

突破经济增长跨国实证研究的困境

——基于机器学习方法的新视角*

刘岩[†]

谢天[‡]

武汉大学

厦门大学

第一版：2019年3月15日

摘要：自上世纪90年代开始，随着跨国经济增长数据数据库的建立与普及，跨国回归分析成为了检验经济增长理论与假设的基准实证方法。经过近30年的发展，跨国回归分析的方法论困境日益明显：样本有限性、模型不确定性、结论因果性等3个方面的挑战，对传统跨国实证分析方法的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。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；处理变量选择问题；克服样本非线性问题。与此同时，机器学习方法的这些特性又能够有效缓解传统分析中普遍存在的内生性问题。我们在标准的跨国经济增长数据集上，对LASSO、回归树、套袋法、随机森林等4种机器学习方法的实证效果进行了检验，并与传统分析方法进行了对比，从而进一步说明机器学习方法的优势。研究结果显示，套袋法与随机森林法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。

关键词：跨国经济增长；实证方法；机器学习；模型不确定性

* 本研究受到国家自然科学基金（项目号：71661137003、71503191、71701175、91646206）资助。

[†] 武汉大学经管学院金融系；联系电话：18062100728；电子邮箱：yanliu.ems@whu.edu.cn。

[‡] 厦门大学王亚南经济研究院；联系电话：18659256281；电子邮箱：xietian@xmu.edu.cn。

一、引言

解释世界上不同国家与地区经济增长表现的差异，预测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。随着上世纪 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的经济增长研究纲领，迄今已经推进了近 30 年，取得了异常丰硕的研究成果，并且目前仍然是经济增长实证研究的最主要方法^①。相应的研究焦点，初期以 Solow 模型资本积累机制与经济增长收敛性为核心；接下来在 90 年代蓬勃发展的内生经济增长理论指引下，迅速拓展到教育与人力资本积累方面；在 2000 年前后，延伸到地理、制度、全球一体化等“深层次”决定因素中（Spolaore and Wacziarg 2013, Johnson and Papageorgiou 2019）。一系列的研究成果，极大的丰富了学术界对经济增长浅层（proximate）与深层（deep）决定因素的理解。相应结论，特别是制度等深层决定因素等的作用，在 2000 年初与“华盛顿共识”（Washington consensus）高度融和，一度被认为是完成了对经济增长过程的客观规律性认识（Rodrik 2006）。

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，进而极大制约了跨国实证研究的进一步发展。过去 30 年来，经济增长跨国实证研究的标准范式是跨国增长回归（cross-country growth regression）：以一国人均实际产出 y_i 为被解释变量，根据所考察的理论引入一系列解释变量作为回归变量 \mathbf{x}_i （黑体表示列向量），进而建立回归方程

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \epsilon_i, \quad (*)$$

并可根据样本情况，将上述截面回归拓展为面板回归模型。每个经济增长理论，都对特定解释变量的系数有对应的理论判断，研究者因此可以通过对回归模型系

^① 最新的跨国回归分析代表为 Acemoglu et al. (2019) 等人在 JPE 的待刊文章，用跨国面板数据分析民主化对经济增长的促进作用。尽管使用了最新的因果推断技术，但该文的基本模型仍然是跨国线性回归。

数估计的统计检验，得出支持或不支持相应理论观点的结论。由于跨国增长回归通常呈现为一个约化（reduced form）的线性回归模型，因而此类模型所固有的问题无一例外，均会对跨国增长回归结果的意义产生影响。

在 Brock and Durlauf（2001）的综述中，总结了主流的增长回归实证研究面临的 3 方面突出挑战：

1. 理论开放性（theory open-endedness）——研究者可以想象多种多样的增长理论，选择越来越多的解释变量加入增长回归（*）中；
2. 参数异质性——不同国家各异的增长经验，可能难以通过（*）中一组各国同质的参数来解释；
3. 因果性——约化线性回归模型中，解释变量 x_i 普遍存在与残差项 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（p. 608 及 Appendix B），总结了 43 类共计 145 个文献中使用过的回归变量。因此每一个跨国实证研究，首先要解决的就是变量选择问题。上述挑战 2 可以等价的解释为潜在增长决定因素与经济增长表现之间可能存在显著的非线性性，变量 x 在一个国家对经济增长有促进作用，在另一个国家可能具有抑制作用。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性（见 Durlauf 2009 的评述）。对于挑战 3，在传统回归分析范式下，上述变量选择问题与非线性问题，更是导致跨国回归出现内生性的一个基本原因。由于样本有限，每个跨国回归模型中不可能吸纳所有的潜在解释变量^①，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。类似的，忽略解释变量的可能具有的非线性效应，以及解释变量之间的交叉（非线性）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。

^① Sala-i-Martin et al.（2004）指出，跨国回归分析中解释变量的数量很少超过 20 个，平均而言在 10 个以内。

Brock and Durlauf (2001) 将上述挑战 1 与 2 并称为模型不确定性 (model uncertainty) 问题^①。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, 简称为 BMA) 方法^②。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性带来的挑战。对于非线性性, 传统研究范式的主要应对方法是诉诸于非参数 (包括半参数) 方法^③, 但由于估计过程的复杂性, 非参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。

如果我们跳出传统线性增长回归模型 (*) 的局限, 考虑更一般的生长决定因素模型, 我们就可以从全新的角度来理解上述模型不确定问题给传统增长回归带来的困境, 同时也便于我们引入新的、基于机器学习的研究范式来突破这一困境。具体而言, 我们可以将经济增长表现 y 与其决定因素通过下述方程联系起来:

$$y = f(\mathbf{x}, \epsilon), \quad (**)$$

其中 \mathbf{x} 是一组研究者事前确定的、对经济增长有潜在影响的变量, 而 ϵ 包括了所有未指定因素的影响; 函数 $f(\cdot)$ 可以具有任意的函数形式。与传统跨国回归 (*) 式相比, (**) 式并不需要向量 \mathbf{x} 包括的变量数小于样本数, 同时也可以涵盖 \mathbf{x} 对 y 的任意非线性影响。不可否认, 传统研究框架下, 我们也可以通过非参数方法来估计 $f(\cdot)$ 。但正如上面讨论已经指出, 当样本有限而 \mathbf{x} 又包含很多变量时, 传统的非参数估计存在很多技术障碍; 此外, 非参数模型中变量选择问

^① Brock 与 Durlauf 进一步将该问题的理论根源, 归结到实证模型是否能够使得样本联合分布满足可交换性 (exchangeability) 这一重要概率性质上来。简而言之, 可交换性可理解为对于一组国家增长经验适用的模型, 是否对其他国家的增长经验同样具有适用性。

^② 经典文献包括 Fernandez et al. (2001), Sala-i-Martin et al. (2004), 以及 Durlauf et al. (2008) 等。

