

基于回归树集成学习方法的工业增长预测和分析

陈磊^{1,2}, 李丽娟¹

(1. 东北财经大学经济学院, 大连 116025; 2. 东北财经大学经济计量分析与预测研究中心, 大连 116025)

摘要 本文从众多变量中筛选出 59 个相关经济指标, 分别考查疫情前后传统时间序列模型和几种回归树集成学习模型对中国工业增加值增速的预测效果, 并结合 Shapley additive explanations (SHAP) 方法对相关预测变量的作用进行解释分析。研究发现, 随着预测步长的增加和新冠疫情的暴发, 传统时间序列模型的预测性能明显减弱, 而集成学习模型的预测表现则相对较好, 其中梯度提升树模型在较长预测步长中更加稳健和准确。基于 SHAP 方法的分析发现, 作为预测变量的经济指标在不同时期的重要性有所不同, 除生产、投资等指标外, 金融类变量在高风险时期也具有一定的预测作用, 需结合具体时间和预期目标来选择合适的经济指标进行工业增长预测。基于预测的视角可在一定程度上说明新冠疫情冲击可能不会改变工业增长未来走势的基本面。

关键词 工业增加值预测; 回归树集成学习; Shapley additive explanations (SHAP) 方法; 梯度提升树模型

Forecasting and Analysis of Industrial Growth Based on Regression Tree Integrated Learning Methods

CHEN Lei^{1,2}, LI Lijuan¹

(1. School of Economics, Dongbei University of Finance and Economics, Dalian 116025, China; 2. Center for Econometric Analysis and Forecasting, Dongbei University of Finance and Economics, Dalian 116025, China)

Abstract The paper selects 59 related economic indicators from numerous variables, and compares the forecast effects of traditional time series models with several regression tree integrated learning models on the growth rate of China's industrial added value under different scenarios, and the Shapley additive explanations (SHAP)

收稿日期: 2023-06-16

基金项目: 辽宁省社会科学规划基金重点建设学科项目 (L22ZD054)

Supported by Key Construction Discipline Project for Social Science Planning Foundation of Liaoning Province (L22ZD054)

作者简介: 陈磊, 教授, 博士, 研究方向: 宏观经济分析与预测、计量经济学, E-mail: chenlei@dufe.edu.cn; 李丽娟, 博士研究生, 研究方向: 宏观经济分析与预测、大数据与机器学习, E-mail: 15232193817@163.com.

method is combined for interpretation. Our results show that, with the increase of the forecast step and the outbreak of the COVID-19 epidemic, the forecast performance of the traditional time series model is significantly weakened, while the integrated learning model is relatively better, among which the gradient boosting decision tree model is more robust and accurate in the longer forecast step. Based on the analysis of SHAP method, we find that the importance of economic indicators as predictors in different periods is different. In addition to indicators such as production and investment, financial variables also play a certain early warning role in high-risk periods, and appropriate economic indicators should be selected according to specific time and expected goals for industrial growth forecasting and analysis. From the perspective of forecasting, the impact of the COVID-19 pandemic may not change the fundamentals of the future trend of industrial growth to some extent.

Keywords industrial value added forecasting; regression tree integrated learning; Shapley additive explanations (SHAP) method; gradient boosting decision tree model

1 引言

2020年伊始暴发的新冠疫情不仅对我国和世界经济造成了严重冲击,也对宏观经济监测和预测提出了新挑战和新要求。《中共中央 国务院关于新时代加快完善社会主义市场经济体制的意见》明确提出“强化经济监测预测预警能力”,并将其作为构建宏观调控新机制的重要组成部分。在百年未有之大变局和构建新发展格局的新形势下,提高宏观经济监测预测的准确性对保证经济在合理区间高质量发展的重要性日益凸显。

宏观经济监测预测通常以国内生产总值(简称GDP)增速为核心,但其公布频率低且存在一定发布时滞,不利于及时了解宏观经济走势,在经济遭遇重大冲击从而具有较大不确定性时尤其如此。因此,考虑到GDP增速在宏观经济监测预测方面的不足,本文重点关注月度工业增加值增速这一指标。众所周知,工业对国民经济发展和全面建设社会主义现代化强国具有不可替代的重要作用。2023年政府工作报告明确指出“任何时候中国都不能缺少制造业”。从各行业增加值占比来看,2022年我国工业增加值占GDP比重最高,为33.2%¹,工业生产增长的变化对经济周期波动具有较大影响。因此,月度工业增加值增速(或工业生产指数)作为衡量工业增长的核心指标,在大多数国家的经济景气监测以及GDP的实时和短期预测中均是一个至关重要的变量(Golinelli and Parigi, 2007; Baumeister and Guérin, 2021; 陈诗一和于鸿宝, 2021)。因此,准确及时地预测工业增速对把握经济形势和政府的宏观调控均具有重要意义(顾光同和许冰, 2018)。

虽然工业增加值增速对经济运行反映迅速,但其易受季节性和各种冲击的影响而难以预测(Bruno and Lupi, 2004; Dendramis et al., 2020; Galdi et al., 2023),因此选择合适的方法对其预测十分必要。尤其在面对百年变局的复杂环境和世纪疫情的异常冲击时,传统时间序列模型和因子模型等部分计量方法越来越表现出一定局限性。近年来,一些学者尝试将新

¹数据来自于国家统计局和华经产业研究院。其中,批发和零售业、金融业、房地产业、建筑业、农林牧渔业、信息传输、软件和信息技术服务业、交通运输、仓储和邮政业、租赁和商务服务业、住宿和餐饮业及其他行业的增加值占比分别为9.5%、8.0%、6.1%、6.9%、7.7%、4.0%、4.1%、3.2%、1.5%和15.9%。

的方法用于宏观经济预测,其中,基于大数据背景的机器学习方法显示出了一定预测优势和运算潜力 (Varian, 2014; 洪永淼和汪寿阳, 2021)。一方面,传统的线性模型难以适应复杂多变的经济环境,机器学习方法则提供了非线性的建模方案,它们可以捕捉预测变量和目标变量之间的复杂关系 (Yoon, 2021; Coulombe et al., 2022; Buckmann et al., 2021); 另一方面,机器学习方法在预测精度和稳健性方面均具有一定优势,尤其是当复杂性增加或发生异常冲击等结构性突变时更是如此 (Mullainathan et al., 2017; Chuku et al., 2019; Vrontos et al., 2021; Coulombe et al., 2021)。

很多研究表明,在众多机器学习模型中,以回归树为基础的集成学习方法 (包括装袋法、随机森林和梯度提升树等模型) 在处理复杂问题时更有优势,相比于深度学习方法,其运算速度更快,参数调优更具稳健性 (Januschowski et al., 2022; Godahewa et al., 2023), 因此,越来越受到学者们的青睐。在公司管理和预测方面,马晓君等 (2019) 使用一种优化后的加权随机森林算法对上市公司的信用评级情况进行分析,其评级准确率明显提高;周卫华等 (2022) 将极度梯度提升树模型应用到了上市公司的财务舞弊分析中,取得了较好的预测结果。在投资和金融分析方面, Yang et al. (2020) 使用决策树、自适应提升树、梯度提升树和极度梯度提升树等多种集成学习方法研究股市的收益与风险,研究表明梯度提升树模型的预测准确性更高;张鹏等 (2022) 通过比较多种集成学习方法,发现极度梯度提升树模型在投资组合优化模型中表现最优。在宏观经济预测领域,集成学习模型同样表现突出。其中,随机森林模型在经济衰退预警方面有较大潜力 (Alessi and Detken, 2011; Nyman and Ormerod, 2017); Gawthorpe (2021) 的研究表明使用随机森林模型对捷克经济的预测效果优于银行家的预判;熊景华和茹璟 (2021) 使用改进的随机森林模型对汇率做了更精准的预测;Yoon (2021) 对日本的 GDP 进行预测,研究发现梯度提升树模型和随机森林模型的预测精度均高于传统模型,且前者的预测效果更好;Aras and Lisboa (2022) 的研究发现,在波动性和不确定性较大的新兴国家 (如土耳其),基于回归树的集成学习模型可以较准确地预测其通胀情况。因此,本文重点关注此类模型在宏观经济方面的预测效果。

虽然集成学习模型的预测准确率相对较高,但是其结构复杂且不透明,解释能力有限,难以让人信服。因而,如何揭示此类机器学习模型背后的经济逻辑是学者们一直致力解决的问题。在对此问题的研究中,SHAP 方法因其优良的统计性质而备受关注 (Štrumbelj and Kononenko, 2010), 并不断用于各种机器学习模型的解释分析中 (肖争艳等, 2022; Tschora et al., 2022), 为打开机器学习黑箱提供了有益参考。因此,本文借鉴该方法尝试对集成学习模型进行解释,以深入分析影响工业增长的主要因素,从而更好地把握宏观经济运行的基本态势。

本文基于筛选出的 59 个月度经济指标,比较了几种回归树集成学习模型在新冠疫情前后对月度工业增加值增速的预测效果,并试图解释模型背后的经济逻辑。本文的边际贡献主要体现在以下三方面:第一,探究回归树集成学习模型在宏观经济预测方面是否具有优势。研究发现并不是所有此类模型均优于传统模型,且预测效果随预测步长的改变而有所差别;第二,分析集成学习模型的预测优势是否与所处经济环境有关。研究表明在疫情前,集成学习模型预测的优势不甚明显,但当扩大样本区间至新冠疫情后,其预测效果则明显提高;第三,试图解释模型背后的逻辑。本文将 SHAP 方法应用到集成学习模型中并提出了指标相对贡献率的测算方法,可据此分析预测变量在不同时期的重要性。

论文的余下结构安排如下: 第二部详细介绍了所用数据和相关模型设定; 第三部分是实证结果, 具体分析了疫情前后各模型的预测效果; 第四部分基于 SHAP 方法对模型进行解释分析; 第五部分为研究结论.

2 数据及模型设定²

2.1 指标筛选与数据处理

本文所用数据的样本区间为 2000 年 1 月 - 2022 年 6 月, 共 270 个月, 主要来自于中国经济信息网和 Wind 数据库. 经济系统复杂交错, 存在高度的非线性和不确定性, 各经济指标之间也相互影响, 因此可充分利用这些指标之间的经济关系进行预测分析. 然而, 使用过多的指标并不一定能提高预测性能, 反而会因为虚假关联关系而增加噪音 (Meng, 2018; Calude and Longo, 2017). 因此, 本文以规模以上工业增加值增长率为预测目标, 根据经济系统构成和运行机制, 通过时差相关性分析选取该指标的先行和一致指标, 并舍弃了相关系数绝对值小于 0.5 的指标, 最终从 129 个变量中选择了符合条件的 59 个, 共包括企业经营、宏观经济和金融三个领域. 其中, 企业经营包括效益类的收入和利润两个指标, 宏观经济包括投资、消费、财政、贸易、房地产、物价、生产、能源、流通共九类, 金融包括汇率、货币信贷和股票三类, 具体见表 1.

在数据处理方面, 针对部分指标存在 1 月份数据缺失和春节因素问题, 本文用 1 至 2 月的累计增长率数据代替 1 月和 2 月的当月增速; 对于部分指标个别月份数据的缺失值使用该指标年内平均值代替; 对部分企业经营类数据在 2000-2010 年时期存在的较多缺失值采用插值方法进行填补. 另外, 本文针对所选取的经济指标统一计算其同比增长率或直接采用公布的同比增速³.

因为不同经济指标的变动在时间上通常具有一定的超前滞后性, 某一个经济指标的变动不但可能受到其他指标当期和若干滞后期的影响, 也可能受到自身之前变动的影响, 所以在进行预测时, 本文加入了被预测变量和预测变量滞后 1~3 期的历史数据 (Medeiros et al., 2021), 时差相关性分析表明, 绝大部分指标在超过滞后 3 期后与被预测变量的相关性小于 0.6, 个别指标在滞后 12 期仍具有较大相关性, 但是考虑到样本容量, 集成学习模型中的预测变量只选择滞后 1~3 期, 即 \mathbf{x}_t 可表示为 $\mathbf{x}_t = \{y_{t-i}, x_{1t-i}, x_{2t-i}, \dots, x_{mt-i}\}$ ($i = 0, 1, 2, 3; m = 59$), 其中 x_{mt-i} 为具体的经济指标, 共计 240 个预测变量, 在传统计量模型中预测变量的滞后阶数依据 BIC 准则确定.

