DOI：10.19581/j.cnki.ciejournal.2019.12.001 网络首发时间：2019-12-20 11:40:55
网络首发地址：http://kns.cnki.net/kcms/detail/11.3536.F.20191219.1123.002.html

跨国增长实证研究的模型不确定性问题：机器学习的视角

刘 岩， 谢 天

[摘要] 过去 30 年间跨国经济增长实证研究领域提出了近 150 个增长决定因素，而全球 200 余个国家(地区)的样本限制意味着在总结跨国增长经验时必须考虑模型不确定性问题。有别于该领域经典文献所使用的传统计量方法，本文探索了新近的机器学习方法对该问题的分析所可能有的贡献。本文从小样本、变量排序、非线性特征三个角度说明具有一定特征的机器学习方法较传统计量方法可以更有效地处理模型不确定性问题。利用标准的跨国经济增长数据集，本文考察了 10 种常见机器学习方法的应用表现，并与 3 种传统计量方法作了比较。结果显示，套袋法与随机森林法及两者的拓展均能在小样本条件下对经济增长决定因素进行有效排序，灵活捕捉数据的非线性特征，让模型不确定性问题化繁为简，得出更为清晰、稳健的结论。本文旨在说明，机器学习方法的应用有助于跨国增长经验事实的归纳与理解，对于补充传统计量方法的局限与不足具有一定的潜力。

[关键词] 跨国经济增长； 机器学习； 模型不确定性； 变量排序； 非线性特征
[中图分类号]F014 [文献标识码]A [文章编号]1006-480X(2019)12-0005-18

一、问题提出

解释世界上不同国家与地区经济增长表现的差异，推断一个国家或地区未来的经济增长状况，寻求恰当的经济增长政策，一直以来都是经济学的核心问题。随着 20 世纪 80 年代末、90 年代初覆盖全球的跨国经济增长与国民经济核算数据库的建立与完善，经济学家得以对“二战”后长期发展的各类经济增长理论进行有效的实证检验，并提出进一步的理论改进意见或经济增长政策建议。以跨国回归分析为核心范式的经济增长研究纲领迄今已推进了近 30 年，取得了丰硕的研究成果，并

[收稿日期] 2019-07-05
[基金项目] 国家自然科学基金青年项目“银行竞争与银行资本对银行信贷标准制定的交互影响”(批准号 71503191)；国家自然科学基金青年项目“大数据环境下模型平均法对金融市场波动率预测的影响研究”(批准号 71701175)；教育部人文社会科学一般项目“稳健异质自回归法对金融市场波动率预测的影响”(批准号 17YJC790174)。
[作者简介] 刘岩，武汉大学经济发展研究中心、经济与管理学院助理教授，经济学博士；谢天，上海财经大学商学院助理教授，经济学博士。通讯作者：谢天，电子邮箱：xietian001@hotmail.com。感谢国家自然科学基金国际合作项目“法、金融与经济增长之再考察”的资助。感谢《中国工业经济》“机器学习在经济学和管理学中的应用”研讨会、浙江工商大学国际商学院讲座与会学者的讨论，感谢操玮、李剑、倪禾、欧卫、潘越的评议，感谢匿名审稿专家和编辑部的宝贵意见，当然文责自负。

且目前仍然是经济增长实证研究的最主要方法[①]。然而，随着跨国实证研究范围与深度的快速增长，学术界也很快意识到这一研究范式实际上存在着诸多不易克服的问题与缺陷，制约了跨国实证研究的进一步发展。在跨国增长回归范式中，一国人均实际产出 y_i 为被解释变量，根据一定的理论引入一系列解释变量 $\boldsymbol{x}_i$（黑体表示列向量），建立回归方程：

$$y_i=\boldsymbol{x}_i^T\boldsymbol{\beta}+\epsilon_i \tag{1}$$

根据样本情况，上述截面回归可拓展为面板回归模型。每个经济增长理论都对特定的解释变量的系数有一个理论推测，研究者进而通过系数的估计值来检验相应的理论推测。由于跨国增长回归通常呈现为上述约化（Reduced Form）线性回归的形式，因而此类模型所固有的问题均会对增长理论的检验产生影响。

Brock and Durlauf（2001）总结了主流的增长回归实证研究面临的 3 方面突出挑战。挑战 1：理论开放性——研究者可以任意添加解释变量到增长回归（1）中。挑战 2：参数异质性——不同国家各异的增长经验难以通过（1）中一组各国同质的参数来解释。挑战 3：因果性——约化线性回归模型中解释变量 $\boldsymbol{x}_i$ 很容易存在与残差项 ϵ_i 的相关性，导致系数估计的内生性问题。上述挑战 1 的实质在于跨国样本的有限（全球只有 200 余个国家和地区）与可能选取的解释变量庞大数量这一基本事实。Durlauf et al.（2005）在《Handbook of Growth Economics》的综述章节中总结了 43 类共计 145 个文献中出现过的作为经济增长决定因素的解释变量，而不同的解释变量及其组合又代表了不同的经济增长理论或者模型。一般而言，具体的跨国增长实证研究都不包括所有文献中考察过的解释变量，而是通过一个先验变量选择过程，聚焦于少数几个解释变量。上述挑战 2 除了传统意义上的解释变量异质性之外，还提示了潜在经济增长决定因素 $\boldsymbol{x}$ 与经济增长之间可能存在的非线性特征。大量的实证文献也反复验证了跨国增长数据样本中存在显著的非线性特征与异质性（Cohen-Cole et al.，2012）。对于挑战 3，在传统回归分析范式下，上述小样本与多变量问题（挑战 1）与非线性问题都很容易导致跨国回归出现内生性问题。由于样本有限，每个跨国回归模型中不可能包含所有的潜在解释变量[②]，但这些变量之间很容易存在相关性，因此，每个特定的跨国回归总是可能出现遗漏变量问题，从而产生内生性。类似地，忽略解释变量可能具有的非线性效应以及解释变量之间的交互效应，也同样会导致约化的线性回归模型出现遗漏变量导致的内生性问题。

Brock and Durlauf（2001）将上述挑战 1 与挑战 2 并称为模型不确定性问题，包括变量排序与选取和变量边际作用设定（即函数形式设定）两方面不确定性。上述讨论表明，模型不确定性是传统的跨国回归分析范式的一个根本性制约因素。过去十余年来学术界不乏克服这一问题的尝试，最具代表性的是 Bayes 模型平均（Bayesian Model Averaging，BMA）方法。BMA 方法的基本思路是给定一组潜在的回归变量，将不同变量组合定义为一个模型，然后考虑大量模型回归估计结果的平均。然而，BMA 方法的基础仍然是线性回归模型，因此，并不能有效克服非线性特征带来的挑战。而对于非线性特征，传统研究范式的主要应对方法是借助于非参数（包括半参数）方法，但由于估计过程的复杂性，非参数方法又无法有效解决大量潜在解释变量带来的变量选择问题。

如果跳出传统线性增长回归模型（1）的局限，考虑更一般的增长决定因素模型，则可以从新的角度来理解上述模型不确定问题，并引入基于机器学习的框架来重新考虑传统增长回归的研究困

① 最新的跨国回归分析的代表性文章为 Acemoglu et al.（2019），其用跨国面板数据分析了可民主化对经济增长的促进作用。尽管使用了新近的因果推断技术，但该文的基本模型仍然是跨国线性回归。

② 根据 Sala-i-Martin et al.（2004）的统计，跨国回归分析中解释变量很少超过 20 个，平均而言少于 10 个。

境。具体而言,可以将经济增长表现 y 与其决定因素之间的关系表示为一个函数:

$$y=f(\boldsymbol{x};\epsilon) \tag{2}$$

式中 $\boldsymbol{x}$ 是一组研究者事前选定并对经济增长具有潜在作用的解释变量。本文将函数 $f(\cdot)$ 称为推测(Prediction)函数①,可以具有任意的形式,代表了研究者所能设想到 $\boldsymbol{x}$ 对 y 的所有可能作用方式。误差项 ϵ 表示所有研究者未能掌握的其他可能影响因素。与传统跨国回归模型(1)相比,(2)式不要求 $\boldsymbol{x}$ 包括的变量数小于样本数,同时也允许 $\boldsymbol{x}$ 对 y 的作用可以呈现任意形式。在传统计量分析框架及特定的样本条件下,研究者有可能通过非参数方法来估计 $f(\cdot)$。但正如上面讨论所指出的,当样本有限而 $\boldsymbol{x}$ 又包含很多变量时,传统的非参数估计存在难以克服的技术障碍;此外,非参数模型中变量选择问题如何处理仍是尚待解决的问题(Henderson et al.,2012)。

机器学习将(2)式看作是对数据特征进行函数拟合的问题:研究者的目标是从数据中确定推测函数 $f(\cdot)$ 的拟合值 $\hat{f}(\cdot)$。机器学习提供了很多新的方法②,可以统一、高效地处理高维度变量排序及非线性拟合问题。本文具体考虑了 10 个监督学习算法:LASSO、回归树(Regression Tree,RT)、套袋法(RT with Bagging,BG)、随机森林法(Random Forest,RF)、神经网络法(Neural Network,NN)、支持向量回归(Support Vector Regression,SVR)、弹性网络(Elastic Net,EN)、最小二乘提升法(Least Square Boosting,LSB),以及 M5P 的两个拓展版本(M5P-BG,M5P-RF)。

为了说明机器学习方法对跨国增长实证分析中模型不确定性问题的可能用途与意义,本文使用 Sala-i-Martin et al.(2004)构建的一个标准跨国增长数据集,对上述 10 种机器学习方法的实证表现进行系统对比。该截面数据集包括 88 个国家的 12 类共 67 个经济增长的解释变量。通过对比这几种方法在经济增长解释变量选择、增长表现解释效能以及增长现象非线性特征刻画等方面的差异,本文力图说明机器学习方法可以有效克服传统方法在应对模型不确定性方面的困境。通过对多维度的解释变量进行全面、一致的排序,并且灵活、全面捕捉经济增长经验数据中广泛存在的非线性特征,基于机器学习的分析框架能够提高研究者对经济增长特征事实的整体认识,帮助研究者分解出经济增长的主要决定因素与其主要作用特征。