^③ 经典文献包括 Liu and Stengos (1999), Henderson et al. (2012) 等。

题带来的模型不确定性如何处理，目前仍是尚待解决的问题（Henderson et al. 2012）。

与传统分析范式不同，我们可以将（**）式看做是在给定样本 $S = \{y_i, x_i\}$ 基础上的预测（prediction）问题：给定样本 S ，研究者的目标是掌握变量 x 的变动如何预测 y 的变动。实现这一目标的关键，在于充分挖掘出样本集合所包含的 x 与 y 的关联信息，包括任何可能存在的非线性关联。对于这一类预测问题，机器学习提供了一套全新的方法论^①，可以灵活、有效的处理多变量、非线性的预测问题。在本文中，我们具体考虑了 4 个较为常用的监督学习算法：LASSO，回归树（Regression Tree，简称为 RT），套袋回归树（RT with Bagging，简称为 BG），以及随机森林法（Random Forest，简称为 RF）。其中，LASSO 主要应用于线性回归形式的变量选择问题，而其他三种方法可以非常深入的刻画数据样本存在的非线性特征，因此我们将分析重点放在后三种方法上。RT、Bagging 与 RF 三种方法都是非常主流的处理小样本数据集的算法，其中 RT 是后两种方法的基础^②。与线性回归模型相比，RT 可以解决多变量与非线性性两个问题，但其自身存在较强的过度拟合倾向。与此不同，Bagging 和 RF 两个方法，通过更有效的样本间交叉学习，可以有效克服 RT 的过度拟合问题，让预测结果具有更强的稳健性。

为了详细说明机器学习在方法论上的对传统跨国实证分析困境的突破意义，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 4 种机器学习方法的实证表现进行了系统对比，并进一步阐述了样本数据存在的非线性性对传统分析方法的制约。该截面数据集包括 12 类共 67 个经济增长的解释变量^③。通过对比这几种方法在增长变量选择、增长表现预测（解释）效能以

^① 具体而言，我们考虑的都是监督学习（supervised learning）方法；另外一大类机器学习方法称为非监督学习，主要用途是对数据进行分类。每一类方法中又各自包含了大量的具体算法。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）提供了全面的综述；中文综述可见陈硕与王宣艺（2018）、黄乃静与于明哲（2018）。

^② RT 在跨国增长回归问题中的应用最早可以追溯到 Durlauf and Johnson（1995），但长期以来经济增长实证研究领域的学者并未将 RT 作为机器学习方法的一种来加以考虑。

^③ 这一数据集仍然是目前最常用的、包括变量最多的跨国增长数据集。

及增长现象非线性特征刻画等方面的差异，我们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，捕捉经济增长经验中广泛存在的非线性性，从而改进我们对经济增长经验机理的认识，有助于实现更准确的因果关系推断。

本文的贡献集中在方法论的层面。就我们掌握的研究现状而言，本文是第一篇系统探讨机器学习方法对跨国增长实证研究意义的文献。我们从机器学习——特别是本文重点考虑的套袋法与随机森林法——的理论特性出发，系统论述了这一新的方法论对于突破传统跨国回归研究范式之困境的原因与意义。在应用层面，本文也是首篇将套袋法与随机森林法引入跨国增长实证研究的文献，并突出了这些方法在处理多变量与非线性性方面的灵活性与有效性。从更宽泛的角度而言，本文意图阐述机器学习方法对解决其他社会科学领域存在的模型不确定性问题所具有的普适性意义^①。

本文的结构安排如下：第二节中，我们对跨国增长实证研究的方法论进行简要的回顾，并对涉及机器学习的相关文献进行梳理；第三节中，我们对机器学习的一般特征，以及本文所应用的四个具体算法，进行较为细致的介绍，特别突出其对于克服模型不确定性问题的原理与直观解释；第四节中，我们汇报上述方法在 Sala-i-Martin et al. (2004) 跨国增长数据集上的应用；第五节是简要的结论。

二、文献综述

跨国增长实证领域最初均使用截面回归模型，如 Kormendi and Meguire (1985)、Barro (1991)、Barro and Sala-i-Martin (1992)、Barro and Lee (1994)。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。Islam (1995) 将跨国回归由截面拓展到面板模型，从而可以通过固定效应来捕捉国家层面不随时间改变的异质性因素。与此同期，Durlauf and Johnson (1995) 强调了样本分组的重要性，并具体说明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

^① 在商业管理领域，参见 Liu and Xie (2019)、Lehrer and Xie (2019) 应用机器学习技术提升电影票房预测的准确性与稳健性。

本存在的门限效应，包括 Hansen（2000）、Canova（2004）、Masanjala and Papageorgiou（2004）、Kourtellos et al.（2007）、Minier（2007a/b）及 Tan（2010）等文章。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）等文章。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。Sala-i-Martin（1997）随后开始利用 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；后续研究还包括 Durlauf et al.（2008）、Ley and Steel（2009/2012）。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。

在机器学习方法的应用方面，就我们所知，仅有 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（pp. 16-17）。Brock and Durlauf（1995）使用了 RT 的方法，作为其分样本估计的稳健性说明；Johnson and Takeyama（2001）利用 RT 方法来说明初始条件对经济增长的影响；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^①。但这些文献均为从机器学习的角度来考察 RT 方法的特征^②，并将其与更一般的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 Bagging 与 RF。

^① 其使用的 RT 算法变种为 GUIDE。

^② 这些文献在使用 RT 方法时，几乎都考虑分段线性回归作为预测函数，从而继承了线性回归模型对解释变量数目的限制，无法在跨国有有限样本条件下考虑大量的回归变量。唯一的例外是 Tan（2010）；该文在分段线性预测之外，考虑了分段常数预测，这与机器学习文献中对 RT 的设定类似。但该文并未深入利用这一特点，从而克服多变量问题对传统分析的局限。

三、机器学习：新的方法论

1. 机器学习方法概要

根据 Mitchell (1997) 的定义，机器学习是指从数据中识别出规律并以此完成预测、分类及聚类的算法总称。从被解释变量 y 是否已知的角度来看，机器学习方法可以被分为监督学习 (supervised learning) 和非监督学习 (unsupervised learning)。其中在监督学习中，被解释变量是已知的；而在非监督学习中，被解释变量是未知的。经济学数据特性决定了我们所研究的被解释变量 y 往往是已知的，因此本文主要考虑监督学习在经济数据环境下的应用。

根据被解释变量 y 的属性，其又可以被细分为类型变量 (categorical variable) 和数值变量 (numerical variable) 两种。类型变量，如颜色、种群、国家等，在机器学习中有很广泛的应用。与之对应的典型机器学习方法包括 Breiman et al. (1984) 提出的分类决策树 (Classification Tree)，Vapnik 等人提出的支持向量机 (Support Vector Machine) 等^①。数值变量在经济学中应用范围更加广阔。和类型变量具有的离散性不同，数值变量往往被假设在连续的实数域之中，因此更加适合回归类型的机器学习方法。比如 Breiman et al. (1984) 提出的回归决策树 (Regression Tree)^②，Tibshirani (1996) 所提出的 LASSO (Least Absolute Shrinkage and Selection Operator) 方法等。

