



中国研究生创新实践系列大赛
“华为杯”第二十届中国研究生
数学建模竞赛

学 校 北京交通大学

参赛队号 23100040017

1.张雯瑗

队员姓名 2.吕欣桐

3.孔文康

中国研究生创新实践系列大赛

“华为杯”第二十届中国研究生

数学建模竞赛

题 目： 区域双碳目标与路径规划研究

摘 要：

中国碳排放问题一直是一个备受关注的议题，我国的双碳目标为力争在 2030 年前实现碳达峰，努力在 2060 年前实现碳中和。本研究通过多维度的分析和建模，利用 LMDI 模型，STIRPAT 模型和情景分析法，深入研究了区域碳排放的动因、趋势，建立了区域碳排放量预测模型，为如何实现区域双碳目标提供了路径规划。

对于问题一，我们首先建立了区域碳排放量的指标与指标体系，其中经济影响因素的指标为人均 GDP，人口影响因素的指标为人口总量，能源消费量影响因素的指标包括非化石能源消费比重、非化石能源发电比重、单位 GDP 能耗、电力消费比重、人均能耗、单位能耗二氧化碳排放量。由于居民消费部门不存在 GDP 产出，居民消费部门与其余五个部门指标体系有所差异。之后对指标数据进行了数据分析，计算出指标环比和同比值，分析得到该区域十二五和十三五期间的碳排放状况。为了得各指标对区域碳排放量产生影响的贡献，我们将 **Kaya 模型** 进行扩展，利用 **LMDI 模型** 建立了各指标与碳排放量之间的关联关系模型，衡量出各部门指标对碳排放量的贡献，比较指标贡献之后得到人均 GDP、单位能耗二氧化碳排放量和非化石能源消费比重是实现双碳目标的主要挑战。之后通过计算各指标之间的皮尔逊相关系数，对于指标进行了筛选优化。

对于问题二，其主要目标是构建区域碳排放量以及经济、人口、能源消费量的预测模型。结合国家相关要求，我们可以得到各指标在中国式现代化的两个时间节点的增速，由此预测出 2021-2060 年的人口总量、GDP 和能源消费量。利用多元线性回归模型，拟合数据后得到了能源消费量关于人口和 GDP 的预测模型。为了建立区域碳排放量预测模型，引入了能够衡量区域生产力的新指标第三产业 GDP 占比。之后利用对 **IPAT 模型** 进行改进的 **STIRPAT 模型** 来对区域碳排放量进行预测。我们根据具体选用的指标，将 **STIRPAT 模型** 进行扩展，得到碳排放量预测模型，再通过多种回归方法，如 Ridge 回归、LASSO 回归、多元线性回归和弹性网络回归方法对模型进行拟合计算，经过结果对比分析发现 Ridge 回归拟合得到的结果更符合实际情况，因此得到碳排放量预测模型时最终采用 **Ridge 回归**。

对于问题三，主要目标是进行区域双碳目标与路径规划。首先进行了**情景设计**，设计了无人干预的自然情景、按时碳达峰与碳中和的基准情景和率先碳达峰与碳中和的雄心情景。之后针对不同情景，设置得到不同情境参数设定表，参数即为指标在一定时间内的

复合增长率，由此可以得到未来 2021-2060 年的指标预测值。因为单位能耗二氧化碳无法通过情景设定表进行复合增长率设定，因此采用 **STIRPAT 模型** 对其进行预测得到未来的值。之后我们结合问题二所计算得到的碳排放量预测模型，得到了不同情景下的碳排放量预测模型。分析不同情景下的 2021-2060 碳排放量可以得知，自然情景因为无人约束，碳排放量持续增长，无法实现双碳。基准情景在 2030 年按时实现碳达峰，但在 2060 年之前无法实现碳中和。雄心情景在 2025 年提前实现碳达峰，在 2049 年提前实现碳中和。

关键词：双碳目标；碳排放；LMDI 模型；STIRPAT 模型；情景分析法

目 录

一、 问题重述	4
1.1 问题背景.....	4
1.2 问题提出.....	4
二、 问题分析	5
2.1 问题一的分析.....	5
2.2 问题二的分析.....	6
2.3 问题三的分析.....	7
三、 模型假设	8
四、 符号说明	9
五、 模型的建立与求解.....	10
5.1 问题一模型的建立与求解.....	10
5.1.1 数据来源及预处理.....	10
5.1.2 模型建立.....	10
5.1.3 模型求解.....	13
5.1.4 问题回答.....	14
5.2 问题二模型的建立与求解.....	23
5.2.1 数据来源.....	23
5.2.2 模型的建立.....	24
5.2.3 模型的求解.....	27
5.3 问题三模型的建立与求解.....	30
5.3.1 数据来源.....	30
5.3.2 模型的建立.....	30
5.3.3 模型的求解.....	31
六、 模型的评价、改进与推广.....	33
6.1 模型的优点.....	33
6.2 模型的缺点.....	33
6.3 模型的改进.....	34
6.4 模型的推广.....	34
七、 参考文献	34

一、问题重述

1.1 问题背景

2020 年 9 月 22 日，习近平总书记在第七十五届联合国大会上首次提出我国的双碳目标，即力争在 2030 年前实现碳达峰，努力在 2060 年前实现碳中和。此后，习近平总书记在国内外会议中多次强调我国实现双碳目标的决心和动力，相关政府机关也出台了一系列规划、纲要、意见及建议，确保双碳政策的贯彻落实，这充分体现了我国对于全球环境保护的责任与担当^[1]。

为达到双碳目标，关键任务在于解决发展与碳减排之间的矛盾。碳减排是指在一定时期内与正常商业活动相比某个国家或地区减少的二氧化碳排放量，主要指标包括碳排放强度或碳生产力^[2]。碳减排的相关措施可以从两个方面着手：首先，提高能源利用效率，即开展能效工程和产业升级工程；其次，增加非化石能源消耗的比例，即开展能源脱碳工程和能源消费电气化工程。能效工程的核心在于节能减排，即通过资源的合理配置和有效利用，降低能源消耗和减少污染排放，以促进经济社会的可持续发展^[2]。产业升级工程的核心在于科技创新，即研发低碳零碳等前沿技术，推动颠覆性低碳技术的发展^[3]。能源脱碳工程的核心在于能源系统转型，即通过新能源发电、火力脱碳、新型电网等多种措施，优化能源结构。能源消费电气化工程的核心在于电能替代化石能源，以提高非化石能源发电占比。

因此，应通过分析、评价、预测等手段充分考虑四个重点工程(能效工程、产业升级工程、能源脱碳工程和能源消费电气化工程)对碳排放的影响，进而确定双碳目标和路径。

1.2 问题提出

为了科学规划双碳目标路径，需要完成以下问题：

问题一：区域碳排放量以及经济、人口、能源消费量的现状分析

(1) 建立指标与指标体系，并满足下列要求：①指标应能描述该区域经济、人口、能源消费量与碳排放量的关系；②指标应能描述能源供应部门、工业消费部门、建筑消费部门、交通消费部门、居民生活消费和农林消费部门的碳排放情况；③指标体系应能描述各主要指标之间的相互关系；④部分指标的变化(同比或环比)可以为碳排放量的预测提供依据。

(2) 分析该区域碳排放量以及经济、人口、能源消费量的现状，并满足下列要求：①将 2010 年作为基期，分析该区域在十二五(2011-2015 年)和十三五(2016-2020 年)期间的碳排放总量、碳排放变化趋势等情况；②分析对该区域碳排放量的影响因素及其贡献。③研判该区域实现双碳目标需要应对的主要挑战，进而为双碳目标中差异化的路径选择提供依据。

(3) 建立该区域碳排放量与经济、人口、能源消费量等指标的关联关系模型，并满足下列要求：①利用环比、同比分析相关指标的变化；②建立各指标间的关联关系模型；③基于关联关系模型指标的变化，并结合双碳政策与技术进步等多重效应，确定碳排放预测模型参数取值。

问题二：区域碳排放量以及经济、人口、能源消费量的预测模型

(1) 基于人口和经济变化的能源消费量预测模型，并满足下列要求：①将 2020 年作为基期，预测该区域十四五(2021-2025 年)至二十一五(2056-2060 年)期间人口、经济(GDP)和能源消费量的变化；②能源消费量应与人口预测相关；③能源消费量应与经济(GDP)预

测相关。

(2) 建立区域碳排放量预测模型，并满足下列要求：①碳排放量应与人口、GDP 和能源消费量预测相关联；②碳排放量应与各能源消费部门(工业消费部门、建筑消费部门、交通消费部门、居民生活消费、农林消费部门)以及能源供应部门的能源消费量相关联；③碳排放量应与各消费部门的能源消费品种(一次能源中化石能源消费与非化石能源消费以及二次能源中电或热消费)以及能源供应部门的能源消费品种(化石能源发电与非化石能源发电)相关。

问题三：区域双碳(碳达峰与碳中和)目标与路径规划方法

(1) 设计不少于三种情景，如无人干预的自然情景、按时碳达峰与碳中和的基准情景、率先碳达峰与碳中和的雄心情景等，并满足下列要求：①设计的情景应与碳达峰、碳中和的时间节点相关；②设计的情景应与能效提升和非化石能源消费比重提升相关。

(2) 确定多场景下碳排放量核算方法，并满足下列要求：①该区域的碳排放与多情景假设一致；②该区域的碳排放与各部门碳排放量的总和一致；③碳排放量核算模型与问题二中预测模型一致。

(3) 确定双碳(碳达峰和碳中和)目标和路径，并满足下列要求：①根据情景设置确定 GDP、人口和能源消费量在目标时间节点(2025 年、2030 年、2035 年、2050 年和 2060 年)的取值；②根据情景设置确定在目标时间节点提高能源利用效率和提高非化石能源消费比重的值；③完成能效提升、产业(产品)升级、能源脱碳和能源消费电气化的定性与定量分析。

二、问题分析

2.1 问题一的分析

问题一需要对区域碳排放量以及经济、人口、能源消费量的现状进行分析，思路如图 2.1 所示。

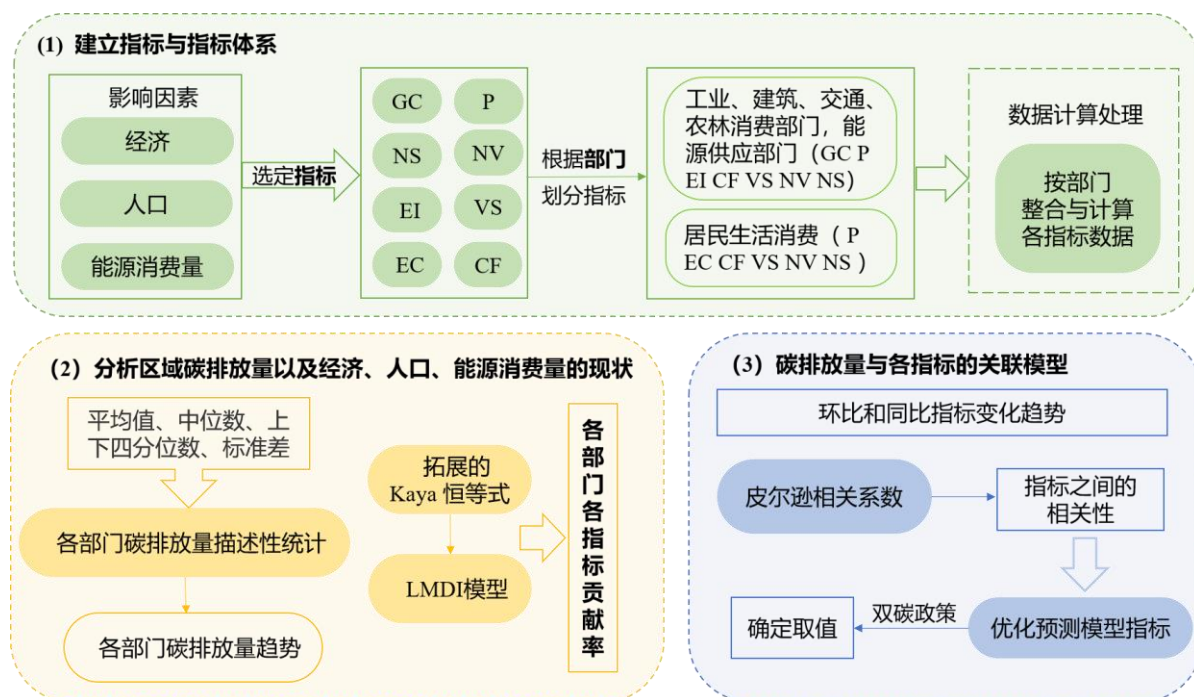


图 2.1 问题一思路流程图

在第(1)问中, 需要根据相关要求建立区域碳排放量的指标与指标体系。对于要求 1, 综合考虑经济、人口和能源消费量这三大类影响因素, 确定各类下面合适的指标。经济影响因素的指标为人均 GDP(GC), 人口影响因素的指标为人口总量(P), 能源消费量影响因素的指标包括非化石能源消费比重(NS)、非化石能源发电比重(NV)、单位 GDP 能耗(EI)、电力消费比重 (VS)、人均能耗(EC)、单位能耗二氧化碳排放量(CF)。对于要求 2, 根据各部门的特点对指标进行划分, 体现各部门指标的差异性, 从而更准确地描述各部门碳排放情况。工业消费部门、建筑消费部门、交通消费部门、农林消费部门和能源供应部门的指标为 GC 、 P 、 NS 、 NV 、 EI 、 VS 、 CF 。由于居民生活消费没有 GDP 产出, 该部门的指标为 P 、 NS 、 NV 、 EI 、 VS 、 EC 、 CF 。对于要求 3, 构建逐步细化的指标体系, 从而反映出各指标之间的相互关系。对于要求 4, 分析各指标的变化, 为后续的碳排放量预测奠定基础。

在第(2)问中, 需要分析区域碳排放量以及经济、人口、能源消费量的现状。对于要求 1, 将各部门碳排放量的变化趋势可视化, 并结合平均值、中位数、上下四分位数、标准差这 5 个描述性统计量, 综合分析该区域十二五和十三五期间的碳排放状况。对于要求 2, Kaya 模型通常被应用于研究区域内碳排放量与该区域人口、社会经济发展水平、能源利用效率以及碳排放因子之间的相互关系, 在其基础上发展出的对数平均迪氏指数法 (Logarithmic Mean Divisia Index, LMDI) 具有消除残差项、分解结果直观的特点^[4]。因此, 我们选用 LMDI 方法建立各指标与碳排放量之间的关联关系模型, 衡量各部门指标对碳排放量的贡献率并将结果可视化。对于要求 3, 根据贡献率确定对部门碳排放量影响较大的指标, 将其视作实现双碳目标需要面对的主要挑战, 为后续该区域双碳路径规划中差异化的路径选择提供依据。

在第(3)问中, 需要建立区域碳排放量以及经济、人口、能源消费量各指标及其关联模型。对于要求 1, 利用环比分析指标的年变化趋势, 利用同比分析指标在十二五期间和十三五期间的变化。对于要求 2, 由于各指标是具有线性关系的定距变量, 故选用皮尔逊相关系数衡量指标之间的相关性并将结果可视化。对于要求 3, 相关性较大的指标之间存在冗余, 应对指标进一步筛选, 确定优化后的模型指标, 确定这些指标在碳达峰、碳中和、中国式现代化关键时间点的取值。

2.2 问题二的分析

问题二需要构建区域碳排放量以及经济、人口、能源消费量的预测模型, 思路流程如图 2.2 所示。

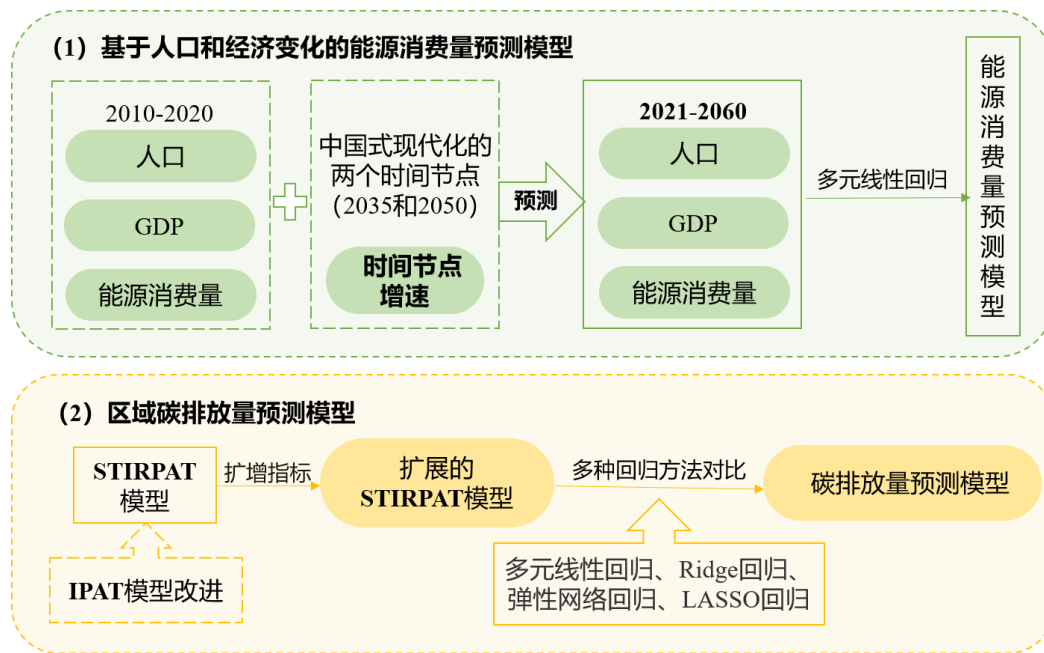


图 2.2 问题二思路流程图

在第(1)问中，需要建立人口和经济变化的能源消费量预测模型。对于要求 1，根据 2010-2020 年数据，结合中国式现代化的两个时间节点(2035 年和 2050 年)可以得到这两个关键时间节点各指标的增速，以此可以预测出 2021-2060 年的人口总量、GDP 和能源消费量。对于要求 2 和要求 3，建立能源消费量与人口总量、GDP 的多元线性回归模型，并根据预测数据进行拟合。

