, 研究方向为能源战略、电碳耦合技术;

杨至元(1993), 男, 工程师, 博士, 研究方向为电力系统分析、电碳耦合技术及人工智能应用。