数值型变量的监督学习算法，可视作在给定样本 $\{y_i, \mathbf{x}_i\}_{i=1, \dots, n}$ 之下，对被解释变量 y 与解释变量 (向量) \mathbf{x} 间 (未知) 函数关系 $y = f(\mathbf{x})$ 寻找最佳近似函数 $\hat{y} = \hat{f}(x)$ 的问题。不同的算法可选取不同的最优近似标准，亦可选取不同的近似函数 \hat{f} 构造方式。不同算法之间的比较，一般是通过比较近似函数所给出的预测精度，如均方预测误差，或其他测度。机器学习 (监督学习) 方法关注的重点，在于如何快速、有效获取近似函数，实现针对特定样本数据集的优良

^① SVM 的理论最早由 Vapnik 与 Chervonenkis 在 1963 年提出；现在广泛使用的 SVM 算法，建立在 Boser, Guyon, and Vapnik (1992) 与 Cortes and Vapnik (1995) 等后续工作上。

^② Breiman et al. (1984) 同时提出了分类和回归决策树 (Classification and Regression Tree, CART)。尽管 CART 经常被同时提及，但它其实包含了两个应用场景截然不同的决策树方法。

预测表现。正是因为其数据导向特征，机器学习方法较少受到（先验）模型设定限制，从而可以灵活应对数据本身的特性，如非线性性等。在本文讨论的跨国经济增长实证研究问题上，机器学习方法的这些特性，正巧可以用来处理传统方法所面对的模型不确定性困难。

2. 本文考虑的具体方法

2.1. 算法介绍

本文考虑了 4 个具体的机器学习算法：LASSO、回归树、套袋法与随机森林。这 4 个方法也是目前数值型监督学习领域应用最广泛的四类方法。

LASSO 方法与传统的线性回归模型最为接近，其近似预测函数同样是解释变量的线性函数 $\hat{y} = \mathbf{x}'\hat{\boldsymbol{\beta}}$ 。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

$$\min_{\boldsymbol{\beta}} \sum_{i=1, \dots, n} (y_i - \mathbf{x}'_i \boldsymbol{\beta})^2$$

来实现；而 LASSO 方法在上述目标函数上再添加了系数的绝对值之和：

$$\min_{\boldsymbol{\beta}} \sum_{i=1, \dots, n} (y_i - \mathbf{x}'_i \boldsymbol{\beta})^2 + \lambda(|\beta_1| + \dots + |\beta_K|)$$

其中 K 为解释变量的数目^①， $\lambda > 0$ 为权重系数。LASSO 方法的最大特征，在于系数绝对值之和项倾向于减少非零回归系数的个数。因此，当潜在解释变量较多，需要确定变量有效性时，LASSO 可以实现便捷的变量选择功能。但由于 LASSO 产生的近似方程仍然是解释变量的线性形式，因此当数据样本可能存在显著的非线性特征时，LASSO 的预测表现会受到很大限制。在我们下文的实证分析中，我们将主要侧重于说明 LASSO 的变量选择功能。

Breiman et al. (1984) 首先提出了分类和回归决策树 (Classification and Regression Tree, CART) 的概念。决策树的结构类似于流程图，其中每个内部节

^① 与 LASSO 紧密联系的还有岭回归 (ridge regression)，其差别在于目标函数添加的不是系数绝对值之和，而是平方和。岭回归主要性质在于克服线性回归系数估计的不稳定性，但不具有 LASSO 方法的变量选取特性。

点表示对数据属性的测试与衡量，每个分支（或叶子）代表衡量的结果，并且每个叶节点代表类标签（即在考虑了所有属性特征之后做出的决定）。从 **root**（起始）到 **leaf**（末端）的路径则代表了分类规则。在本文的应用中，我们专注于使用回归树，因为我们实证应用使用的都是实数数据而非分类数据。

回归树递归地将数据划分为不同的本地数组，并将每个组的平均值作为最终预测值。在回归树中，节点 τ 包含 n_τ 个观测值（视作观测值指标 i 的集合）。每个节点只能分为两个叶子，由 τ_L 和 τ_R 分别代表。每片叶子分别包含 n_L 和 n_R 个观测值，并且 $n_L + n_R = n_\tau$ 。我们将节点内的平方和定义为

$$SSR(\tau) = \sum_{i=1}^{n_\tau} (y_i - \bar{y}_\tau)^2,$$

其中 \bar{y}_τ 为平均值。我们将节点 τ 的 n_τ 个观察值拆分为 τ_L 和 τ_R 两个部分，拆分点的选择遵循最大化以下残差平方和（SSR）差值的原则^①

$$\Delta = SSR(\tau) - SSR(\tau_L) - SSR(\tau_R). \quad (1)$$

之后，树叶 τ_L 或者 τ_R 可以视为新的节点并继续拆分过程。我们从树的顶部开始（完整样本）并将相同的方法应用于所有后续节点。

一个完成的回归树会包含很多叶子，而每个叶子包含完整样本的数据子集，并且所有叶子的合集是完整样本。也就是说，与逐步回归类似，回归树每一次拆分类似于模型变量的选择。回归树在每个子集内继续拆分，直到触发某些预先设置好的停止规则。而这很可能导致过度拟合。因此，在实践中我们往往会使用某些标准（或信息准则等）去修剪回归树。这些标准往往会在拟合优度以及模型复杂性（节点数量、叶子数量等）之间寻求一个均衡，以降低过度拟合带来的危害。

^① 具体而言，假设解释（预测）向量 \mathbf{x} 包括 K 个变量 $x^k, k = 1, \dots, K$ 。拆分点的选择是按照如下步骤。首先，对每一个变量 x^k ，将节点 τ 所包括的样本取值 $X(\tau, k) \equiv \{x_i^k\}_{i \in \tau}$ 从小到大排序，然后依次考虑一系列样本分组 $\tau_S^j(k), \tau_B^j(k), j = 1, \dots, |\tau|$ ，其中 $\tau_S^j(k)$ 为 $X(\tau, k)$ 中前 j 小取值 x_i^k 所对应的指标集合， $\tau_B^j(k) = \tau \setminus \tau_S^j(k)$ 。其次，对所有分组，计算 $\Delta^j(\tau, k) = SSR(\tau) - SSR(\tau_S^j(k)) - SSR(\tau_B^j(k))$ 。最后，找到实现 $\max_{j,k} \Delta^j(\tau, k)$ 的分组 $\tau_S^{j^*}(k^*), \tau_B^{j^*}(k^*)$ ，进而将前者记为 τ_L ，后者记为 τ_R ，相应的 $X(\tau, k^*)$ 中第 j 小的取值，记为该节点的拆分点。