在第(2)问中，需要建立区域碳排放量预测模型。对于要求 1，根据问题一第(3)问分析结论，确定初步的区域碳排放量预测指标。对于要求 2，为考虑各部门能源消费量的影响，引入第三产业占比指标。对于要求 3，确保指标能体现各部门能源消费品种的影响。根据上述 3 个要求，确定最终的区域碳排放量预测指标。可拓展的随机性的环境影响评估模型 (Stochastic Impacts by Regression on Population, Affluence and Technology, STIRPAT)是对 IPAT 模型的改进，可以定量分析各因素对环境负荷的影响，该方法已被广泛地应用于环境保护研究中^[6]。因此，我们选用 STIRPAT 模型来对区域碳排放量进行预测。根据具体选用指标，将 STIRPAT 模型进行扩展，得到碳排放量预测模型，再通过多种回归方法对比计算，得到预测模型中各个参数的值。

2.3 问题三的分析

问题三需要得到区域双碳(碳达峰与碳中和)目标与路径规划方法，思路如图 2.3 所示。

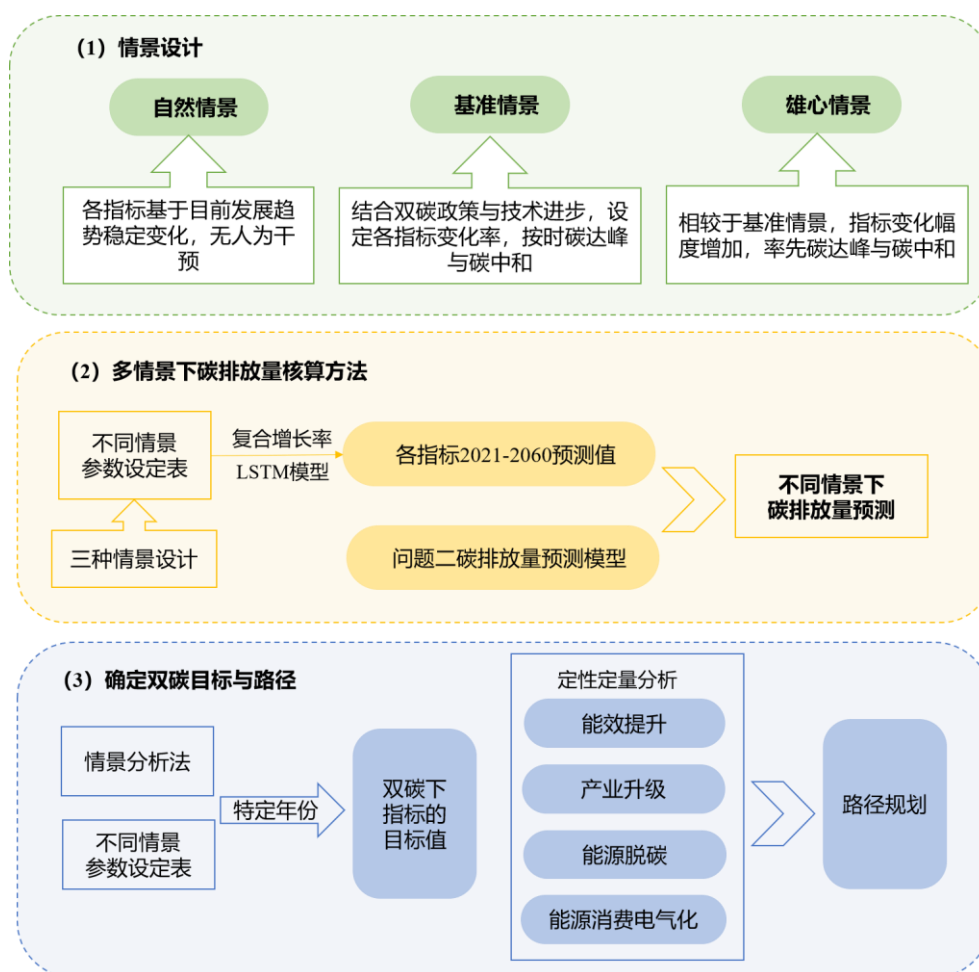


图 2.3 问题三思路流程图

在第(1)问中, 需要进行情景设计。对于要求 1 和要求 2, 考虑情景的设计时, 要与碳达峰、碳中和的时间节点以及能效提升和非化石能源消费比重提升相关联。因此, 我们设计了三种情景: 无人为干预的自然情景、按时碳达峰与碳中和的基准情景和率先碳达峰与碳中和的雄心情景。

在第(2)问中, 需要获得多情景下碳排放量核算方法。对于要求 1, 根据第(1)问的三种情景假设, 结合国家碳排放规划, 可以得到不同情景下各指标变化的参数设定表, 再通过复合增长率计算和 LSTM 模型得到不同情景下各指标 2021 年-2060 年的预测值。对于要求 2 和要求 3, 碳排放量核算模型选用问题二中的区域碳排放预测模型, 带入各指标的预测值可以得到不同情景下 2021 年-2060 年碳排放量的预测值。

在第(3)问中, 需要确定双碳(碳达峰与碳中和)目标与路径。对于要求 1 和要求 2, 结合第(2)问所计算出的各指标在 2021 年-2060 年的预测值, 可以确定基准情景和雄心情景下 GDP、人口、能源消费量、能源利用效率和非化石能源消费比重在 2025 年、2030 年、2035 年、2050 年和 2060 年的目标值。对于要求 3, 结合 2021 年到 2060 年碳排放量预测值, 完成能效提升、产业升级、能源脱碳和能源消费电气化的定性与定量分析。

三、模型假设

1. 根据题目描述其他能源的碳排放因子为 0, 假设其他能源归为非化石能源。
2. 假设原始数据中的离群点为异常值, 需要自行处理。

- 假设经济、人口、能源消费量以外的因素对碳排放量影响有限。
- 假设多场景条件下 2035 年的 GDP 比基期(2020 年)翻一番，即为 2020 年的两倍；2060 年比基期翻两番，即为 2020 年的四倍。
- 假设题目所给数据中的碳排放量为净排放量，当碳排放量不再增加时达到碳达峰，碳排放量为 0 时达到碳中和。

四、符号说明

符号	说明	单位
$Q(i)$	部门 i 的碳排放量, $i \in \{1, 2, 3, 4, 5, 6\}$	万 tCO ²
$G(i)$	部门 i 的 GDP, $i \neq 6$	万元
$P(i)$	部门 i 的人口总量, $i \in \{1, 2, 3, 4, 5, 6\}$	万人
$E(i)$	部门 i 的能源消费总量, $i \in \{1, 2, 3, 4, 5, 6\}$	万 tce
$E_N(i)$	部门 i 的非化石能源消费量, $i \in \{1, 2, 3, 4, 5, 6\}$	万 tce
$E_F(i)$	部门 i 的化石能源消费量, $i \in \{1, 2, 3, 4, 5, 6\}$	万 tce
$E_V(i)$	部门 i 的发电量, $i \in \{1, 2, 3, 4, 5, 6\}$	万 tce
$GC(i)$	部门 i 的人均 GDP, $i \neq 6$	元
$EI(i)$	部门 i 的单位 GDP 能耗, $i \neq 6$	万 tce/亿元
$VS(i)$	部门 i 的电力消费比重, $i \in \{1, 2, 3, 4, 5, 6\}$	-
$NV(i)$	部门 i 的非化石能源发电占比, $i \in \{1, 2, 3, 4, 5, 6\}$	-
$NS(i)$	部门 i 的非化石能源消费比重, $i \in \{1, 2, 3, 4, 5, 6\}$	-
$CF(i)$	部门 i 的单位能耗二氧化碳排放量, $i \in \{1, 2, 3, 4, 5, 6\}$	tCO ² /tce
$EC(i)$	部门 i 的人均能耗, $i \in \{1, 2, 3, 4, 5, 6\}$	tce/人
t	年份	-
TI	第三产业 GDP 占比	-
r_P	人口总量增长率	-
r_G	GDP 增长率	-
r_E	人口总量增长率	-
r_F	化石能源消费量增长率	-
r_V	发电量增长率	-
r_{TI}	第三产业 GDP 占比增长率	-

注:① i 是分类变量,用于区分不同部门。 $i=1$ 时对应能源供应部门, $i=2$ 时对应工业消费部门, $i=3$ 时对应建筑消费部门, $i=4$ 时对应交通消费部门, $i=5$ 时对应农林消费部门, $i=6$ 时对应居民生活消费。

②对于上表变量,不标注 i 时表示所有部门的数据。例如, $Q(i)$ 表示部门 i 的碳排放量, Q 表示区域总碳排放量。

③对于上表变量,上角标 t 表示第 t 年的取值,上角标0表示基期的取值。例如, Q^t 表示第 t 年的碳排放量, Q^0 表示基期的碳排放量

④对于上表变量,符号 Δ 表示变化量。例如, ΔQ 表示碳排放量变化量。

⑤“-”表示该变量无单位。

五、模型的建立与求解

5.1 问题一模型的建立与求解

5.1.1 数据来源及预处理

(1)**数据来源**: 按能源供应部门、工业消费部门、建筑消费部门、交通消费部门、居民生活消费和农林消费部门对题目所给的原始数据进行整合, 并根据假设 1 将其他能源相关数据归入非化石能源数据。

(2)**数据预处理**: 分析整合后的数据, 绘制箱线图, 发现能源供应部门其他转换项目的煤炭碳排放因子存在 2 个离群点, 天然气碳排放因子存在 1 个离群点, 如图 5.1 所示。根据假设 2, 将这 3 个离群点作为异常值, 利用均值来进行替代。



图 5.1 能源供应部门其他项目的煤炭碳排放因子和天然气碳排放因子箱线图

5.1.2 模型建立

● 建立各部门碳排放量与经济、人口、能源消费量各指标的关联关系模型: LMDI 模型

① 建立指标与指标体系

对于居民生活消费以外的部门, 选择人均 GDP 作为经济描述指标, 选择人口总量作为人口描述指标。由于单位 GDP 能耗是表示能效的指标, 单位能耗二氧化碳排放量(碳排放因子)是表示低碳排放的指标, 非化石能源消费比重和非化石能源发电占比是表示能源脱碳的指标, 电力消费比重是表示能源消费电气化的指标, 这 5 个指标均对碳排放量具有影响, 故将其选作能源消费量描述指标。因此, 将人均 GDP、人口总量、单位 GDP 能耗、单位能耗二氧化碳排放量、非化石能源消费比重、非化石能源发电占比、电力消费比重这 7 个指标作为居民生活消费以外的部门碳排放量的影响指标, 各指标计算方式如下:

$$GC(i) = G(i) / P(i) \quad (1)$$

$$EI(i) = E(i) / G(i) \quad (2)$$

$$VS(i) = E_V(i)/E(i) \quad (3)$$

$$NV(i) = E_N(i)/E_V(i) \quad (4)$$

$$NS(i) = E(i)/E_N(i) \quad (5)$$

$$CF(i) = E(i)/E_N(i) \quad (6)$$

其中, i 表示部门, 当取值不为 6 时表示居民生活消费以外的部门, $G(i)$ 为部门 i 的 GDP, $P(i)$ 为部门 i 的人口总量, $E(i)$ 为部门 i 的能源消费总量, $E_N(i)$ 为部门 i 的非化石能源消费量, $E_V(i)$ 为部门 i 的发电量, $GC(i)$ 为部门 i 的人均 GDP, $EI(i)$ 为部门 i 的单位 GDP 能耗, $VS(i)$ 为部门 i 的电力消费比重, $NV(i)$ 为部门 i 的非化石能源发电占比, $NS(i)$ 为部门 i 的非化石能源消费比重, $CF(i)$ 为部门 i 的单位能耗二氧化碳排放量。

由于居民生活消费只耗能而不产出 GDP, 相较其他部门不考虑与 GDP 相关的指标的影响, 即人均 GDP 和单位 GDP 能耗。因此, 保留人口总量、单位能耗二氧化碳排放量、非化石能源消费比重、非化石能源发电占比指标、电力消费比重, 并增加人均能耗指标, 将这 6 个指标作为居民生活消费碳排放量的影响指标, 计算方式如下:

$$EC(i) = E(i)/P(i) \quad (7)$$

其中, i 表示部门, 取值为 6 时表示居民生活消费, $EC(i)$ 为部门 i 的电力消费比重, $E(i)$ 为部门 i 的能源消费总量, $P(i)$ 为部门 i 的人口总量。

②建立各部门碳排放量的 LMDI 模型

LMDI 模型的基本思路^[5]如图 5.2 所示, 首先, 将目标变量分解为不同指标的组合并构建 Kaya 恒等式; 接着, 将 LMDI 的指数进行分解; 然后, 计算目标变量的变化; 最后, 计算出各指标对目标变量的贡献。

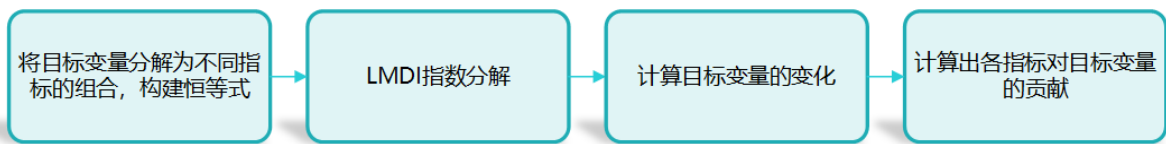


图 5.2 LMDI 模型的基本思路图

(a)构建扩展的 Kaya 恒等式

居民生活消费以外的部门 $i(i \neq 6)$ 具有 $GC(i)$ 、 $P(i)$ 、 $NS(i)$ 、 $NV(i)$ 、 $EI(i)$ 、 $VS(i)$ 、 $CF(i)$ 这 7 个指标, 构建的 Kaya 恒等式如下:

$$Q(i) = P(i) \times \frac{G(i)}{P(i)} \times \frac{E(i)}{G(i)} \times \frac{E_V(i)}{E(i)} \times \frac{E_N(i)}{E_V(i)} \times \frac{E(i)}{E_N(i)} \times \frac{Q(i)}{E(i)} \quad (8)$$

其中, i 表示部门, 对于居民生活消费以外的部门 $i \neq 6$, $Q(i)$ 为部门 i 的碳排放量, $G(i)$ 为部门 i 的 GDP, $P(i)$ 为部门 i 的人口总量, $E(i)$ 为部门 i 的能源消费总量, $E_N(i)$ 为部门 i 的非化石能源消费量, $E_V(i)$ 为部门 i 的发电量。 $G(i)/P(i)$ 为部门 i 的人均 GDP 指标 $GC(i)$,

$E(i)/G(i)$ 为部门 i 的单位 GDP 能耗指标 $EI(i)$, $E_V(i)/E(i)$ 为部门 i 的电力消费比重 $VS(i)$, $E_N(i)/E_V(i)$ 为部门 i 的非化石能源发电占比 $NV(i)$, $E(i)/E_N(i)$ 为非化石能源消费比重 $NS(i)$ 的倒数 $NS'(i)$, $C(i)/E(i)$ 为单位能耗二氧化碳排放量 $CF(i)$ 。因此, 公式(6)可以进一步表示为:

$$Q(i) = P(i) \times GC(i) \times EI(i) \times VS(i) \times NV(i) \times NS'(i) \times CF(i) \quad (9)$$

居民消费部门 $i(i=6)$ 具有 $P(i)$ 、 $EC(i)$ 、 $NS(i)$ 、 $NV(i)$ 、 $VS(i)$ 、 $CF(i)$ 这 6 个指标, 同理, 构建 Kaya 恒等式, 公式如下:

$$Q(i) = P(i) \times \frac{E(i)}{P(i)} \times \frac{E_V(i)}{E(i)} \times \frac{E_N(i)}{E_V(i)} \times \frac{E(i)}{E_N(i)} \times \frac{Q(i)}{E(i)} \quad (10)$$

由于 $E(i)/P(i)$ 为部门 i 的人均能耗 $EC(i)$, 该公式可以进一步表示为:

$$Q(i) = P(i) \times EC(i) \times VS(i) \times NV(i) \times NS'(i) \times CF(i) \quad (11)$$

(b)LMDI 指数分解

根据公式(9), 对居民生活消费以外($i \neq 6$)的部门碳排放量 Kaya 恒等式进行分解, 公式如下:

$$\begin{aligned} \Delta Q(i) = Q^t(i) - Q^0(i) = & \Delta Q_P(i) + \Delta Q_{GC}(i) + \Delta Q_{EI}(i) + \Delta Q_{VS}(i) + \Delta Q_{NV}(i) \\ & + \Delta Q_{NS'}(i) + \Delta Q_{CF}(i) \end{aligned} \quad (12)$$