2.2 模型介绍

考虑以下的一般性预测模型:

$$y_{T+h} = F_h(\mathbf{x}_t) + u_{T+h}, \quad h = 1, \dots, H, \quad (1)$$

²本文所使用的数据和代码请参见科学数据银行 (ScienceDB) 期刊社区, DOI: 10.57760/sciencedb.j00214.00017 和 CSTR: 31253.11.sciencedb.j00214.00017. 若使用文中数据信息, 请注明引文和数据出处.

³其中, 同比指数类指标未做增长率处理, 比如, 国房景气指数、居民消费价格指数等.

表1 预测指标分类

领域	分类	数量	细分指标
企业经营	效益	2	营业收入 (OR) 利润总额 (profit)
	投资	3	固定资产投资完成额 (inv) 第二产业固定资产投资 (inv2) 第三产业固定资产投资 (inv3)
	消费	1	社会消费品零售总额 (cons)
	财政	2	公共预算收入 (income) 税收收入 (tax)
	贸易	7	进口额 (import) 出口额 (export) 进出口总额 (im-ex) 出口商品价值增长指数 (VGIE) 进口商品价值增长指数 (VGII) 出口商品数量增长指数 (QGIE) 进口商品数量增长指数 (QGII)
	房地产	7	国房景气指数 (CERCI) 商品房销售额 (HSV) 商品房销售面积 (HSA) 新增固定资产投资 (NREI) 开发投资 (CREI) 开发计划总投资 (PREI) 本年新开工房屋施工面积 (NHC)
	物价	2	居民消费价格指数_衣着 (CPI_C) 商品零售价格指数_服装、鞋帽 (RPI_C)
	生产	15	水泥产量 (cement) 发动机产量 (engine) 焦炭产量 (coke) 生铁产量 (iron) 工业锅炉产量 (boiler) 汽车产量 (car) 饮料产量 (drink) 饲料产量 (fodder) 纯碱产量 (soda) 布产量 (cloth) 磷矿石产量 (phosO) 化学纤维产量 (fiber) 铁矿石产量 (ironO) 塑料制品产量 (plastics) 十种有色金属产量 (metal)
	能源	2	发电量 (electr) 原煤产量 (coal)
	流通	6	货物周转量 (TOV) 民航货物周转量 (TOVF) 货运量 (FV) 民航货运量 (FVF) 港口货物吞吐量 (port) 客运量 (passer)
宏观经济	汇率	5	外汇储备 (FER) 港币对人民币汇率 (H2R) 美元对人民币汇率 (D2R) 实际有效汇率指数 (RERI) 名义有效汇率指数 (NERI)
	金融	6	M1 M2 单位活期存款 (BCD) 单位定期存款 (BTD) 个人存款 (PTD) 人民币存款余额 (RMBD)
	股票	1	深证综指 (SZCI)

其中, y_{T+h} 是在时刻 $h+T$ 的工业增加值增速, $\mathbf{x}_t = (x_{1t}, \dots, x_{nt})'$ 是 n 维的特征变量⁴, H 是最大预测步长, T 为训练模型时对应的样本容量, 随预测步长 (h) 增加而减小, $F_h(\cdot)$ 表示特征变量与未来工业增加值之间的函数关系, 并根据预测步长 (h) 的改变而变化, u_{T+h} 是零均值随机误差.

本文采用的直接预测方程如下:

$$\hat{y}_{T+h|T} = \hat{F}_{h,T-R_h+1:T}(\mathbf{x}_t), \quad T \geq R_h, \quad (2)$$

其中, $\hat{F}_{h,T-R_h+1:T}(\mathbf{x}_t)$ 是根据 $T-R_h+1$ 到 T 的数据估计得到的函数, 且随模型估计样本的不同而需不断重新估计. R_h 是模型训练的窗口大小, 其大小由模型的预测步长 (h) 和特征变量的滞后阶数 (p) 决定. 具体而言, 预测步长 (h) 越长, 在训练数据集中损失的信息越多, R_h 则越小; 模型中特征变量的滞后阶数 (p) 越多, 在训练数据集中可使用的样本越少, 则 R_h 越小, 具体数值可参考下文的实证部分.

因未对特征变量进行预测, 本文参考 Medeiros et al. (2021) 采取直接预测 (direct fore-

⁴此处的特征变量与上文提到的预测变量等价, 如未特别说明, 两者在本文中不做区分.

cast) 方法⁵, 其预测误差不会随着预测步长的增加而累积, 且更适用于机器学习方法建模. 在模型训练时, 本文采用固定窗口的滚动预测, 窗口大小取决于训练数据集中观测到的样本数目和预测步长, 即在不同的预测情景下, 窗口大小略有差异. 本文选择固定窗口的滚动预测主要是基于以下两个原因: 一是为了减弱潜在的结构突变和异常值的影响, 二是可以更好地检验预测性能. 例如, 在滚动窗口预测中, 无条件的 Giacomini-White (GW) 检验相当于传统的 Diebold-Mariano (DM) 检验, 具体检验过程可参考 Giacomini and White (2006).

2.2.1 单变量自回归模型

与通常一样, 以如下的单变量自回归模型 (autoregressive model, AR) 作为基准模型:

$$\hat{y}_{t+h|t} = \hat{\beta}_0 + \hat{\beta}_1 y_t + \hat{\beta}_2 y_{t-1} + \cdots + \hat{\beta}_p y_{t-p}, \quad (3)$$

其中, $\hat{\beta}_p$ 可通过 OLS 估计求得, $0 \leq p \leq 12$, 对应于 BIC 准则取最小值时的滞后阶数.

2.2.2 多变量目标因子模型

作为传统计量模型中处理大量预测变量的典型模型, 因子模型 (factor model) 利用降维思想, 可在损失较少信息的前提下把多个指标转化为几个不相关的因子, 并利用上述因子进行预测分析. 然而, 若不加区分地将所有指标进行因子分析, 可能会影响预测性能. 因此, 为进一步提高预测准确性, 文本借鉴 Bai and Ng (2008) 提出的目标因子模型 (target factor model, 简称为 T.Factor), 首先以被预测变量的滞后变量为控制变量对各经济指标进行筛选, 即通过阈值法选择显著性水平大于 5% 的经济指标, 然后再进行因子提取. 因此, 最终预测模型可表示为:

$$\hat{y}_{t+h|t} = \hat{\beta}_0 + \alpha y_{t-j} + \hat{\beta}_1 f_{1t-j} + \hat{\beta}_2 f_{2t-j} + \cdots + \hat{\beta}_p f_{pt-j}, \quad (4)$$

其中, f_{it-j} 为提取的第 i 个因子, $0 \leq j \leq 12$, 根据 BIC 准则选择最优滞后阶数 j . 因子个数 p 的选择以其累积方差贡献比例大于 80% 为基准, 此时可基本上代表所有相关指标的信息.

2.2.3 装袋法

集成学习的基本原理是通过训练几个性能较差的弱学习器, 然后通过适当的集成方法形成一个性能较好的强学习器. 自助法聚集 (bootstrap aggregation) 或简称装袋法 (Bagging) 是集成学习中使用较早的模型, 在降低模型预测方差方面表现良好. 原理如下: 首先从初始训练集中有放回地提取 n 个样本, 其大小与原始样本一致; 然后将上述抽样进行 B 轮⁶, 生成 B 个相互独立的样本训练集. 本文以回归树为弱学习器, 给定因变量 y_{t+h} 和一组特征变量 \mathbf{x}_t , 回归树 b 可将该特征变量空间分割成 J_b 个彼此间不重叠的区域 $R_{1,b}, R_{2,b}, \cdots, R_{J,b}$, 每

⁵ 本文使用的直接预测方式不对其他协变量进行预测, 仅基于已有的样本信息估计模型参数, 并根据预测步长的改变重新估计模型, 无需将前一期的预测值代入模型外推本期的预测值, 因此预测误差不会随预测步长的增加而增大.

⁶ 本文令 $B = 100$ 进行实证分析.

个区域 $R_{j,b}$ 上有一个最优输出值 $\hat{c}_{j,b}$, 最终结果是所有回归树预测的平均值, 即

$$\hat{y}_{t+h|t} = \frac{1}{B} \sum_{b=1}^B \left[\sum_{j=1}^{J_b} \hat{c}_{j,b} I_{j,b}(\mathbf{x}_t \in R_{j,b}(\theta_{j,b})) \right], \quad (5)$$

其中, $I_{j,b}(\mathbf{x}_t \in R_{j,b}(\theta_{j,b}))$ 和 $\hat{c}_{j,b}$ 分别表示属于区域 $R_{j,b}$ 的示性函数和在该区域上的最优输出值, $\theta_{j,b}$ 表示回归树 b 在第 j 个区域的参数集.

2.2.4 随机森林模型

随机森林模型 (random forests, 简称 RF) 是 Breiman (2001) 对 Bagging 的优化, 与其类似, 同样以回归树为弱学习器, 最后对所有回归树的预测结果求平均, 如 (5) 式. 不过, 随机森林在建立回归树时, 做了去相关 (decorrelate) 处理, 即在分裂回归树时从全部的 m 个特征变量集中选出一个包含 p 个特征变量的随机样本集作为候选变量, 通常 $p \approx \sqrt{m}$, 这会使得回归树的平均值具有更小的方差, 预测可信度更高.

2.2.5 梯度提升树模型

提升方法 (Boosting) 的基本思想与 Bagging 类似, 但其集成方式不同. 在该模型中各弱学习器彼此相关, 即在训练完一个弱学习器后, 会根据预测表现调整训练集的分布, 并在此基础上训练下一个弱学习器, 最后串联起全部的弱学习器从而构成一个强学习器. 梯度提升树 (gradient boosting decision tree, GBDT) 是 Friedman (2001) 提出的一个 Boosting 算法, 以回归树为弱学习器, 利用加法模型与前向分布算法, 将残差作为最小化目标变量. 假设共有 B 棵回归树, 最终模型可表示为:

$$\hat{y}_{t+h|t} = F_h^B(\mathbf{x}_t) = \sum_{b=1}^B \left[\sum_{j=1}^{J_b} \hat{c}_{j,b} I_{j,b}(\mathbf{x}_t \in R_{j,b}(\theta_{j,b})) \right], \quad (6)$$

其中 $R_{j,b}(\theta_{j,b})$ 表示对近似残差 $r_{b,t}$ 拟合时第 b 棵回归树叶节点的区域 $R_{j,b}$, 该近似残差 $r_{b,t}$ 可表示为损失函数负梯度在当前值的大小, 即

$$r_{b,t} = - \left[\frac{\partial L(y_{t+h}, F_h^{b-1}(\mathbf{x}))}{\partial f(\mathbf{x}_t)} \right], \quad (7)$$

其中 $L(y_{t+h}, F_h^{b-1}(\mathbf{x}))$ 表示先前累加模型的损失函数, $\hat{c}_{j,b}$ 是其对应区域 $R_{j,b}$ 上的最优输出值, 即,

$$\hat{c}_{j,b} = \arg \min_{\mathbf{x}_t \in R_{j,b}} \sum L(y_{t+h}, F_h^{b-1}(\mathbf{x}) + k), \quad (8)$$

其中, k 为常数, 通过持续更新回归树得到最终模型.