使用回归树做预测主要依赖于每个叶子中因变量子集的平均值。这些平均值也构成了回归树的整体拟合值。Hastie et al. (2009) 提供的证据表明, 在实践中, 来自回归树的预测具有较低的偏差但是方差往往较大。这种大方差性主要来自于回归树的“不稳定性”, 即观测数据中的较小的变化可能导致差异很大的分割序列, 并因此导致迥然不同预测值。这种不稳定性是由于回归树向下分割的“等级性质”造成的; 一旦分割过程完成便不可逆。

回归树(包括各种变种)已在理论和实证中被证明具有优秀的预测能力, 我们接下来简要概述以回归树为基础的两种流行方法: 套袋回归树法(RT with Bagging)和随机森林法(Random Forest)。

Breiman (1996) 提出的用 bootstrap 聚合(bootstrap aggregation, 简称为 Bagging)方法, 通过构造随机生成的训练集来达到改进 CART 等方法预测精度和稳健性的目的。给定一个原始数据集 $\{y_i, \mathbf{x}_i\}_{i=1, \dots, n}$, 套袋法首先生成 B 个新的训练数据集 $\{y_i^b, \mathbf{x}_i^b\}_{i=1, \dots, n}$, $b = 1, \dots, B$ 。其中每一组新训练集都包含有 n 个样本, 而每一个样本都是原始数据的“有放回”采样。因为是有放回采样, 某些观测值会有重复, 大样本下的 $\{y_i^b, \mathbf{x}_i^b\}$ 预计会有 $(1 - 1/e) \approx 63.2\%$ 的独特样本。我们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。在套袋法的预测中, 我们首先使用每棵回归树(总共 B 棵)计算出一个预测值。最终预测值是所有预测值的简单平均值。许多研究发现, 结合了数百或数千棵树的套袋法相比只使用单一棵树的回归树法, 在预测表现方面有大幅改进。

对套袋法来说, 其包含的每棵树都是相同分布的。特别是当 \mathbf{x} 中包含有解释力极强的变量, 所有的树都将重点使用这些强变量, 进而导致所有的树都很相似, 并且有较大的相关性。这会导致所有树的偏差与单个树木的偏差相差无几, 然而相关树的数量 B 增加会导致方差下降。

为了减弱树之间的相关性, Breiman (2001) 进一步提出了随机森林法。随机森林法类似于套袋法, 因为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。但对于随机森林, 在构建每棵树时, 每一个节点上的拆分, 我们会从总共 K^{total} 个预测变量中随机抽取 q ($q < K^{total}$) 个预测变量(不

放回采样)。其中, q 的默认值为 $K^{total}/3$ 。而如果 $q = K^{total}$, 该随机森林等同于套袋法。依照此法, 我们最终生成了 B 棵树, 而最终的预测值仍然是每棵树的预测的简单平均值。

研究发现随机森林在预测变量数较大时, 预测表现提升度很明显。如果预测变量中存在许多无关变量, 则有相关变量参与分割的概率会变得较低。同时, 通过随机选择预测因子, 随机森林生成的 B 棵树的相关性要低于套袋法。

总结上述介绍, 我们可以明显看出, LASSO 方法主要着眼于变量选择, 对变量间的非线性关系不予考虑, 也并不着眼最大程度提高预测准确性。与此不同, 回归树方法着眼于获得尽可能好的样本内预测准确性, 同时在预测树的构造过程中, 也连带考虑了变量选择问题。回归树的过度拟合倾向, 是套袋法与随机森林法所希望解决的问题。因此, 后两种方法在继承回归树方法优异的非线性预测性质的同时, 也能够有效解决近似函数的不稳定性, 从而实现预测的稳健性。

2.2. 变量重要性的衡量

LASSO 方法在衡量变量重要性时, 通常采用如下标准。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。这提供了一种衡量变量重要次序的方法, 即将系数最先达到 0 的变量, 认定为最不重要, 而系数越晚变为 0 的变量, 其重要性越高。因此, LASSO 对变量重要性的衡量是一种排序, 而不是一个连续取值的标准。这与剩下 3 个方法具有较大差别。

回归树、套袋法、随机森林等机器学习方法的主要目的是使用已有数据来进行预测。而通过关注预测变量对最终预测结果的影响程度, 我们同样可以识别出哪些预测变量在进行预测时更重要, 并据此对变量的重要性进行排序。我们首先在单棵树上, 计算出每个预测变量的重要性得分, 然后在所有树(若适用)上对分数进行平均, 以获得变量的全局重要性分数。理念上来讲, 最重要的变量即是对(预测)精度变化影响最大的变量。

对于回归树, 我们计算每一个变量在所有拆分节点上 Δ 值(1 式)的总和并除以节点的数量, 以此来衡量变量的重要性。更高的值意味着所对应的预测因子更加重要。

对于套袋和随机森林，每棵树都由各自随机抽取的自举样本生成，并且每一个自举样本中都会部分观测数据被排除。这部分被排除的 **Out-of-Bag** 样本(OOB)可用于对回归树进行评估，而且没有过度拟合的风险，因为 OOB 样本并没有参与回归树的构建过程。而使用套袋和随机森林对变量的重要性进行排序主要依赖于 OOB 样本的应用。我们首先在第 j 棵树的 OOB 样本中，对某一目标变量 i 计算其预测误差 (MSE, MSFE 等)。然后，我们针对变量 i 在 OOB 样本中随机置换其所包含的数据，并根据新生成的 OOB 样本计算出新的预测误差。我们比较两次预测误差的差值，定义为 Δ_i^j 。对于每一个变量 $i = 1, \dots, K$ ，我们计算出所有树 $j = 1, \dots, B$ 的平均值 $\bar{\Delta}_i$ ，并以 $\bar{\Delta}_i$ 的由大到小的排列顺序决定 K 个变量的重要性。这么做的理念是，如果预测因子 i 更加重要，那么通过置换其观测值，这个预测因子会失去因果关系并丧失它的重要性，进而令 $\bar{\Delta}_i$ 增加更多。

四、 实证应用

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优势，我们使用跨国增长研究领域的标准数据，系统对比上述 4 种广泛使用的数值型监督学习算法及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。

1. 数据样本

我们的数据样本是 Sala-i-Martin et al. (2004) AER 文章所采用的样本^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，如 Ley and Steel (2009)。表 1 包括了所有变量的代码及简要释义，详细说明及来源请见 Sala-i-Martin et al. (2004) 原文。

表 1: 样本代码

变量代码	释义	变量代码	释义
GR6096	1960-96 年人均 GDP 平均增长率	LIFE060	1960 年预期寿命

^① 样本数据可从 AER 官网页面 (<http://www.aeaweb.org/articles?id=10.1257/0002828042002570>) 获得。