其中, $\Delta Q(i)$ 为第 t 年与基期(2020 年)之间碳排放量的变化值, $Q^t(i)$ 为第 t 年的碳排放量, $Q^0(i)$ 为基期(2020 年)的碳排放量, $\Delta Q_P(i)$ 为部门 i 的人口效应, $\Delta Q_{GC}(i)$ 为部门 i 的人均 GDP 效应, $\Delta Q_{EI}(i)$ 为部门 i 的单位 GDP 能耗效应, $\Delta Q_{VS}(i)$ 为部门 i 的电力消费比重效应, $\Delta Q_{NV}(i)$ 为部门 i 的非化石能源发电占比效应, $\Delta Q_{NS'}(i)$ 为部门 i 的非化石能源发电占比效应, $\Delta Q_{CF}(i)$ 为部门 i 的单位能耗二氧化碳排放量效应。根据 LMDI 分解方法对公式(12)中各效应进行分解, 公式如下:

$$\Delta Q_P(i) = W \times \ln(P^t(i)/P^0(i)) \quad (13)$$

$$\Delta Q_{GC}(i) = W \times \ln(GC^t(i)/GC^0(i)) \quad (14)$$

$$\Delta Q_{EI}(i) = W \times \ln(EI^t(i)/EI^0(i)) \quad (15)$$

$$\Delta Q_{VS}(i) = W \times \ln(VS^t(i)/VS^0(i)) \quad (16)$$

$$\Delta Q_{NV}(i) = W \times \ln(NV^t(i)/NV^0(i)) \quad (17)$$

$$\Delta Q_{NS'}(i) = W \times \ln(NS'^t(i)/NV'^0(i)) \quad (18)$$

$$\Delta Q_{CF}(i) = W \times \ln(CF^t(i)/CF^0(i)) \quad (19)$$

其中, W 为对数均值, 所有带上角标 t 的符号为该变量在第 t 年的取值, 所有带上角标 0 的符号为该变量在基期(2020 年)的取值。

$$W = (Q^t(i) - Q^0(i)) / (\ln Q^t(i) - \ln Q^0(i)) \quad (20)$$

根据公式(11), 对居民生活消费($i = 6$)的碳排放量 Kaya 恒等式进行分解, 公式如下:

$$\begin{aligned} \Delta Q(i) = Q^t(i) - Q^0(i) = & \Delta Q_P(i) + \Delta Q_{EC}(i) + \Delta Q_{VS}(i) + \Delta Q_{NV}(i) \\ & + \Delta Q_{NS'}(i) + \Delta Q_{CF}(i) \end{aligned} \quad (21)$$

同理, 对公式(21)中各效应进行分解, 对于增加的人均能耗效应 $\Delta Q_{EC}(i)$, 分解公式如下:

$$\Delta Q_{EC}(i) = W \times \ln(EC^t(i)/EC^0(i)) \quad (22)$$

● 建立各部门碳排放量的影响指标之间的关联模型: 皮尔逊相关性模型

各部门碳排放量指标是关于时间的定距变量且彼此之间具有一定的线性关系, 适用皮尔逊相关系数 r 来进行衡量, 公式如下:

$$r = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^n \frac{(X_i - E(X))}{\sigma_X} \frac{(Y_i - E(Y))}{\sigma_Y}}{n} \quad (23)$$

其中, $Cov(X,Y)$ 为变量 X 和变量 Y 的协方差, σ_X 为变量 X 的标准差, σ_Y 为变量 Y 的标准差, n 为样本个数。

5.1.3 模型求解

● 求解各部门碳排放量的 LDMI 模型

根据附录 3 中的代码对各部门碳排放量的 LDMI 模型进行求解, 算法时间复杂度为 $O(n)$, 平均运行时间约为 0.00109s。该模型的求解结果为各指标对部门碳排放量的逐年贡献率, 可以反映出各指标的贡献率变化趋势, 从而反映出各指标对碳排放量影响程度的变化趋势, 为之后碳排放量的预测提供依据。居民生活消费之外的部门以建筑消费部门为例, 各指标对该部门碳排放量的逐年贡献率如表 5.1 所示, 居民生活消费部门逐年贡献率如表 5.2 所示。

表 5.1 建筑消费部门各指标对碳排放量的逐年贡献率

年份	人口总量	人均 GDP	单位 GDP 能耗	单位能耗 二氧化碳 排放量	电力消费 比重	非化石能 源占比	非化石能 源消费比 重倒数
2011	59.193	339.115	59.549	16.153	-59.865	18.280	41.586
2012	40.860	371.317	-51.323	-79.072	-49.578	52.149	-2.571

2013	35.006	484.542	-254.925	93.650	175.118	-38.002	-137.116
2014	26.205	274.367	-406.012	-320.151	110.726	19.083	-129.810
2015	22.992	470.096	-308.426	213.071	112.403	29.242	-141.645
2016	35.776	473.089	-286.680	106.079	214.293	-25.020	-189.272
2017	30.774	399.647	-57.529	80.295	164.161	-47.625	-116.536
2018	27.603	489.564	94.916	-13.044	92.437	-14.150	-78.287
2019	22.670	458.673	-124.963	47.948	38.849	3.523	-42.372
2020	-3.623	134.819	-262.112	-121.198	65.787	14.564	-80.351

表 5.2 居民生活消费各指标对碳排放量的逐年贡献率

年份	人口总量	人均能耗	单位能耗二氧化碳排放量	电力消费比重	非化石能源占比	非化石能源消费比重倒数
2011	59.193	89.345	30.777	-12.604	56.093	-43.489
2012	40.860	379.873	-59.021	1.0672	6.341	-7.409
2013	35.006	225.250	126.223	291.804	140.152	-431.956
2014	26.205	-68.809	-457.346	-330.492	120.203	210.289
2015	22.992	228.622	110.440	84.882	-3.618	-81.264
2016	35.776	308.607	72.174	232.341	31.825	-264.166
2017	30.774	369.873	-2.759	198.287	-149.892	-48.394
2018	27.603	439.164	-36.449	54.628	113.977	-168.605
2019	22.670	144.463	3.462	20.756	128.401	-149.157
2020	-3.623	150.856	-146.999	-169.826	195.203	-25.377

● 求解各部门碳排放量影响指标之间的皮尔逊相关性模型

根据附录 3 中代码对各部门碳排放量影响指标之间的皮尔逊相关模型进行求解，算法时间复杂度为 $O(n^2)$ ，运行时间约为 0.00112s。对计算得出的皮尔逊系数进行评估，根据表 5.3 可以确定各部门碳排放量影响指标之间的相关强度等级。

表 5.3 居民生活消费各指标对碳排放量的逐年贡献率

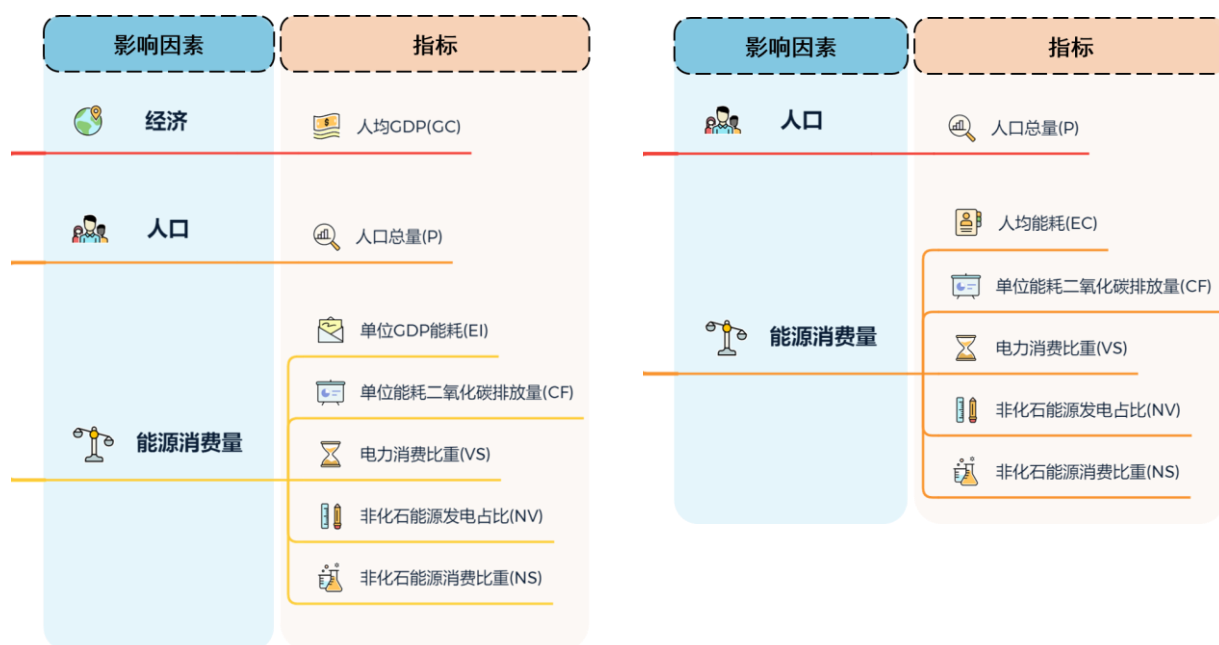
皮尔逊相关系数 r	相关强度
0	零相关
0~0.3	微相关
0.3~0.5	低度相关
0.5~0.8	显著相关
1	完全相关

5.1.4 问题回答

● 第(1)问：建立指标与指标体系

根据假设 3，只考虑经济、人口、能源消费量类因素对碳排放量的影响，居民生活消费以外的部门与居民生活消费的指标体系结构对比如图 5.3 所示。居民生活消费以外的部门包含人均 GDP(GC)、人口总量(P)、单位 GDP 能耗(EI)、单位能耗二氧化碳排放量(CF)、电力消费比重(VS)、非化石能源发电占比(NV)、非化石能源比重(NS)共 7 个指标，其中，人均 GDP(GC)体现经济影响，人口总量(P)体现人口影响，单位 GDP 能耗(EI)、单位能耗二氧化碳排放(CF)、非化石能源发电占比(NV)、电力消费比重(NS)这 5 个指标体现能源消费量影响。居民生活消费包含人口总量(P)、人均能耗(EC)、单位能耗二氧化碳排放量(CF)、电力消费比重(VS)、非化石能源发电占比(NV)、非化石能源比重(NS)共 6 个指

标, 其中, 人口总量(P)体现人口影响, 人均能耗(EC)、单位能耗二氧化碳排放(CF)、非化石能源发电占比(NV)、电力消费比重(NS)这 5 个指标体现能源消费量影响, 由于该部门不产生 GDP, 无须考虑经济影响因素。



图(a) 居民生活消费以外的部门

图(b) 居民生活消费

图 5.3 各部门的体系结构图

● 第(2)问：分析区域碳排放量以及经济、人口、能源消费量

①绘制 2011-2020 年碳排放量趋势图,如图 5.4 所示。并利用平均值、中位数、上四分位数、中位数、下四分位数、标准差对 2011-2020 年碳排放量进行描述性统计,结果如表 5.4 所示。

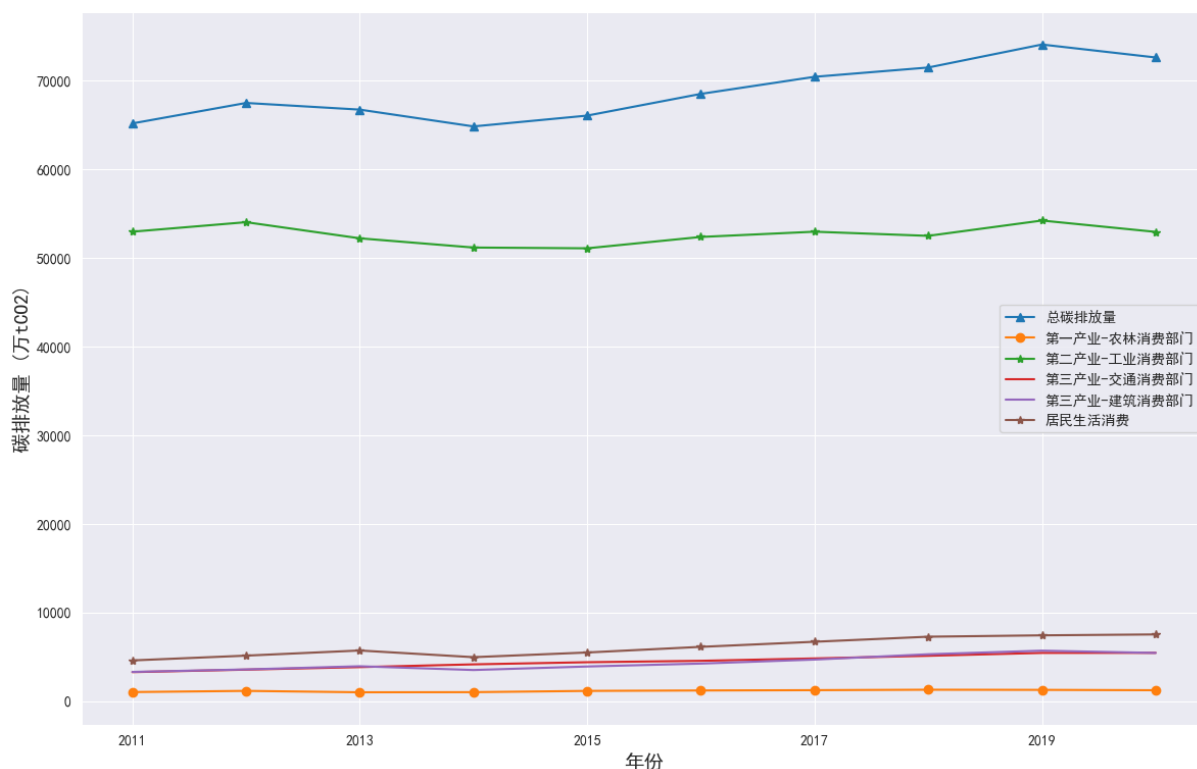


图 5.4 2011-2020 年碳排放量趋势图

表 5.4 2011-2020 年碳排放量的描述性统计

时期	类别	平均值	中位数	上四分位数	下四分位数	标准差
十二五	总量	1077.46	1031.18	1162.51	1020.85	79.36
	农林消费部门	52308.60	52229.08	52975.79	51187.98	1243.98
	工业消费部门	7502.88	7676.08	7791.47	7147.77	659.83
	交通消费部门	3848.92	3847.16	4157.35	3561.73	447.92
	建筑消费部门	3653.96	3586.05	3916.46	3518.72	273.13
	居民生活消费	5185.74	5141.28	5495.89	4968.37	439.67
十三五	总量	1253.74	1245.02	1278.38	1238.76	33.45
	农林消费部门	53010.89	52954.05	52975.85	52506.89	733.69
	工业消费部门	10160.83	10422.11	10906.03	9524.01	982.16
	交通消费部门	5082.81	5125.17	5449.65	4826.11	393.50
	建筑消费部门	5078.02	5296.95	5449.19	4697.91	393.50
	居民生活消费	7016.42	7277.52	7431.55	6706.67	588.92

结合 2011-2020 年碳排放量趋势图以及描述性统计结果进行分析，可以得出：a)总碳排放量在 2011-2019 年总体呈上升趋势，但在 2020 年略有下降；b)工业消费部门的碳排放量总体呈波动趋势，居民生活消费、交通消费部门和建筑消费部门的碳排放量总体呈小幅上升趋势，农林消费部门的碳排放量总体略微上升且上升速度十分缓慢；c)十三五期间的总碳排放量以及各部门碳排放量总体高于十二五；d)工业消费部门的碳排放量远远高于其他部门，其次是居民生活消费，再次是交通消费部门和建筑消费部门，它们相差很小，最后是农林消费部门，概况来讲第二产业碳排放量远大于居民生活消费和第三产业碳排放量，大于第一产业碳排放量。

②根据各部门碳排放量的 LDMI 模型的求解结果，绘制各部门碳排放量影响指标的贡

献率的逐年变化趋势图，如图 5.5 到图 5.10 所示。以居民生活消费的趋势图为例，通过观察可以发现：从总体上看，人均能耗对碳排放量具有显著拉动作用，非化石能源消费比重、单位能耗二氧化碳排放量对碳排放量具有显著抑制作用，人口总量、非化石能源发电占比、电力消费比重的影响较小。

