微观医疗费用预测模型：从线性回归到机器学习

石 菊 王小倩 王 熙*

摘 要：机器学习模型在微观医疗费用预测方面的潜在价值值得进一步发掘。本文基于微观医疗大数据，对比了线性回归模型与多种机器学习模型的个人医疗费用预测性能。结果显示，大部分机器学习模型的预测性能优于线性回归模型，其中梯度提升模型表现最优。政策模拟发现，将机器学习模型预测结果作为依据以确定按人头付费方式下医保支付额时，医疗机构财务风险低于以线性回归模型作为依据的财务风险。本研究为医保支付方式改革提供了技术支持。

关键词：医疗费用预测；机器学习；医保支付方式改革
DOI：10.13821/j.cnki.ceq.2023.06.11

一、引 言

2009 年新医改以来，我国政府大力提高医疗卫生投入，成功建立了覆盖城乡居民的社会基本医疗保障制度。2020 年年底，参加基本医疗保险的人数超过 13.6 亿，医保覆盖率超过 95%。[①] 然而，逐年增长的医疗费用给政府带来巨大的公共财政负担。如何更有效地使用巨大的财政投入，克服道德风险造成的医疗资源浪费，以确保医疗服务平等惠及每个个体，是我国医疗体系亟待解决的问题。医保支付是基本医保管理和深化医改的重要环节，是调节医疗服务行为、引导医疗资源配置的重要杠杆。[②] 从国际经验来看，医保支付方式设计能显著影响医疗体系的运行效率，更合理的医保支付方式有利于医疗控费（Garber and Skinner，2008；Song et al.，2012）。医保支付方式改革是我国当下医疗改革的重点方向，将以按病种付费、按人头付费等方式逐渐替代当前按服务付费的方式，已成为政府和学术界的共识（Yip and Hsiao，2014；封进，2016；刘国恩，2016；朱恒鹏和彭晓博，2018）。随着多种按病种付费方式的国家试点工作稳步推进，住院层面的医保支付方式改革方向已逐渐明晰。[③] 在门诊和慢病层面，多项政府文件强调，按

* 石菊，北京大学经济学院、北京大学全球健康发展研究院；王小倩，南开大学金融学院；王熙，北京大学经济学院。通信作者及地址：王熙，北京市海淀区颐和园路 5 号北京大学经济学院，100871；电话：(010) 62753635；E-mail：wang.x@pku.edu.cn。作者感谢第四届中国健康经济发展论坛和北京大学劳动-健康经济学工作坊参与人的有益建议，感谢匿名审稿专家的宝贵意见，感谢吴赵远志的优秀助研工作。本文受到国家社会科学基金项目（23BJL025）、国家自然科学基金项目（72241419）、北京市社会科学基金项目（21JCC082）的资助。文责自负。

① 国家医疗保障局，《2020 年全国医疗保障事业发展统计公报》，2021 年。
② 国务院，《国务院办公厅关于进一步深化基本医疗保险支付方式改革的指导意见》，2017 年。
③ 国家医疗保障局，《关于印发疾病诊断相关分组（DRG）付费国家试点技术规范和分组方案的通知》，2019 年。

人头付费是未来的改革方向。[①] 根据患者个体风险进行调整的按人头付费（risk-adjusted capitation payment）的模式已广泛应用于美国、荷兰等发达国家，并且被证实有良好的控费效果（Newhouse et al.，2012；Buchner et al.，2013；Van Kleef et al.，2013）。

经风险调整的按人头付费模式的执行面临良好的时代契机，在控制总体费用增长的同时还兼顾了民众对于医疗保险的基本需求。该模式中的风险调整环节需要更准确精细的微观医疗费用预测模型作为基础，这对预测模型提出了更高的要求。随着我国医疗体系信息化的逐步推广以及信息技术在近年的大幅进步，获取越来越丰富的医疗机构和患者信息变得更为便利，为复杂统计模型的构建，比如机器学习模型，提供了数据可得性这一基本保障。然而，我国在医疗费用预测方面的研究起步较晚，特别是基于微观数据预测个人医疗费用的研究较少，且目前应用于医保支付的预测模型多是直接采用其他国家的（线性）模型。由于各国在医疗体制、居民体质、患者疾病谱等方面有着显著差异，其他国家所使用的费用预测模型在我国的适应性有待商榷。