ABSLATIT	绝对维度	LT100CR	通航水域附近土地面积的比例
AIRDIST	距大城市的空中距离	MALFAL66	60年代疟疾虚拟变量
AVELF	民族语言分化	MINING	采矿业的GDP比例
BRIT	英国殖民地虚拟变量	MUSLIM00	穆斯林比例
BUDDHA	佛教徒比例	NEWSTATE	独立时间
CATH00	天主教徒比例	OIL	石油生产国家虚拟变量
CIV72	公民自由	OPENDEC1	开放度量 1965-74
COLONY	殖民地虚拟变量	ORTH00	东正教比例
CONFUC	儒教比例	OTHFRAC	说外语人口比例
DENS60	人口密度 1960	P60	小学教育 1960
DENS65C	60年代沿海人口密度	PI6090	1960-90 平均通货膨胀率
DENS65I	60年代人口密度	SQPI6090	1960-90 平均通货膨胀率平方
DPOP6090	1960-90 人口增长率	PRIGHTS	政治权利
EAST	东亚虚拟变量	POP1560	15岁以下人口比例
ECORG	资本主义	POP60	1960年人口
ENGFRAC	说英语人口	POP6560	65岁以上人口比例
EUROPE	欧洲虚拟变量	PRIEXP70	1970年主要出口
FERTLDC1	60年代的生育率	PROT00	新教徒比例
GDE1	国防开支分享	RERD	实际汇率扭曲
GDPCH60L	1960年国内生产总值 (log)	REVCoup	革命和政变
GEEREC1	60年代公共教育支出占GDP的份额	SAFRICA	非洲虚拟变量
GGCFD3	公共投资份额	SCOUT	经济外向性
GOVNOM1	60年代名义政府GDP份额	SIZE60	经济规模
GOVSH61	60年代政府占GDP的份额	SOCIALIST	社会主义虚拟变量
GVR61	60年代政府消费占GDP的份额	SPAIN	西班牙殖民地虚拟变量
H60	1960高等教育衡量	TOT1DEC1	60年代的贸易增长
HERF00	宗教衡量	TOTIND	贸易条件排名
HINDU00	印度教徒比例	TROPICAR	热带地区面积占比
IPRICE1	投资品价格	TROPPop	热带地区人口占比
LAAM	拉丁美洲虚拟变量	WARTIME	在1960-90战争花费比例
LANDAREA	土地面积	WARTORN	在1960-90战争参与虚拟变量
LANDLOCK	内陆国家虚拟变量	YRSOPEN	1950-94年经济开放年数
LHPC	油气储量 1993	ZTROPICS	热带地区气候虚拟变量

2. 实证结果

2.1. 变量选择

我们首先汇报 4 种机器学习方法的变量选择结果，结果见表 2。为简洁期间，我们只列出了按照第三节 2.2 中介绍的重要性测度排序前 10 位的变量。注意，不同的方法其重要性衡量标准是不同的。作为对比，我们同时列出了 Sala-i-Martin et al. (2004) 使用 BMA 方法得到的前 10 位显著解释变量。BMA 方法中，

变量显著性的通用衡量指标是后验纳入概率（posterior inclusion probability）；Sala-i-Martin et al.（2004）一共发现 18 个变量具有显著性^①。

表 2：变量选取对比

排序	回归树	套袋法	随机森林	LASSO	BMA
1	MALFAL66	MALFAL66	MALFAL66	<u>YRSOPEN</u>	EAST
2	<u>BUDDHA</u>	<u>BUDDHA</u>	<u>BUDDHA</u>	TROPPOP	P60
3	ABSLATIT	P60	LIFE060	P60	IPRICE1
4	LANDAREA	EAST	EAST	EAST	GDPCH60L
5	GDE1	ABSLATIT	P60	CONFUC	TROPICAR
6	<u>GVR61</u>	LIFE060	ABSLATIT	MALFAL66	DENS65C
7	IPRICE1	CONFUC	H60	RERD	MALFAL66
8	AIRDIST	<u>YRSOPEN</u>	<u>YRSOPEN</u>	IPRICE1	LIFE060
9	SIZE60	<u>SPAIN</u>	IPRICE1	<u>BUDDHA</u>	CONFUC
10	POP6560	IPRICE1	DENS65C	<u>GVR61</u>	SAFRICA

注：代码涵义见表 1；加黑表示该变量出现在 BMA 选取的前十显著变量；下划线表示该变量是 BMA 选取的非前十但显著的变量

为了说明机器学习方法变量选择的有效性，我们在表 2 中，将标准的 BMA 方法变量选取结果作为对比基础。第 2 到 5 列中，我们将机器学习方法选择出来的前十个重要变量中，与 BMA 前十变量重复的变量均加黑标识；此外，如果该变量是 BMA 方法选取的前十之外的显著变量，我们以下划线标识。在对比结果之前，我们再次指出，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、具有显著预测效应的变量，但其基础是线性模型，无法应对非线性性。

从该表结果可见，回归树选取的前十变量中，只有 4 个是 BMA 显著变量。这反映了回归树方法过度拟合数据的倾向，使其在有限样本下容易关注预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预测效能的稳健性，表 2 的结果也显示出这一点，一共识别出 9 个 BMA 显著变量，且其中有 6 个都是 BMA 前十显著的。随机森林与套袋法接近，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预测能力；但当截面相关性有限时，预测效能未必一定超过套袋法。表 2 的结果清晰的显示了这

^① Sala-i-Martin et al.（2004）将变量显著性定义为后验纳入概率超过先验纳入概率，亦即数据样本的信息提高了该变量的纳入概率。其文章中具体选用的先验纳入概率为 10.4%。故这 18 个显著变量的后验纳入概率均大于等于 10.4%。

一点，选取的前十变量中有 8 个都是 BMA 显著变量，且 6 个是 BMA 前十显著变量。表 2 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 是为一个以线性预测函数为基础的，最接近 BMA 的基础设定；同时系数绝对值之和的增加，大幅提高了其变量选取的稳健性。与随机森林方法表现类似，LASSO 选取的前十变量中有 8 个是 BMA 显著变量，其中 5 个是 BMA 前十显著变量。

图 1 进一步绘制了回归树、套袋法与随机森林 3 种方法选择前十变量时各自的量化标准，即第三节 2.2 中描述的重要性指标。

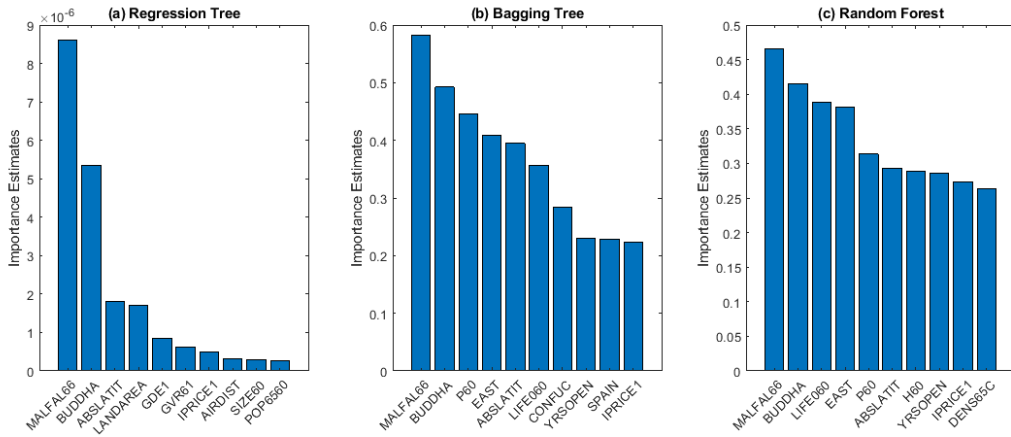


图 1: 变量选择重要性指标

由图可见，回归树倾向于对不同变量给出差异明显的重要性判断。与此不同，套袋法与随机森林法对各自前十大解释变量的重要性评估较为一致，且后者较前者更为均衡。这反映出套袋法与随机森林法通过 Bootstrap 进行模拟交叉验证后，其产生的近似预测函数稳定性更高，变量选择过程不会由于微弱的样本异常导致结果的大幅改变。