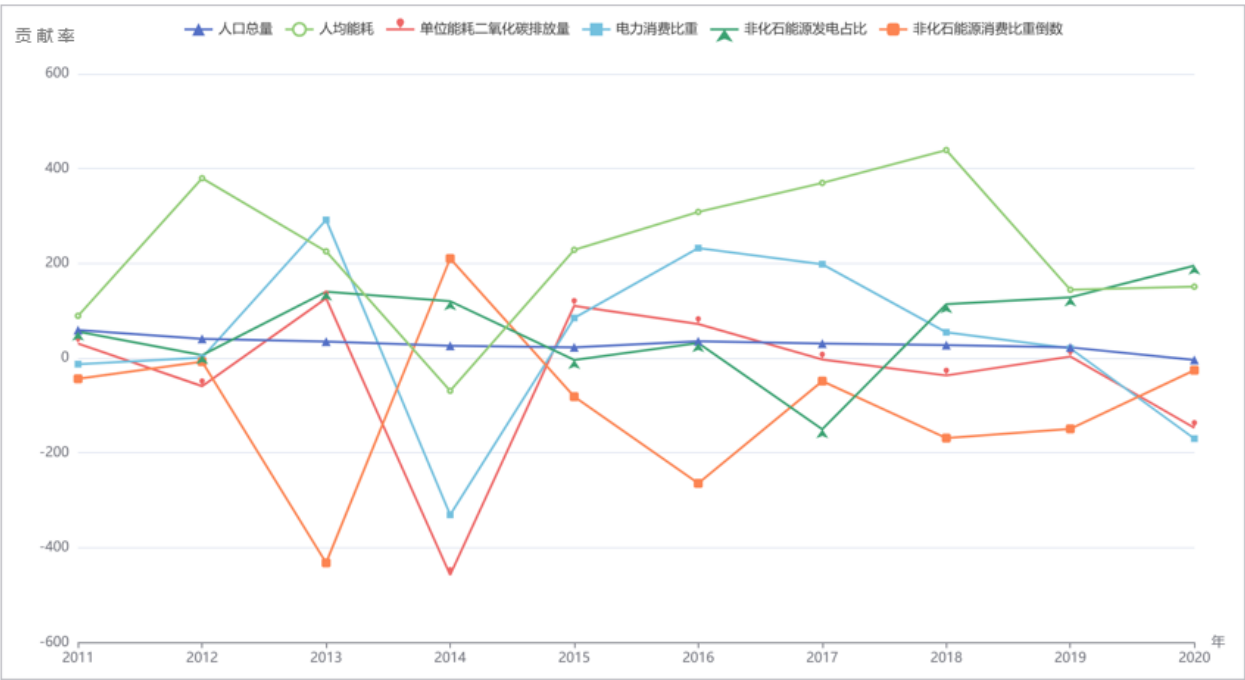


图 5.5 居民生活消费碳排放量影响指标的贡献率的逐年变化趋势图

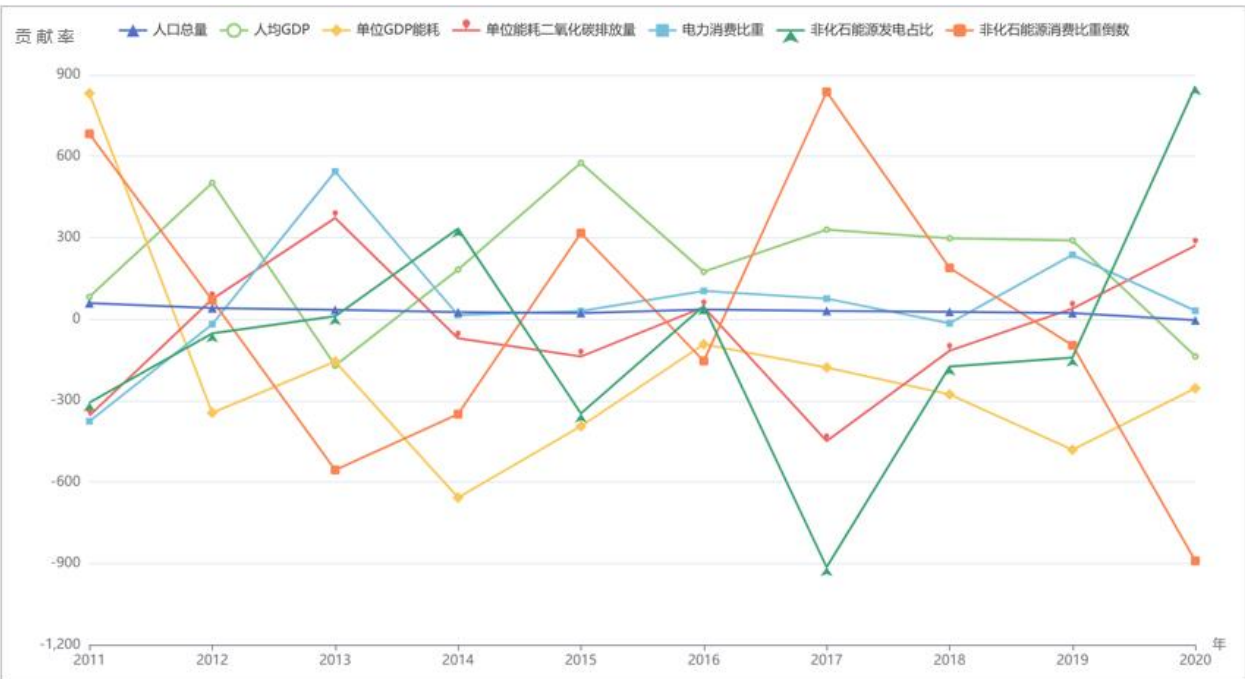


图 5.6 能源供应部门碳排放量影响指标的贡献率的逐年变化趋势图

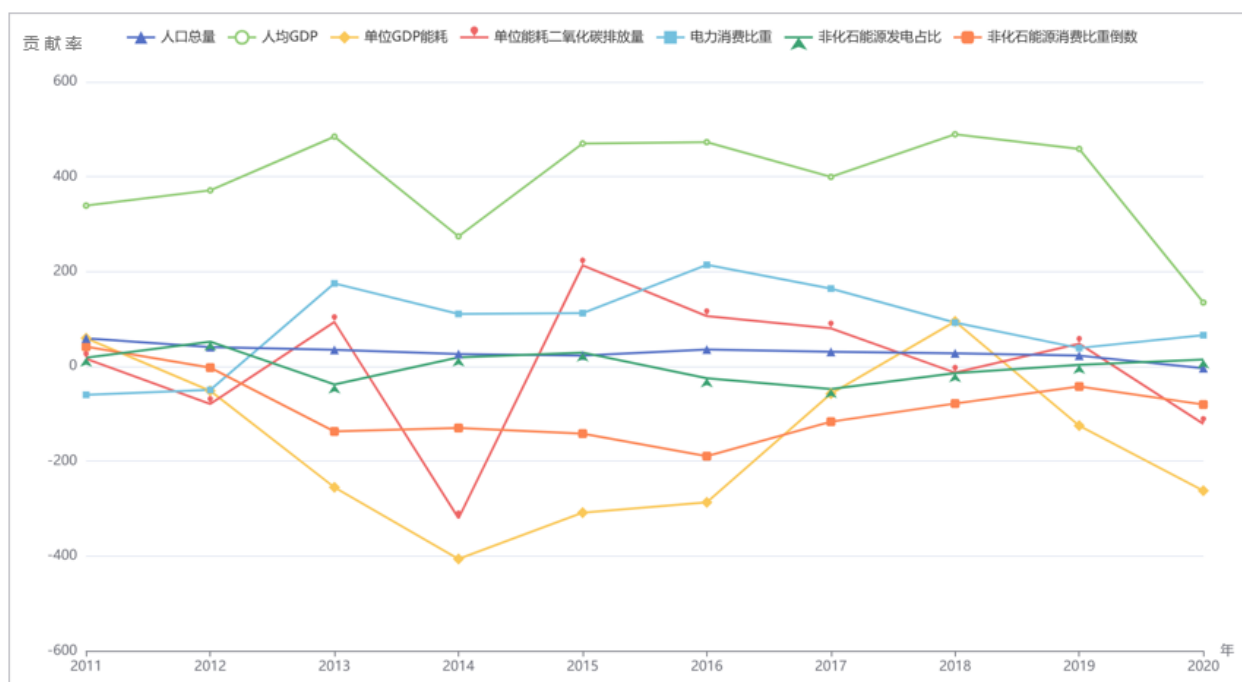


图 5.7 建筑消费部门碳排放量影响指标的贡献率的逐年变化趋势图

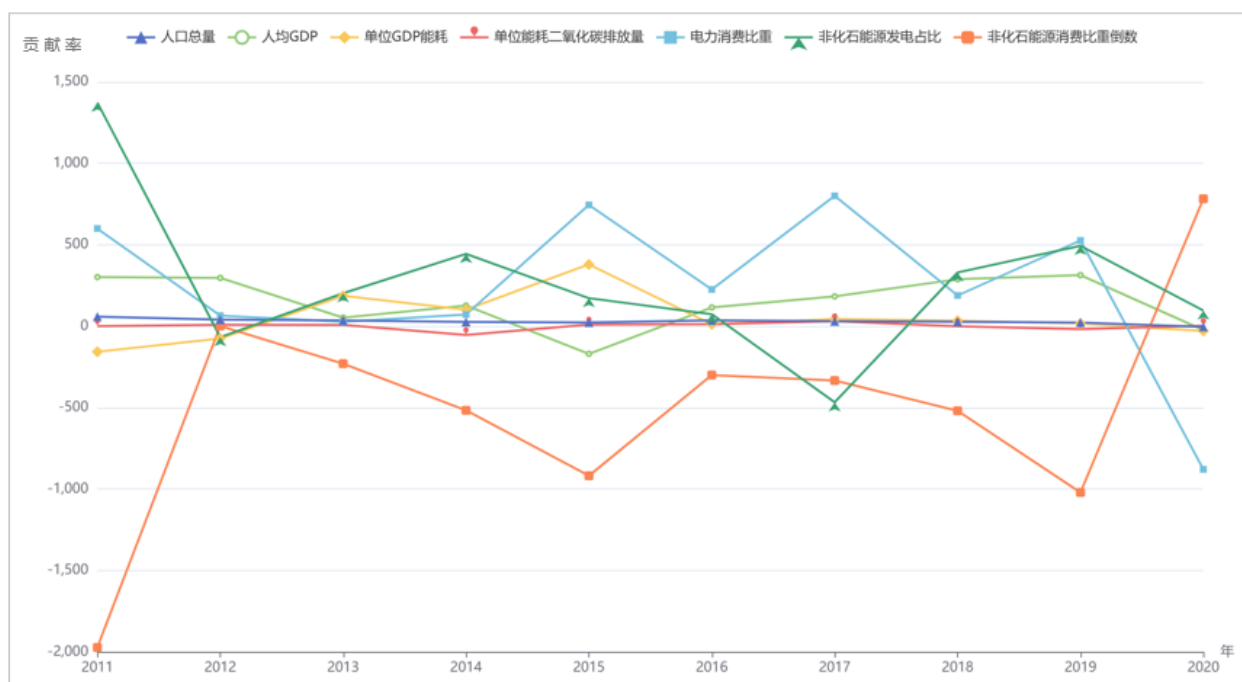


图 5.8 交通消费部门碳排放量影响指标的贡献率的逐年变化趋势图

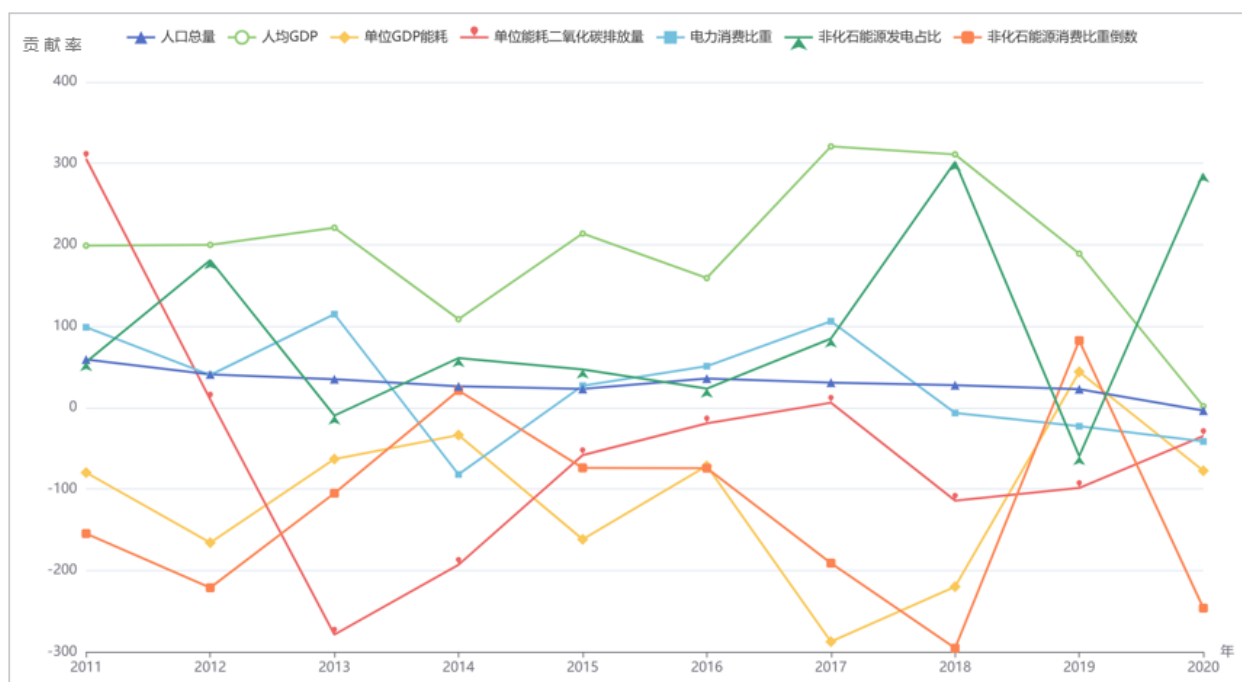


图 5.9 工业消费部门碳排放量影响指标的贡献率的逐年变化趋势图

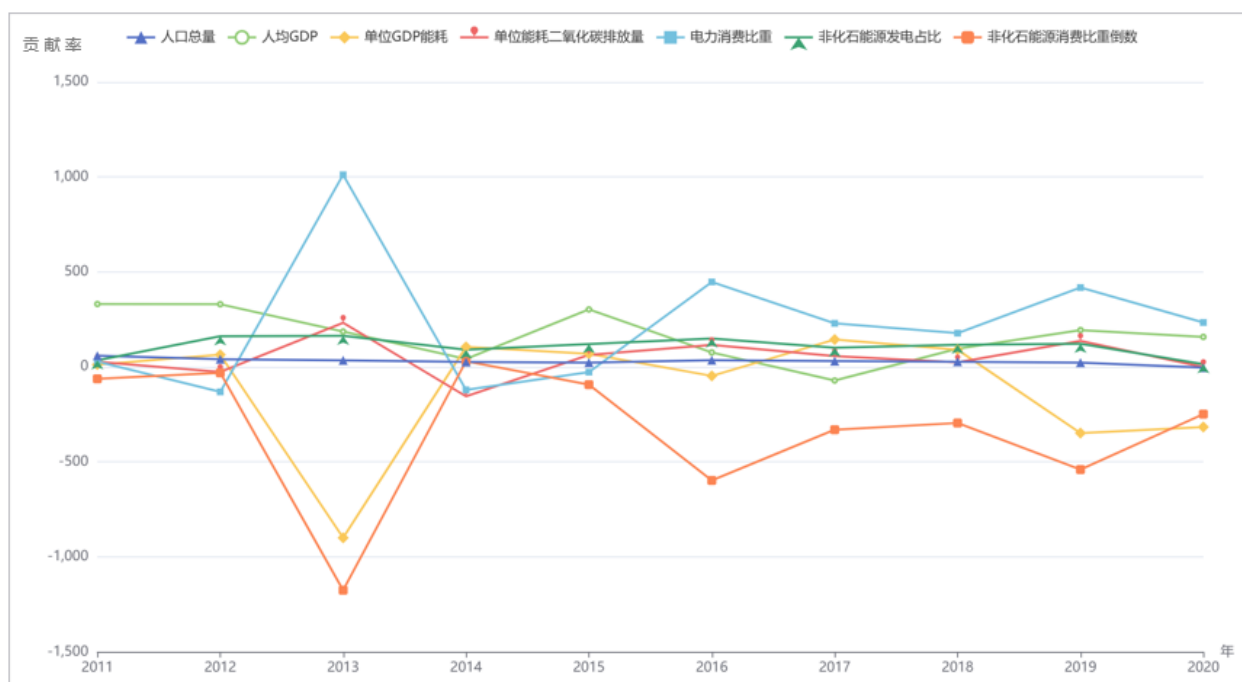


图 5.10 农林消费部门碳排放量影响指标的贡献率的逐年变化趋势图

③通过观察各部门碳排放量影响指标的贡献率情况，可以发现各部门人均 GDP、单位能耗二氧化碳排放量和非化石能源消费比重对碳排放量的影响较大。因此，实现碳达峰与碳中和需要面对的主要挑战为综合考虑这 3 个指标对碳排放量影响。

● 第(3)问：区域碳排放量以及经济、人口、能源消费量各指标及其关联模型。

① 根据上一问的分析，人均 GDP、单位能耗二氧化碳排放量和非化石能源消费比重

是实现双碳目标的主要挑战，因此重点分析这 3 个指标在总体级别和部门级别上(以建筑消费部门为例)的环比和同比。将相关指标的同比趋势可视化，如图 5.11、图 5.12 所示。通过观察图 5.11 可以发现，总体级别的非化石能源消费比重的环比均为正值且存在较大的波动，说明总体级别的非化石能源消费比重逐年以波动性的速率增长；总体级别的人均 GDP 的环比均为正值且呈现略微的下降趋势，说明总体级别的人均 GDP 的增长速率放慢；总体级别的单位能耗二氧化碳排放量环比在 2013-2016 年以及 2019 年取值为负，说明总体级别的单位能耗二氧化碳排放量在该时间实现负增长。通过观察图 5.12 可以发现，建筑消费部门人均 GDP 的环比均为正值且存在较大的波动，说明建筑消费部门人均 GDP 逐年以波动性的速率增长；建筑消费部门非化石能源消费比重除 2011 年外均为正值且总体呈现先上升后下降的趋势，说明建筑消费部门非化石能源消费比重的增长速率先快后慢；建筑消费部门单位能耗二氧化碳排放量环比呈现较大的正负波动，说明建筑消费部门二氧化碳排放量出现较大的增减波动。