2.2.6 极度梯度提升树模型

极度梯度提升树 (extreme gradient boosting, 简称 XGBoost) 是 Chen and Guestrin (2016) 基于梯度提升树发展而来的. 假设有 T 个样本 ($t = 1, \dots, T$), 生成 B 棵树, 即

$\hat{y}_{t+h|t} = \sum_{b=1}^B f_h^b(\mathbf{x}_t)$, $f_h^b(\mathbf{x}_t)$ 为第 b 棵树的预测值, 则损失函数可表示为:

$$L_h^b(\varphi) = \sum_{t=1}^T L[y_{t+h}, F_h^{b-1}(\mathbf{x}_t) + f_h^b(\mathbf{x}_t)] + \Omega(f_h^b), \quad (9)$$

其中, y_{t+h} 是真实数值, $F_h^{b-1}(\mathbf{x}) = \sum_{i=1}^{b-1} f_h^i(\mathbf{x}_t)$, φ 为待求解参数, Ω 为正则项, 进一步对损失函数 $L_h^b(\varphi)$ 进行二阶泰勒展开, 则

$$L_h^b(\varphi) \approx \sum_t \left[g_t f_h^b(\mathbf{x}_t) + \frac{1}{2} h_t [f_h^b(\mathbf{x}_t)]^2 \right] + \gamma N + \frac{1}{2} \lambda \sum_{j=1}^N \omega_j^2, \quad (10)$$

其中, g_t 和 h_t 分别为损失函数的一、二阶导数, γ 和 λ 表示正则系数, N 为回归树的叶节点数, ω_j 为叶节点 j 的权值, 通过最小化上式, 可得到 ω_j 的最优值和最小损失函数.

3 实证结果

由图 1 可知, 工业增加值增长率在新冠疫情前后的波动明显不同, 疫情给经济带来了较大冲击, 后疫情时期的预测变得愈加困难. 因此, 本文根据新冠疫情暴发的时间考虑两个不同的样本区间. 第一个样本区间不包含新冠疫情的影响, 即 2000 年 1 月 - 2019 年 12 月, 建模的实际样本个数为 $R_h = 240 - N - h - p - 1$, 其中, N 是测试数据集的样本大小, h 是预测步长, p 是模型中变量的最长滞后阶数. 第二个样本区间包含新冠疫情的影响, 即 2000 年 1 月 - 2022 年 6 月, 建模的实际样本个数为 $R_h = 270 - N - h - p - 1$.

本节主要对比分析不同模型在疫情前后对工业增加值增速的预测效果, 以验证机器学习方法在异常波动背景下是否具有预测优势. 依照前述设定, 本文采取预测步长为 h 个月的直接预测, 并分别考虑模型在 $h = 1, 3, 6, 9, 12$ 等不同预测步长下的预测效果. 在每个样本区间内, 将数据集分为训练集和测试集, 测试集采用样本末端 36 个月的数据, 即 $N = 36$, 用于检验各模型的预测效果⁷, 剩余样本列入训练集用于模型拟合. 因此, 分别基于两个训练集样本数据进行模型训练, 从而可得到第一个区间内在其测试集 (2017 年 1 月 - 2019 年 12 月) 上的样本外预测结果和第二个区间内在其测试集 (2019 年 7 月 - 2022 年 6 月) 上的样本外预测结果, 并针对不同预测步长, 分别比较各模型的预测性能.

在评估模型预测准确性时, 本文采用均方根误差 (RMSE) 和平均相对误差 (MRE) 两个通用评价指标, 其定义如下:

$$\text{RMSE}_{m,h} = \sqrt{\frac{1}{N} \sum_{t=T+1}^{T+N} (y_t - \hat{y}_{t,m,h})^2}, \quad (11)$$

$$\text{MRE}_{m,h} = \frac{1}{N} \sum_{t=T+1}^{T+N} \frac{|y_t - \hat{y}_{t,m,h}|}{y_t}, \quad (12)$$

⁷在固定窗口的滚动向前预测过程中, 测试集中的部分数据亦可能会参与模型拟合. 比如, 进行向前 1 个月预测 2022 年 6 月工业增加值增速时, 截止到 2022 年 5 月的信息均会用于模型训练.



图 1 工业增加值增长率 (%)

其中 $\hat{y}_{t,m,h}$ 是指模型 m 在预测步长为 h 时对 t 期工业增加值增速的预测结果. 为清晰比较不同模型之间的预测准确性, 可将对应模型的平均预测误差与基准模型的平均预测误差做比值, 构造相对平均预测误差, 即 rRMSE 和 rMRE.

为进一步比较机器学习模型与传统方法的预测差异, 本文使用 Diebold and Mariano (1995) 提出的方法 (以下简称 DM 方法) 对两个模型的样本外预测误差是否存在显著差异进行统计检验, 采用的统计量为:

$$DM_h = \frac{\bar{d}_h}{\sqrt{\frac{\hat{V}(d_{t,h})}{N}}} \sim N(0, 1), \quad (13)$$

其中, $\bar{d}_h = \frac{1}{N} \sum_{t=T+1}^{T+N} d_{t,h}$, $d_{t,h} = g(\hat{e}_{t,h}^{(1)} - \hat{e}_{t,h}^{(2)})$, $\hat{e}_{t,h}^{(1)}$ 和 $\hat{e}_{t,h}^{(2)}$ 分别表示模型 1 和模型 2 在预测步长为 h 时对应的预测误差, $g(\cdot)$ 为预测误差的函数, 取平方或者绝对值, 分别对应 rRMSE 和 rMRE 的分析, $\hat{V}(d_{t,h})$ 是 $d_{t,h}$ 的方差估计.

3.1 疫情前的预测效果分析

基于疫情前各模型的训练, 分别对其测试集 (2017 年 1 月 - 2019 年 12 月) 区间进行预测效果检验. 表 2 列出了疫情前各模型与基准 AR 模型的的样本外相对预测误差. 总体来看, 在不同的预测步长中, 各模型的预测表现不尽相同.

具体来看, 在向前 1 个月的预测中, 传统时间序列模型的预测效果整体较好, 其中, RMSE 显示目标因子模型 (T.Factor) 略优于自回归模型 (AR), 但 MRE 的结论则相反, 不过两者均统计上不显著, 说明这两种模型的预测效果相近. 此时集成学习模型的 rRMSE 和 rMRE 几乎均大于 1, 只有随机森林模型 (RF) 的 rRMSE 略小于 1, 但统计上不显著, 即集成学习模型未体现出预测优势. 这主要是因为在经济平稳期, 经济波动小, 依照变量自身的历史信息和基本的线性模型即可较好地进行下个月的预测. 但在向前 3 个月的预测中,

表 2 疫情前各模型相对 AR 模型的平均预测误差

模型	rRMSE					rMRE				
	1 月	3 月	6 月	9 月	12 月	1 月	3 月	6 月	9 月	12 月
T.Factor	0.991	0.869	0.770*	0.891	1.105	1.041	0.993	0.741*	0.821	1.020
Bagging	1.032	1.582*	0.986	1.208	0.697***	1.077	1.706**	0.911	1.256	0.693***
RF	0.971	1.194	0.908	1.270**	0.772***	1.083	1.454**	0.971	1.345	0.799**
GBDT	1.019	1.318	1.015	0.906	0.668***	1.160	1.495**	0.969	0.887	0.604***
XGBoost	1.043	1.007	0.961	1.157	0.843	1.166	1.165	1.030	1.215	0.839

注: 表中结果是各模型相对于 AR 模型的平均预测误差比率, 当某个模型对应的 rRMSE 或 rMRE 值大于 1 时, 说明该模型的预测精度低于 AR 模型; 当其值小于 1 时则相反。***、** 和 * 分别表示相应模型与 AR 模型的 DM 检验统计量在 1%、5% 和 10% 上的显著性水平。

T.Factor 模型的 rRMSE 和 rMRE 均略小于 1, 说明其预测性能已相当于 AR 模型, 但统计上不显著。此时, 集成学习模型的预测效果依然不及 AR 模型, 可能因为其易出现模型过拟合而增加了预测误差。其中, 随机森林 (RF) 和极度梯度提升树 (XGBoost) 模型的预测效果略优于装袋法 (Bagging) 和梯度提升树 (GBDT) 模型, 这可能因为 RF 模型具有较好的特征选择功能, 而 XGBoost 模型可将损失函数拓展到二阶泰勒展开式从而较准确地捕捉到误差的细微变化。然而, 在向前 6 个月的预测中, 集成学习模型的预测优势逐渐显现, 整体上好于 AR 模型。在向前 9 个月的预测中, T.Factor 模型依然具有最优的预测性能, 而梯度提升树 (GBDT) 模型在集成学习模型中表现最佳, 且优于 AR 模型。在向前 12 个月的预测中, T.Factor 模型的预测效果最差, 其 rRMSE 和 rMRE 均略大于 1, 不过其在统计上不显著, 说明此时 T.Factor 模型和 AR 模型预测效果相当, 未表现出预测优势; 与此相反, 集成学习模型则显著优于传统模型, 其中 GBDT 模型的预测误差相对更小。此外, 从图 2 中各模型预测的 RMSE 和 MRE 可知, 整体上传统模型的预测误差随预测步长的增加而增大, 与此不同, 集成学习模型的预测误差与预测步长不存在正相关性, 其在向前 6 个月和 12 个月的预测表现更好, 这可能是因为集成学习模型主要以数据驱动, 更容易从数据中挖掘如季节性变动等规律性和相关性信息。

总体而言, 在向前 1 个月的疫情前预测中, AR 模型对工业增加值增速的预测效果具有一定的优势, 表明在经济平稳期进行短期预测时, 工业增加值增速的近期历史变动已经包含了未来短期变化的主要信息, 因而结构简单和较少特征变量的 AR 模型更利于短期预测, 过多的变量信息容易造成噪音干扰; 在 3 个月、6 个月和 9 个月的预测中, T.Factor 模型的预测优势更明显, 各集成学习模型的预测效果有所不同; 在向前 12 个月的预测中, 集成学习模型的预测效果最佳, 其预测精度显著高于传统 AR 和 T.Factor 模型, 其中, GBDT 模型的预测误差最小。这说明随着预测步长的增加, 经济的不确定性和波动性通常随之增加, 此时可捕捉复杂关系和结构变化的集成学习模型具有相对优势。因此, 在经济波动较小时, 集成学习方法比传统时间序列模型更适用于向前 12 个月的预测。

进一步, 本文考查了疫情前 AR 模型和 GBDT 集成学习模型在测试集上向前 12 个月的预测趋势, 从而更好地分析集成学习模型的预测优势。如图 3 所示, 基于 AR 模型的工业增加值的预测值明显高于真实值, 且在部分时期的变动趋势与真实值不同; 而 GBDT 模型的预

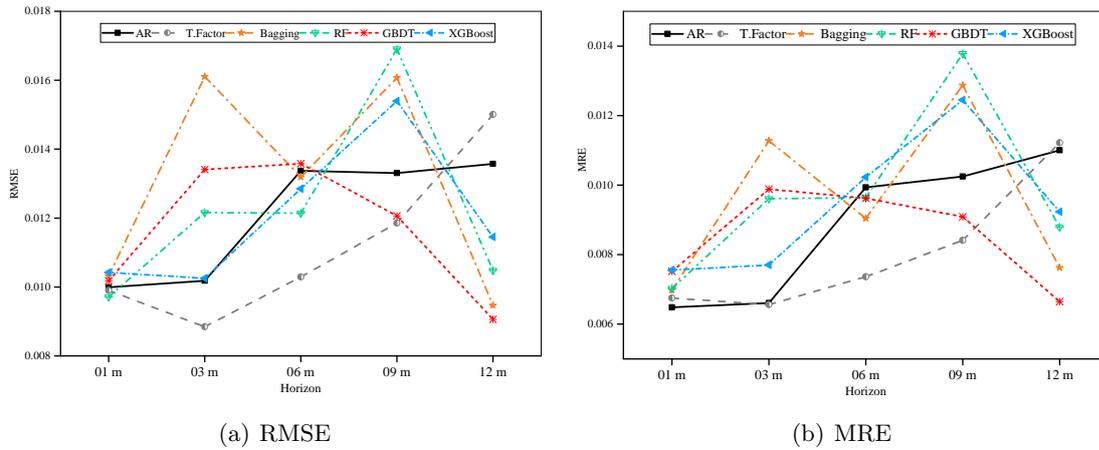
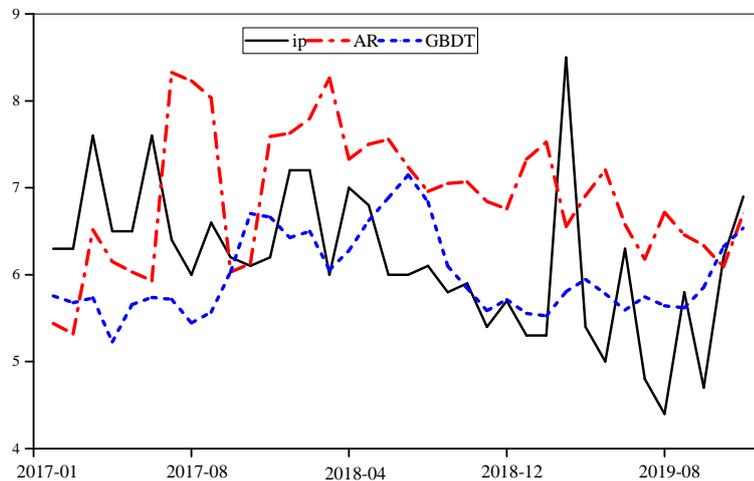


图 2 疫情前各模型的预测误差

图 3 2017–2019 年 AR 和 GBDT 模型向前 12 个月的预测比较⁸

测值更接近真实值, 尽管在工业增加值波动较大时的偏差较大, 但预测结果的走势与其基本保持一致, 可见, 集成学习模型向前 12 个月的预测效果总体更佳。

3.2 疫情后的预测结果分析

为了考查疫情冲击后的模型预测效果, 将样本区间扩展到 2022 年 6 月. 表 3 中列出了在对应测试集 (2019 年 7 月 – 2022 年 6 月) 上各模型的样本外相对预测误差结果. 总体来看各模型的相对预测效果与疫情前有所不同.