2.2. 模型有效性

为了评估机器学习方法给出的近似模型的有效性，我们系统对比不同方法预测值对应的 R^2 。具体而言，给定样本数据 $\{y_i, \mathbf{x}_i\}_{i=1, \dots, n}$ 以及一个机器学习方法所得的近似预测函数 \hat{f} ，我们首先计算相应的样本内中心 R^2 ：

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{f}(\mathbf{x}_i))^2}{\sum_i (y_i - \bar{y})^2},$$

其中 \bar{y} 为样本均值。注意，由于回归树及其衍生的套袋法与随机森林，近似函数（预测树）的构造过程中，均是以样本估计 SSR 为判断基础，因此本质上均是最小化 L^2 -风险的预测方程。相应的，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，因此其产生的预测方程 \hat{f} 可能导致 R^2 小于 0。注意，对于 LASSO 之外的 3 个机器学习方法，其产生的近似预测函数 \hat{f} 一般而言均会呈现出高度非线性的特征。

注意上述中心 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——我们暂时将非线性特征搁置，利用 4 个机器学习方法所挑选出的 4 组前十预测变量，在线性回归框架下，考察其样本内预测功效。这样做的一个好处，在于我们可以有效对比回归树、套袋法、随机森林与 LASSO 四个方法在线性约束下的预测能力，避免 LASSO 方法本身 R^2 评价的不适用性。此外，我们还可以对比包括所有 67 个变量的全样本线性回归模型的预测表现，从而更全面的考察机器学习方法的有效性。

表 3：模型有效性（样本内预测）评估

	回归树	套袋法	随机森林	LASSO	GUM
无约束（非线性）样本内预测					
中心 R^2	0.8696	0.8769	0.8644	-0.3128	0.9135
调整 R^2					0.6236
有约束（前十大变量线性）样本内预测					
中心 R^2	0.5799	0.6838	0.7121	0.7043	
调整 R^2	0.5254	0.6427	0.6747	0.6660	

注：GUM 表示一般无约束线性预测模型，此处即指包含全部 67 个变量的线性回归模型

表 3 汇报了按照上述两种方式所进行的模型预测效力评估。其中上半部分对应了使用各个机器学习方法产生的原始（非线性）近似预测函数的 R^2 表现。总体而言，除 LASSO 之外的 3 种方法，均能较好的捕捉 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 3 下半部分的结果将预测模型约束到线性范围内，并且只考虑 4 个机器学习方法各自挑选的前十大变量。如此我们可以暂时忽略样本非线性性，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。回归树因为过度拟合数据局部特征，而忽略了数据整体的线性趋势特征，造成其中心 R^2 相比无约束非线性情况大幅下降。套袋法与随机森林法由于更加稳健的预测特征，能够较好的捕捉数据的线性趋势特征，因此对应的中心 R^2 明显超过回归树。LASSO 由于模型本身的线性特征，其约束预测效果也具有优势。比较四个方法在约束预测时的调整 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，随机森林所选变量呈现出最好的线性预测能力，甚至超过基准模型为线性的 LASSO 方法。同时，考虑到参数个数后，套袋法、随机森林与 LASSO 所选变量的线性预测能力，均明显高于包含所有 67 个解释变量的 GUM，进一步说明了机器学习方法较传统线性模型在克服变量不确定性问题的明显优势。

2.3. 非线性特征

最后，我们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。由于 LASSO 本身是线性模型，而前面汇报的结果已经说明回归树本身的重要局限，因此我们下面只考虑套袋法与随机森林两个方法。

注意这两个方法最后都产生了近似预测函数 $\hat{f}(\mathbf{x})$ ，所以我们可以通过 \hat{f} 的性质来考察数据样本中的非线性性质。然而解释变量 \mathbf{x} 包括 67 个分量，因此 \hat{f} 是一个高维、非线性的对象，没有简单的方法描述其特征。考虑到这个问题，我们将关注点放在套袋法与随机森林方法各自选择的前十大解释变量 $x^{(k)}, k = 1, \dots, 10$ 之上。具体而言，对每个方法的每个前十大解释变量 $x^{(k)}$ ，我们将除此变量之外的其余 66 个解释变量，依次固定在其样本分布的 25%、50%和 75%三个分点取值上，进而考察 $x^{(k)}$ 在样本取值范围内变动时，相应的经济增长预测值 $\hat{y} = \hat{f}(x^{(k)}, \dots)$ 的变化。

图 2 和 3 分别报告了套袋法与随机森林法近似预测函数的结果。注意这些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曲线的斜率上。以图 2 中子图 (a) 为例，解释变量 1960 年代