本文的贡献集中在方法论的层面。本文从机器学习的理论特性出发,系统论述了这一分析框架对于改善传统跨国增长研究范式所面对的模型不确定性困境的原因与意义。在应用层面,本文将 10 种常用的机器学习方法系统引入跨国增长实证研究中,突出了这些方法在处理多变量与非线性特征方面的灵活性与有效性,并与传统方法进行了多方面的对比,从而说明机器学习方法对传统方法的改进与完善作用。从更宽泛的角度看,本文希望说明机器学习方法对处理社会科学领域广泛存在的模型不确定性问题具有普遍的适用性,机器学习方法也值得研究者将其纳入常用的实证分析工具箱中。

从方法论方面看,跨国增长实证最初均使用回归模型,如 Barro(1991)、Barro and Sala-i-Martin(1992)。其潜在假设是模型设定的解释变量已经可以考虑各国经济增长差异的所有异质性特征,但显然这样的假设是过于严苛的。Durlauf and Johnson(1995)强调了样本分组的重要性,并具体说明

① 为了强调与一般时间序列模型中预测(Forecasting)的区别,本文专门将 Prediction 一词译为推测,以突出其理论、模型推定方面的涵义。经济学中 Prediction 一词强调的是对变量间逻辑关系的判断,而 Forecasting 仅指对变量未来取值的预先测算而已,并不需要这一测算有理论逻辑基础。

② 关于机器学习在社会科学中的应用与前景,Varian (2014)、Athey and Imbens (2017)、Mullainathan and Spiess(2017)、陈硕和王宣艺(2018)、黄乃静和于明哲(2018)提供了全面的综述。Hastie et al.(2009)、Murphy(2012)等提供了更多的方法介绍。

了广泛存在的组间异质性，从而引领一系列后续文献分析跨国增长样本存在的门限效应，包括Canova(2004)、Minier(2007)等。针对广泛存在的非线性特征，一系列文献使用了非参数(包括半参数)方法，如Liu and Stengos(1999)、Durlauf et al.(2001)及Henderson et al.(2012)等。但这类非线性方法并未讨论变量选择问题带来的模型不确定性。

几乎从跨国增长回归研究的兴起之时，研究者对通过线性回归模型所得结果的稳健性就持谨慎态度。Levine and Renelt(1992)应用极值界限(Extreme Bound)方法，指出当时大部分跨国回归系数均不显著。随后的文献开始利用Bayes方法评估跨国回归系数的稳健性，并最终形成了以Fernandez et al.(2001)与Sala-i-Martin et al.(2004)为代表的Bayes模型平均法。但BMA方法几乎都只在线性回归框架下讨论变量选择带来的模型不确定性问题，并未涉及非线性特征与变量选择并存的模型不确定性。

已有文献中，将机器学习方法应用到经济增长研究中的工作较少。Varian(2014)在其述评文章中讨论了应用LASSO等方法进行跨国回归变量选择的可能性。Durlauf and Johnson(1995)使用了RT的方法，作为其分样本估计的稳健性说明；Tan(2010)使用了RT方法的一个变种来分析地理、制度等因素对经济增长俱乐部效应的影响。但这些文献均未从机器学习的角度来考察RT方法的特征，并将其与模型不确定性问题相联系。此外，这些文献无一例外没有考虑RT方法近期在机器学习领域的重要拓展，如BG与RF及其衍生的方法。

二、模型不确定性的机器学习分析框架

从是否存在已知的被解释变量y的角度看，机器学习方法可以被分为监督学习和非监督学习，前者中y是已知拟合目标。经济学问题的特性决定了研究者关注的被解释变量往往是已知的，因此，主要考虑监督学习方法。解释变量的常见类型包括类别型变量和数值型变量两种。针对本文研究的问题，此处仅考虑数值型监督学习方法。

数值型变量的监督学习算法，可视作在给定样本$X=\{y_i,\boldsymbol{x}_i\}_{i=1}^{N}$之下对被解释变量$y$与解释变量$\boldsymbol{x}$间的(未知)推测函数关系$y=f(\boldsymbol{x};\epsilon)$，寻找最优拟合$\hat{y}=\hat{f}(\boldsymbol{x})$的问题。其中，解释变量$\boldsymbol{x}_i$包括$K$个分量，可写为$\boldsymbol{x}_i=(x_{1i},\cdots,x_{Ki})^T$的形式，并且解释变量个数$K$可以接近或超过样本量$N$。不同的机器学习算法使用不同的最优近似标准，对应不同的拟合函数$\hat{f}$构造方式。机器学习方法较少受到传统计量方法的模型设定限制，可以灵活捕捉数据本身的特性，处理变量不确定性与数据非线性特征等。本文力图说明机器学习方法的这些特性正好可以用来处理经济增长的模型不确定性问题。

推测函数拟合误差可以做进一步分解，从而便于系统对比不同机器学习方法的特性。具体而言，依照Hastie et al.(2009)的方法，本文将样本点$\boldsymbol{x}_0$处的拟合误差做如下分解：

$$\mathrm{E}[(f(\boldsymbol{x}_0;\epsilon)-\hat{f}(\boldsymbol{x}_0))^2\mid\boldsymbol{x}=\boldsymbol{x}_0]=$$

$$\mathrm{E}[(f(\boldsymbol{x}_0;\epsilon)-f(\boldsymbol{x}_0;0))^2\mid\boldsymbol{x}=\boldsymbol{x}_0]+(\mathbb{E}\hat{f}(\boldsymbol{x}_0)-f(\boldsymbol{x}_0;0))^2+\mathbb{E}(\hat{f}(\boldsymbol{x}_0)-\mathbb{E}\hat{f}(\boldsymbol{x}_0))^2=$$

$$\text{Residual error}+\text{Bias}^2(\hat{f}(\boldsymbol{x}_0))+\text{var}(\hat{f}(\boldsymbol{x}_0)) \tag{3}$$

其中，残差误差是由于不可观测的残差项ϵ引起，属于无法消除的因素；偏差(Bias)代表拟合推测值与样本真实值之间的差距；而方差(Var)表示由于样本随机性造成了推测函数拟合值的不确定性。为了减小偏差，一般需要机器学习方法能够灵活捕捉数据的非线性特征；而为了缩小方差，则需

要机器学习方法具有较强的稳健性，能够适应样本数据集的变动。每个机器学习方法在进行推测函数拟合时，通常面对偏差与方差的权衡取舍问题：如果模型越复杂，通常而言拟合精度更高、偏差更小，但此时往往伴随了推测函数拟合不确定性增加、稳健性较差的问题。此外，小样本条件下，机器学习方法需要较强的变量排序与选择功能，能够从大量解释变量中识别出最有效的变量，进而在控制偏差的同时，有效削减方差。

为了行文便利，本文将前 5 种算法（LASSO，回归树、套袋法、随机森林法、神经网络法）划分为常规方法，进行重点分析；将后 5 种算法（支持向量回归、弹性网络、最小二乘提升、M5P-BG，以及 M5P-RF）划分为进阶方法，主要作为常规方法的对比，突出本文结论。除此之外，本文还考虑了 3 种传统计量方法：Bayes 线性回归（Bayes Linear Regression，BLR）、频率模型平均（Frequentist Model Averaging，FMA）、逐步回归（Stepwise Regression，SR）。详细说明请见 Qiu et al.（2019）。

1. 常规方法

（1）LASSO。LASSO 方法与传统的线性回归模型最为接近，其近似推测函数同样是解释变量的线性函数 $\hat{y}=\boldsymbol{x}^T\hat{\boldsymbol{\beta}}$。不同之处在于，传统线性回归确定回归系数 $\hat{\boldsymbol{\beta}}$ 是直接通过最小化残差平方和来实现；而 LASSO 方法在上述目标函数上又增加了系数的绝对值之和（L^1 范数）：$\min_\beta \sum_{i=1}^{N}(y_i-\boldsymbol{x}_i^T\boldsymbol{\beta})^2+\lambda(|\beta_1|+\cdots+|\beta_K|)$，其中，$K$ 为解释变量的数目，$\lambda>0$ 为权重系数。LASSO 的最大特征在于目标函数中 β 的 L^1 范数项会迫使重要性较低的变量系数取值为 0。因此，当备选解释变量较多时，LASSO 可以通过便捷的变量选择功能，减小拟合值的方差，提高推测函数的稳健性。

（2）回归树。Breiman et al.（1984）首先提出了回归树的概念。回归树递归地将数据划分为不同的子样本，并将每个子样本下 y_i 的平均值作为最终推测值。在回归树中，结点 τ 包含 n_τ 个观测值。每个结点能进一步分为两个结点 τ_L 和 τ_R，分别包含 n_L 和 n_R 个观测值。定义结点内的 y_i 离差平方和：$SSR(\tau)=\sum_{i=1}^{n_\tau}(y_i-\bar{y}_\tau)^2$，其中 $\bar{y}_\tau$ 为平均值。在将结点 τ 的 n_τ 个样本拆分为 τ_L 和 τ_R 两部分时，拆分变量及拆分阈值的选择遵循最大化离差平方和（SSR）缩减值的原则（Hastie et al.，2009）：

$$\Delta=SSR(\tau)-SSR(\tau_L)-SSR(\tau_R) \tag{4}$$

结点 τ_L 或者 τ_R 可以视为新的结点并继续拆分过程。该方法从树的顶部开始（完整样本）并将相同的拆分方法应用于所有后续结点，直到触发预先设置好的结点拆分停止规则。拆分过程终止时的结点称为叶子结点。每个叶子结点 l 都对应了一组子样本 X_l，两两不相交且并集为全样本 X。最终的推测拟合值为变量 $\boldsymbol{x}$ 所属子样本 X_l 中 y 的均值 $\bar{y}_l$。Hastie et al.（2009）说明，回归树具有较低的偏差但是方差往往较大，原因在于观测数据中微小的变化可能导致差异很大的回归树分割序列。套袋法和随机森林法纠正了这个问题。