由表 3 并结合图 4 可知, 在向前 1 个月的预测中, AR 模型仍具有一定预测优势, 但不及 T.Factor 模型, 甚至 T.Factor 模型的 rMRE 显著小于 1, 因为后者充分利用了其他经济指标的信息, 更有助于经济波动期的短期预测, 集成学习模型的预测表现整体上与 AR 模型相近,

⁸ip 表示工业增加值增速, 下同.

表 3 疫情后各模型相对 AR 模型的平均预测误差

模型	rRMSE					rMRE				
	1 月	3 月	6 月	9 月	12 月	1 月	3 月	6 月	9 月	12 月
T.Factor	0.940	0.641***	0.723***	0.838	0.756	0.697**	0.5181***	0.637***	0.791	0.850
Bagging	1.038	0.600***	0.745***	0.722**	0.747*	0.953	0.502***	0.677**	0.736*	0.818*
RF	1.052	0.652**	0.816*	0.896	0.812	1.007	0.624**	0.806	0.937	0.895
GBDT	1.008	0.601***	0.741**	0.804	0.727**	0.824	0.525***	0.647***	0.741*	0.778*
XGBoost	1.138	0.737**	0.701***	0.887	0.888	1.097	0.745*	0.634***	0.770	0.879

注: 表中结果是各模型相对于 AR 模型的平均预测误差比率, 当某个模型对应的 rRMSE 或 rMRE 值大于 1 时, 说明该模型的预测精度低于 AR 模型; 当其值小于 1 时则相反. ***、** 和 * 分别表示相应模型与 AR 模型的 DM 检验统计量在 1%、5% 和 10% 上的显著性水平.

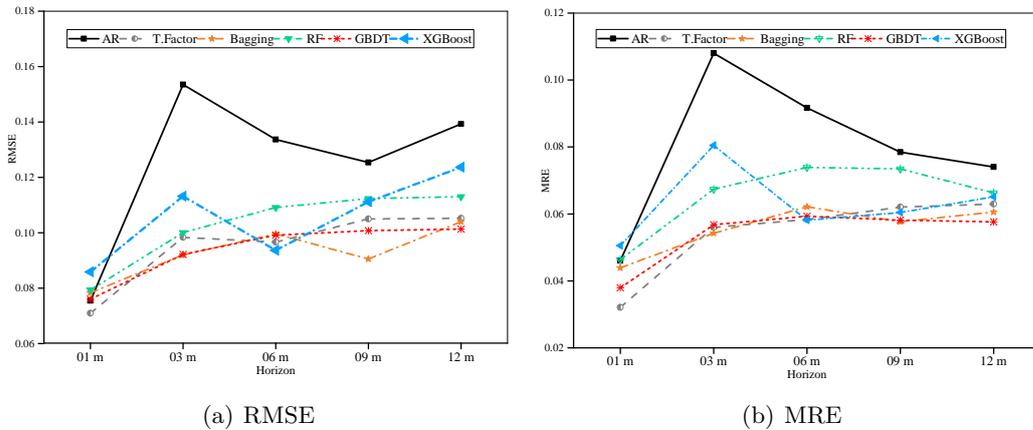


图 4 疫情后各模型的预测误差

但不及 T.Factor 模型, 其中 GBDT 模型的预测结果相对较好, 其 rMRE 小于 1, 但统计不显著. 在向前 3 个月的预测中, 集成学习模型的效果均显著优于 AR 模型, 其中, Bagging 和 GBDT 模型的预测误差最小. T.Factor 模型亦显著优于 AR 模型, 但由于其进行因子提取时会遗漏部分信息且主要为线性拟合, 预测效果稍逊于 Bagging 和 GBDT 模型⁹. 在向前 6 个月的预测中, T.Factor 模型仍具有一定预测优势, 与 GBDT 模型和 XGBoost 模型的预测水平相当, 其中, XGBoost 模型具有较小的相对误差, 且显著低于 AR 模型, 这可能与可准确拟合细微变化相关. 随着预测步长的增加, 在进行向前 9 个月和 12 个月预测时, T.Factor 模型的预测误差逐渐增大, 而集成学习模型的 RMSE 和 MRE 在向前 12 个月的预测中基本未明显增加, 与疫情前类似. 其中, Bagging 和 GBDT 模型的表现较好, 预测精度高, 稳定性强; RF 模型整体上优于 AR 模型, 但不及上述两个集成学习模型, 可能与其变量选择有关, 此时 XGBoost 模型的预测效果与 RF 模型相近, 优于 AR 模型却统计不显著.

总体而言, AR 模型在疫情后的预测中不具有优势, 在疫情前表现较好的 T.Factor 模型在预测步长增加时亦无显著优势, 相反, 集成学习模型的预测效果在疫情后明显提升, 尤其

⁹其中, Bagging 和 GBDT 的 rRMSE 均小于 T.Factor, 只有 GBDT 的 rMRE 略大于 T.Factor.

是 Bagging 和 GBDT 模型在各种预测步长下均具有较高的预测精度, 后者的预测表现总体更佳.

通过进一步考查疫情后 AR 模型和 GBDT 模型在整个测试集内向前 1 个月的预测走势情况 (见图 5) 可以发现¹⁰, AR 模型的预测值波动更为明显, 上升和下滑幅度较大, 且其变化趋势与工业增加值增速真实值走势的吻合度较差, 预测效果不佳. 而 GBDT 模型的预测结果波动性较小, 除在个别波动性较大时点的预测值与真实值有一定差距外, 整体变动趋势与真实值较为一致, 从而更利于预判和把握工业增长的短期未来走势.

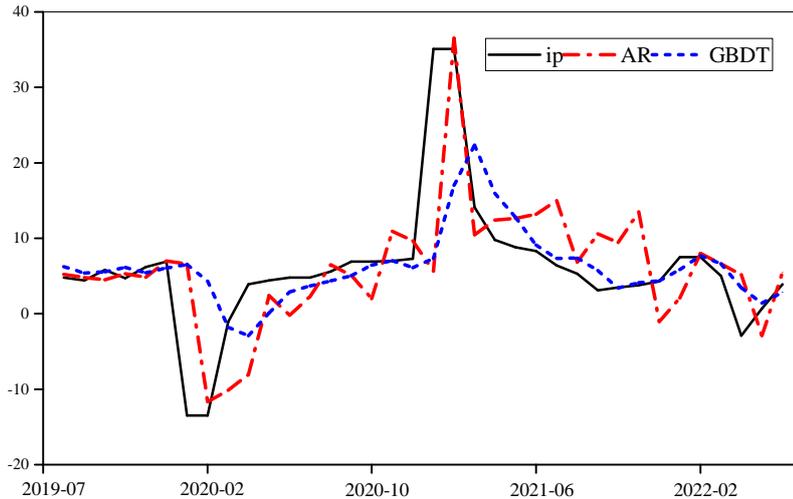


图 5 2019 年 1 月 - 2022 年 6 月 AR 和 GBDT 模型向前 1 个月的预测比较

3.3 稳健性检验

为探究上述几种集成学习模型的预测性能是否具有一定的稳定性, 本节进一步从三方面对疫情前后的样本进行稳健性检验. 第一, 分别增加和减少测试数据集中的样本数量, 即分别设为 48 个月和 30 个月, 并进行类似的固定窗口滚动预测分析, 发现以上结论未发生较大改变. 第二, 改变预测窗口的形式, 将固定窗口改为递归窗口, 即建模时保留所有相关变量的历史数据信息, 结果表明各模型预测误差的相对大小未明显变化. 第三, 分别比较了其他两类机器学习方法在疫情前后的预测效果, 一类是具有特征选择功能的惩罚线性回归模型, 这里分别考虑了 LASSO 回归模型、岭回归模型 (ridge regression, 简写 RR) 和弹性网络回归模型 (elastic net, 简写 Elnet); 另一类是可以处理复杂关系的 BP 神经网络模型. 从表 4 可以看出¹¹, 在疫情前, 这三种惩罚线性回归模型未表现出明显的预测优势, 多数情况下不及 AR 模型; 另外, BP 神经网络模型的相对误差远大于 1, 其预测效果显著差于 AR 模型, 均无法与集成学习模型相比. 不过, 在疫情后, 上述模型的预测效果有所提升, 其中, 三种惩罚线性模

¹⁰由上述分析可知, 在疫情后的多种预测步长评估中, AR 模型向前 1 个月的预测表现较好, 以此为例进行分析更有助于捕捉集成学习模型相较于该基准模型的预测优势.

¹¹前两种稳健性检验中使用的模型与前文相同, 其结果也与前文相近, 受篇幅限制, 此处未列出, 若有需求可向作者索要.

表 4 疫情前后其他机器学习模型相对 AR 模型的平均预测误差

模型	rRMSE					rMRE				
	1 月	3 月	6 月	9 月	12 月	1 月	3 月	6 月	9 月	12 月
疫情前										
LASSO	0.967	1.094	1.209*	1.433**	1.879**	1.192	1.358*	1.322**	1.414**	1.575**
RR	1.332*	1.304	0.971	0.893	0.931	1.716***	1.644**	1.010	0.894	0.945
ElNet	0.962	1.109	1.233*	1.438**	1.817**	1.195	1.379*	1.355**	1.439***	1.541**
BP	6.016***	5.750***	4.482***	4.479***	4.347***	8.635***	8.224***	5.639***	5.387***	4.969***
疫情后										
LASSO	0.938	0.642***	0.820	1.145	0.809	0.745*	0.575***	0.740*	1.121	0.928
RR	1.042	0.582***	0.727**	0.812	0.688**	0.879	0.479***	0.627***	0.750*	0.736**
ElNet	0.926	0.627***	0.825	1.142	0.800*	0.750*	0.564***	0.754*	1.131	0.905
BP	1.349*	0.656**	0.772*	0.818	0.728	1.648***	0.699**	0.855	0.988	1.012

注: 表中结果是各模型相对于 AR 模型的平均预测误差比率, 当某个模型对应的 rRMSE 或 rMRE 值大于 1 时, 说明该模型的预测精度低于 AR 模型; 当其值小于 1 时则相反。***、** 和 * 分别表示相应模型与 AR 模型的 DM 检验统计量在 1%、5% 和 10% 上的显著性水平。

型整体上均优于 AR 模型, 甚至在向前 3 个月的预测中略优于上述集成学习模型; 同时 BP 神经网络模型的预测效果亦明显提升, 甚至在一些预测步长下优于 AR 模型。但是, 在多数情况下, 这两类机器学习模型的相对平均误差仍大于前文中集成学习模型的相对误差, 因此, 其预测效果整体上均不及 Bagging 和 GBDT 模型。