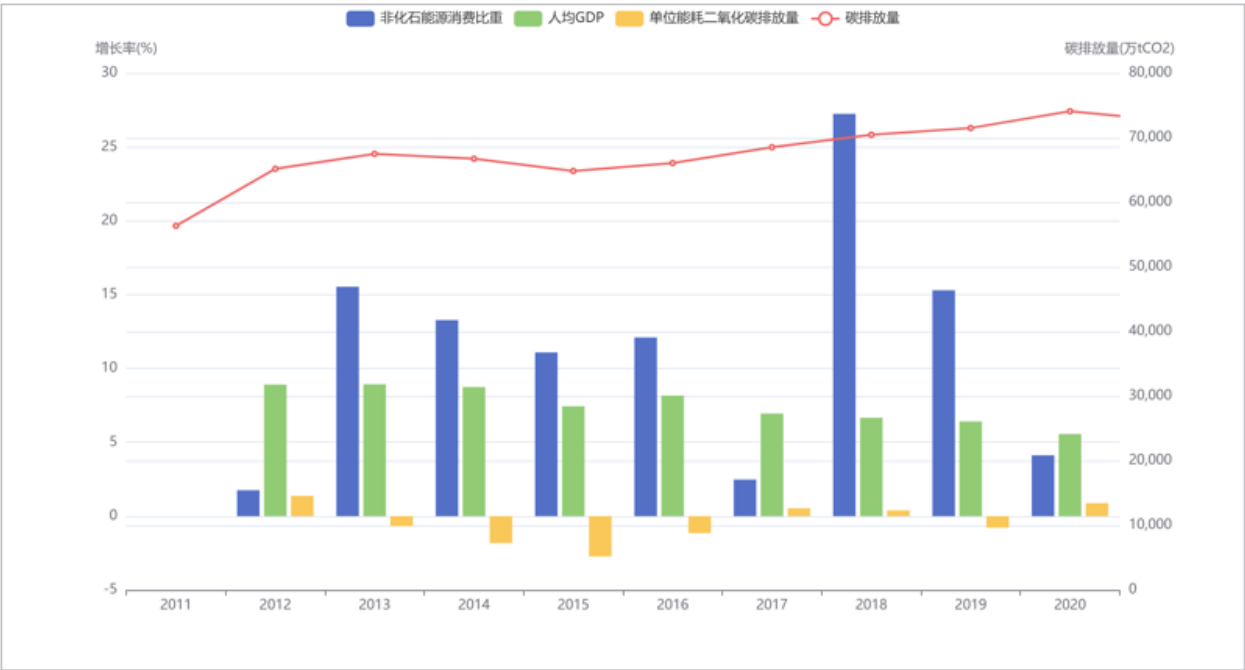


图 5.11 总体级别重点指标的环比趋势图

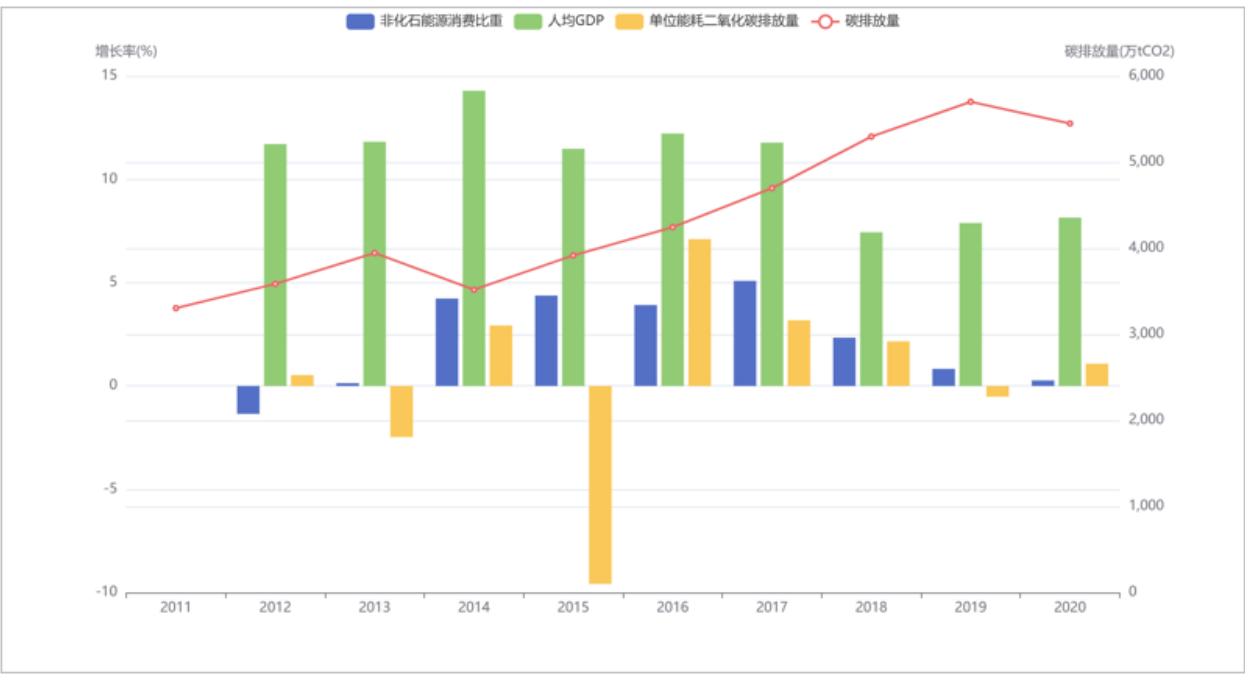


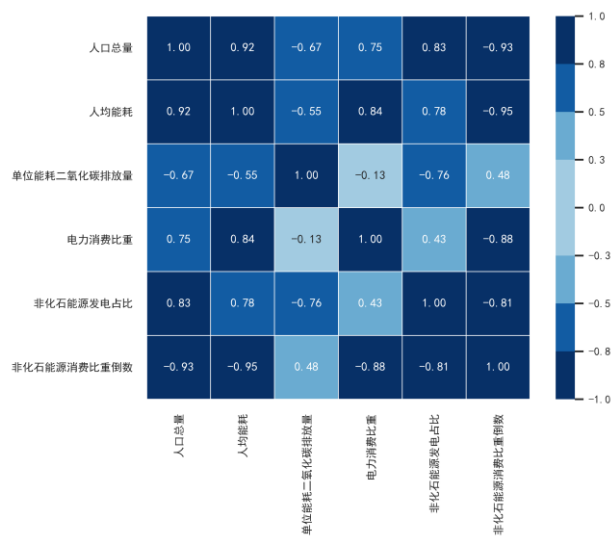
图 5.12 部门级别(以建筑消费部门为例)重点指标的环比趋势图

然后，以 5 年作为时间划分，计算十二五(2011-2015 年)和十三五期间(2016-2020 年)人均 GDP、单位能耗二氧化碳排放量和非化石能源消费在总体级别和部门级别上(以建筑消费部门为例)的同比，如表 5.5 所示和表 5.6 所示。

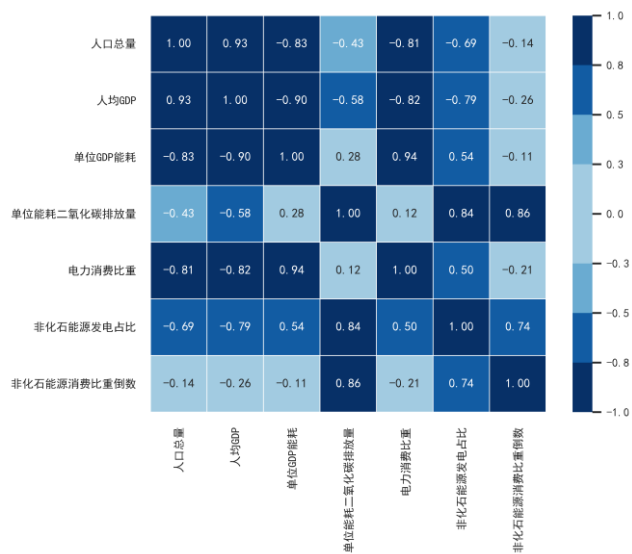
表 5.5 总体级别重点指标的同比			
时期	非化石能源消费比重	人均 GDP	单位能耗二氧化碳排放量
十二五	0.12	0.79	-0.02
十三五	0.11	0.48	0.029

表 5.6 部门级别(以建筑消费部门为例)重点指标的同比			
时期	非化石能源消费比重	人均 GDP	单位能耗二氧化碳排放量
十二五	0.66	0.50	-0.05
十三五	0.61	0.33	0.02

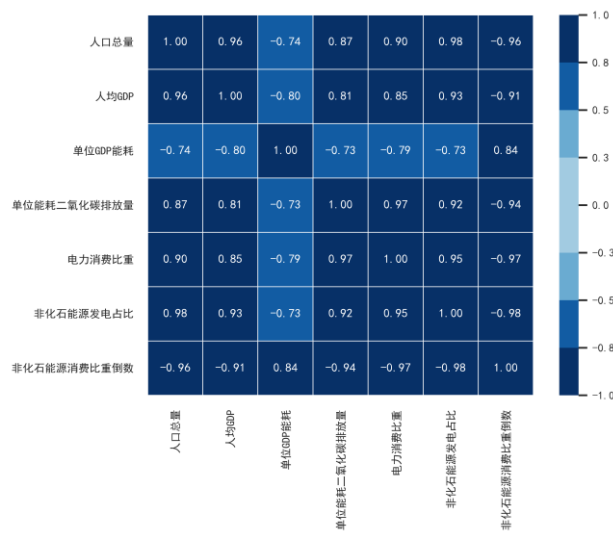
②根据各部门碳排放量影响指标之间的皮尔逊相关性模型的求解结果，利用热力图的颜色深浅对相关强度等级进行可视化，如图 5.13 所示。



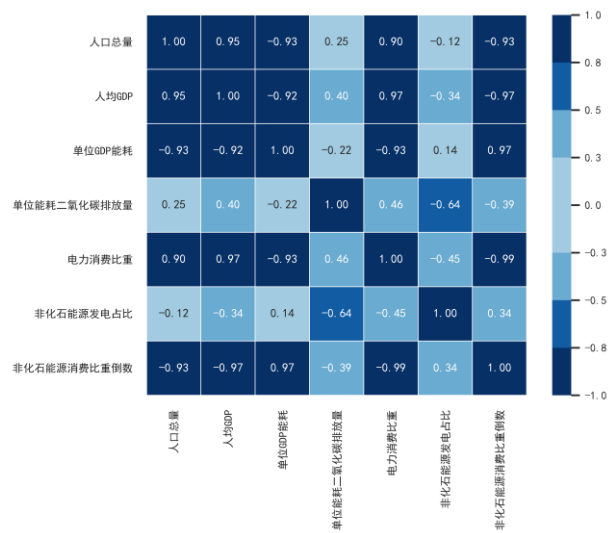
图(a) 居民生活消费



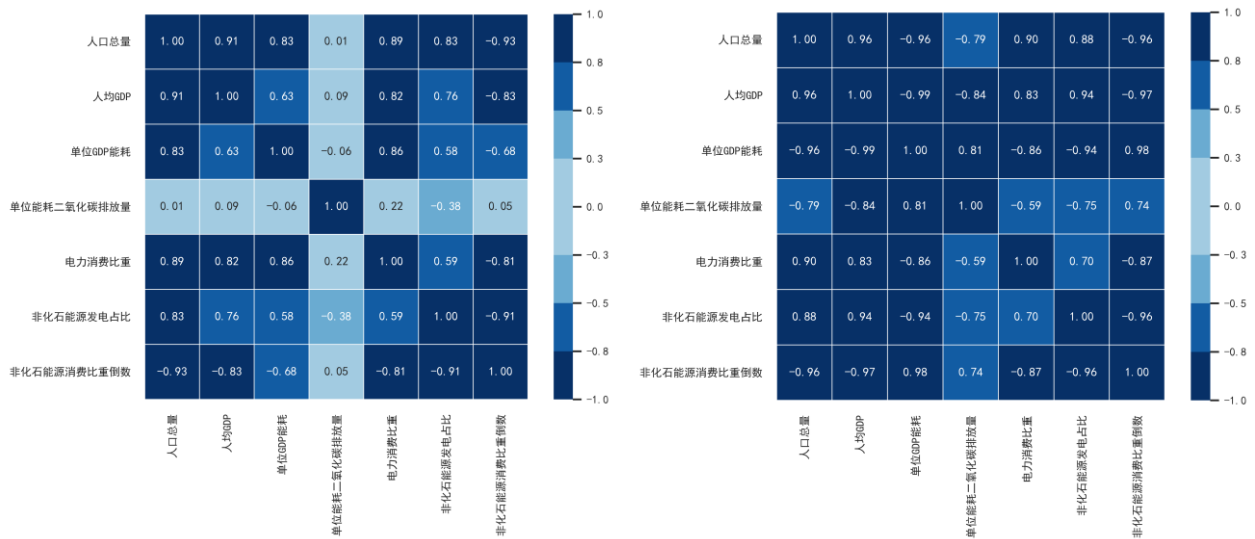
图(b) 能源供应部门



图(c) 农林消费部门



图(d) 建筑消费部门



图(e) 交通消费部门

图(f) 工业消费部门

图 5.13 各部门碳排放量影响指标之间的相关强度热力图

③ 皮尔逊相关系数较大的指标说明存在一定的冗余，需要进行筛选优化。由于能源消费量的指标较多，因此将其作为重点优化对象。通过观察图 5.12 可以发现，非化石能源发电占比与其他能源消费量指标整体相关强度较大，说明该指标对碳排放量的影响可以在其他能源消费量指标中体现，因此可以对该指标进行删减。结合双碳政策并参照《中国 2030 年前碳达峰研究报告》^[7]《中国 2060 年前碳中和研究报告》^[8]《中国 2030 年能源电力发展规划研究及 2060 年展望》^[9]中的政策要求，确定关键时间节点(2030 年、2060 年)碳排放量预测模型中各指标的取值，结果如表 5.7 所示。

表 5.7 关键时间节点的碳排放预测模型指标取值

时间	人均 GDP (GC)	人口总量 (P)	单位 GDP 能耗 (EI)	单位能耗二氧化碳排放量 (CF)	电力消费比重 (VS)	非化石能源消费比重 (NS)
2030	165240.96	8783.03	0.27	2.23	0.20	0.45
2060	660855.40	5897.37	0.10	2.31	0.32	0.88

5.2 问题二模型的建立与求解

5.2.1 数据来源

问题二数据包括题目原始数据和参照政策文件^[7-9]的补充数据两部分，题目原始数据为各指标在 2010 年到 2020 年的数据，补充数据如表 5.8 所示。

表 5.8 问题二补充数据

时间	人口总量 增长率 (r_P)	GDP 增长率 (r_G)	能源消费总量 增长率 (r_E)	化石能源消费量 增长率 (r_F)	发电量 增长率 (r_V)	第三产业 GDP 占比 增长率 (r_{TI})
2025	0.0043	0.055	0.0286	-0.0169	0.042	0.0122
2030	0.0028	0.046	0.0143	-0.0127	0.03	0.005
2035	/	/	/	/	/	/

2040	/	/	/	-0.0466	/	/
2050	-0.0021	0.035	-0.0008	-0.048	0.02	0.0053
2060	-0.005	0.025	-0.0017	-0.0578	0.58	0.0043

注：①指标增长率表示该时间节点与上一时间节点之间的年平均增速。

② “/” 表示缺失值。

5.2.2 模型的建立

● 建立基于人口和经济变化的能源消费量预测模型：多元线性回归模型

①利用复合增长率推算 2021-2060 年人口总量(P)、GDP(G)和能源消费量(E)

从表 5.8 中获得人口总量增长率(r_P)、GDP 增长率(r_G)和能源消费量增长率(r_E)，取值情况如下：

$$r_P = \begin{cases} 0.0043, & t \in [2020, 2025] \\ 0.0028, & t \in (2025, 2030] \\ -0.0021, & t \in (2030, 2050] \\ -0.005, & t \in (2050, 2060] \end{cases} \quad (24)$$

$$r_G = \begin{cases} 0.055, & t \in [2020, 2025] \\ 0.046, & t \in (2025, 2030] \\ -0.035, & t \in (2030, 2050] \\ -0.025, & t \in (2050, 2060] \end{cases} \quad (25)$$

$$r_E = \begin{cases} 0.0286, & t \in [2020, 2025] \\ 0.0143, & t \in (2025, 2030] \\ -0.0008, & t \in (2030, 2050] \\ -0.0017, & t \in (2050, 2060] \end{cases} \quad (26)$$