（3）套袋法。Breiman（1996）提出用套袋法（BG），通过构造随机生成的训练集来达到改进 RT 拟合值稳健性的目的。给定一个原始数据集 $\{y_i,\boldsymbol{x}_i\}$，套袋法首先生成 B 个新的训练数据集 $\{y_i^b,\boldsymbol{x}_i^b\}$，$b=1,\cdots,B$，其中每一组新训练集都包含有 N 个样本，且每一个样本都是原始数据的“有放回”抽样。大样本下的 $\{y_i^b,\boldsymbol{x}_i^b\}$ 预计会有 $(1-1/e)\approx63.2\%$ 的独特样本；随后将回归树法应用到每个训练数据集，并构建一个（不剪枝）回归树。套袋法使用每棵回归树计算出一个推测值，而最终推测值是所有 B 个推测值的简单平均。套袋法一方面继承了回归树的非线性特征，能够保证较小的拟合偏差，同时又通过 Boostrap 缩减了拟合值的方差，提高了推测函数的稳健性。

（4）随机森林法。套袋法一个潜在的弱点在于，生成每棵树的模拟样本都有相同的分布，这容易

导致各树之间的相关性较高，从而削弱了 Bootstrap 方法降低拟合值方差的作用。针对这一问题，Breiman(2001)进一步提出了随机森林法(RF)。随机森林法类似于套袋法，两者都依靠对原始数据使用 Bootstrap 方法，进而构建 B 个随机训练集。但随机森林法在构建每棵树的每一个结点拆分时，都从 K 个解释变量中随机抽取 q 个(不放回抽样，默认值为 $K/3$)，并仅从这 q 个变量中寻找拆分点。最终的推测值仍然是每棵树推测值的简单平均值，不过由于每棵树构造过程中的随机性，B 棵树之间的相关性就得到了控制，从而可能进一步增强推测函数拟合值的稳健性。

(5)神经网络法。神经网络法(NN)包括异常多的变种，也是近年来流行的人工智能深度学习方法的基础。本文需要处理的推测函数拟合问题较为简单，因此仅考虑最基本的单层神经网络方法，以及通用的转换函数设定。算法流程如下：①根据确定好的中间神经元层结点数目 L，将输入的 K 个解释变量组合为 L 个不同的线性组合 $\boldsymbol{\theta}_l^T\boldsymbol{x}$，$l=1,\cdots,L$；②通过 Sigmoid 函数将这些线性组合变量转换为中间变量 $z_l=\sigma(\boldsymbol{\theta}_l^T\boldsymbol{x})$；③由 z_l 的线性组合得到最终的推测值 $y=\boldsymbol{\phi}^T\boldsymbol{z}$。算法会根据样本$\{y_i,\boldsymbol{x}_i\}$自动学习并确定相关参数，从而得到推测函数。在后续分析中，本文将 L 设为10。

2. 进阶方法

(1)支持向量回归。支持向量回归方法(SVR)将拟合函数设定为 $y_i=f(\boldsymbol{x}_i)+e_i=\beta_0+\sum_{m=1}^{M}\beta_m h_m(\boldsymbol{x}_i)+e_i$，其中 $h_m(\cdot)$为一组基函数，由选定的核函数生成。系数向量 $\boldsymbol{\beta}$ 通过最小化如下目标函数得到：$H(\boldsymbol{\beta})=\sum_{i=1}^{N}V_\epsilon(y_i-f(\boldsymbol{x}_i))+\frac{\lambda}{2}\|\boldsymbol{\beta}\|^2$，其中，$V_\epsilon(\cdot)$为“$\epsilon$-不敏感误差度量”；当$|r|<\epsilon$ 时 $V_\epsilon(r)=0$，当$|r|\geqslant\epsilon$ 时 $V_\epsilon(r)=|r|-\epsilon$。本文对比了常用的核函数，最终选择了表现效果最好的 Gauss 核函数。

(2)弹性网络法。弹性网络法(EN)可以看做 LASSO 与岭回归(Ridge Regression)的结合，其推测函数仍然为线性形式$\hat{y}_i=\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}$，但系数的确定是通过最小化如下目标函数：$\min_\beta\sum_i(y_i-\boldsymbol{x}_i^T\boldsymbol{\beta})^2+\lambda(\alpha\sum_k|\beta_k|+\frac{1-\alpha}{2}\sum_k\beta_k^2)$，即在 LASSO 的 L^1 惩罚项外增加了 L^2 惩罚项。考虑到对比的简明性，本文在后续分析中仅考虑 $\alpha=0.5$ 这一基本设定。

(3)最小二乘提升法。最小二乘提升法(LSB)以 RT 为基础，通过迭代计算，对样本局部特征不断进行 RT 拟合。每一步迭代中，上一步拟合过程的残差项都自动获得更高的权重，因此进一步拟合过程中会更加偏向对应的样本变量，不断提高整体的拟合精度。最终输出的推测值是多次迭代拟合结果的平均值。尽管形式类似，但这与 BG 或 RF 中通过 Bootstrap 得到多次推测值求平均的方法有本质差异；LSB 的计算过程中，样本集都是固定不变的，多次推测值来自迭代拟合，因此，其本质上是对固定样本特征的不断逼近，而这也导致 LSB 容易出现过度拟合问题①。

(4)M5P-BG 与 M5P-RF。M5P 是由 Wang and Witten(1997)在 Quinlan(1992)的 M5 算法基础上改进得到的，可视作 RT 方法的拓展。M5P 与 RT 有两方面差异。一方面是 M5P 在进行样本切割时，目标不是 RT 之下的最小化两分支的残差平方和，而是最小化经过结点(子)样本量调整的样本标准差之和，即在每个结点拆分时，最大化下列目标函数：

$$\max_{\tau_L,\tau_R} sd(\tau)-\frac{|\tau_L|}{|\tau|}sd(\tau_L)-\frac{|\tau_R|}{|\tau|}sd(\tau_R) \tag{5}$$

其中，$|\tau|$表示该结点之下子样本的样本量，$sd(\tau)$表示对应子样本下 y_i 的样本标准差。另一方面是当 M5P 回归树生成以后，每个最终结点(子样本)处的推测值不再简单地定义为该子样本下 y_i 的

① 关于 LSB 以及更一般的提升法(Boosting)的详尽讨论，见 Hastie et al.(2009)及 Murphy(2012)。

均值，而是由该子样本下的 y_i 与 $\boldsymbol{x}_i$ 的线性回归方程拟合得到。显然，对比 RT，M5P 的这一特征将提升算法的拟合精度。M5P-BG 和 M5P-RF 分别是基础的 M5P 算法与套代法和随机森林法下的结合。通过 Bootstrap 自助抽样，拟合函数的稳健性将得到大幅的提升。

3. 解释变量的排序与筛选

LASSO 方法在进行变量排序时，通常采用以下标准（如 Varian，2014）：随着权重系数 λ 的增大，LASSO 在计算系数向量 β 的取值时会将越来越多的分量设为 0。在这一过程中，本文将系数最先达到 0 的变量认定为最不重要，而系数越晚达到 0 的变量，其重要性越高。弹性网络方法下的变量排序 LASSO 一致。

对于回归树、套袋法、随机森林法以及 M5P-BG、M5P-RF 五个方法，由于其推测函数的构建均是基于决策树，因而各分支结点对应的解释变量自然的具有一个重要性排序，越重要的变量越是靠近树的根部。由于一棵决策树中同一个变量可能出现在多个结点，以及套袋法和随机森林法中同一个变量在各（随机）样本下所生成决策树中对应的位置可能不同，因此，本文仍然需要一个量化的指标来衡量各变量的重要性。

对于回归树，本文计算每一个变量在所有拆分结点上对应的 Δ 值（式 1）的总和并除以结点的数量，以此来衡量变量的重要性（Breiman et al.，1984）。更高的值意味着所对应的推测因子更加重要。由于最小二乘提升法每次迭代均使用 RT 进行拟合，RT 的排序方法可进行自然扩展。

对于套袋法和随机森林法，每棵树都由各自随机抽取的样本生成，并且每一个自助法样本中都会有部分观测数据被排除。这部分被排除的样本（Out-of-Bag，OOB）可用于对回归树进行评估，并且由于没有参加树的构造，不存在过度拟合的风险。基于此，本文通过如下方法进行变量排序（Breiman，1996，2001）：①计算第 b 棵树产生的推测函数在其 OOB 样本中的推测误差；②对所关心的变量 x_k，在 OOB 样本中随机置换该变量观测值的顺序，并根据新生成的样本计算出新的推测误差；③比较两次推测误差的差值，定义为 Δ_k^b。对于每一个变量 $k=1,\cdots,K$，本文计算出所有树 $b=1,\cdots,B$ 的平均值 $\bar{\Delta}_k$，并以 $\bar{\Delta}_k$ 的大小顺序决定 K 个变量的重要性。如果解释变量 k 更加重要，那么通过置换其观测值顺序，这个解释变量会因失去正确的样本对应关系（特别是 x_{ki} 与 y_i 之间的对应关系），导致拟合精度更大的下降，即更高的 $\bar{\Delta}_k$。作为套袋法与随机森林法的拓展，M5P-BG 和 M5P-RF 的变量选择方法与上述方法一致。