综合以上不同视角的稳健性分析, 发现 Bagging 和 GBDT 模型等集成学习模型在经济运行平稳期和异常冲击下均具有预测优势, 实证结果相对稳健。

4 对典型集成学习模型的解释和分析

尽管前述分析表明, 回归树集成学习模型的预测性能与传统时间序列模型相比具有一定优势, 但其经济意义和解释能力常常相对薄弱。在很多实证研究中, 模型可解释性与预测精度同样重要, 因为这有助于人们更好地理解经济运行背后的逻辑。通常, 传统的线性回归模型可以通过分析每个自变量或输入变量的边际效应进行解释, 但由集成算法构造的 Bagging 和 Boosting 模型则不容易给出直接的解释。因此, 学术界和业界也一直在摸索打开机器学习黑箱的方法。尽管 RF 和 Boosting 等模型也提供了反映特征重要性的简单度量, 但是该方法比较单一和笼统, 无法提供特征变量对预测结果在影响方向和程度上的具体分析, 而且对于单个样本或者不同时点来说, 各个特征变量与预测变量之间的相互关系很难解释 (肖艳丽和向有涛, 2022)。因此, 本文尝试采用 SHAP 方法对集成学习模型进行解释和分析, 以弥补该类模型的不足, 从而加深对其预测优势的理解和认识。

4.1 SHAP 方法简介

SHAP 是 Shapley additive explanations 的缩写, 该方法基于 Shapley (1953) 在合作博弈研究中提出的 Shapley 值, 即根据玩家 (players) 对总支付 (payoff) 的各自贡献将该支付分配给玩家的一种度量, 具有一系列理想的理论性质。近年来, SHAP 方法被用于解释机器学习模型并取得了一定效果, 此时, 所有的特征变量类似于参与的玩家, 模型预测结果类似于总支付, 相应地, 对于其中一个特征变量的 Shapley 值而言, 可以表示该特征变量为模型预测

结果提供的边际贡献. 该方法与其他方法相比具有一定优势: 第一, 解释力强. 它是一种全局解释方式, 假设预测结果与其基准值之间的差值可完美地分布在各个特征变量之间, 这个属性将 SHAP 方法与其他解释方法 (如 LIME¹²) 区分开来, 更加接近人们的认知, 且在基于回归树的集成学习模型中有较好的解释力 (Lundberg et al., 2017, 2020); 第二, 适用性广. 该方法假设所有输入的特征变量的子集均是相互扰动的, 即考虑了不同特征变量之间的交互作用, 适用于任何类型的模型, 可减少在比较模型时反复替换解释变量的麻烦 (Štrumbelj and Kononenko, 2014). 此外, 通过将所有时点上特征变量的贡献值 (以 Shapley 值表示) 进行绝对值平均, 可以计算每个特征变量在整个样本内的贡献. 因此, 这种解释性方法近期已在宏观经济预测中得到一定应用 (Buckmann et al., 2021; Chapman and Desai, 2022).

用 S 表示所有解释变量 (\mathbf{x}_t) 的集合, 其中的元素称为特征变量, 记 $|S|$ 为特征集 S 中元素的个数. 机器学习模型 $F_h(\cdot)$ 的预测值 $\hat{y}_{t+h|t}(\mathbf{x}_t)$ 可以分解成:

$$\hat{y}_{t+h|t}(\mathbf{x}_t) = \varphi_0 + \sum_{j=1}^{|S|} \varphi_t^j, \quad (14)$$

其中, φ_0 是预测模型的基准值, 一般将预测的均值 $E[\hat{y}_{t+h|t}(\mathbf{x}_t)]$ 作为基准. φ_t^j 是第 j 个特征变量对预测结果的影响大小, 即 Shapley 值. $|\varphi_t^j|$ 越大表示该特征变量对目标预测结果的影响越大, $\varphi_t^j > 0$ 和 $\varphi_t^j < 0$ 分别表示对预测结果产生正向和负向影响. 第 j 个特征的 Shapley 值 φ_t^j 的计算公式如下:

$$\varphi_t^j = \sum_{Q \subseteq S \setminus \{j\}} \frac{|Q|!(|S| - |Q| - 1)!}{|S|!} (v(Q \cup j) - v(Q)), \quad (15)$$

其中, Q 是不包括第 j 个特征变量的集合 S 的子集, 其元素个数为 $|Q|$, 基于特征集 S 可构建 $\sum_{Q \subseteq S \setminus \{j\}} \frac{|Q|!(|S| - |Q| - 1)!}{|S|!}$ 个不包含第 j 个特征变量的子集 Q . $v(Q \cup j)$ 表示子集 Q 与第 j 个特征变量的并集对预测结果期望值的影响大小, $v(Q)$ 表示子集 Q 对预测结果期望值的影响大小, 则 $v(Q \cup j) - v(Q)$ 表示子集 Q 中第 j 个特征变量对预测结果期望值的影响大小.

随着特征变量的增加, Shapley 值的计算时间呈指数级增长, 当变量较多时该方法难以应用于实际, 因此, 本文借鉴 Štrumbelj and Kononenko (2010, 2014) 的做法, 采用基于准随机和自适应采样的近似算法降低计算复杂度和运行时间.

根据上述实证结果, Bagging 模型、RF 模型和 GBDT 模型在疫情前后的不同预测步长中表现各有优劣, 因此, 本文进一步借助 SHAP 方法来分析这三种模型中不同特征变量的重要性.

4.2 向前 1 个月预测的主要特征变量分析

首先考查在疫情前三个典型集成学习模型中不同特征变量的重要程度, 即将各特征变量在不同时点上的 Shapley 绝对值求平均后排序. 从图 6 中可以看出, 在疫情前向前 1 个月预测时, 作为特征变量的当期工业增加值增速 (ip) 的平均 Shapley 绝对值在三个预测模型

¹²LIME 的全称是 local interpretable model-agnostic explanations, 是一种机器学习模型的局部解释方法, 具体算法可参考 Ribeiro et al. (2016).

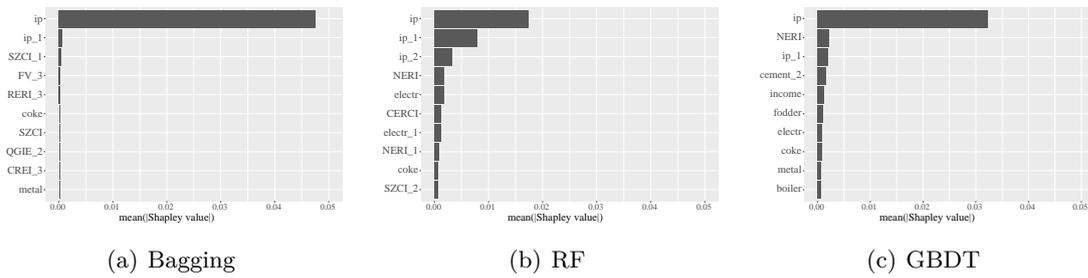


图 6 疫情前三个典型集成学习模型的平均 Shapley 绝对值 ($h = 1$)¹³

中均最大且显著大于其他变量 (尤其是 Bagging 和 GBDT 模型), 即 t 期工业增加值增速对其 $t + 1$ 期的预测贡献远大于其他变量. 对于 RF 模型, 滞后 1 至 2 个月的工业增加值增速 (ip_1 和 ip_2) 也有助于其向前 1 个月的预测, 其他特征变量的平均 Shapley 绝对值相对较小, 这进一步印证了前述分析中 AR 模型在向前 1 个月预测中表现较好的情况, 即依据工业增加值增速滞后期的信息即可进行较准确的预测. RF 模型由于具有特征选择功能, 进入模型的变量减少而分配在每个入选变量上的 Shapley 绝对值相对较均匀, 因此, 当期工业增加值增速的平均 Shapley 绝对值比其他模型小一些, 但由于较多地捕捉了人民币名义有效汇率指数 (NERI)、发电量 (electr) 和国房景气指数 (CERCI) 等经济指标的信息, 从而在向前 1 个月的预测中表现更佳.

在分析不同特征变量的静态重要性基础上, 进一步考查在疫情前向前 1 个月预测模型中几个重要特征变量的 Shapley 值的动态变化情况, 此处以预测效果较好的 RF 模型为例. 图 7 显示, 尽管当期工业增加值增速的 Shapley 值波动性较大, 但其变动趋势与工业增加值增速 (ip) 和向前 1 个月的预测值 (ip_f) 的走势保持了较高的一致性, 在预测值出现波峰和波谷时, 其当期的 Shapley 值也随之增加和下降, 对预测结果具有较好的解释性. 与此类似, 另一个较重要的特征变量——人民币名义有效汇率指数 (NERI) 的 Shapley 值也有大体相近的变化趋势, 但其 Shapley 值相对较小. 从图 7 可以看出, 2012 年之后, 随着中国经济增长进入结构性减速的新常态阶段, 工业增加值增速逐步放缓且波动性减弱, 当期工业增加值增速和人民币名义有效汇率指数的 Shapley 值总体转为负值, 较好地反映了经济新常态阶段的减速换挡特征, 但在一些时点上的解释力不甚明显, 还需结合其他经济指标进行预测分析. 整体而言, 重要特征变量 Shapley 值的动态变化趋势与被预测变量的走势基本一致.

上一节对疫情后各模型预测效果的分析表明, 此时 AR 模型的预测性能有所下降, 而集成学习模型则有所提升. 从图 8 中三个模型主要特征变量的平均 Shapley 绝对值的计算结果容易看出, 尽管当期工业增加值增速的平均 Shapley 绝对值仍然保持最大, 但是该指标其他滞后期的影响力则明显减弱, 而发电量 (electr) 和生铁产量 (iron) 等指标对预测结果的影响却出现了较大提升, 与疫情前有所不同, 这也是集成学习模型预测效果优于 AR 模型的重要原因.

鉴于 GBDT 模型在疫情后向前 1 个月的预测中表现较好, 下面以该模型为例分析几个

¹³变量名称后缀的数字表示滞后期数, 此处展示了平均 Shapley 绝对值较大的前 10 个特征变量, 下同.

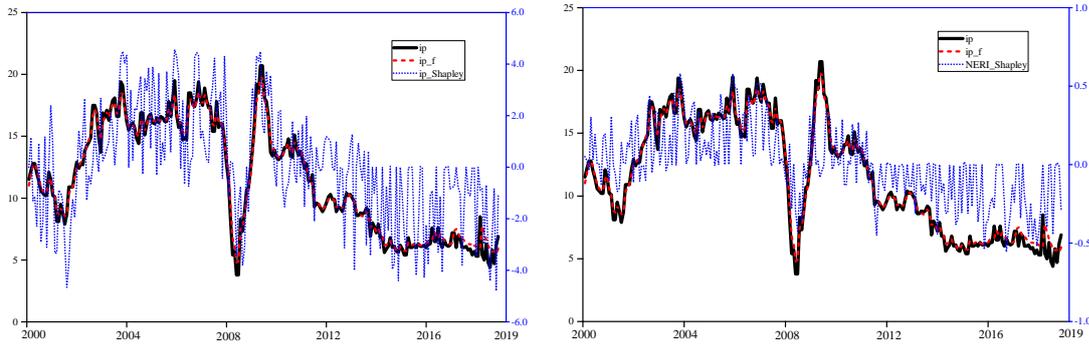
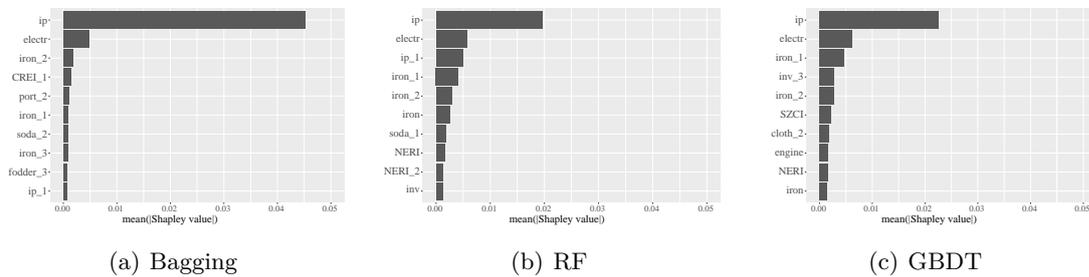