其中， t 为年份。然后，利用复合增长率与基期(2020 年)数据对之后的数据进行推算，公式如下：

$$\begin{cases} P^t = P^{t-1}(1 + r_P) & \text{当 } t > t_0 + 1 \text{ 时} \\ P^t = P^0(1 + r_P) & \text{当 } t = t_0 + 1 \text{ 时} \end{cases} \quad (27)$$

$$\begin{cases} G^t = G^{t-1}(1 + r_G) & \text{当 } t > t_0 + 1 \text{ 时} \\ G^t = G^0(1 + r_G) & \text{当 } t = t_0 + 1 \text{ 时} \end{cases} \quad (28)$$

$$\begin{cases} E^t = E^{t-1}(1 + r_E) & \text{当 } t > t_0 + 1 \text{ 时} \\ E^t = E^0(1 + r_E) & \text{当 } t = t_0 + 1 \text{ 时} \end{cases} \quad (29)$$

其中, t 表示年份, t^0 表示基期, 在本问中 $t^0=2020$, P^t 、 G^t 、 E^t 分别为第 t 年的人口总量、GDP 和能源消费量, P^0 、 G^0 、 E^0 分别为基期(2020 年)的人口总量、GDP 和能源消费量, r_P 、 r_G 、 r_E 分别为人口总量、GDP 和能源消费量的增长率。

②建立能源消费量与人口总量、GDP 的多元线性回归模型

根据公式(24)到公式(26)结合基期(2020 年)人口总量、GDP 能源消费量推算出 2021-2060 年的人口总量、GDP, 据此, 建立能源消费量与人口总量、GDP 的多元线性回归模型, 公式如下:

$$E = \alpha P + \beta G + \lambda \quad (30)$$

其中, α 为人口总量的系数, β 为 GDP 的系数, λ 表示误差项。

● 建立区域碳排放量预测模型: 扩展的 STIRPAT 模型

①确定预测模型指标

根据问题一第(3)问的分析结果, 确定区域碳排放量的影响指标, 包括人均 GDP(GC)、人口总量(P)、单位 GDP 能耗(EI)、单位能耗二氧化碳排放量(CF)、电力消费比重(VS)、非化石能源消费比重(NS)。在此基础上, 引入第三产业 GDP 占比(TI)指标, 以体现产业部门结构对碳排放量的影响。

②利用复合增长率推算 2021-2060 年人口总量(P)、GDP(G)、能源消费量(E)、非化石能源消费量(E_N)、发电量(E_V)、第三产业 GDP 占比(TI)。

从表 5.8 中获得人口总量增长率(r_P)、GDP 增长率(r_G)、能源消费量增长率(r_E)、化石能源消费量增长率(r_F)、发电量增长率(r_V)、第三产业 GDP 占比增长率(r_{TI}), r_P 、 r_G 、 r_E 增长率取值情况如公式(24)到公式(26)所示, 其余取值情况如下:

$$r_F = \begin{cases} -0.0169, & t \in [2020, 2025] \\ -0.0127, & t \in (2025, 2030] \\ -0.0466, & t \in (2030, 2040] \\ -0.048, & t \in (2040, 2050] \\ -0.0578, & t \in (2050, 2060] \end{cases} \quad (31)$$

$$r_V = \begin{cases} 0.042, & t \in [2020, 2025] \\ 0.03, & t \in (2025, 2030] \\ 0.02, & t \in (2030, 2050] \\ 0.58, & t \in (2050, 2060] \end{cases} \quad (32)$$

$$r_{TI} = \begin{cases} 0.0122, & t \in [2020, 2025] \\ 0.005, & t \in (2025, 2030] \\ 0.0053, & t \in (2030, 2050] \\ 0.0043, & t \in (2050, 2060] \end{cases} \quad (33)$$

据此，结合基期(2020 年)数据对之后的数据进行推算，人口总量、GDP、能源消费量推算公式如公式(27)到公式(29)所示，其余指标推算公式如下：

$$\begin{cases} E_F^t = E_F^{t-1}(1 + r_F) & \text{当 } t > t_0 + 1 \text{ 时} \\ E_F^t = E_F^0(1 + r_F) & \text{当 } t = t_0 + 1 \text{ 时} \end{cases} \quad (34)$$

$$\begin{cases} E_V^t = E_V^{t-1}(1 + r_V) & \text{当 } t > t_0 + 1 \text{ 时} \\ E_V^t = E_V^0(1 + r_V) & \text{当 } t = t_0 + 1 \text{ 时} \end{cases} \quad (35)$$

$$\begin{cases} TI^t = TI^{t-1}(1 + r_{TI}) & \text{当 } t > t_0 + 1 \text{ 时} \\ TI^t = TI^0(1 + r_{TI}) & \text{当 } t = t_0 + 1 \text{ 时} \end{cases} \quad (36)$$

其中， t 表示年份， t^0 表示基期，在本问中 $t^0=2020$ ， E_F^t 、 E_V^t 、 TI^t 分别为第 t 年的化石能源消费量、发电量和第三产业 GDP 占比， E_F^0 、 E_V^0 、 TI^0 分别为基期(2020 年)的化石能源消费量、发电量和第三产业 GDP 占比， r_P 、 r_G 、 r_E 分别为化石能源消费量、发电量和第三产业 GDP 占比的增长率。

结合公式(1)到公式(3)、公式(5)、公式(6)计算人均 GDP(GC)、单位 GDP 能耗(EI)、单位能耗二氧化碳排放量(CF)、电力消费比重(VS)和非化石能源消费比重(NS)，其中非化石能源消费量的计算公式如下：

$$E_N = E - E_F \quad (37)$$

③建立区域碳排放量(Q)的扩展 STIRPAT 模型

STIRPAT 模型源于 IPAT 等式，IPAT 等式^[10]表示如下：

$$I = P \times A \times T \quad (38)$$

其中，I(Impact)为区域碳排放量；P(Population)为区域人口规模；A(Affluence)为区域人均富裕程度；T(Technology)为区域总体技术水平。STIRPAT 模型基本思路如图 5.14 所示，首先，将基础的 IPAT 等式扩展为经典的 STIRPAT 模型；然后，结合具体指标对经典的 STIRPAT 模型进行扩展和改进；最后，对模型取对数，将其改写为加法形式，以此对碳排放量进行预测。

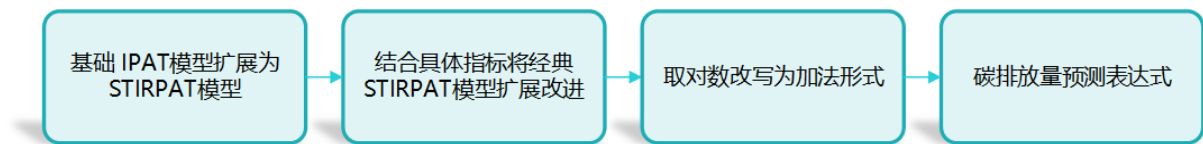


图 5.14 STIRPAT 模型基本思路图

根据确定的指标建立扩展的 STIRPAT 模型，公式如下：

$$Q = a \times GC^\alpha \times P^\beta \times EI^\gamma \times CF^\delta \times VS^\varepsilon \times NS^\theta \times TI^\lambda \times e \quad (39)$$

其中， Q 为区域碳排放量， GC 为人均 GDP， P 为人口总量， EI 为单位 GDP 能耗， CF 为单位能耗二氧化碳排放量， VS 为电力消费比重， NS 为非化石能源消费比重， TI 为第三产业 GDP 占比， α 、 β 、 γ 、 δ 、 ε 、 θ 、 λ 分别为指标 Q 、 GC 、 P 、 CF 、 VS 、 NS 、 TI ， a 、 e 为常数项。对公式(39)取对数，可以得到：

$$\ln Q = \ln a + \alpha \ln GC + \beta \ln P + \gamma \ln EI + \delta \ln CF + \varepsilon \ln VS + \theta \ln NS + \lambda \ln TI + \ln e \quad (40)$$

5.2.3 模型的求解

● 求解能源消费量与人口总量、GDP 的多元线性回归模型

人口总量和 GDP 总量的预测趋势如图 5.15 所示。通过观察可以发现，人口总量在 2050 年之前较为平稳，但在 2050 年突然下降，这是因为 2050 年是中国式现代化的关键时间节点，人口总量的趋势在 2050 年受政策影响明显。GDP 总量呈明显的上升趋势，说明在预测时间段内我国经济将持续高速发展。

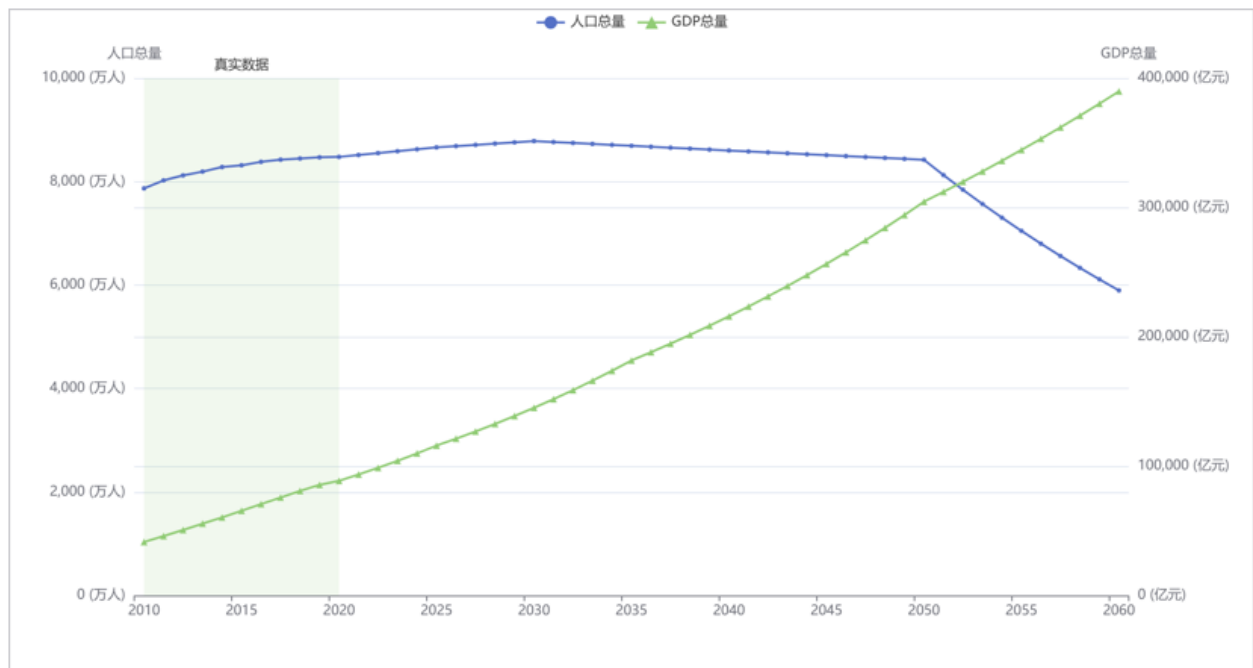


图 5.15 能源消费量预测趋势图

根据附录 4 中代码求解能源消费量与人口总量、GDP 的多元线性回归模型，求解方法为线性回归，原理是最小二乘法。对模型效果进行评估， R^2 为 0.94，说明拟合程度较高，算法时间复杂度为 $O(n^3)$ 。模型求解结果如下：

$$E = 10.16P + 0.02G - 56613.86 \quad (41)$$

该表达式体现了能源消费量与人口、GDP 的关联关系，满足要求 2 和要求 3。

绘制能源消费量的预测趋势图，如图 5.16 所示。通过观察可以发现，能源消费量在 2030 年和 2050 年具有明显拐点，总体呈现“倒 U”型。这是因为用于拟合模型的人口数据和 GDP 数据是根据相关政策文件中要求的增长率计算得出，考虑了中国式现代化关键时间节点的影响，满足要求 1。



图 5.16 能源消费量预测趋势图

● 求解区域碳排放量的扩展 STIRPAT 模型

在对该模型进行求解时，对比 Ridge 回归、经过交叉验证的 Ridge 回归、LASSO 回归、经过交叉验证的 Ridge 回归、弹性网络回归和多元线性回归这 6 种方法的预测情况，如图 5.17 所示。通过观察可以发现 LASSO 回归、弹性网络回归和多元线性回归预测值趋势和真实值趋势相距过近，说明存在过拟合现象；Ridge 回归和经过交叉验证的 Ridge 回归在预测过程中逐渐趋近真实值且趋于稳定。绘制岭迹图，如图 5.18 所示，通过观察可以发现 STIRPAT 模型变量间的交叉较多，说明特征之间具有较高的多重共线性。因此，在求解时采用岭回归法，如附录 5 展示的代码所示，以牺牲部分拟合精度为代价来获取更符合实际情况的系数，克服因素本身存在的相关性问题^[10]。算法时间复杂度为 $O(n^2)$ 。

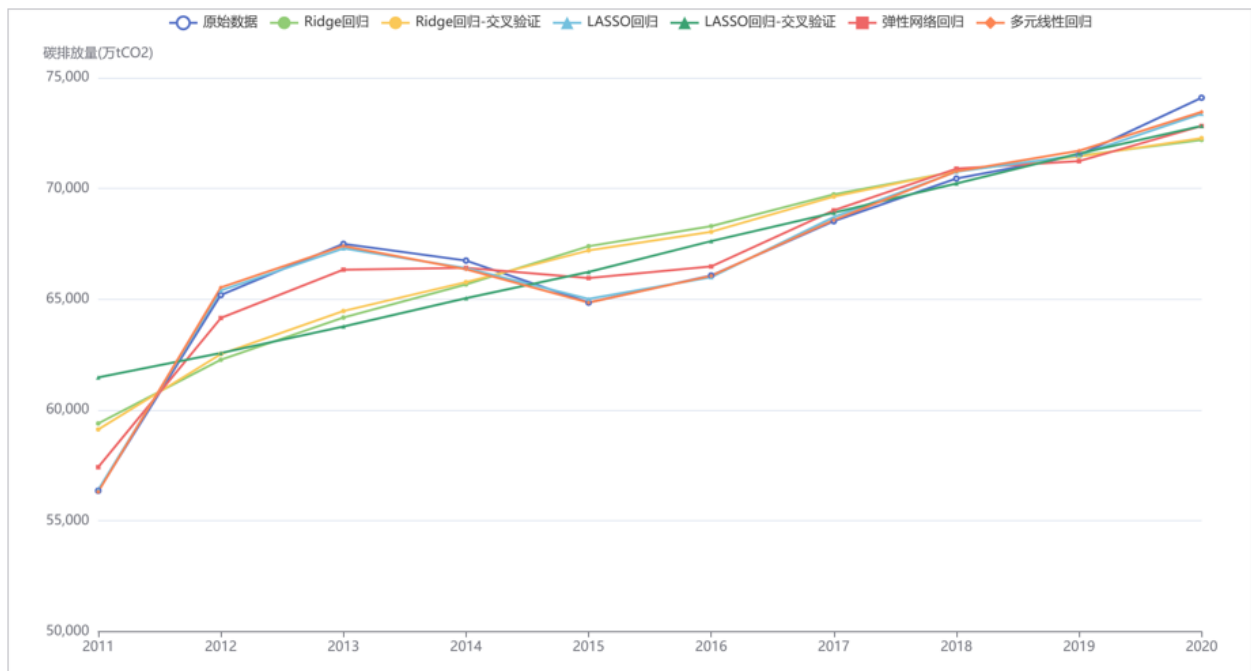


图 5.17 能源消费量多种回归方法预测对比图

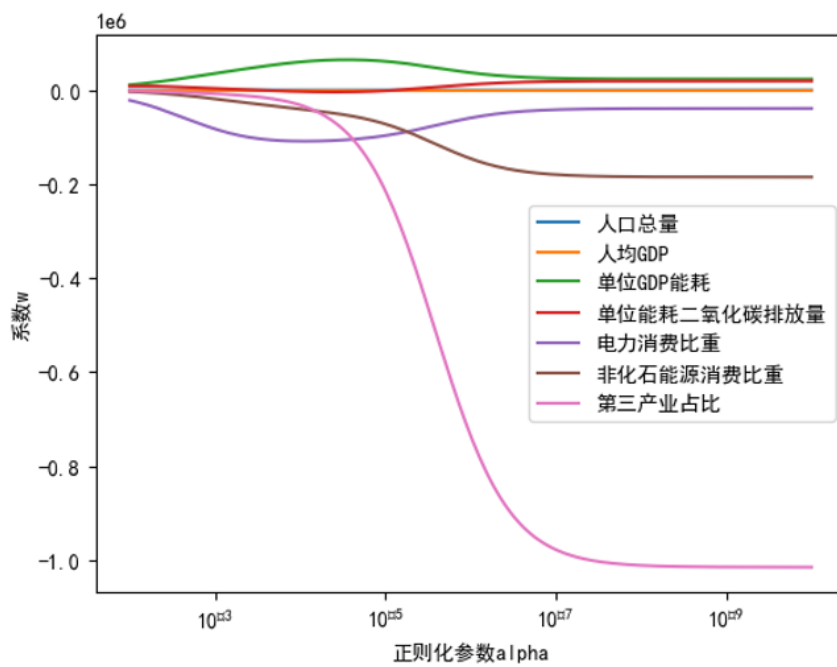


图 5.18 岭迹图

模型求解结果如下：

$$\begin{aligned} \ln Q = & \ln a + 0.094 \ln GC + 0.01 \ln P - 0.019 \ln EI - 0.003 \ln CF \\ & + 0.004 \ln VS + 0.016 \ln NS + 0.005 \ln TI + 9.972 \end{aligned} \quad (42)$$