神经网络法并不具有一个自然的变量选择结构，所有解释变量通过中间层神经元都进入到最终推测目标中。因此，该方法下的变量选择标准具有一定的任意性。注意到本文考虑的单层神经网络方法中，解释变量到各个神经元需经过一次线性组合（忽略转换函数 σ），而从各个神经元到最终推测值，又需要经过一次线性组合。因此，一个简便的变量重要性测度方法就是将一个变量在两次线性变换中所涉及的系数（标准化后）相乘并求和。具体而言，本文使用如下公式计算一个变量的重要性：$w_k=\sum_{l=1,\cdots,L}\varpi_{kl}\cdot\varpi_l$，其中 $\varpi_l=|\phi_l|/\sum_m|\phi_m|$ 为 $\boldsymbol{\phi}$ 中各系数标准化得到，而 $\varpi_{kl}=|\theta_{kl}|/\sum_m|\theta_{ml}|$ 为 $\boldsymbol{\theta}_l$ 中各系数标准化得到。

与神经网络类似，支持向量回归也不具有自然的变量选择结构。本文仿照 Wang and Witten（1997）的思路衡量每个变量对于推测函数的重要性。具体方法如下：通过 Bootstrap 构造自助抽样样本，在每组样本中打乱除去变量 x_k 之外的其他变量顺序，并用已得到的拟合函数计算该样本下的 R^2。这一 R^2 与未打乱顺序样本的 R^2 的差距，就体现了该组样本下变量 x_k 的重要性。最后，将自助抽

样得到的各组样本下变量 x_k 对应的 R^2 差距取平均，即可衡量该变量的重要性。

4. 方法小结

以回归树为基础的套袋法与随机森林，其算法特征最符合处理模型不确定性问题的内在需求，既能灵活捕捉样本的非线性特征从而减少偏差，又能通过 Bootstrap 减少推测函数拟合值的方差；相应地，M5P–BG 和 M5P–RF 分别是套袋法与随机森林法的拓展，并进一步提高了拟合精度。同时，这一类算法均具有自然的变量排序功能。最小二乘提升是回归树的直接拓展，但容易出现拟合值方差较大的问题。LASSO 及 EN 具有很好的变量排序与选择功能，但仅考虑线性模型。神经网络法理论上具有很好的推测函数拟合灵活性，但缺少内在的变量排序性质，并在小样本条件下容易出现过度拟合与忽略样本非线性特征的问题。SVR 方法可以处理非线性性，但拟合灵活性相对较低。

5. 推测函数关键特征的度量

推测函数的拟合值$\hat{f}$通常是一个高维度的非线性函数，为了尽可能直观、全面地理解该推测函数的特性，本文采取一定的方法，对各个解释变量在$\hat{f}$所起作用进行一个分解、测算。这一分解有 3 方面目标：①测算单个变量的解释效力；②测算多个变量组合的联合解释效力；③分解推测函数的线性趋势与非线性特征[①]。

对单个变量解释效力的测算，本文采取一个“减法”思维，从样本推测值$\hat{f}(x_i)$中减去变量 x_{ki} 的作用，再求和度量总的解释效能损失程度，从而获得单变量 x_k 的解释效力。本文使用如下公式测算单变量解释力度的水平值（Average Explained Prediction Variation）：

$$AEPV(\hat{f};k)=\frac{1}{N}\sum_{i=1}^{N}\left(\hat{f}(\boldsymbol{x}_i)-\frac{1}{N}\sum_{j=1}^{N}\hat{f}(x_{1i},\cdots,x_{k-1,i},x_{kj},x_{k+1,i},\cdots,x_{Ki})\right)^2 \tag{6}$$

其中，$\frac{1}{N}\sum_{j=1}^{N}\hat{f}(x_{1i},\cdots,x_{k-1,i},x_{kj},x_{k+1,i},\cdots,x_{Ki})$剔除了变量 x_k 的样本变动在样本点 i 处对推测值 $\hat{f}(x_i)$的贡献，故两者之差反映了 x_k 在该样本点处对推测值的贡献。进一步定义所有变量联合的解释力度（Average Prediction Variation）：$APV(\hat{f})=\frac{1}{N}\sum_{i=1}^{N}(\hat{f}(\boldsymbol{x}_i)-\frac{1}{N}\sum_{j=1}^{N}\hat{f}(\boldsymbol{x}_i))^2$，则可计算单个变量的百分比解释力度（Fraction of EPV）：

$$FEPV(\hat{f};k)=\frac{AEPV(\hat{f};k)}{APV(\hat{f})} \tag{7}$$

类似地，本文定义两个变量的联合解释力度为：

$$AEPV(\hat{f};k,m)=\frac{1}{N}\sum_{i=1}^{N}\left(\hat{f}(\boldsymbol{x}_i)-\frac{1}{N}\sum_{j=1}^{N}\hat{f}(\cdots,x_{kj},\cdots,x_{m,j},\cdots)\right)^2 \tag{8}$$

再减去 x_k，x_m 两个变量各自的单独解释力度，得到其交互作用解释力度（Interactive EPV）：

$$IEPV(\hat{f};k,m)=AEPV(\hat{f};k,m)-AEPV(\hat{f};k)-AEPV(\hat{f};m) \tag{9}$$

并计算交互作用的百分比解释力度：

① 由于这部分内容技术性较强，本文在正文中直接给出相关公式，更详细的推导与讨论见《中国工业经济》网站（http://www.ciejournal.org）附件。

$$FIEPV(\hat{f};k)=\frac{IEPV(\hat{f};k,m)}{APV(\hat{f})} \tag{10}$$

最后，本文将 $\Delta\hat{y}_{[k]i}\equiv\hat{f}(\boldsymbol{x}_i)-\frac{1}{N}\sum_j\hat{f}(\cdots,x_{kj},\cdots)$ 视作变量样本值 x_{ki} 对推测值 $\hat{y}_i=\hat{f}(\boldsymbol{x}_i)$ 的贡献，并通过下述辅助回归来刻画变量 x_k 对推测值的贡献中，线性趋势项与非线性项各自的贡献。上述回归的 R^2 捕捉了线性趋势的解释贡献，对应的 $1-R^2$ 则测算了变量非线性特征的解释贡献。

$$\Delta\hat{y}_{[k]i}=\alpha+\beta_k x_{ki}+e_i,\ \ i=1,\cdots,N \tag{11}$$

三、实证分析

为了具体说明机器学习方法在经济增长跨国实证分析中的价值，本文使用跨国增长研究领域的标准数据集，系统对比上述 10 种数值型监督学习算法及 3 种传统分析方法的实证表现，以此说明机器学习方法可以有效突破传统方法范式所面对的困境。具体而言，本文详细讨论 5 种常规方法的实证表现，将 5 种进阶方法和 3 种传统计量方法作为对比，以此突出研究重点。

1. 数据样本

本文的数据样本是 Sala-i-Martin et al.(2004)所采用的样本[①]，包括 88 个国家的人均实际 GDP 增速(被解释变量)与 67 个潜在解释变量。这一数据集也是后续 BMA 方法文献所使用的标准数据集。表 1 包括了下文分析中用到的部分变量的代码及释义。表中第一个变量 *GR*6096 是被解释变量。为了能够更充分地讨论后续的实证结果及其经济意义，本文借鉴 Ciccone and Jarociński(2010)的分类方法，将上述 67 个解释变量，依据其理论归属，分为 12 组。具体分组及编号见表 2。

本文所使用的计算程序为 MATLAB 自带的程序包。套袋法、随机森林、神经网络等方法，因为使用了 Bootstrap 抽样，因此，结果具有一定的随机性。为了控制随机性，本文将所有程序的 Bootstrap 数量设置为 B=10000，以保证结果的稳健性[②]。

2. 实证结果

(1)变量排序。本文首先汇报 5 种常规机器学习方法的变量排序与选择结果(见表 3)。为简洁起见，本文只列出了按照第三部分中介绍的重要性测度排序前 10 的变量。作为对比，本文同时列出了 Sala-i-Martin et al.(2004)使用 BMA 方法得到的前 10 显著解释变量。BMA 方法中，变量显著性的通用衡量指标是后验纳入(Posterior Inclusion)概率；Sala-i-Martin et al.(2004)一共发现 18 个变量具有显著性。

表 3 第 2—6 列将机器学习方法选择出来的前 10 重要变量中，与 BMA 前 10 变量重复的变量均加黑标识；此外，如果该变量是 BMA 方法选取的前 10 之外的显著变量，则以下划线标识。在对比结果之前，本文再次指出，BMA 方法从设计上能够有效处理变量选择带来的模型不确定性，识别出不同模型设定下反复出现、具有显著推测效应的变量，但其基础是线性模型，无法应对样本的非线性特征。

从该表结果可见，回归树选取的前 10 变量中只有 4 个是 BMA 显著变量。这反映了回归树方法

① 详见 American Economic Review 网站，67 个变量的完整列表见《中国工业经济》网站(http://www.ciejournal.org)附件。

② 本文尝试了 5000—100000 的 Bootstrap 抽样次数，发现 3 个相应算法的变量选择与推测函数拟合在 10000 次 Bootstrap 抽样下已经基本稳定。更详细的计算程序说明见《中国工业经济》网站(http://www.ciejournal.org)附件。