图7 工业增加值增速(左轴)与RF模型重要特征变量的Shapley值变化(右轴)¹⁴



(a) Bagging

(b) RF

(c) GBDT

图8 疫情后三个典型集成学习模型的平均Shapley绝对值($h=1$)

重要特征变量 Shapley 值随时间的变动情况. 图 8 和图 9 显示, 当期工业增加值增速 (ip) 的 Shapley 值整体最大, 约为其他特征变量的 3 倍以上, 其波动趋势与向前 1 个月的预测结果的变动基本一致, 但对新冠疫情等重大冲击的响应不够迅速, 还需结合其他经济指标进行预测. 当期发电量 (electr) 和滞后 1 期的生铁产量 (iron_1) 的 Shapley 值波动态势与工业增加值增速走势亦保持了高度的一致性, 尤其在经济遭遇剧烈冲击时期表现更为明显. 在 2020 年新冠疫情冲击下, 工业增加值增速陡降后又快速反弹, 发电量和生铁产量的 Shapley 值亦呈现出类似走势, 说明这两个指标分别作为工业生产的重要能源和原料对经济变动相对敏感, 可为预测工业增长提供重要参考信息. Galdi et al. (2023) 利用高频电力市场数据对工业生产进行预测, 发现其在新冠疫情等高波动时期的预测效果较好, 同样说明电力数据在预测工业生产中的重要作用. 另一个特征变量——滞后 3 期的固定资产投资 (inv_3) 的 Shapley 值在部分时间的波动性较大, 但在很多时间几乎是零, 即固定资产投资增速变动只在个别时期对工业增加值增速具有预测作用, 这可能与政府宏观调控和固定资产投资自身的运行周期相关, 比如, 在 2008 年金融危机时, 固定资产投资的 Shapley 值为正, 即投资对工业的作用是增强的而非减弱; 同时在经济新常态时期, 该指标的 Shapley 值整体为正, 说明投资在此期间对工业增长起到一定的支撑作用; 相反, 在新冠疫情发生后, 该指标的 Shapley 值为负, 表明此时投资减速对工业增加值增速产生了向下拉动的负向影响, 这与隔离管控政策和企业预期

¹⁴为更清晰地比较各特征变量的 Shapley 值大小, 将其扩大 100 倍显示, 下同.

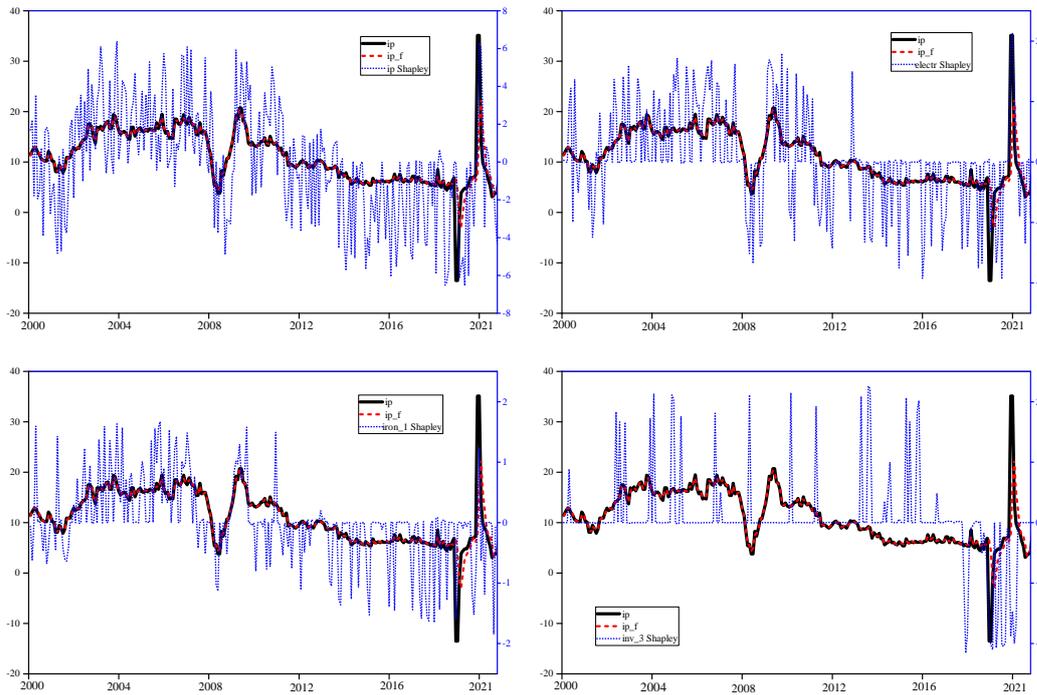


图 9 工业增加值增速 (左轴) 与 GBDT 模型重要特征变量的 Shapley 值变化 (右轴) ($h = 1$)

减弱相关, 固定资产投资增速的降低会直接传导到工业增加值的增速变化上. 总体上, 当经济运行遇到较大冲击时, 依靠集成学习方法和多种经济指标来预测工业增长会获得更好的结果.

上述图 6 和 8 显示的是主要特征变量在样本期内所有时点上 Shapley 绝对值平均后的对比情况. 为了便于对比各时点或某阶段内各特征变量相对贡献的动态变化情况, 本文进一步提出并测算了不同特征变量的相对贡献率 (relative contribution ratio, 简写 RCR), 即特征变量 j 在 t 时的 Shapley 值与对应时点上所有特征变量的 Shapley 绝对值之和的比率, 公式如下:

$$\text{RCR}_t^j = \frac{\varphi_t^j}{\sum_{i=1}^{|S|} |\varphi_t^i|}, \quad (16)$$

其中, $|\varphi_t^i|$ 表示第 i 个特征变量在 t 时的 Shapley 值的绝对值, $|S|$ 仍然表示特征变量的个数. 当该贡献率大于 0 时, 表示特征变量 j 的 Shapley 值对预测结果的影响方向为正, 反之则反之.

为考查不同时期主要特征变量影响的动态变化情况, 本节参照陈磊等 (2019) 对宏观经济景气周期的测定并考虑疫情冲击的影响, 将样本期分为八个阶段¹⁵, 分别计算各阶段特征

¹⁵ 主要根据经济周期和重大事件划分阶段. 其中, 依据“波峰——波峰”的测定标准判定经济周期, 参照陈磊等 (2019) 的研究, 2000 年 8 月、2004 年 2 月、2006 年 6 月、2007 年 11 月、2010 年 2 月、2013 年 9 月、2017 年 5 月均为宏观经济一致景气指数的波峰, 据此可划分为六个阶段, 鉴于 2000 年 5 月是样本初始时间且与第一个波峰 (2000 年 8 月) 接近, 故以此为起点划分为第一阶段. 考虑到 2020 年 1 月暴发的新冠疫情对经济影响较大, 故将此作为第七和第八阶段的分界点.

变量的平均相对贡献率,并分别找到每个阶段上除当期工业增加值增速外平均相对贡献率最大和最小的特征变量¹⁶,具体结果见表5。

从表5容易看出,在金融危机前的3个阶段中,发电量和生铁产量增速对向前1个月的工业增加值增速预测结果具有较大的正向影响,而深证综指与工业生产的走势相悖,表现为负向作用,因为此时正值中国金融市场的摸索阶段,存在较多非理性波动。处于金融危机时期的第4个阶段,社会消费品零售总额对稳定工业生产具有较大正向贡献,但饮料产量的贡献率却为负值,而且第三产业投资也有较大的负向影响,在一定程度上反映金融危机对这一时期我国第三产业和居民生活产生一定负面冲击。在金融危机后的第5个阶段,经济和工业增加值增速呈阶段性下滑态势,此时深证综指总体呈现负向影响,反映此时股票市场的波动总体上对预判未来工业增速走势已具有一定作用。在第6和第7阶段,经济运行进入新常态,工业增加值增速缓慢下滑,生铁产量和发电量的平均相对贡献率均为负且重要性突出,反映出能源和原材料在此阶段对工业生产仍然具有重要影响;但固定资产投资在第6个阶段的平均相对贡献率较大且为正向影响,说明投资在此阶段对稳定工业增长起到一定的助推作用。另外,发动机产量在第7阶段的平均相对贡献率亦表现为正向影响,进一步证明实体经济在经济平稳运行阶段的重要支撑作用。与这两个阶段相反,在新冠疫情暴发以来的第8个阶段,固定资产投资增速的下滑则对工业增长产生了较大的负向作用,证实了上文中关于投资对工业生产影响的分析;而此阶段深证综指的较大相对贡献率反映出我国股票市场已具有一定的经济“晴雨表”作用。整体而言,生铁产量和发电量在对工业增加值增速向前1个月预测时具有重要作用,在遭遇金融危机或新冠疫情的严重冲击时,还需结合消费、投资和金融市场等指标的变动以提高预测效果。

表5 向前1个月预测时不同阶段主要特征变量的平均相对贡献率(%)

时间阶段	平均相对贡献率最大		平均相对贡献率最小	
	特征变量	平均相对贡献率	特征变量	平均相对贡献率
阶段1: 2000年05月-2004年02月	发电量(electr)	2.04	深证综指(SZCI)	-1.23
阶段2: 2004年03月-2006年06月	生铁产量(iron_1)	5.63	深证综指(SZCI)	-1.66
阶段3: 2006年07月-2007年11月	发电量(electr)	7.36	开发投资(CREI_1)	-0.96
阶段4: 2007年12月-2010年02月	社零额(cons)	1.43	饮料产量(drink_1)	-0.93
阶段5: 2010年03月-2013年09月	布产量(cloth_2)	1.73	深证综指(SZCI)	-1.08
阶段6: 2013年10月-2017年05月	固定资产投资(inv_3)	3.90	发电量(electr)	-3.25
阶段7: 2017年06月-2019年12月	发动机产量(engine)	1.13	生铁产量(iron_1)	-3.42
阶段8: 2020年01月-2022年06月	深证综指(SZCI)	3.12	固定资产投资(inv_3)	-4.44

注:平均相对贡献率最大可给出某阶段总体正向影响最大的特征变量,平均相对贡献率最小(负值)可分析某阶段总体负向影响最大的特征变量。

4.3 向前12个月预测的主要特征变量分析

在分析了向前1个月的短期预测情况后,本节进一步考查疫情前后几种集成学习模型在向前12个月预测时主要特征变量的贡献情况。

¹⁶由于当期工业增加值增速的Shapley值整体最大,排除其影响可更好地发现不同阶段内其他重要的特征变量。

首先分析疫情前的预测情况. 对比图 10 与图 6 可以看出, 主要特征变量与向前 1 个月预测时发生了很大变化. 三个集成学习模型在向前 12 个月的预测中平均 Shapley 绝对值最大的前两个特征变量均为滞后 3 个月和 2 个月的深证综指 (SZCI), 当期工业增加值增速的重要性已不明显, 其它重要特征变量则包括人民币存款余额 (RMBD)、外汇储备 (FER) 等金融类指标, 说明在进行较长步长的预测时, 预测结果更依赖股票市场等与金融相关的变量, 而不是工业增加值增速自身的滞后信息. 这说明股票价格波动在一定程度上可作为经济活动的晴雨表提前为工业增长变动提供预警信号. 因此, 仅依靠工业增加值增速历史信息的 AR 模型在向前 12 个月的预测中并没有得出较好的预测结果. 虽然 T.Factor 模型可以利用多种经济指标的信息, 但是在较长预测步长中的表现仍然不乐观, 这可能受限于该模型不能像集成学习模型一样反映错综复杂的非线性经济关系. 除深证综指等金融类指标外, 不同模型还选择了衣着类 CPI 或者 RPI 等价格类指标, 尽管价格类指标通常是滞后指标, 但是价格作为反映经济活动最直接和最敏感的指标之一, 更容易影响政府、企业和居民的经济决策, 因此与经济波动密切相关, 在较长预测步长中能够发挥一定的预测作用.