由于人口、人均 GDP 和能源消费量是由公式(27)到公式(29)推算预测而来,满足要求 1。公式(42)体现了区域碳排放量与人均 GDP(GC)、人口总量(P)、单位 GDP 能耗(EI)、单位能耗二氧化碳排放量(CF)、电力消费比重(VS)和非化石能源消费比重(NS)、第三产业 GDP 占比(TI)这 7 个指标的关联关系。由于单位 GDP 能耗(EI)、单位能耗二氧化碳排放量(CF)、电力消费比重(VS)和非化石能源消费比重(NS)这 4 个指标可以体现能源消费量对碳排放量的影响,且第三产业 GDP 占比(TI)可以体现产业部门结构对碳排放量的影响,满足要求 2。由于非化石能源消费比重(NS)可以体现能源消费品种结构对碳排放量的影响,且第三产业 GDP 占比(TI)可以体现产业部门结构对碳排放量的影响,满足要求 3。

5.3 问题三模型的建立与求解

5.3.1 数据来源

问题三数据来源于问题二相同,包括题目原始数据和参照政策文件^[7-9]的补充数据两部分,题目原始数据为各指标在 2010 年到 2020 年的数据,补充数据如表 5.8 所示。

5.3.2 模型的建立

● 情景设计：自然情景、基准情景和雄心情景

设计自然情景、基准情景、雄心情景三类情景,根据假设 4、假设 5 以及表 5.8,设置各指标的情景参数值,结果如表 5.9 所示。在自然情景下,各指标基于目前发展趋势稳定变化,无人干预;在基准情景下,结合双碳政策与技术进步,设定各指标变化率,可以按时碳达峰与碳中和;在雄心场景下,相较于基准情景,指标变化幅度增加,可以率先碳达峰与碳中和。

表 5.9 情景设置参数表

参数	自然情景		基准情景		雄心情景	
	2030	2060	2035	2060	2035	2060
GDP 增长率	0.059~0.046	0.033~0.023	0.055~0.046	0.035~0.025	0.046~0.035	0.025~0.018
第三产业 GDP 占比增长率	0.01~0.005	0.003~0.001	0.0122~0.005	0.0053~0.0043	0.005~0.0053	0.0043~0.003
人口总量增长率	0.0046~0.0052	0.0055~0.0058	0.0043~0.0028	-0.0021~-0.005	0.0028~-0.0021	-0.035~-0.05
能源消费增长率	0.00195~0.00196	0.00196~0.00197	0.0286~0.0143	-0.0008~-0.0017	0.0143~-0.0008	-0.0017~-0.0026
化石能源消费总量增长率	-0.01~-0.008	-0.006~-0.05	-0.0169~-0.00127	-0.048~-0.0578	-0.0169~-0.0127	-0.0578~-0.073
用电量增长率	0.086~0.085	0.084~0.083	0.042~0.03	0.02~0.006	0.03~0.02	0.006~0.004

注：碳排放情景设置中数据表示该影响因素的复合增长率取值范围。

● 建立多情景下碳排放量核算模型

①利用复合增长率对 2021-2060 年人口总量(P)、GDP(G)、能源消费量(E)、非化石能源消费量(E_N)、发电量(E_V)、第三产业 GDP 占比(TI)进行推算,见公式(27)到公式(29)以及公式(34)到公式(36)。

②预测 2021-2060 年单位能耗二氧化碳排放量。

由于表 5.8 中的补充数据不包括单位能耗二氧化碳排放量增长率，因此，需要通过预测来获得 2021-2060 年单位能耗二氧化碳排放量数据。考虑到单位能耗二氧化碳排放量与碳排放量的相关性，采用长短时记忆网络(Long Short-Term Memory,LSTM)进行预测。该模型主要用于处理具有时间序列特征的数据，工作机理如图 5.19 所示。该模型的核心是记忆细胞和门机制，记忆细胞允许网络可以学习何时遗忘历史信息以及何时用新信息更新记忆单元，并受门机制控制，即输入门、遗忘门和输出门。

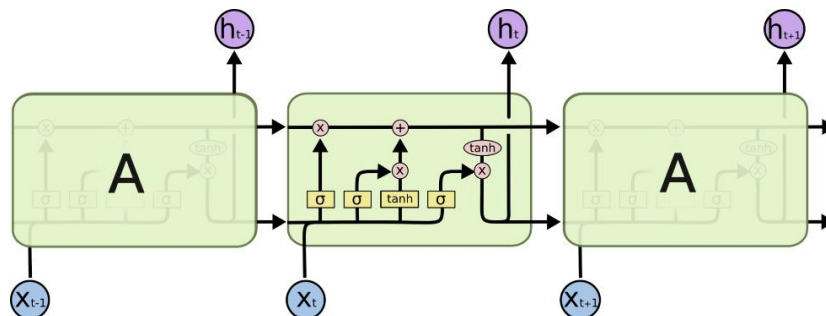


图 5.19 LSTM 模型工作机理

在对 2021-2060 年单位能耗二氧化碳排放量数据进行预测时，需要利用①中推算得出的 2021-2060 年人口总量(P)、GDP(G)、能源消费量(E)、非化石能源消费量(E_N)、发电量(E_V)、第三产业 GDP 占比(TI)数据来对模型进行训练。

③在多情景下，沿用问题二中的碳排放量预测模型，如公式(40)所示，并利用①②算出的数据进行拟合。当碳排放量不再上升，判断达到碳达峰目标；根据假设 5，当非化石能源消费比重大于 80%时，判断达到碳中和目标。

5.3.3 模型的求解

● 第(1)问：情景设计

共设计 3 种情景，包括无人干预的自然情景、按时碳达峰与碳中和的基准情景、率先碳达峰与碳中和的雄心情景。由于情景参数值的设置参照政策文件^[7-9]，情景设计与碳达峰和碳中和的时间节点相关联，满足要求 1。由于情景参数设置中包括单位 GDP 能耗和非化石能源消费比重这两个参数，情景设计与能效提升和非化石能源比重提升相关联，满足要求 2。

● 第(2)问：多情景下碳排放量核算方法

对多情景下碳排放量核算模型进行求解并对比各种回归方法，如图 5.20 所示。以基准场景为例，通过观察可以发现，弹性网络回归过拟合情况严重，Ridge 回归和经过交叉验证的 Ridge 回归随着预测碳排放量会从波动趋于稳定。因此，最终采用 Ridge 回归对多情景下碳排放量模型进行求解。

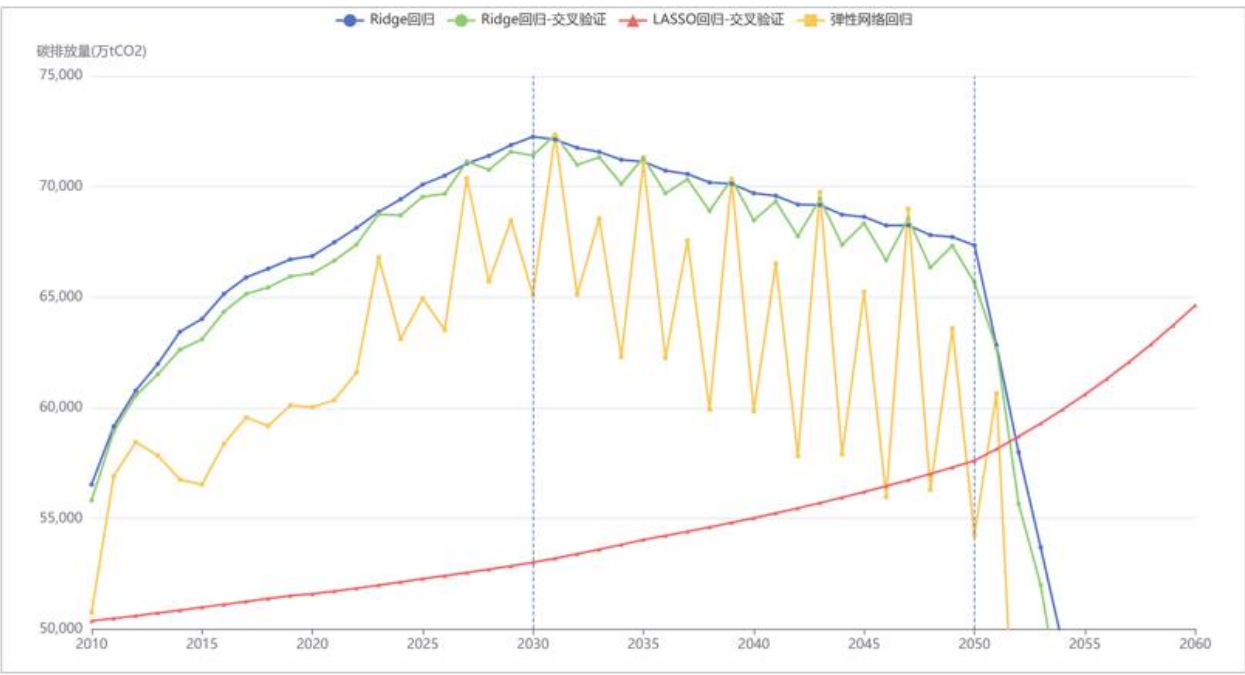


图 5.20 多情景碳排放量核算模型多种回归方法对比（以基准情景为例）

根据多情景碳排放量核算模型的求解结果,绘制多情景碳排放量预测趋势图,如图 5.21 所示。通过观察可以看出,自然情景下碳排放量会持续上升,无法实现碳达峰目标和碳中和;基准情景下,在 2030 年可以实现碳达峰目标,但无法在 2060 年之前实现碳中和目标;雄心情景下,在 2030 年可以实现碳达峰目标并在 2050 年之前实现碳中和目标。

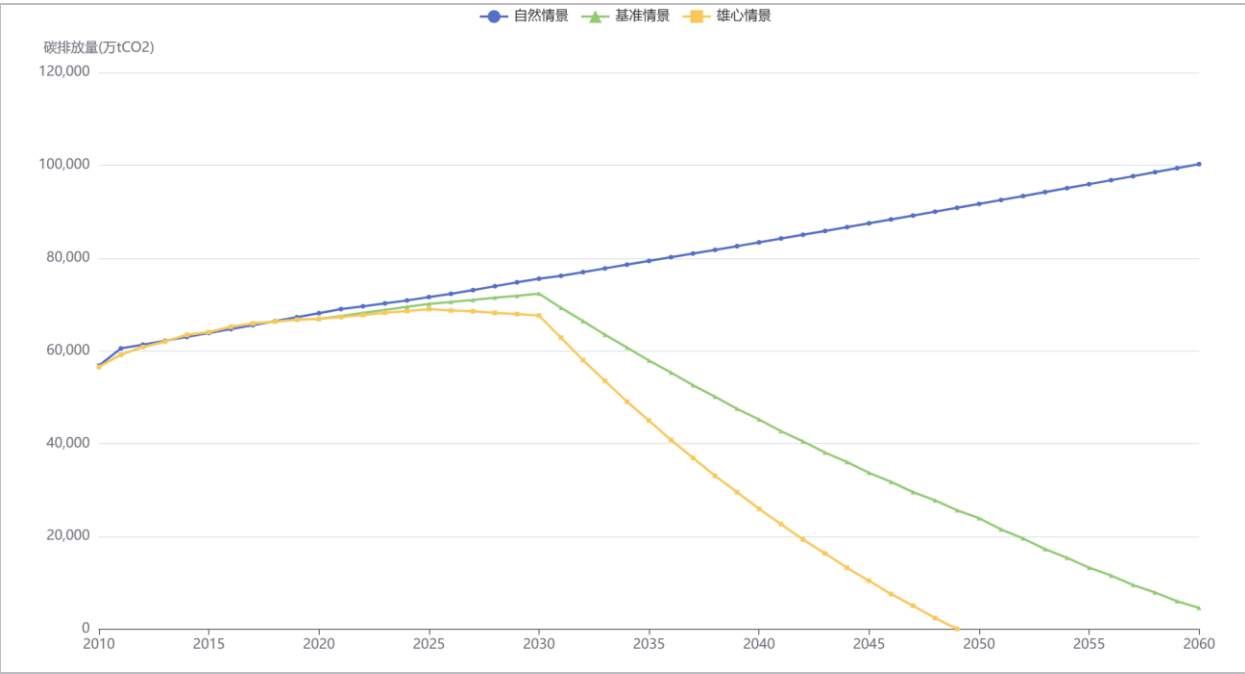


图 5.21 多情景碳排放量预测趋势图

● 第(3)问：确定双碳(碳达峰与碳中和)目标与路径

①对于要求 1 和要求 2，根据情景参数设置表，确定 GDP、人口和能源消费量、单位 GDP 能耗、非化石能源消费比重在 2025 年、2030 年、2035 年、2050 年和 2060 年的取值，结果如表 5.10 所示。

表 5.10 实现双碳指标目标取值

时间	GDP	人口总量	能源消费量	单位 GDP 能耗	非化石能源消费比重
2025	111045.22	8596.60	33751.02	0.30393	0.305344
2030	131886.88	8506.72	33616.23	0.25488	0.45060
2035	156640.24	7118.66	33331.47	0.21278	0.56672
2050	226861.77	4171.65	32491.55	0.14322	0.80837
2060	271168.41	2497.72	31656.59	0.11674	0.90783

③对于要求 3，第一，发展能效工程，即通过推广和采用更为节能的生产技术与工艺，提高单位 GDP 能耗的效率。这可以通过持续的科技创新、能源管理体系的建立、能效标准的制定与强化来实现，从而减少能源消耗，降低碳排放。第二，发展产业升级措施，即促进产业结构升级，大力发展第三产业，将经济的重心转向服务业、高科技产业和文化创意产业等低碳和高附加值领域。政府可通过税收激励、金融支持以及政策引导来鼓励投资于绿色产业，减少传统重工业的比重，提高产业链附加值，发展产业升级工程。能源脱碳：实施措施以减少对高碳化石燃料的依赖。这包括大力发展可再生能源如太阳能、风能和水电，提高其在总能源消耗中的比例。同时，推广低碳能源如核能，采用碳捕获与封存技术来降低高碳燃料的碳排放。第四，发展消费电气化工程，即通过电气化来替代传统的燃料消耗。这可通过推广电动交通、使用电热水器、电暖气等电力供热设备，以及发展智能电网来实现。电力消费比重的提高将减少传统燃料的使用，有助于碳排放的降低。

六、模型的评价、改进与推广

6.1 模型的优点

(1) **多维度分析：**模型综合考虑经济、人口和能源消费量类影响因素的多个指标，建立了较为完整的指标体系，有助于更全面地分析碳排放情况。

(2) **扩展经典 LMDI 和 STIRPAT 模型：**结合碳排放量指标体系，对经典的 LMDI 模型和 STIRPAT 模型进行扩展，使模型更适用于题目场景；量化各指标对碳排放量的贡献，为政策制定提供了科学依据，定量化体现指标目标值。

(3) **多方法对比分析：**在对 STIRPAT 模型进行拟合时，对比多种回归方法，结合实际情况选取拟合效果最好的回归方法——岭回归分析，使得最终预测碳排放量结果更具有可靠性。

(4) **设计三种情景科学分析：**设计自然情景、基准情景和雄心情景，使决策者能够在不同条件下灵活进行决策和规划。情景分析时，结合真实的国家双碳发展规划，使得模型更具可信度和实用性。

(5) **丰富的可视化分析：**通过多种图表对数据进行可视化，可以直观地体现数据的分布情况和发展趋势等。例如，通过碳排放量预测曲线图可以直观地看出碳达峰和碳中和年份。

6.2 模型的缺点

(1) **数据局限性：**题目所给数据只包括 2010-2020 年这 11 年的数据，在进行回归预测

时具有一定的局限性。

(2) 指标确立局限性：题目只讨论经济、人口和能源消费量这 3 类影响因素对碳排放量的影响，指标确立具有一定的局限性。

(3) 情景设计局限性：设计情景时，只考虑自然情景、基准情景和雄心情景这 3 大类常见场景，未对其进行进一步的分类。

6.3 模型的改进

(1) 针对第一个缺点：由于模型的有效性依赖于可用数据的质量和完整性，可以自行收集 2010 年以前的数据对题目原始数据进行补充，从而提高模型预测的准确性。

(2) 针对第二个缺点：建立指标体系时可以考虑引入更多可能影响碳排放的因素，以提升模型的准确性。

(3) 针对第三个缺点：设计更为细致的情景分类，建立更为完善的多场景碳排放量预测模型。

6.4 模型的推广

(1) 扩展模型应用地区：将该模型推广到其他地区，用于不同地区碳排放情况的分析预测以及双碳目标路径的科学化规划。政府和决策者可以根据该模型的分析结果，制定更具针对性的碳减排政策，从而推动促进双碳目标的实现。

(2) 扩展模型应用场景：在科研领域中对模型进行推广，将其应用于碳排放量预测、双碳目标路径规划等研究，为环境保护领域提供更多实证研究支持。科研工作者可以在本模型的基础上进行优化改进，从而推动相关领域的科研进展。

七、参考文献

- [1]袁学良,杨月,盛雪柔等.碳达峰碳中和政策解析与对策建议[J/OL].山东大学学报(工学版):1-11[2023-09-22].<http://kns.cnki.net/kcms/detail/37.1391.T.20230613.1729.004.html>
- [2]王明喜,胡毅,郭冬梅等.低碳经济:理论实证研究进展与展望[J].系统工程理论与实践,2017,37(01):17-34.
- [3]中华人民共和国工业和信息化部.中华人民共和国国家发展和改革委员会.中华人民共和国生态环境部.关于印发工业领域碳达峰实施方案的通知[EB/OL].(2022-08-01)[2023-9-22]. https://www.gov.cn/zhengce/zhengceku/2022-08/18/content_5705865.htm
- [4]彭军霞,李新爱,梅玉坤.基于 LMDI 模型的城市碳排放影响因素分析[J].能源与节能,2023(07):61-64.DOI:10.16643/j.cnki.14-1360/td.2023.07.013.
- [5]Xiang X , Ma X , Ma Z ,et al.Python-LMDI: A tool for index decomposition analysis of building carbon emissions[J]. 2022.DOI:10.13140/RG.2.2.16088.21767.
- [6]刘茂辉,翟华欣,刘胜楠等.基于 LMDI 方法和 STIRPAT 模型的天津市碳排放量对比分析[J].环境工程技术学报,2023,13(01):63-70.
- [7]全球能源互联网发展合作组织. 中国 2030 年前碳达峰研究报告[M]. 中国电力出版社:, 202103.
- [8]全球能源互联网发展合作组织. 中国 2060 年前碳中和研究报告[M]. 中国电力出版社:, 202103.
- [9]全球能源互联网发展合作组织. 中国 2030 年能源电力发展规划研究及 2060 年展望[M]. 中国电力出版社:, 202103.
- [10]黄蕊,王铮,丁冠群等.基于 STIRPAT 模型的江苏省能源消费碳排放影响因素分析及趋势预测[J].地理研究,2016,35(04):781-789.