表 1　　样本变量

变量代码	释义	分组	变量代码	释义	分组
*GR*6096	1960—1996 年人均 GDP 增速				
ABSLATIT	纬度绝对值	4	***LIFE*060**	1960 年预期寿命	2
AIRDIST	距大城市的空中距离	7	***MALFAL*66**	20 世纪 60 年代疟疾流行程度	4
BUDDHA	佛教徒比例	6	*MINING*	采矿业占 GDP 比例	11
CONFUC	儒教比例	6	*OPENDEC*1	1965—1974 年开放度	3
*DENS*60	1960 年人口密度	7	***P*60**	1960 年小学教育	1
DENS*65*C	20 世纪 60 年代沿海人口密度	7	*PRIGHTS*	政治权利	8
EAST	东亚虚拟变量	5	*POP*60	1960 年人口	7
*GDE*1	国防开支	3	*POP*6560	65 岁以上人口比例	2
*GVR*61	20 世纪 60 年代政府消费占比	3	*RERD*	实际汇率扭曲	3
*H*60	1960 年高等教育水平	1	*SIZE*60	经济规模	7
*HINDU*00	印度教徒比例	6	***TROPICAR***	热带地区面积占比	4
***IPRICE*1**	投资品价格	3	*TROPPOP*	热带地区人口占比	4
LANDAREA	土地面积	7	*YRSOPEN*	1950—1994 年开放年数	3

注：加粗变量为 Sala-i-Martin et al.(2004)识别出的前 10 显著变量，加下划线变量表示非前 10 但显著的变量。

资料来源：根据 Sala-i-Martin et al.(2004)。

表 2　　变量分组及编号

编号	分组名称	编号	分组名称	编号	分组名称	编号	分组名称
1	新古典增长理论	4	自然地理特征	7	区位及贸易	10	殖民地历史
2	人口结构	5	地区异质性	8	制度	11	自然资源
3	宏观政策	6	宗教	9	战争及冲突	12	贸易条件

表 3　　常规方法变量排序结果

	LASSO	回归树	套袋法	随机森林法	神经网络法	BMA
1	*YRSOPEN*	***MALFAL*66**	***MALFAL*66**	*BUDDHA*	***IPRICE*1**	***EAST***
2	***EAST***	*BUDDHA*	*BUDDHA*	***MALFAL*66**	*MINING*	***P*60**
3	*TROPPOP*	*ABSLATIT*	***EAST***	***EAST***	***EAST***	***IPRICE*1**
4	***P*60**	*LANDAREA*	*ABSLATIT*	***LIFE*060**	***CONFUC***	***GDPCH*60*L***
5	***MALFAL*66**	*GDE*1	***P*60**	***P*60**	***P*60**	***TROPICAR***
6	***CONFUC***	*GVR*61	***LIFE*060**	*ABSLATIT*	*OPENDEC*1	***DENS*65*C***
7	*RERD*	***IPRICE*1**	*YRSOPEN*	*YRSOPEN*	***DENS*65*C***	***MALFAL*66**
8	***IPRICE*1**	*AIRDIST*	***CONFUC***	***TROPICAR***	*POP*60	***LIFE*060**
9	*BUDDHA*	*SIZE*60	*AIRDIST*	***CONFUC***	*HINDU*00	***CONFUC***
10	*GVR*61	*POP*6560	***TROPICAR***	*H*60	*DENS*60	***SAFRICA***

注：加黑表示该变量出现在 BMA 选取的前 10 显著变量；下划线表示该变量是 BMA 选取的非前 10 但显著的变量。表 4 同。

过度拟合数据的倾向，使其在有限样本下容易关注推测效果并不特别稳健的变量。套袋法在设计时有意增强模型解释效能的稳健性，表 3 的结果也显示出这一点，一共识别出 8 个 BMA 显著变量，且其中有 6 个都是 BMA 前 10 显著的。随机森林法与套袋法接近，理论上在数据呈现高截面相关性时有更强的稳健性，使其选取的解释变量能在更多的情况下显示出良好的解释能力；但当截面相关性有限时，解释效能未必一定超过套袋法。表 3 的结果清晰地显示了这一点，选取的前 10 变量中有 8 个都是 BMA 显著变量，且 6 个是 BMA 前 10 显著变量。表 3 中 LASSO 的结果同样在意料之中。与其他 3 个机器学习方法相比，LASSO 是为一个以线性推测函数为基础的，最接近 BMA 的基础设定；

表 4　　进阶方法与传统计量方法变量排序结果

	M5P-BG	M5P-RF	SVR	EN	LSB	BLR	FMA	SR
1	*SQPI*6090	*BUDDHA*	*SQPI*6090	***EAST***	***MALFAL66***	*DPOP*6090	*ORTH*00	***EAST***
2	*BUDDHA*	***MALFAL66***	*PI*6090	***P60***	*BUDDHA*	*POP*6560	*DENS*65I	***MALFAL66***
3	***MALFAL66***	***EAST***	***SAFRICA***	*YRSOPEN*	*LANDAREA*	*GEEREC*1	*HINDU*00	*SPAIN*
4	***EAST***	***LIFE060***	*PRIGHTS*	***CONFUC***	*ABSLATIT*	*GDE*1	*BUDDHA*	***P60***
5	*ABSLATIT*	***P60***	*ECORG*	*RERD*	*POP*6560	*H*60	*SQPI*6090	***GDPCH60L***
6	***P60***	*ABSLATIT*	*NEWSTATE*	***IPRICE1***	***IPRICE1***	*TOT1DEC*1	*SOCIALIST*	***IPRICE1***
7	***LIFE060***	*YRSOPEN*	*LANDAREA*	*BUDDHA*	*DPOP*6090	*GGCFD*3	*PROT*00	*OTHFRAC*
8	*YRSOPEN*	*H*60	*PROT*00	*GVR*61	*PRIGHTS*	*POP*1560	*EUROPE*	***LIFE060***
9	***CONFUC***	***TROPICAR***	*POP*1560	***DENS65C***	*OPENDEC*1	*GOVNOM*1	*ENGFRAC*	*MINING*
10	*AIRDIST*	***CONFUC***	*ZTROPICS*	***TROPICAR***	*AIRDIST*	*GVR*61	*POP*60	*GGCFD*3

同时增加系数 L^1 范数作为惩罚项，大幅提高了其变量选取的稳健性。与随机森林法表现类似，LASSO 选取的前 10 变量中有 8 个是 BMA 显著变量，其中 5 个是 BMA 前 10 显著变量。与这 4 个方法相对比，神经网络法结果有较大区别，所选择的前 10 变量中，6 个是 BMA 显著变量，其中 5 个是前 10 显著。如图像与语音识别中的广泛应用所说明的，神经网络法的特征是深入挖掘样本间横向或纵向的固定关联特性，从而获取优良的解释能力。但是在跨国经济增长的样本中，样本本身具有的横向固定相关性有限，因此，其数据挖掘的能力无法凸显出来。

作为常规方法的对比与参照，表 4 汇报了 5 种进阶方法与 3 种传统计量方法的变量排序结果。从中可以看出，EN、M5P-BG 和 M5P-RF 总体表现最佳。EN 法选取的 9 个变量都是 Sala-i-Martin et al.(2004)应用 BMA 方法获得的显著变量，而 M5P-BG 与 M5P-RF 分别选择了 7 个和 8 个显著变量，其中分别有 5 个和 6 个前 10 显著变量。与此不同，SVR、LSB 两个机器学习方法仅选取出 1 个与 3 个显著变量，说明其变量排序与稳健性方面有较大的欠缺。传统计量方法上，BLR 与 FMA 均只选择出 1 个显著变量，甚至远差于最基础的逐步回归法，后者排序选择出了 8 个显著变量，其中有6 个是前 10 显著的。

(2)推测函数有效性。为了评估机器学习方法给出的拟合函数推测能力的有效性，本文系统对比不同方法推测值对应的 R^2。具体而言，给定样本数据$\{y_i, \boldsymbol{x}_i\}_{(i=1,\cdots,N)}$以及一个机器学习方法所得的近似推测函数$\hat{f}$，本文首先计算相应的样本内中心 R^2：$1-\sum_i (y_i-\hat{f}(\boldsymbol{x}_i))^2/\sum_i (y_i-\bar{y})^2$，其中$\bar{y}$为样本均值。由于回归树及其衍生的套袋法与随机森林法，拟合推测函数的构造过程中，均是以样本估计 SSR 为判断基础，因而本质上均是最小化 L^2-风险的推测函数。相应地，这些方法计算出的 R^2 应当在 0 和 1 之间。与此不同，LASSO 考虑的不是纯粹的 L^2-风险，其产生的推测函数$\hat{f}$可能导致R^2 小于0。对于 LASSO 之外的 4 个机器学习方法，其产生的近似推测函数$\hat{f}$均可能会呈现出高度非线性的特征。

上述中心 R^2 反映的是对数据样本整体的拟合情况，从模型不确定性问题角度看，主要着眼于样本中的非线性特征。为了更加突出模型不确定性的另一个来源：变量选择问题，本文暂时将非线性特征搁置，利用 5 个机器学习方法所挑选出的 5 组前 10 解释变量，在线性回归框架下，考察其样本内推测功效。这样做的一个好处在于，本文可以有效对比回归树、套袋法、随机森林法与 LASSO 四个方法在线性约束下的解释能力，避免 LASSO 方法本身 R^2 评价的不适用性。此外，本文还可以对比包括所有 67 个变量的全样本线性回归模型的推测表现，从而更全面地考察机器学习方法的有效性。

表 5 汇报了按照上述两种方式所进行的模型推测效力评估。其中，上半部分报告了使用各个机器学习方法产生的原始（非线性）近似推测函数的 R^2 表现以及推测残差绝对值的平均值（Mean Absolute Error）。总体而言，除 LASSO 之外的 4 种方法，均能较好地捕捉 88 个国家样本数据的总体特征，其中心R^2 没有明显差别。GUM 所代表的包含 67 个解释变量的线性回归模型产生了高于机器学习方法的中心 R^2，但这仅只是由于解释变量数目接近样本量本身这一事实。如果将参数个数考虑进来，GUM 调整 R^2 自然大幅下降。