基于我国当前正在大力推行医保支付方式改革这一时代背景，和缺乏针对我国国情的医疗费用预测模型的现实背景，为减少执行医保支付方式改革的技术障碍，本文专注于利用现有前沿统计技术，设计并估计更贴合我国居民特征的费用预测模型。本文基于大样本医保微观数据，构建了多个针对个体的医疗费用预测模型，着重比较传统预测模型与机器学习模型对个人医疗费用的跨期样本外预测能力；构建根据风险调整的按人头付费的医保支付模式，展示了机器学习模型的应用如何降低医疗机构所面临的财务不确定性，从而使医保支付方式改革更能得到微观主体的支持，进而可以更顺利地推广并深化改革。具体而言，首先，我们通过多个样本外预测评价指标对比了线性回归模型与梯度提升（gradient boosting decision tree，GBDT）、前馈神经网络（feedforward neural network，FNN）、随机森林（random forest）等多种经典机器学习模型的预测性能。其次，我们将医疗费用分为门诊、住院和慢病费用，分别使用不同模型进行预测，对比了不同模型在子样本内的样本外预测能力。最后，本文将所估计模型预测应用于政策模拟：构建根据风险调整的按人头付费的医保支付方式，采用不同模型预测结果模拟医疗机构的医保收入，以衡量医疗机构所面临的财务风险。

实证结果表明，大部分机器学习模型的跨期样本外预测性能优于线性回归模型。对于年度总医疗费用的预测，表现最好的梯度提升模型相对线性回归模型的样本外 R^2 提升了 9.52%。同时，无论是对于门诊、住院还是慢病费用，机器学习模型相较于线性回归模型都保持优势。在稳健性方面，在增加或改变预测变量以及更换样本的情况下，机器学习模型依然维持了优势。最后，通过模拟发现，在经风险调整的按人头付费的支付方式下，若根据机器学习模型的预测结果进行付费，医疗机构的财务风险比基于线性回归模型结果进行付费降低了 4.19%。

本文的学术贡献主要在以下三个方面。第一，本文首次将机器学习模型应用于我国医保支付领域，有助于医保支付方式改革的推进。近年来，我国医疗花费增长较快，医保基金支出压力增加。已被国际经验证实的先进医保支付模式亟需个人医疗费用预测模

① 国务院，《国务院办公厅关于进一步深化基本医疗保险支付方式改革的指导意见》，2017 年；《中共中央、国务院关于深化医疗保障制度改革的意见》，2020 年。

型作为政策支持工具。构建准确的跨期预测模型，能够降低医疗机构收入不确定性，降低医保基金运行风险，是顺利推行医保支付方式改革的必要保障。基于机器学习模型在预测方面的优势，目前我国已有少量研究采用其预测医疗费用（夏涛等，2019；赵颖旭等，2020；张宁等，2020），然而，已有研究或是采用某一家医疗机构的数据，或是采用某一种特定的机器学习方法，且大多关注特定疾病的住院费用，对如何将大数据的机器学习技术应用于医保支付方式改革中，以及这一应用会带来多大的增益等问题提供的启发有限。如赵颖旭等（2020）和张宁等（2020）分别探索了多种机器学习模型对糖尿病和老年痴呆症带病人群住院费用的预测能力。与之前文献不同的是，本文采用某县级市的医保参保和报销数据，不但研究了总医疗费用和多类别医疗费用，且着力比较了多种机器学习方法在跨期样本外预测方面的优劣，并将其应用于医保支付方式的政策模拟。据我们所知，本文是首篇系统性探索机器学习模型在我国微观医疗费用跨时样本外预测方面的研究，为机器学习模型在医保支付领域的应用提供了理论支持。