[11]郭海兵,孟晨.基于 STIRPAT 模型的碳排放预测——以连云港市为例[J].城市建设理论研究(电子版),2023(23):196-198.DOI:10.19569/j.cnki.cn119313/tu.202323065.

附录

附录 1
支撑材料列表
<div><div>└─问题一</div><div><div>└─代码</div><div><div>LMDI计算各部门.m</div><div>pearson.py</div></div></div><div><div>└─数据</div><div><div>原始总结所有数据.xlsx</div><div>各能源消费部门的能源消费品种结构.csv</div><div>各部门-碳排放量-描述性统计量.xlsx</div><div>各部门的碳排放量状况.csv</div><div>指标同比数据.xlsx</div><div>某区域的经济、人口、能源消费量和碳排放量的总体状况.csv</div></div><div><div>└─各部门LMDI</div><div><div>交通消费部门.xlsx</div><div>农林消费部门.xlsx</div><div>居民生活消费.xlsx</div><div>工业消费部门.xlsx</div><div>建筑消费部门.xlsx</div><div>能源供应部门.xlsx</div></div></div></div><div><div>└─问题三</div><div><div>└─代码</div><div><div>碳排放预测.ipynb</div></div><div><div>└─数据</div><div><div>各种情景下的假设数据.xlsx</div><div>基准情景目标值.xlsx</div><div>自然情况.xlsx</div><div>雄心情景目标值.xlsx</div></div></div></div><div><div>└─问题二</div><div><div>└─代码</div><div><div>STIRPAT模型.ipynb</div></div><div><div>└─数据</div><div><div>各部门碳排放量相关数据分析.xlsx</div><div>基准情景下的假设数据.xlsx</div><div>总数据.xlsx</div></div></div></div></div></div></div>

└─各部门能源消费种类与发电种类对排放量

交通消费部门.xlsx

农林消费部门.xlsx

居民生活消费.xlsx

工业消费部门.xlsx

建筑消费部门.xlsx

能源供应部门.xlsx

附录 2

问题一各部门碳排放量的 LDMI 模型求解代码

```
%load 建筑消费.mat
%load 交通部门.mat
%load 工业消费.mat
%load 农林消费.mat
%load 能源供应部门.mat
load 居民消费.mat
load Y.mat

% 数据定义，Y是观测数据，X是因素矩阵
% 矩阵Y和X，Y包含一系列年份的数据，X包含多个参数
Years = (2010 : 2020)'; % 创建包含年份的向量

dertX = []; % 创建一个空矩阵来存储计算结果
for i = 2 : length(Y)
    curdertX = []; % 用于存储当前参数的计算结果
    for j = 1 : size(X, 2)
        % 计算dert值
        curdert = (Y(i) - Y(1)) / (log(Y(i) / Y(1))) * log(X(i, j) / X(1, j));

        % 处理0值
        if Y(i) * Y(1) * X(i, j) * X(1, j) == 0
            curdert = 0;
        end
        if Y(i) > 0 && Y(1) == 0 && X(i, j) > 0 && X(1, j) == 0
            curdert = Y(i);
        end
        if Y(i) == 0 && Y(1) > 0 && X(i, j) == 0 && X(1, j) > 0
            curdert = -Y(1);
        end

        curdertX = [curdertX, curdert]; % 将当前计算结果添加到curdertX中
    end
end
```

```

    dertX = [dertX; curdertX];% 将当前年份的结果添加到dertX矩阵中
end
% 将结果存储在outresult矩阵中，包括年份和dertX值
outresult = [Years(2 : length(Years)), dertX];% 创建包含结果的矩阵

```

附录 3

问题一各部门碳排放量影响指标之间的皮尔逊相关性模型求解代码

```

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import matplotlib.colors as colors
rc = {'font.sans-serif': 'SimHei',
      'axes.unicode_minus': False}
sns.set(font_scale=0.7,rc=rc) # 设置字体大小
# 设置部门变量地址
base_config = '建筑消费部门'
# 读取部门数据
data_pearson = pd.read_excel(f'/home/kwk/math_model/pearson/{base_config}.xlsx')
# 计算斯皮尔曼相关性系数矩阵
pearson_corr = data_pearson.corr(method='pearson')
# 定义我的指标的颜色区间
bins = [-1, -0.8, -0.5, -0.3, 0, 0.3, 0.5, 0.8, 1]
# 定义每个区间的颜色
color_list = ["#073067", "#125CA3", "#6AABD1", "#A2CBE1", "#A2CBE1", "#6AABD1", "#125CA3",
              "#073067"]
cmap2 = colors.ListedColormap(color_list)
# BoundaryNorm对象会根据bins和N来将数据值映射到颜色索引，然后再根据颜色映射来确定颜色。
norm = colors.BoundaryNorm(bins, cmap2.N)
sns.heatmap(pearson_corr,
            annot=True, # 显示相关系数的数据
            fmt='.2f', # 只显示两位小数
            linewidth=0.5, # 设置每个单元格的距离
            square=True, # 每个方格都是正方形
            cmap= cmap2, # 设置热力图颜色
            norm= norm,
            )
plt.savefig(f"{base_config}-pearson.png",dpi=800, bbox_inches = 'tight')#保存图片，分辨率为600
plt.ion() #显示图片

```

附录 4

问题二能源消费量与人口总量、GDP 的多元线性回归模型求解代码

```

# 引入pandas库处理数据
import pandas as pd

# 从scikit-learn库引入线性回归模型
from sklearn.linear_model import LinearRegression

# 从scikit-learn库导入线性回归模型
from sklearn.linear_model import LinearRegression
# 导入R平方值计算函数系数及其显著性
from sklearn.metrics import r2_score
# 为了简化，我们将使用statsmodels库中的标准误差和t值
import statsmodels.api as sm

# 读取数据
data = pd.read_excel('/home/kwk/math_model/Q3/总数据.xlsx', sheet_name='总数据')

# 选择需要的指标
filter_columns = ['年份', '人口总量', '人均GDP', '能耗/总能源消费']
cal_data = data[filter_columns]
cal_data['生产总值'] = cal_data['人口总量'] * cal_data['人均GDP']

# 获取需要的数据，X是特征（人口总量和GDP总量），y是目标变量（能源消耗量）
X = cal_data[['人口总量', '生产总值']]
y = cal_data['能耗/总能源消费']

# 创建一个线性回归模型
linear = LinearRegression().fit(X, y)

# 获取线性回归方程的系数
# 线性回归方程的一般形式：  $y = \alpha * x_1 + \beta * x_2 + \gamma$ 
# 其中，alpha和beta是特征的系数，gamma是截距
alpha = linear.coef_ # 获取特征的系数
gamma = linear.intercept_ # 获取截距

# alpha表示人口总量的系数，beta表示GDP总量的系数，gamma表示常数截距
alpha, gamma

# 计算R平方值
r2 = r2_score(y, predicted_values)

from SALib.sample import saltelli
from SALib.analyze import sobol
from SALib.test_functions import Ishigami
import numpy as np

```



```

num_vars = len(X.columns)
bounds = []
for v in X.columns:
    bounds.append([min(X[v]), max(X[v])])

problem = {
    'num_vars': num_vars,
    'names': X.columns,
    'bounds': bounds
}
param_values = saltelli.sample(problem, 1024)
Y = np.zeros([param_values.shape[0]])
for i, x in enumerate(param_values):
    Y[i] = linear.predict([x])

Si = sobol.analyze(problem, Y)

# 输出结果
print(Si['S1'])
print(Si['ST'])

```

附录 5

问题二区域碳排放量的 STIRPAT 预测模型求解代码

```

from sklearn.linear_model import Lasso, Ridge
import pandas as pd

data_Y = pd.read_excel('homekwkmath_modelQ3各部门碳排放量相关数据分析.xlsx', sheet_name='总数')
data_Y.head()

data_X = pd.read_excel('homekwkmath_modelQ3总数据.xlsx', sheet_name='总数据')
X = data_X
del X['年份']
del X['非化石能源消费比重倒数']
del X['非化石能源发电占比']
Y = data_Y['碳排放']

# 岭回归
rg = Ridge()
rg.fit(X, Y)
rg.score(X, Y)

```

```

# lasso
ls = Lasso()
ls.fit(X, Y)
ls.score(X,Y)

# 弹性网络
from sklearn.linear_model import ElasticNet
from sklearn.linear_model import ElasticNet

# 初始化弹性网络回归器
reg = ElasticNet(alpha=0.001, l1_ratio=0.5, fit_intercept=True)
# 拟合线性模型
reg.fit(X, Y)
# 评分
reg.score(X,Y)

# 多元线性回归
# 从scikit-learn库导入线性回归模型
from sklearn.linear_model import LinearRegression

# 创建一个线性回归模型
linear = LinearRegression().fit(X, Y)

# lasso-交叉验证
from sklearn.linear_model import LassoCV, RidgeCV
lsCV = LassoCV().fit(X,Y)
lsCV.score(X,Y)

# ridge-交叉验证
rgCV = RidgeCV(cv=5).fit(X,Y)
rgCV.score(X,Y)

# 岭回归下的岭迹图

X = X[X.columns].apply(np.log1p)
# 计算横坐标
n_alphas = 200
alphas = np.logspace(-10,-2,n_alphas)

# 建模。获取每一个正则化去之下的系数组合
coefs = []
for a in alphas:
    ridge = linear_model.Ridge(alpha=a,fit_intercept=False)
    ridge.fit(X,np.log1p(Y))

```

```
coefs.append(ridge.coef_)
```

```
# 可视化
```

```
# 当前的图表和子图可以使用plt.gcf()和plt.gca()获得，分别表示Get Current Figure和Get Current Axes。
```

```
# 在pyplot模块中，许多函数都是对当前的Figure或Axes对象进行处理，比如说：plt.plot()实际上会通过plt.gca()获得当前
```

```
# 的Axes对象ax，然后再调用ax.plot()方法实现真正的绘图。
```

```
ax = plt.gca()
```

```
ax.plot(alphas,coefs,label=X.columns)
```

```
ax.set_xscale('log') # 将 y轴 和 x轴 的比例设置为对数比例
```

```
ax.set_xlim(ax.get_xlim()[::-1]) # 将横坐标逆转，再设置坐标轴尺度范围
```

```
plt.xlabel('正则化参数alpha')
```

```
plt.ylabel('系数w')
```

```
# 显示legend在正右方
```

```
plt.legend(loc = 'center right')
```

```
plt.title('岭回归下的岭迹图')
```

```
plt.axis('tight')
```

```
plt.show()
```

附录 6

问题二模型单位能耗二氧化碳排放量的 LSTM 预测模型求解代码

```
# 导入必要的库
```

```
import numpy as np
```

```
from keras.models import Sequential
```

```
from keras.layers import Dense, LSTM
```

```
import pandas as pd
```

```
import os
```

```
# 设置使用1卡
```

```
os.environ['CUDA_VISIBLE_DEVICES'] = '1'
```

```
# 读取数据，筛选出需要的指标
```

```
data= pd.read_excel('/home/kwk/math_model/Q3/总数据.xlsx', sheet_name='总数据')
```

```
# del data['年份']
```

```
del data['非化石能源消费比重倒数']
```

```
del data['非化石能源发电占比']
```

```
data['人均GDP'] = data['人均GDP']
```

```
# 目标变量
```

```
target_columns = ["单位能耗二氧化碳排放量"]
```

```
# 输入特征
```

```
feature_columns = [
```

```

"人口总量",
"人均GDP",
"单位GDP能耗",
"电力消费比重",
"非化石能源消费比重",
"第三产业占比"
]

# 提取所选的特征列和目标列
filter_data = data[["年份"] + feature_columns + target_columns]

filter_data[feature_columns] = filter_data[feature_columns].apply(np.log1p)
filter_data[target_columns] = filter_data[target_columns].apply(np.log1p)

# 导入必要的库
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler

# 使用MinMaxScaler对特征进行归一化
MM_Scaler_X = MinMaxScaler(feature_range=(0, 1))
MM_Scaler_Y = MinMaxScaler(feature_range=(0, 1))

# 定义测试集的大小
test_size = 3
train_size = len(filter_data) - test_size

# 将数据分为训练集和测试集
train_data = filter_data.iloc[:train_size]
test_data = filter_data.iloc[train_size:]

# 分离特征和目标变量
X_train = train_data.drop(target_columns+['年份'], axis=1) # 训练集特征
y_train = train_data[target_columns] # 训练集目标变量
X_test = test_data.drop(target_columns+['年份'], axis=1) # 测试集特征
y_test = test_data[target_columns] # 测试集目标变量

# 使用MinMaxScaler对特征和目标变量进行归一化
X_Train_Scaled = MM_Scaler_X.fit_transform(X_train) # 归一化训练集特征
X_test_scaled = MM_Scaler_X.transform(X_test) # 归一化测试集特征
y_train_scaled = MM_Scaler_Y.fit_transform(y_train) # 归一化训练集目标变量
y_test_scaled = MM_Scaler_Y.transform(y_test) # 归一化测试集目标变量

# 将输入数据调整为LSTM模型所需的形状 [samples, time steps, features]
X_train_resaped = X_Train_Scaled.reshape(X_Train_Scaled.shape[0], 1, X_Train_Scaled.shape[1]) #

```

```

调整训练集特征形状
X_test_reshaped = X_test_scaled.reshape(X_test_scaled.shape[0], 1, X_test_scaled.shape[1]) # 调整测试集特征形状

# 定义LSTM模型
model = Sequential() # 创建一个序贯模型

# 添加LSTM层，110个神经元，输入形状为(input_shape)，这里是(1, 特征数量)
model.add(LSTM(110, input_shape=(X_train_reshaped.shape[1], X_train_reshaped.shape[2])))

# 添加输出层，一个神经元，用于回归问题
model.add(Dense(1))

# 编译模型，指定优化器和损失函数
model.compile(optimizer='adam', loss='mean_squared_error')

# 训练模型
# X_train_reshaped: 训练数据特征
# y_train_scaled: 训练数据目标变量
# epochs: 迭代次数
# batch_size: 每个批次的样本数
# validation_data: 验证数据，用于评估模型性能
# verbose: 显示训练过程的详细程度（1表示显示进度条，0表示不显示）
# shuffle: 是否在每个时期之前打乱数据
history = model.fit(X_train_reshaped, y_train_scaled, epochs=100, batch_size=1, validation_data=(X_test_reshaped, y_test_scaled), verbose=1, shuffle=False)

# 使用模型进行预测
y_pred_scaled = model.predict(X_test_reshaped) # 预测结果在归一化的范围内

# 将归一化的预测结果转换回原始比例
y_pred = MM_Scaler_Y.inverse_transform(y_pred_scaled) # 使用逆变换将结果还原

# 定义MAPE（平均绝对百分比误差）函数
def mape(actual, pred):
    actual, pred = np.array(actual), np.array(pred)
    return np.mean(np.abs((actual - pred) / actual)) * 100

# 计算MAPE
mapes = mape(y_test.values, y_pred)

# 打印MAPE值
print(f"Mean Absolute Percentage Error (MAPE): {mapes}")

```

```

# 使用所有数据进行训练
X_all = filter_data.drop(target_columns+['年份'], axis=1) # 所有数据的特征
y_all = filter_data[target_columns] # 所有数据的目标变量

# 对数据进行归一化处理
X_all_scaled = MM_Scaler_X.fit_transform(X_all) # 归一化特征
y_all_scaled = MM_Scaler_Y.fit_transform(y_all) # 归一化目标变量

# 调整数据形状以适应LSTM模型
X_all_reshaped = X_all_scaled.reshape(X_all_scaled.shape[0], 1, X_all_scaled.shape[1])

# 继续训练之前定义的LSTM模型
model.fit(X_all_reshaped, y_all_scaled, epochs=100, batch_size=1, verbose=1)

```

附录 6

问题二模型单位能耗二氧化碳排放量的 LSTM 预测模型求解代码