表 5　　常规方法推测函数有效性评估

	LASSO	回归树	套袋法	随机森林法	神经网络法	GUM
	拟合推测函数样本内解释力					
MAE	0.0336	0.0047	0.0047	0.0050	0.0047	
中心 R^2	−2.3774	0.8696	0.8820	0.8674	0.8886	0.9135
	排序前 10 解释变量样本内线性推测解释力					
中心 R^2	0.5449	0.6215	0.5659	0.5719	0.2141	
调整 R^2	0.4858	0.5723	0.5095	0.5163	0.1121	0.6236

注：GUM 表示一般无约束线性推测模型，此处指包含全部 67 个变量的线性回归模型。表 6 同。

表 5 下半部分的结果将推测模型约束到线性范围内，并且只考虑 5 个机器学习方法各自挑选的前十大变量。由此本文可以暂时忽略样本非线性特征，而只聚焦变量选择不确定性问题。此时中心 R^2 已经呈现出明显差别。尽管回归树倾向于过度拟合数据局部特征，但调整 R^2 的结果说明其所选取的变量仍然能够捕捉样本变动趋势。套袋法与随机森林法由于更加稳健的推测特征，能够较好地捕捉数据的线性趋势，因此，对应的中心 R^2 依然稳健较优。LASSO 由于模型本身的线性特征，其约束推测效果也具有优势。而神经网络法由于较突出的过度拟合问题以及变量选择缺陷，导致所选变量的调整 R^2 与其他方法形成较大差距。比较 5 个方法在受约束线性推测时的调整 R^2，可以完全消除参数数目的影响，并且可以直接与 GUM 的调整 R^2 进行对比。结果显示，除神经网络法外，其他方法所选变量均呈现出较好的线性解释能力，回归树、套袋法及随机森林法所选变量解释效果甚至接近 GUM 中 67 个变量的解释效果。

表 6 进一步报告了 5 种进阶方法与 3 种传统计量方法所得到的推测函数的有效性，以及各方法所选取前十变量的线性趋势解释力。仅考察全样本拟合推测函数的解释力，能够看到拟合效果最好的是最小二乘提升法。这与本文前面对该方法特征的讨论一致：在逐步迭代中，LSB 不断改进对上一步残差值的拟合效果，最终得到的样本内拟合结果自然较高。但这一特征其实造成了 LSB 的过度拟合倾向，这从其变量选择结果及前 10 解释变量的样本内拟合效果就可看出。其余 4 种机器学习方法中，M5P–BG 与 M5P–RF 的样本内解释力保持在较高水平，而 SVR 与 EN 方法的样本内解释力明显较弱。这说明两种方法在解释变量作用受限时（过多的解释变量会导致额外的惩罚），对于数据非线性特征的捕捉均存在较大的局限。三种传统计量方法较好的样本解释力并不意外，因为在小样本条件下，充足的解释变量确保了线性推测函数对样本的拟合精度。

当限制在所选前 10 变量时，全样本拟合精度最高的 LSB 方法并不能很好地捕捉数据的总体趋势特征。相反，EN 方法由于推测函数的线性性，此时并未明显丧失样本内拟合水平；同时，上下两个 R^2 的对比也说明 EN 方法最终所得到的推测函数，几乎只用了其前 10 解释变量。具有类似特征的，还有 SR 方法。与 LSB 明显不同，M5P–BG 与 M5P–RF 对样本的线性趋势捕捉能力仍然很强，甚至接近 EN 方法。这再次说明了通过 Bootstrap 提升回归树推测函数拟合稳健性在小样本下能够有效克服模型不确定性问题。同时，得益于叶子结点推测值拟合效能的提高，M5P–BG 与 M5P–RF 不论

表 6　　进阶方法与传统计量方法推测函数有效性评估

	M5P-BG	M5P-RF	SVR	EN	LSB	BLR	FMA	SR	GUM
	拟合推测函数样本内解释力								
MAE	0.0047	0.0051	0.0057	0.0081	0.0024	0.0060	0.0051	0.0066	
中心 R^2	0.8779	0.8617	0.7031	0.6942	0.9749	0.8270	0.8764	0.7952	0.9135
	排序前 10 解释变量样本内线性推测解释力								
中心 R^2	0.6678	0.7084	0.4168	0.7270	0.5717	0.3311	0.3222	0.7847	
调整 R^2	0.6247	0.6706	0.3410	0.6915	0.5160	0.2442	0.2342	0.7568	0.6236

在整体的推测函数解释力还是在前 10 变量线性推测解释力方面，都要优于基本的 BG 与 RF 方法。最后需要注意，SVR、BLR 与 FMA 三个方法下，前 10 变量样本内线性推测拟合效能均大幅下降。这与这三个方法的变量排序与选择结果是一致的。

(3)推测函数特征分解。上述变量排序和拟合效果的分析已经说明特定的机器学习方法在处理跨国增长模型不确定性问题时具有较强的优势。为了进一步理解机器学习方法的特性，本文利用第三部分中的推测函数关键特征分解、测度方法，对常规机器学习方法所得的推测函数进行分析，突出各方法所选取前十变量的关键特征。①

本文首先汇报 4 个常规方法选取的前 10 解释变量单一解释力度，结果见表 7②。表中 FEPV 一列，报告了每个方法下前 10 变量所具有的解释力度，以百分比表示。4 种方法所选取的前十解释变量基本上同时也是单一解释力度排名前 10 的变量。各方法选取的前 10 变量所具有的单一平均解释力度相差较大，从回归树的最高值 4.73%到随机森林法的最低值 0.48%，说明各种方法的推测函数对变量间的交互作用捕捉程度不尽相同。套袋法与随机森林法下，前 10 变量的平均单一解释力度分别为 0.85%与 0.47%。对比表 5 所述这两个方法良好的总体推测效能，说明套袋法与随机森林法通过捕捉变量间的交互效应，大幅提高了各自的总体推测解释力。

与表 1 中各个变量所属理论分组类别相对照可见，LASSO、回归树、套袋法与随机森林法所识别出的主要推测解释变量，集中在自然地理环境、宗教、人口结构、区位和贸易等方面，更多地反映经济增长的深层次根源(Spolaore and Wacziarg，2013)。归属于新古典增长理论类别的变量，主要为 *P*60，即 1960 年小学教育程度。特别地，1960 年人均 GDP(*GDPCH*60*L*)不但没有进入各方法选取的前 10 解释变量行列，甚至都不属于各方法下具有前 10 解释力度的变量序列③。这说明新古典增长模型的主要理论推测，即经济增长的绝对收敛或者俱乐部收敛在样本数据中很难获得支持。

表 7 中 NL 的一列报告了各个变量在相应推测函数中非线性特征，单位同为百分比，测度方法为 100%减去(11)式辅助回归所得的 R^2。除去 LASSO 方法总是线性推测函数，故非线性性总为 0 之外，其他方法均呈现出一定的非线性特征。其中，回归树的非线性特征最高，平均而言其前 10 解释变量的推测效果中 84%来自非线性变动。与之相比，套袋法与随机森林法通过使用 Bootstrap 抽样方法，平滑了回归树方法频繁出现的局部非线性拟合，因此，变量的平均非线性程度有明显下降。此外，神经网络法产生的推测函数并不具有突出的非线性特征，变量的平均非线性特征仅 20%有余。

① 对进阶方法拟合推测函数的分解、测度详见《中国工业经济》网站(http://www.ciejournal.org)附件。

② 67 个变量的完整结果见《中国工业经济》网站(http://www.ciejournal.org)附件。根据变量排序及推测函数有效性的结果，为节约篇幅，表 7 中将神经网络的结果略去。

③ 详见《中国工业经济》网站(http://www.ciejournal.org)附件。*GDPCH*60*L* 仅在神经网络方法下进入解释力前十行列，在回归树方法中位列 12，在其他 3 种方法中均在解释力度前 20 之外。

本文用热力图的形式考察了 4 种常规方法下前 10 变量的两两交互作用对推测值的贡献。[①]从整体特征看，回归树方法下，存在比较突出的正、负向双重交互作用。与此不同，LASSO、套袋法与随机森林法下，双重交互作用均为正向，且两两变量组合件交互作用差异较大。从幅度看，套袋法与 LASSO 均存在交互作用大于 2%的情况，联系表 7 给出的单一变量解释力度，可见交互作用对推测值的贡献可以非常显著。观察套袋法、随机森林法与 LASSO 三种方法，一个突出的特征是显著的交互作用更容易出现在同类别或相近类别变量之间。如儒家文化(*CONFUC*)、佛教比例(*BUDDHA*)与东亚地区(*EAST*)三者之间，就呈现出显著的两两正向交互作用。类似地，疟疾流行程度(*MALFAL*60)与预期寿命(*LIFE*60)之间，也存在很强的正向交互作用。这些特征说明，在经济增长的深层次决定因素中，不同类别深层次因素或同类别不同因素之间的交互作用对经济增长具有重要的推测意义。同时，这也说明拘泥于单纯线性模型的跨国实证研究会容易忽略掉深层次决定因素间通过交互效应带来的重要影响。