对比分析图 10 和图 11 可知, 在三个模型向前 12 个月的预测中, 疫情前后重要特征变量未发生明显变化. 如图 11 所示, 金融类指标的平均 Shapley 绝对值依然较大, 比如滞后 3 期的深证综指 (SZCI_3) 在三个模型中均占据首位. 此外, 人民币名义有效汇率 (NERI)、外汇储备 (FER) 和个人存款 (PTD) 等指标也为预测做出较大贡献. 可见, 疫情后金融类指标仍对工业增加值增速向前 12 个月的预测具有重要作用. 这也在一定程度上说明金融是现代市场经济的核心, 随着金融工具的发展和金融化程度的提高, 金融市场的波动更容易传导到实体经济, 影响宏观经济运行 (邓创等, 2018).

下面以预测效果较好的 GBDT 模型为代表, 对主要金融类特征变量 Shapley 值的动态变化进行详细分析. 图 12 显示, 滞后 3 期的深证综指 (SZCI) 和个人存款 (PTD) 的 Shapley 值变化与向前 12 个月的工业增加值增速预测值的变动趋势较为一致, 特别是在 2008 年金融危机和 2020 年新冠疫情时期, 金融类指标与以工业增长为代表的宏观经济的相关性有所增强, 对宏观经济走势具有一定预警作用. 然而, 在 2021 年工业增加值增速触底反弹时, 深证综指的 Shapley 值并未出现明显正向上升, 个人存款的 Shapley 值则出现上升趋势, 这与人们应对不确定性的预防性储蓄上升有关. 人民币有效汇率 (NERI) 和外汇储备量 (FER) 这两个指标不但与国内经济运行相关, 还会受到全球经济形势和重要国际事件的影响. 在 2015 年汇率改革后, 前者的 Shapley 值除部分时点外总体呈现负向波动, 一定程度上反映了复杂多变的国际环境可能会加大汇率的波动性和不确定性, 对工业增长产生一定不利影响. 但国内经济韧性强, 外汇储备充足, 使得外汇储备的 Shapley 值波动相对平稳. 因此, 在进行经济预测时可综合考虑国际和国内因素, 充分发挥金融类指标的预警作用.

虽然金融类指标对工业增长的未来变动具有一定的预警作用, 但在不同时期的作用可能有所不同. 与前文类似, 本节进一步分析不同经济阶段各主要特征变量的平均相对贡献率. 由表 6 可知, 金融危机前的 3 个阶段, 深证综指的平均相对贡献率较大从而支撑其在整个样本期间保持最大的 Shapley 平均值. 与前文向前 1 个月预测的情况有所不同, 深证综指在向前 12 个月的预测中总体具有正向作用, 可能因为 2001 年中国加入 WTO 后, 金融市场提前释放了更多反映经济高速增长的利好信息. 第 4 阶段是金融危机和超强刺激经济时期, 人民

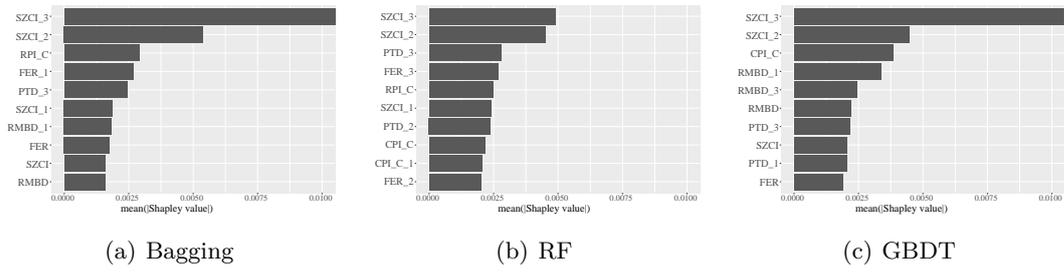


图 10 疫情前三个典型集成学习模型的平均 Shapley 绝对值 ($h = 12$)

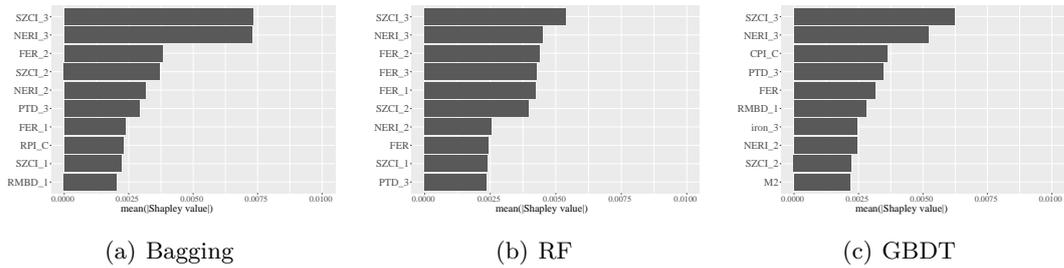


图 11 疫情后三个典型集成学习模型的平均 Shapley 绝对值 ($h = 12$)

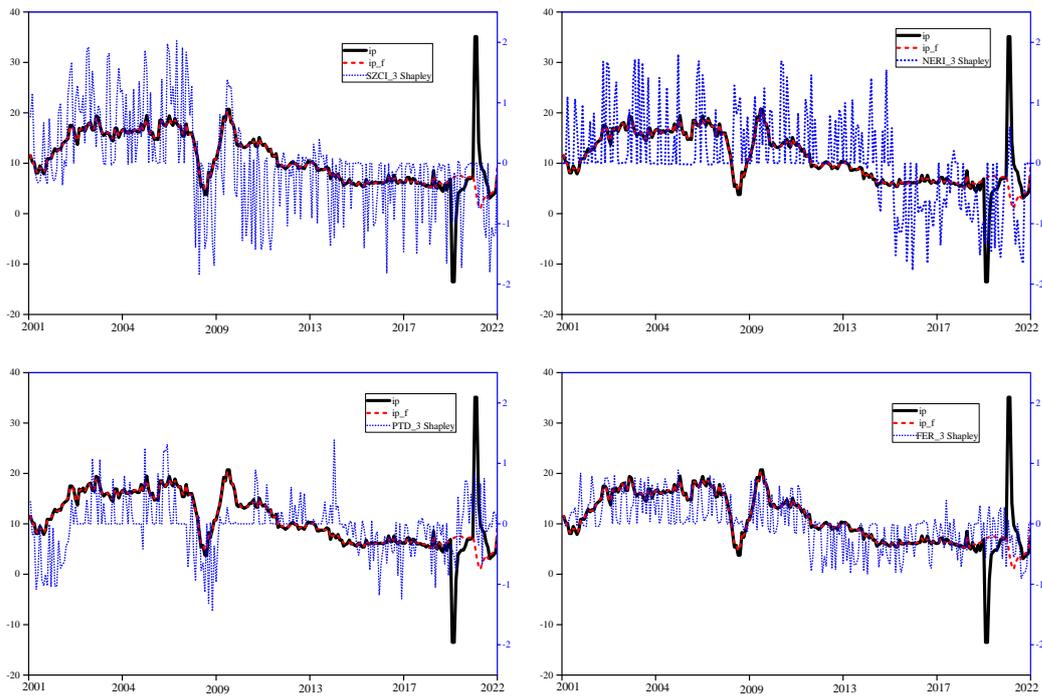


图 12 工业增加值增速 (左轴) 与 GBDT 模型重要特征变量的 Shapley 值变化 (右轴) ($h = 12$)

表 6 向前 12 个月预测时不同阶段主要特征变量的平均相对贡献率 (%)

时间阶段	平均相对贡献率最大		平均相对贡献率最小	
	特征变量	平均相对贡献率	特征变量	平均相对贡献率
阶段 1: 2001 年 04 月 – 2004 年 02 月	深证综指 (SZCI_3)	4.82	个人存款 (PTD_3)	-2.53
阶段 2: 2004 年 03 月 – 2006 年 06 月	深证综指 (SZCI_3)	6.22	发动机产量 (engine)	-0.93
阶段 3: 2006 年 07 月 – 2007 年 11 月	深证综指 (SZCI_3)	5.67	第二产业投资 (inv2_3)	-1.08
阶段 4: 2007 年 12 月 – 2010 年 02 月	名义汇率 (NERI_3)	3.00	个人存款 (PTD_3)	-2.67
阶段 5: 2010 年 03 月 – 2013 年 09 月	名义汇率 (NERI_3)	3.74	深证综指 (SZCI_3)	-3.18
阶段 6: 2013 年 10 月 – 2017 年 05 月	生铁产量 (iron_3)	1.65	衣着价格 (CPI_C)	-4.52
阶段 7: 2017 年 06 月 – 2019 年 12 月	生铁产量 (iron_3)	1.75	名义汇率 (NERI_3)	-5.20
阶段 8: 2020 年 01 月 – 2022 年 06 月	第二产业投资 (inv2_3)	1.86	名义汇率 (NERI_3)	-4.46

注: 与向前 1 个月的预测不同, 当预测步长扩大至 12 个月时, 会损失更多初始样本, 因此, 第 1 阶段的开始时间延后至 2001 年 4 月。

币名义有效汇率表现为较大的正向平均相对贡献率, 这可能由国际金融危机下多数国家的货币贬值引起人民币相对升值所致; 此时个人存款表现为负向影响, 可能是金融危机冲击导致了居民储蓄增长趋缓。金融危机后的第 5 阶段, 金融环境依然不稳定, 深证综指的平均相对贡献率为负, 在一定程度上反映宏观经济经过危机冲击和强刺激后进入调整阶段; 相反, 该时期较稳定的人民币名义有效汇率使其平均相对贡献率表现为正。进入经济新常态的第 6 个阶段, 金融类指标的重要性减弱, 生铁产量和价格水平成为主导变量, 表明实体经济的发展在经济稳定期发挥着关键作用。与此类似, 在疫情前的第 7 阶段, 生铁产量对未来 12 个月的工业增加值增速仍具有较强的预测性, 但受中美贸易摩擦等国际环境影响, 人民币名义有效汇率表现为负向影响, 并延续至疫情发生后的第 8 阶段, 但此时第二产业固定资产投资的平均相对贡献率为正, 反映政策刺激下的基建投资等提速拉动了未来工业增速的回升。整体而言, 在向前 12 个月的工业增加值增速预测中, 金融类指标的平均重要性比较明显, 对经济走势具有一定预警作用, 但不同指标的重要性随时间发生变化。经济运行进入新常态以来, 金融指标的总体作用有所下降, 但 2017 年以后汇率的负向作用有所加大, 这与中美贸易摩擦和汇率波动对国内工业经济的负向影响密切相关, 因而强化金融市场和外部环境的监测预警意义重大。

通过分析上述四种经济情景 (即疫情前预测步长分别为 1 个月和 12 个月与疫情后预测步长分别为 1 个月与 12 个月) 下集成学习模型中重要特征变量的贡献情况可以发现, 疫情冲击导致向前 1 个月预测中的重要特征变量发生明显改变, 除当期工业增加值增速外, 还需

结合生铁产量和发电量增速等指标来提高预测效果。然而,疫情冲击并未明显改变向前12个月预测时的重要特征变量,从该视角来看,新冠疫情对中国工业经济的冲击可能是暂时性的,不会改变工业增长未来走势的基本面,表明中国经济具有较大的韧性。

5 结论

工业增加值增速是反映宏观经济运行的核心指标之一,在面对新冠疫情等重大突发事件时,提高其预测的准确性对宏观调控具有重要意义。因此本文系统评估了AR、T.Factor两种传统计量模型和Bagging、Boosting等回归树集成学习模型在新冠疫情冲击前后对工业增加值增速的预测效果,并采用SHAP方法和本文提出的相对贡献率,识别和分析了不同情形下预测的主要影响指标,得出如下结论:

第一,回归树集成学习模型在工业增加值增速预测中具有一定优势。在疫情前向前1个月和3个月的预测中,此类模型未表现出明显优势,但在向前12个月的预测中,其预测误差则显著降低。在疫情后3至12个月的预测中,回归树集成学习模型的预测准确性均相对较高,其中,GBDT模型表现最佳。

第二,回归树集成学习模型的预测优势与所处的经济环境密切相关。新冠疫情冲击影响了模型的预测效果,整体而言,各模型在疫情后的预测误差均显著增大;在不同预测步长下的效果有所不同,回归树集成学习模型在向前6个月和12个月预测时的误差均相对较小,尤其是向前12个月的预测效果明显优于传统模型。

第三,基于SHAP方法的分析发现,不同特征变量的预测贡献率随预测步长和经济环境的改变而有所差异。在新冠疫情前后向前1个月的预测中,主要特征变量发生明显变化,疫情前使用工业增加值增速的滞后变量即可得到较高的预测精度,但疫情后生铁产量、发电量和投资增速等变量均对预测工业增加值增速有较大影响;然而,在疫情前后的向前12个月预测中,几个主要特征变量则未发生明显改变,但其相对贡献率在不同阶段有所不同,其中,金融类指标在高风险时期的重要性更为突出。从疫情前后不同预测步长下主要特征变量的变化来看,新冠疫情对中国经济的冲击可能是暂时性的,不会影响中国工业增长未来走势的基本面。

未来还有待从以下几个方面进行深入研究:一、结合其他宏观经济指标对机器学习模型中一些更加复杂的集成学习模型和其他非线性机器学习模型进行全面考查;二、筛选相关的高频指标并结合混频计量模型对工业增加值增速进行实时预测,以提高其在异常冲击下的时效性和准确性;三、加强对机器学习模型的解释和深入分析,从而更准确地反映复杂经济环境下变量间的内在联系和变动规律。

参 考 文 献

- 陈磊,孟勇刚,咸金坤,(2019).我国宏观经济景气的实时监测与预测[J].数量经济技术经济研究,36(2):86-102.
- Chen L, Meng Y G, Xian J K, (2019). Real-time Monitoring and Forecasting of China's Macroeconomic Conditions[J]. Journal of Quantitative & Technological Economics, 36(2): 86-102.