第二，本文将机器学习模型应用于我国个人医疗费用的预测，有利于进行该领域研究的国际比较。国外已有研究将机器学习应用于医疗费用预测，但所得结论并不一致（Duncan et al.，2016；Rose，2016；Morid et al.，2017；Iommi et al.，2022）。比如，Duncan et al.（2016）对线性回归模型与多种机器学习模型进行性能分析，认为线性回归模型的预测性能不如机器学习模型。而 Rose（2016）采用美国医保数据，发现机器学习模型的预测能力并不优于线性回归模型。本文采用机器学习领域广泛应用的标准模型，与文献中的模型具有高度可比性。将这些模型应用于我国的医保实践，有助于我们进一步进行国际比较，有利于推动统计模型在微观医疗费用领域应用的进一步改进。

第三，本文补充了我国在微观医疗费用预测领域的研究。微观医疗费用预测一直是健康经济学的重要研究方向（Jones，2000；Mihaylova et al.，2011；Ellis et al.，2013）。目前有关我国医疗费用预测的研究大多集中于宏观层面，即对医疗总费用进行预测（周绿林等，2008；温小霓等，2014），较少研究直接关注微观医疗费用预测。与本文最相关的研究是 Shi et al.（2018），该研究使用了八种传统计量模型预测了个人医疗费用，但并未着眼于跨期样本外预测。与前述文献不同的是，本文着力于对各类机器学习模型在微观医疗费用跨期样本外预测方面的探索，补充了我国微观医疗费用预测的研究。

本文剩余部分安排如下：第二部分介绍研究方法，包括预测模型、用于预测的个人特征及评价指标；第三部分介绍数据，并展示数据的描述性统计；第四部分比较线性回归模型和多种机器学习模型在个人医疗费用预测方面的性能，并进行异质性分析与稳健性检验；第五部分模拟模型预测结果如何应用于医保支付方式改革，对比线性回归模型和机器学习模型下医疗机构的财务风险；最后是对全文的总结。

二、研 究 方 法

（一）预测模型和方法

我们分别利用传统线性回归模型和七种机器学习模型构建预测模型，利用个人特征对个人年度医疗费用进行跨期样本外预测，并通过多种评价指标评估各类模型的预测性能。本文采用线性回归模型与机器学习模型进行对比的理由有三点。首先，已有研究表

明，通过采用两部分模型等方式可提高线性回归模型的拟合度，但随着数据样本量的提升，这些方法带来的预测精度提升程度有限（Mihaylova et al.，2011；Malehi et al.，2015）；而已有研究建议在大样本下首选线性回归模型预测医疗费用（Mihaylova et al.，2011；Mesike et al.，2012；Shi et al.，2018）。其次，线性回归模型是医疗保险市场中最常用的风险调整工具（Pope et al.，2011，Kautter et al.，2014）。最后，对传统计量模型和机器学习模型进行对比的国际研究一般采用线性回归模型为传统模型的基准（Rose，2016；Morid et al.，2017），因此本文的模型选择方式有助于进行该领域的国际比较。①

传统线性回归模型的优势是操作简单，可通过加入虚拟变量拟合变量间的非线性关系。但是，医疗费用的分布具有特殊性，比如存在大量零费用的观测、分布呈左偏和厚尾特征、较强的个体异质性，因此线性回归模型在拟合观测数据时面临明显缺陷（Jones，2010）。机器学习模型在处理结构化和非结构化数据时，不但强调了变量本身非线性关系，还考虑了变量之间的非线性交互影响。更重要的是，机器学习模型关注预测模型的样本外表现，对政策支持具有重要的实践意义。因此，在我们的应用场景中机器学习模型对比传统的线性回归模型理论上有优势。

我们采用的七种机器学习模型可分为三类。第一类为变量选择类模型，包含了拉索回归（least absolute shrinkage and selection operator，Lasso）、岭回归（ridge regression）、弹性网络（elastic net）和正交匹配追踪（orthogonal matching pursuit，OMP）。第二类为基于决策树的模型，包含随机森林和梯度提升。第三类为神经网络模型，我们主要使用了前馈神经网络。各模型的具体介绍和超参数设定见附录Ⅱ。