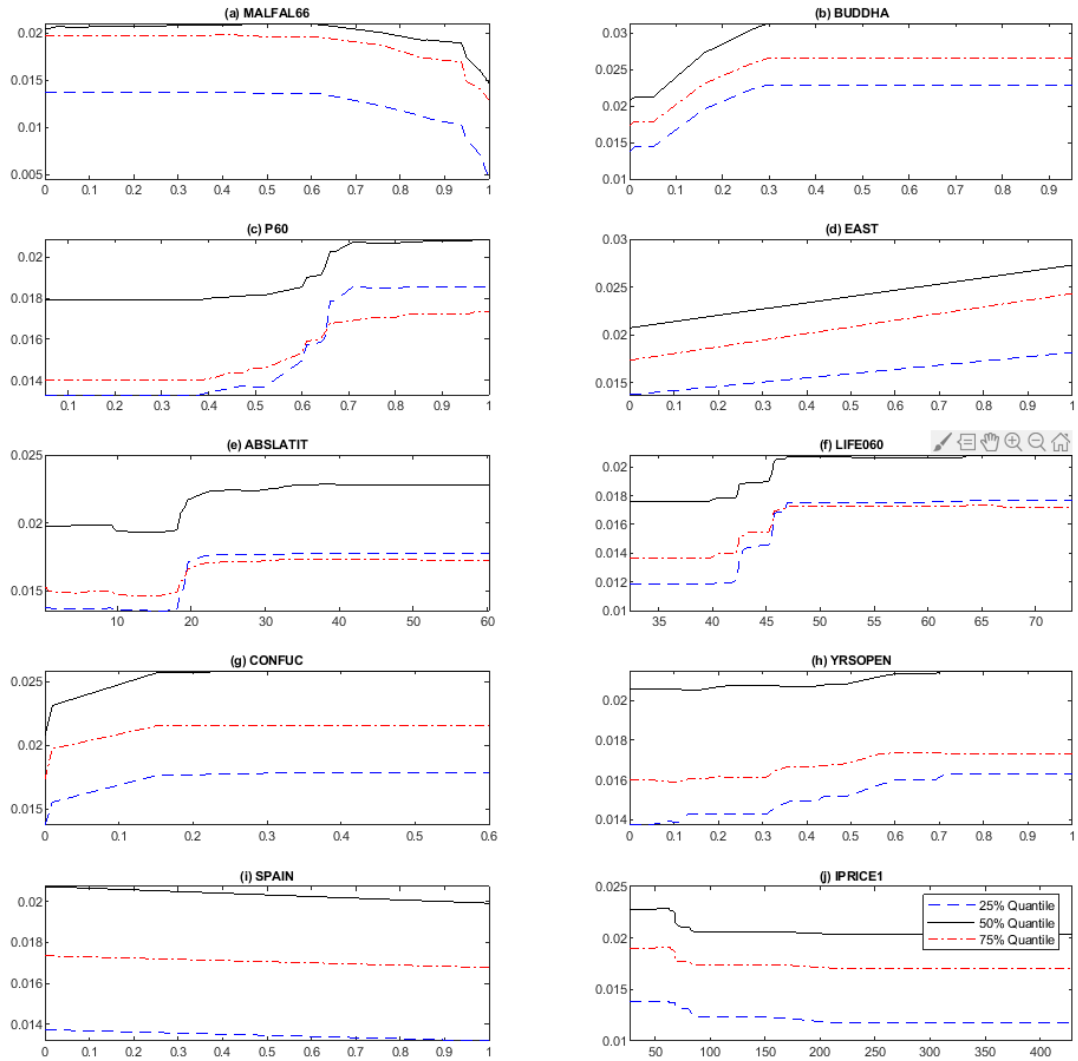


图 2：套袋法近似预测函数

疟疾流行程度对经济增长的作用，仅在其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子图(c) 1960 年代小学教育普及水平与子图 (f) 1960 年代人均预期寿命等解释变量中。观察图中红线、蓝线和黑线的差别，也可以清晰的看出不同样本空间区域对解释变量的作用效果会产生显著影响。例如图 2 和图 3 所示 1950-1994 年经济开放年数占比变量（均对应子图 h），当给定该变量取值（如 0.5），样本取值的变动（如 25%分位点到 50%分位点）可以对经济增长预测值产生近 40%的影响。

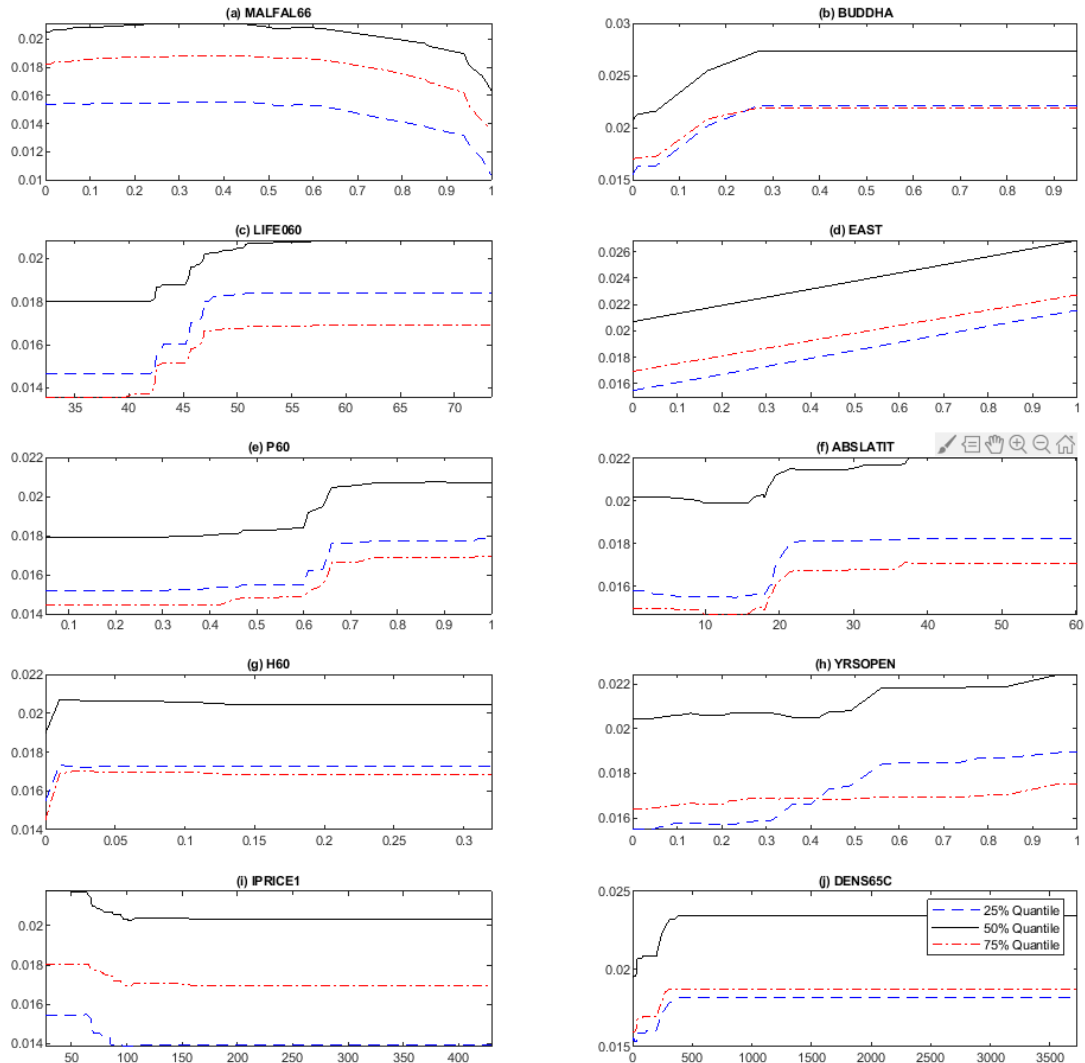


图 3: 随机森林法近似预测函数

五、 结论

经过近 30 年的发展，跨国实证研究中最主流的回归分析范式，日益显示出方法论困境。样本有限性、模型不确定性、结论因果性等 3 个方面的挑战，对传统跨国实证分析方法的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。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；处理变量选择问题；克服样本非线性问题。与此同时，机器学习方法的这些特性又能够有效缓解传统分析中普遍存在的内生性问题。利用 Sala-i-Martin et al. (2004) 标准的跨国经济