- 陈诗一, 于鸿宝, (2021). 基于因子模型的中国宏观经济总量的预测 [J]. 计量经济学报, 1(1): 84–93.
Chen S Y, Yu H B, (2021). Forecast of China's Macroeconomic Aggregate Based on Factor Model[J]. China Journal of Econometrics, 1(1): 84–93.
- 邓创, 张甜, 徐曼, 赵珂, (2018). 中国金融市场风险与宏观经济景气之间的关联动态研究 [J]. 南方经济, (4): 1–19.
Deng C, Zhang T, Xu M, Zhao K, (2018). The Dynamic Association between Financial Market Risks and Business Cycle in China[J]. South China Journal of Economics, (4): 1–19.
- 顾光同, 许冰, (2018). 中国工业增加值的半月预报: 基于宏观月度数据 [J]. 系统工程理论与实践, 38(8): 1983–1993.
Gu G T, Xu B, (2018). Half-month Forecast of China's Industrial and Added Value: Based on the Macro Monthly Data[J]. Systems Engineering — Theory & Practice, 38(8): 1983–1993.
- 洪永淼, 汪寿阳, (2021). 大数据如何改变经济学研究范式?[J]. 管理世界, 37(10): 40–55.
Hong Y M, Wang S Y, (2021). How is Big Data Changing Economic Research Paradigms?[J]. Journal of Management World, 37(10): 40–55.
- 马晓君, 董碧滢, 王常欣, (2019). 一种基于 PSO 优化加权随机森林算法的上市公司信用评级模型设计 [J]. 数量经济技术经济研究, 36(12): 165–182.
Ma X J, Dong B Y, Wang C X, (2019). Design of a Credit Rating Model of Quoted Companies Based on the PSO Optimized Weighted Random Forest Algorithm[J]. Journal of Quantitative & Technological Economics, 36(12): 165–182.
- 肖艳丽, 向有涛, (2022). 金融市场极端风险状态预测模型及其应用 [J]. 金融发展研究, (3): 8–17.
Xiao Y L, Xiang Y T, (2022). Prediction Model of Extreme Risk State in Chinese Financial Market and its Application[J]. Journal of Financial Development Research, (3): 8–17.
- 肖争艳, 陈衍, 陈小亮, 陈彦斌, (2022). 通货膨胀影响因素识别——基于机器学习方法的再检验 [J]. 统计研究, 39(6): 132–147.
Xiao Z Y, Chen K, Chen X L, Chen Y B, (2022). Identifying the Influencing Factors of Inflation: Reexamination Based on Machine Learning Methods[J]. Statistical Research, 39(6): 132–147.
- 熊景华, 茹璟, (2021). 基于随机森林算法和模糊信息粒化的汇率预测组合模型研究 [J]. 数量经济技术经济研究, 38(1): 135–156.
Xiong J H, Ru J, (2021). Research on Exchange Rate Forecasting Integrated Model Based on Random Forest and Fuzzy Information Granulation[J]. Journal of Quantitative & Technological Economics, 38(1): 135–156.
- 张鹏, 党世力, 黄梅雨, (2022). 基于机器学习预测股票收益率的两步骤 M-SV 投资组合优化 [J/OL]. 中国管理科学, 1–14. <https://doi.org/10.16381/j.cnki.issn1003-207x.2021.2308>.
Zhang P, Dang S L, Huang M Y, (2022). Two-stage Mean Semi-variance Portfolio Optimization with Stock Return Prediction Using Machine Learning[J/OL]. Chinese Journal of Management Science, 1–14. <https://doi.org/10.16381/j.cnki.issn1003-207x.2021.2308>.
- 周卫华, 翟晓风, 谭皓威, (2022). 基于 XGBoost 的上市公司财务舞弊预测模型研究 [J]. 数量经济技术经济研究, 39(7): 176–196.
Zhou W H, Zhai X F, Tan H W, (2022). Research on Financial Frauds Prediction Model of Chinese Public Companies with XGBoost[J]. Journal of Quantitative & Technological Economics, 39(7): 176–196.
- Alessi L, Detken C, (2011). Quasi Real Time Early Warning Indicators for Costly Asset Price Boom/Bust Cycles: A Role for Global Liquidity[J]. European Journal of Political Economy, 27(3): 520–533.

- Aras S, Lisboa P J G, (2022). Explainable inflation forecasts by machine learning models[J]. *Expert Systems with Applications*, 207: 117982.
- Bai J, Ng S, (2008). Forecasting Economic Time Series Using Targeted Predictors[J]. *Journal of Econometrics*, 146(2): 304–317.
- Baumeister C, Guérin P, (2021). A Comparison of Monthly Global Indicators for Forecasting Growth[J]. *International Journal of Forecasting*, 37(3): 1276–1295.
- Breiman L, (2001). Random Forests[J]. *Machine Learning*, 45(1): 5–32.
- Bruno G, Lupi C, (2004). Forecasting Industrial Production and the Early Detection of Turning Points[J]. *Empirical Economic*, 29(3): 647–671.
- Buckmann M, Joseph A, Robertson H, (2021). Opening the Black Box: Machine Learning Interpretability and Inference Tools with an Application to Economic Forecasting[M]. *Data Science for Economics and Finance*. Cham: Springer International Publishing, 43–63. https://doi.org/10.1007/978-3-030-66891-4_3.
- Calude C S, Longo G, (2017). The Deluge of Spurious Correlations in Big Data[J]. *Foundations of Science*, 22: 595–612.
- Chapman J, Desai A, (2022). Macroeconomic Predictions Using Payments Data and Machine Learning[J/OL]. arXiv preprint arXiv: 2209.00948.
- Chen T, Guestrin C, (2016). Xgboost: A Scalable Tree Boosting System[C]// *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, California, USA.
- Chuku C, Simpasa A, Oduor J, (2019). Intelligent Forecasting of Economic Growth for Developing Economies[J]. *International Economics*, 159: 74–93.
- Coulombe P G, Marcellino M, Stevanovic D, (2021). Can Machine Learning Catch the COVID-19 Recession?[J]. *National Institute Economic Review*, 256: 71–109.
- Coulombe P G, Leroux M, Stevanovic D, Surprenant S, (2022). How is Machine Learning Useful for Macroeconomic Forecasting?[J]. *Journal of Applied Econometrics*, 37(5): 841–1090.
- Dendramis Y, Kapetanios G, Marcellino M, (2020). A Similarity-based Approach for Macroeconomic Forecasting[J]. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 183(3): 801–827.
- Diebold F X, Mariano R S, (1995). Comparing Predictive Accuracy[J]. *Journal of Business & Economic Statistics*, 13(3): 134–144.
- Friedman J H, (2001). Greedy Function Approximation: A Gradient Boosting Machine[J]. *Annals of Statistics*, 29(5): 1189–1232.
- Galdi G, Casarin R, Ferrari D, Fezzi C, Ravazzolo F, (2023). Nowcasting Industrial Production Using Linear and Non-linear Models of Electricity Demand[J]. *Energy Economics*, 126: 107006.
- Gawthorpe K, (2021). Random Forest as a Model for Czech Forecasting[J]. *Prague Economic Papers*, 30(3): 336–57.
- Giacomini R, White H, (2006). Tests of Conditional Predictive Ability[J]. *Econometrica*, 74(6): 1545–1578.
- Godahewa R, Webb G I, Schmidt D, Bergmeir C, (2023). SETAR-Tree: A Novel and Accurate Tree Algorithm for Global Time Series Forecasting[J]. *Machine Learning*, 112: 2555–2591.
- Golinelli R, Parigi G, (2007). The Use of Monthly Indicators to Forecast Quarterly GDP in the Short Run: An Application to the G7 Countries[J]. *Journal of Forecasting*, 26(2): 77–94.

- Januschowski T, Wang Y, Torkkola K, Erkkilä T, Hasson H, et al. (2022). Forecasting with Trees[J]. *International Journal of Forecasting*, 38(4): 1473–1481.
- Lundberg S M, Erion G, Chen H, DeGrave A, Prutkin J M, et al. (2020). From Local Explanations to Global Understanding with Explainable AI for Trees[J]. *Nature Machine Intelligence*, 2(1): 2522–5839.
- Lundberg S M, Lee S, (2017). A Unified Approach to Interpreting Model Predictions[C]// *Proceedings of the 31st International Conference on Neural Information Processing Systems*: 4768–4777.
- Medeiros M C, Vasconcelos G F R, Veiga Á, Zilberman E, (2021). Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods[J]. *Journal of Business And Economic Statistics*, 39(1): 98–119.
- Meng X L, (2018). Statistical Paradises and Paradoxes in Big Data (I): Law of Large Populations, Big Data Paradox, and the 2016 US Presidential Election[J]. *The Annals of Applied Statistics*, 12(2): 685–726.
- Mullainathan S, Spiess J, (2017). Machine Learning: An Applied Econometric Approach[J]. *Journal of Economic Perspectives*, 31(2): 87–106.
- Nyman R, Ormerod P, (2017). Predicting Economic Recessions Using Machine Learning Algorithms[J/OL]. *ArXiv Working Paper No. arXiv: 1701.01428 [2022-03-21]*. <https://arxiv.org/abs/1701.01428>.
- Ribeiro T M, Singh S, Guestrin C, (2016). “Why Should I Trust You?”: Explaining the Predictions of Any Classifier[C]// *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA*.
- Shapley L S, (1953). *A Value for N-person Games, Contributions to the Theory of Games, Volume II*[M]. Princeton: Princeton University Press.
- Štrumbelj E, Kononenko I, (2010). An Efficient Explanation of Individual Classifications Using Game Theory[J]. *Journal of Machine Learning Research*, 11: 1–18.
- Štrumbelj E, Kononenko I, (2014). Explaining Prediction Models and Individual Predictions with Feature Contributions[J]. *Knowledge and Information Systems*, 41(3): 647–665.
- Tschora L, Pierre E, Plantevit M, Robardet C, (2022). Electricity Price Forecasting on the Day-ahead Market Using Machine Learning[J]. *Applied Energy*, 313: 118752.
- Varian H R, (2014). Big Data: New Tricks for Econometrics[J]. *Journal of Economic Perspective*, 28(2): 3–28.
- Vrontos S D, Galakis J, Vrontos I D, (2021). Modeling and Predicting US Recessions Using Machine Learning Techniques[J]. *International Journal of Forecasting*, 37(2): 647–671.
- Yang J S, Zhao C Y, Yu H T, Chen H Y, (2020). Use GBDT to Predict the Stock Market[J]. *Procedia Computer Science*, 174: 161–171.
- Yoon J, (2021). Forecasting of Real GDP Growth Using Machine Learning Models: Gradient Boosting and Random Forest Approach[J]. *Computational Economics*, 57(1): 247–265.