表 7　　LASSO、回归树、套袋法与随机森林法前 10 解释变量特征测度　　单位：百分比

LASSO	FEPV	NL	回归树	FEPV	NL	套袋法	FEPV	NL	随机森林法	FEPV	NL
YRSOPEN	0.88	0.00	*MALFAL66*	14.63	72.65	*MALFAL66*	2.10	29.15	*BUDDHA*	0.95	24.74
EAST	2.31	0.00	*BUDDHA*	12.68	24.68	*BUDDHA*	1.91	23.48	*MALFAL66*	1.02	26.89
TROPPOP	0.00	0.00	*ABSLATIT*	7.19	99.93	*EAST*	1.32	4.53	*EAST*	0.65	4.10
P60	4.38	0.00	*LANDAREA*	4.02	99.61	*ABSLATIT*	0.56	44.12	*LIFE060*	0.70	46.98
MALFAL66	0.35	0.00	*GDE1*	2.98	93.88	*P60*	0.86	29.42	*P60*	0.60	25.06
CONFUC	1.78	0.00	*GVR61*	0.97	95.97	*LIFE060*	0.89	47.02	*ABSLATIT*	0.24	42.92
RERD	0.57	0.00	*IPRICE1*	1.68	98.52	*YRSOPEN*	0.17	23.77	*YRSOPEN*	0.28	21.87
IPRICE1	2.03	0.00	*AIRDIST*	1.26	71.61	*CONFUC*	0.13	28.86	*TROPICAR*	0.07	25.64
BUDDHA	0.88	0.00	*SIZE60*	1.03	85.36	*AIRDIST*	0.06	47.19	*CONFUC*	0.13	28.07
GVR61	0.29	0.00	*POP6560*	0.88	97.64	*TROPICAR*	0.03	25.32	*H60*	0.12	84.49

(4)非线性特征图示。这一部分着重说明机器学习方法在应对数据中可能存在的非线性特征时的显著优势。鉴于前述分析结果已经说明套袋法与随机森林法两个方法的优良表现，此处仅汇报这两个方法所得推测函数的非线性特征。图 1 和图 2 分别对套袋法和随机森林法所选取的前十解释变量，绘制$\hat{y}=\hat{f}(x_k,\boldsymbol{x}_{K\backslash k})$关于变量$x_k$的函数图像，其中$\boldsymbol{x}_{K\backslash k}$表示所有 K 个变量中除去 k 外的 $K-1$ 的变量。显然，推测值$\hat{y}$不仅依赖于x_k的取值，也依赖于$\boldsymbol{x}_{K\backslash k}$的取值。为了反映$\hat{y}$与$x_k$之间的全样本特征，本文考虑 3 种$\boldsymbol{x}_{K\backslash k}$的取值：所有 $K-1$ 个变量固定在各自样本 25%、50%和 75%分位点。

比较图 1 和图 2，本文首先注意到，对两图中同时出现的变量x_k，如 *MALFAL66*、*BUDDHA*，推测值(平均经济增长率)$\hat{y}$与x_k之间的函数关系均高度相似。由于随机森林法所构造的回归树之间相关性天然弱于套袋法，因此，两者最终所产生推测函数的一致性进一步互相印证了其稳健性，同时也说明数据样本中存在如图所反映的稳健特征。

图中的纵坐标均为经济增长速度的水平值，因此，相应解释变量对经济增长的边际作用反映在图中推测拟合值的整体走势与局部非线性变化上。以图 1(a)为例，解释变量 *MALFAL66* 为 20 世纪 60 年代疟疾流行程度。从图中拟合线明显可见，其对经济增长的作用仅在水平上升到一定程度(0.6 以上)才开始显现，在此之前疟疾流行程度并不影响经济增长。类似的情况还出现在图 1(d)1960 年小学教育普及水平与图 1(f)1960 年人均预期寿命等解释变量中。对照图 1 和图 2，也可以观察到，

① 详细内容见《中国工业经济》网站(http://www.ciejournal.org)附件

部分解释变量呈现出较为一致的增长作用,非线性程度较低。如经济开放程度(*YRSOPEN*),不论在哪种方法与何种样本区间下,总体的线性趋势是其解释效能的主要部分。本文不对各个变量的具体非线性特征进行展开分析,而只是再次指出,如果天然地将推测函数限定为线性形式,则会遗漏客观的样本非线性特征信息。

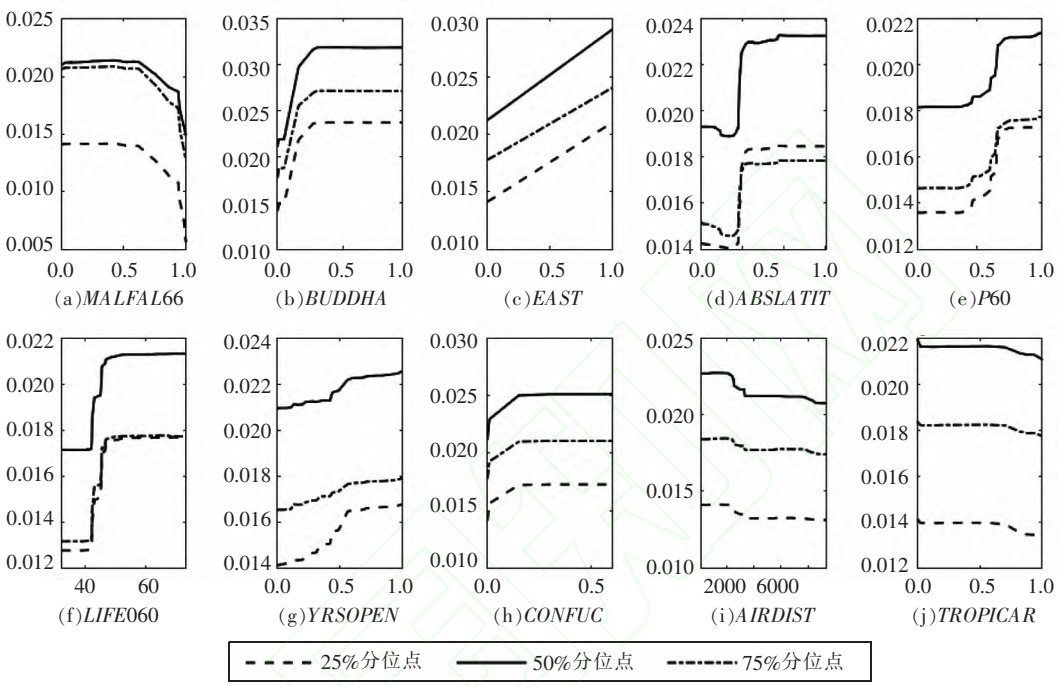

图 1　套袋法推测函数拟合值非线性特征

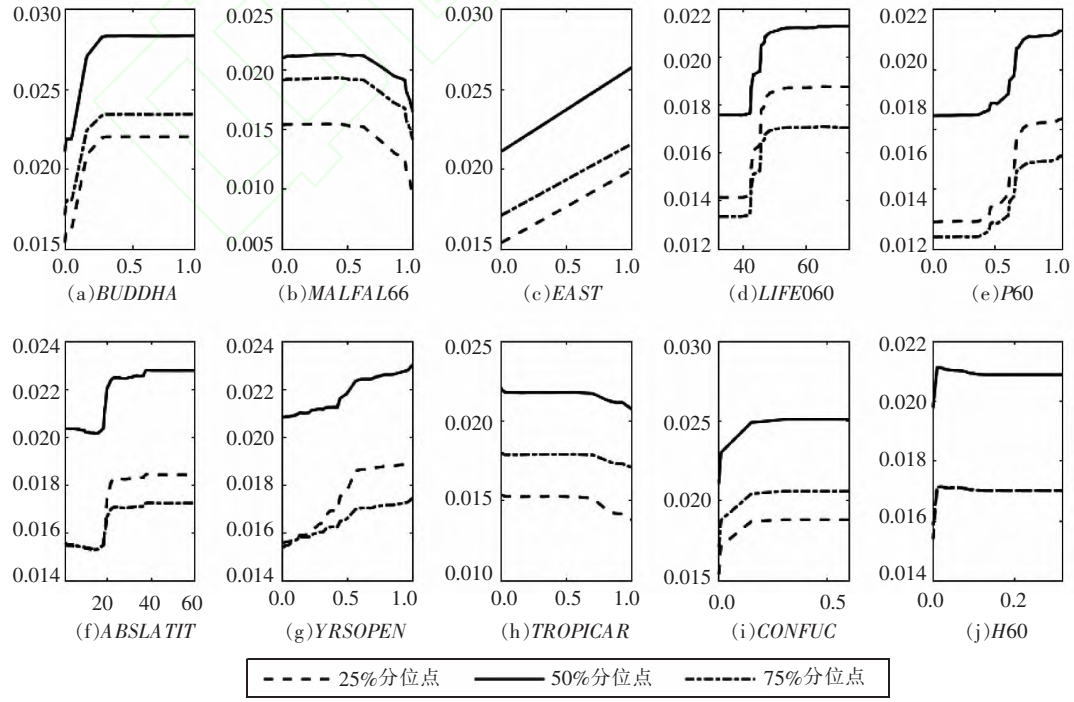

图 2　随机森林法推测函数拟合值非线性特征

四、结论与讨论

经过近30年的发展，跨国实证研究中主流的回归分析范式日益显示出方法论困境。由于样本有限性带来的模型不确定性问题，对传统跨国实证分析方法所得结论的归纳、总结与进一步发展形成了显著的制约。针对这一困境，本文提出可以充分利用新近的机器学习方法，有效缓解或克服传统分析方法存在的局限，从而对传统分析范式形成有益补充。本文系统阐述了机器学习方法的三个优势：适应小样本问题；处理变量排序、选择问题；捕捉样本非线性特征。利用Sala-i-Martin et al.(2004)标准的跨国经济增长数据集，本文研究发现，套袋法与随机森林法及其拓展版本M5P-BG与M5P-RF在跨国经济增长及类似实证问题中具有突出的适用性、有效性、稳健性。利用机器学习方法提供的更高效数据特征提取功能，研究者可以更好地处理跨国增长问题中的模型不确定性问题，从而获得更全面、稳健的经济增长经验典型事实，为未来经济增长理论的发展奠定更好的实证基础。

本文就机器学习方法在经济增长研究中的应用前景提供几点讨论与展望。由于在数据特征识别、提取方面所展现出的优势，机器学习方法近年来在越来越多的学科中得到广泛应用，取得良好效果，经济学也不例外。就经济增长的实证研究而言，机器学习方法的应用或许才刚刚开始。如本文所述，机器学习方法对处理经济增长实证领域的模型不确定性问题有着很好的表现。但就目前初步的应用而言，这个框架也是有其自身局限的。