增长数据集，我们对 LASSO、回归树、套袋法、随机森林等 4 种最广泛使用的数值型监督学习方法进行了实证测试，以说明其相对传统分析方法的优势。研究结果显示，套袋法与随机森林法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。

参考文献

- [1] Athey S, Imbens G W. The State of Applied Econometrics: Causality and Policy Evaluation[J]. *Journal of Economic Perspectives*. 2017, 31(2): 3-32.
- [2] Barro R J. Economic Growth in a Cross Section of Countries[J]. *Quarterly Journal of Economics*. 1991, 106(2): 407-443.
- [3] Barro R J, Lee J W. Sources of Economic Growth[J]. *Carnegie-Rochester Conference Series on Public Policy*. 1994, 40: 1-46.
- [4] Barro R J, Sala-i-Martin X. Convergence[J]. *Journal of Political Economy*. 1992, 100(2): 223-251.
- [5] Boser B E, Guyon I M, Vapnik V N. A Training Algorithm for Optimal Margin Classifiers[C]. *ACM*, 1992.
- [6] Breiman L. Bagging Predictors[J]. *Machine Learning*. 1996, 24(2): 123-140.
- [7] Breiman L. Random Forests[J]. *Machine Learning*. 2001, 45(1): 5-32.
- [8] Breiman L, Friedman J, Olshen R A, et al. *Classification and Regression Trees*[M]. Boca Raton: Chapman and Hall/CRC, 1984.
- [9] Brock W A, Durlauf S N. 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.
- [10] Canova F. Testing for Convergence Clubs in Income Per Capita: A Predictive Density Approach[J]. *International Economic Review*. 2004, 45(1): 49-77.
- [11] Cortes C, Vapnik V. Support-vector Networks[J]. *Machine Learning*. 1995, 20(3): 273-297.
- [12] Durlauf S N. The Rise and Fall of Cross-Country Growth Regressions[J]. *History of Political Economy*. 2009, 41(Suppl_1): 315-333.
- [13] Durlauf S N, Johnson P A. Multiple Regimes and Cross-country Growth Behaviour[J]. *Journal of Applied Econometrics*. 1995, 10(4): 365-384.
- [14] Durlauf S N, Johnson P A, Temple J R W. *Growth Econometrics*[M]. Aghion P, Durlauf S N, Elsevier, 2005: 1A, 555-677.
- [15] Durlauf S N, Kourtellos A, Minkin A. The Local Solow Growth Model[J]. *European Economic Review*. 2001, 45(4): 928-940.
- [16] Durlauf S N, Kourtellos A, Tan C M. Are Any Growth Theories Robust?[J]. *Economic Journal*. 2008, 118(527): 329-346.
- [17] Easterly W, Kremer M, Pritchett L, et al. Good Policy or Good Luck?[J]. *Journal of Monetary Economics*. 1993, 32(3): 459-483.
- [18] Fernández C, Ley E, Steel M F J. Model Uncertainty in Cross-country Growth Regressions[J]. *Journal of Applied Econometrics*. 2001, 16(5): 563-576.
- [19] Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Prediction, Inference and*

- Data Mining[M]. 2 ed. New York: Springer-Verlag, 2009.
- [20] Henderson D J, Papageorgiou C, Parmeter C F. Growth Empirics without Parameters[J]. *Economic Journal*. 2012, 122(559): 125-154.
- [21] Johnson P A, Papageorgiou C. What Remains of Cross-Country Convergence?[J]. *Journal of Economic Literature*. 2019, forthcoming.
- [22] Johnson P A, Takeyama L N. Initial Conditions and Economic Growth in the US States[J]. *European Economic Review*. 2001, 45(4): 919-927.
- [23] Kormendi R C, Meguire P G. Macroeconomic Determinants of Growth: Cross-country Evidence[J]. *Journal of Monetary Economics*. 1985, 16(2): 141-163.
- [24] Kourtellos A, Tan C M, Zhang X. Is the Relationship between Aid and Economic Growth Nonlinear?[J]. *Journal of Macroeconomics*. 2007, 29(3): 515-540.
- [25] Lehrer S F, Xie T. The Bigger Picture: Combining Econometrics with Analytics Improve Forecasts of Movie Success[Z]. 2018.
- [26] Levine R, Renelt D. A Sensitivity Analysis of Cross-Country Growth Regressions[J]. *American Economic Review*. 1992, 82(4): 942-963.
- [27] Ley E, Steel M F J. On the Effect of Prior Assumptions in Bayesian Model Averaging with Applications to Growth Regression[J]. *Journal of Applied Econometrics*. 2009, 24(4): 651-674.
- [28] Ley E, Steel M F J. Mixtures of g -priors for Bayesian Model Averaging with Economic Applications[J]. *Journal of Econometrics*. 2012, 171(2): 251-266.
- [29] Liu Y, Xie T. Machine Learning Versus Econometrics: Prediction of Box Office[J]. *Applied Economics Letters*. 2019, 26(2): 124-130.
- [30] Liu Z, Stengos T. Non-linearities in Cross-country Growth Regressions: A Semiparametric Approach[J]. *Journal of Applied Econometrics*. 1999, 14(5): 527-538.
- [31] Masanjala W H, Papageorgiou C. The Solow Model with CES Technology: Nonlinearities and Parameter Heterogeneity[J]. *Journal of Applied Econometrics*. 2004, 19(2): 171-201.
- [32] Minier J. Nonlinearities and Robustness in Growth Regressions[J]. *American Economic Review*. 2007, 97(2): 388-392.
- [33] Minier J. Institutions and Parameter Heterogeneity[J]. *Journal of Macroeconomics*. 2007, 29(3): 595-611.
- [34] Mitchell T M. *Machine Learning*[M]. McGraw-Hill, 1997.
- [35] Mullainathan S, Spiess J. Machine Learning: An Applied Econometric Approach[J]. *Journal of Economic Perspectives*. 2017, 31(2): 87-106.
- [36] Sala-I-Martin X, Doppelhofer G, Miller R I. Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach[J]. *American Economic Review*. 2004, 94(4): 813-835.
- [37] Spolaore E, Wacziarg R. How Deep Are the Roots of Economic Development?[J]. *Journal of Economic Literature*. 2013, 51(2): 325-369.
- [38] 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.
- [39] Tibshirani R. Regression Shrinkage and Selection Via the Lasso[J]. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1996, 58(1): 267-288.

- [40] Varian H R. Big Data: New Tricks for Econometrics[J]. Journal of Economic Perspectives. 2014, 28(2): 3-28.
- [41] 陈硕, 王宣艺. 机器学习在社会科中的应用: 回顾与展望[Z]. 复旦大学, 2018.
- [42] 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(07): 115-129.