(1)机器学习方法可以有效解决小样本、多变量条件下的增长推测函数拟合问题，从而在一定程度上削减了由于遗漏变量所带来的内生性影响。但机器学习方法本身并不能轻而易举地解决跨国经济增长实证研究中广泛存在的内生性问题。不过，正如Durlauf et al.(2009)关于经济增长计量问题的综述中所指出的，跨国增长回归的内生性问题难以避免，但研究者仍然可以从中得到有用的信息。这些约化回归中的相关性结论，对于指导增长理论的发展、判断基本理论推断是否符合现实方面具有很大的作用。同理，对机器学习方法的初步应用而言，通过更好地处理模型不确定性问题得到更全面、稳健的解释变量及其边际作用的经验规律，同样对经济增长理论的后续发展有着重要的意义。

(2)从模型不确定性的角度看，目前在对经济增长的主要决定因素及其主要作用尚存诸多争论的情况下，研究者甚至不应当寄期望于机器学习方法来解决内生性问题。人工智能领域图灵奖得主Judea Pearl在其关于因果推断的最新著作中反复指出，变量间因果性的建立不可能仅靠数据挖掘得到，而必须依靠含有相应变量间逻辑关系的因果性理论或模型(Pearl and Mackenzie，2018)。具体在经济增长领域，这就要求在识别每个具体解释变量的因果性时，需要有一个明确的理论或模型作支撑。因此，当研究的对象本身就是经济增长理论或模型的不确定性时，讨论某个变量的因果性问题本身在逻辑上就有一定的不一致性。

(3)将机器学习方法应用于因果推断，是当前一个研究热点(Athey and Imbens，2016；Wagner and Athey 2018；Chernozhukovet al.，2018)。但目前这一方面的工作仍然主要集中在更好地测算处理变量的作用效果从而实现因果推断这一框架中。在跨国经济增长的数据中，相关解释变量自身几乎都不满足处置变量的要求，应用最近的机器学习因果推断方法还有不少挑战。

当然，随着机器学习方法的不断进步与推广，未来将会诞生可以直接应用于类似经济增长这类复杂因果关系问题类型的新方法。希望本文的研究内容和结果可以激发更多研究者思考、探索机器学习在经济增长领域以及经济学其他领域的广阔应用前景。

〔参考文献〕

〔1〕陈硕,王宣艺. 机器学习在社会科中的应用:回顾与展望[R]. 复旦大学工作论文, 2018.

〔2〕黄乃静,于明哲. 机器学习对经济学研究的影响研究进展[J]. 经济学动态, 2018,(7):115-129.

〔3〕Acemoglu, D., S. Naidu, P. Restrepo, and J. Robinson. Democracy Does Cause Growth [J]. Journal of Political Economy, 2019,127(1):47-100.

〔4〕Athey, S., and G. Imbens. Recursive Partitioning for Heterogeneous Causal Effects [J]. Proceedings of the National Academy of Sciences, 2016,113(27):7353-7360.

〔5〕Athey, S., and G. Imbens. The State of Applied Econometrics: Causality and Policy Evaluation [J]. Journal of Economic Perspectives, 2017,31(2):3-32.

〔6〕Barro, R. J. Economic Growth in a Cross Section of Countries [J]. Quarterly Journal of Economics, 1991,106(2):407-443.

〔7〕Barro, R. J., and X. Sala-i-Martin. Convergence[J]. Journal of Political Economy, 1992,100(2):223-251.

〔9〕Breiman, L. Bagging Predictors[J]. Machine Learning, 1996,24(2):123-140.

〔10〕Breiman, L. Random Forests[J]. Machine Learning, 2001,45(1):5-32.

〔11〕Breiman, L., J. Friedman, and R. A. Olshen et al. Classification and Regression Trees [M]. Roca Raton: Chapman and Hall/CRC, 1984.

〔12〕Brock, W. A., and S. N. Durlauf. What Have We Learned from a Decade of Empirical Research on Growth? Growth Empirics and Reality[J]. World Bank Economic Review, 2001,15(2):229-272.

〔13〕Canova, F. Testing for Convergence Clubs in Income Per Capita: A Predictive Density Approach [J]. International Economic Review, 2004,45(1):49-77.

〔14〕Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. Double/Debiased Machine Learning for Treatment and Structural Parameters[J]. Econometrics Journal, 2018,21(1):1-68.

〔15〕Ciccone, A., and M. Jarociński. Determinants of Economic Growth: Will Data Tell [J]. American Economic Journal: Macroeconomics, 2010,2(4):222-246.

〔16〕Cohen-Cole, E. B., S. N. Durlauf, and G. Rondina. Nonlinearities in Growth: From Evidence to Policy[J]. Journal of Macroeconomics, 2012,34(1):42-58.

〔17〕Durlauf, S. N., and P. A. Johnson. Multiple Regimes and Cross-country Growth Behaviour [J]. Journal of Applied Econometrics, 1995,10(4):365-384.

〔18〕Durlauf, S. N., P. A. Johnson, and J. R. W. Temple. Growth Econometrics [A]. Aghion, P., and S. N. Durlauf[C]. Handbook of Growth Economics. Elsevier, 2005.

〔19〕Durlauf, S. N., P. A. Johnson, and J. R. W. Temple. The Methods of Growth Econometrics [A]. Mills,T. C., and K. Patterson. Palgrave Handbook of Econometrics[C]. Palgrave Macmillan UK, 2009.

〔20〕Durlauf, S. N., A. Kourtellos, and A. Minkin. The Local Solow Growth Model[J]. European Economic Review, 2001,45(4):928-940.

〔21〕Fernandez, C., E. Ley, and M. Steel. Model Uncertainty in Cross-country Growth Regressions [J]. Journal of Applied Econometrics, 2001,16(5):563-576.

〔22〕Hastie, T., R. Tibshirani, and J. Friedman. The Elements of Statistical Learning: Data Mining, Inference and Prediction(2 ed)[M]. New York: Springer-Verlag, 2009.

〔23〕Henderson, D. J., C. Papageorgiou, C. F. Parmeter. Growth Empirics without Parameters[J]. Economic Journal, 2012,122(559):125-154.

〔24〕Levine, R., and D. Renelt. A Sensitivity Analysis of Cross-Country Growth Regressions[J]. American Economic Review, 1992,82(4):942-963.

〔25〕Liu, Z., and T. Stengos. Non-linearities in Cross-country Growth Regressions: A Semiparametric Approach[J]. Journal of Applied Econometrics, 1999,14(5):527-538.

〔26〕Minier, J. Nonlinearities and Robustness in Growth Regressions [J]. American Economic Review, 2007,97(2): 388-392.
〔27〕Mullainathan, S., and J. Spiess. Machine Learning: An Applied Econometric Approach[J]. Journal of Economic Perspectives, 2017,31(2):87-106.
〔28〕Murphy, K. P. Machine Learning: A Probabilistic Perspective[M]. Cambridge, Massachusetts: MIT Press, 2012.
〔29〕Pearl, J., and D. Mackenzie. The Book of Why: The New Science of Cause and Effect [M]. Cambridge:Basic Books, 2018.
〔30〕Qiu, Y., Y. Ren, and T. Xie. Weighing Asset Pricing Factors: A Least Squares Model Averaging Approach[J]. Quantitative Finance, 2019,19(10):1673-1687.
〔31〕Quinlan, J. R. Learning with Continuous Classes [R]. Proceedings of the 5th Australian Joint Conference on Artificial Intelligence (AI 92),1992.
〔32〕Sala-i-Martin, X., G. Doppelhofer, and R. Miller. Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach[J]. American Economic Review, 2004,94(4):813-835.
〔33〕Spolaore, E., and R. Wacziarg. How Deep Are the Roots of Economic Development [J]. Journal of Economic Literature, 2013,51(2):325-369.
〔34〕Tan, C. M. No One True Path: Uncovering the Interplay between Geography, Institutions, and Fractionalization in Economic Development[J]. Journal of Applied Econometrics, 2010,25(7):1100-1127.
〔35〕Varian, H. R. Big Data: New Tricks for Econometrics[J]. Journal of Economic Perspectives, 2014,28(2):3-28.
〔36〕Wager, S., and S. Athey. Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests[J]. Journal of the American Statistical Association, 2018,113(523):1228-1242.
〔37〕Wang, Y., and I. H. Witten. Inducing Model Trees for Continuous Classes [R]. Proceedings of the 9th European Conference on Machine Learning Poster Papers, 1997.

Model Uncertainty of Cross-country Growth Empirics: Machine Learning Perspective

LIU Yan[1], XIE Tian[2]

(1. Center for Economic Development Research, Wuhan University, Wuhan 430072, China;
2. College of Business, Shanghai University of Finance and Economics, Shanghai 200433, China)

Abstract: In the past three decades, around 150 factors have been proposed in studies on economic growth. Due to the limited sample size, model uncertainty has become one of the major concerns on analyzing determinants of economic growth. Different from classic growth literature which rely on conventional econometric techniques, we explore and discuss potential contribution of machine learning on this topic. From the perspective of small sample size, variable ranking, and nonlinearity, we demonstrate that specific machine learning techniques are capable of dealing with model uncertainty. We contrast 10 machine learning techniques with 3 conventional econometric methods using standard data set in the literature. The results show that bootstrap aggregation tree and random forest are more robust on capturing the nonlinearity in the data, hence, are more effective on ranking variables in small sample size. Our study implies that the machine learning techniques can be useful alternatives on studying economic growth.

Key Words: cross-country growth; machine learning; model uncertainty; variable selection; nonlinearity
JEL Classification: F43 O40 C52

〔责任编辑:覃毅〕