预测方法上，已有文献大多根据 K 折交叉验证（K-fold cross-validation）进行同时段内样本外预测，这忽略了数据在不同时间的分布差异性，从而容易低估样本外误差。因此，我们采取了跨期样本外预测。具体而言，以第一年和第二年样本为训练样本，即以第一年为基准年并利用第一年个体的各类特征预测其第二年年度医疗费用；以第二年和第三年样本为测试样本，即基于用第一年和第二年样本估计出的模型，利用第二年的个体特征对其在第三年的年度医疗费用进行预测。在现实政策制定中，跨期样本外预测比期内样本外预测具有更重要的指导意义。当前我国医保支付由后付制向预付制转变，费用预测模型在医保支付中的应用场景主要是利用已有数据对于未来费用进行预测，并非对当期费用进行预测。另外，进行期内样本外预测的潜在假设是样本的各项特征以及现实环境没有时变特性，不符合实际情况，导致模型缺乏泛化能力。因此，跨期样本外预测对回答本文的研究问题更贴合实际，也更为关键。

（二）预测对象和用于预测的个体特征

本文的主要研究对象是微观层面的个人年度医疗总费用。除此以外，我们将总医疗费用分解为互不重合的三个子项：门诊费用、住院费用和慢病费用。若居民在某一年没有任何医疗费用支出，我们则将其医疗费用及各项分类费用设为 0。

① 我们也使用了线性回归模型、广义线性模型（generalized linear model，GLM）、Tobit 模型、两部分模型等计量模型进行跨期样本外预测，评价结果见附录Ⅰ表Ⅰ1。结果显示，线性回归模型在 R^2 和 RMSE 上优于其他所有模型，在 MAE 与 MAPE 上不差于大部分模型。限于篇幅，附录从略，感兴趣的读者可在《经济学》（季刊）官网（http：//ceq. ccer. pku. edu. cn）下载。

本文采用年龄、性别和疾病诊断作为用于预测的自变量，这三组变量的选择与文献高度吻合。虽然数据中包含其他与医疗费用相关的信息，比如受教育水平，但主要分析结果中未将其作为预测的自变量，原因如下。第一，在费用预测模型应用场景下，其他与医疗费用相关的信息可能会缺失，出于更广泛的模型适用范围的考虑，我们仅使用数据可获得性较大的年龄、性别和疾病诊断作为预测变量；第二，费用预测模型的应用是为了确定医保支付额，使用某些特征作为预测变量可能会影响基本医疗保险的公平性。因此，我们仅在稳健性检验中使用包含其他信息的预测变量。

用于预测的自变量的具体处理如下。依照已有文献的分组法，将 80 岁以下的人群以 5 为间距将年龄分组，将大于 80 岁的人群单独分为一组，形成 17 个组别。性别与年龄组别组成交互项作为模型的自变量，以考虑不同性别在不同年龄人群中的异质性。对于病种变量，即病人所患疾病类别，我们采用 ICD－10 编码（*The International Statistical Classification of Diseases and Related Health Problems* (10th Revision)，《疾病和有关健康问题的国际统计分类》(第 10 次修订本)）。该编码依据疾病的特征将病种分为 21 个大类。

（三）预测性能评价

不同的医疗费用预测模型在应用场景、预测方面的优势不同，而医疗费用预测涵盖了广泛的潜在目标，因此我们采用了多个指标以更加全面地衡量不同预测模型在不同维度上的性能。除了最为常见的评价指标 R^2（拟合优度），我们还采用了 RMSE（root mean square error，均方根误差）、MAE（mean absolute error，平均绝对误差）以及 MAPE（mean absolute percentage error，平均绝对百分误差）。各项评价指标的计算公式如下：

$$R^2=1-\frac{\sum_{i=1}^{n}(\hat{y_i}-y_i)^2}{\sum_{i=1}^{n}(\bar{y}-y_i)^2}, \tag{1}$$

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y_i}-y_i)^2}, \tag{2}$$

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|\hat{y_i}-y_i|, \tag{3}$$

$$MAPE=\frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y_i}-y_i}{y_i}\right|, \tag{4}$$

其中，i 表示个体（$i=1, 2, \cdots, n$），$\hat{y_i}$是模型预测值，y_i是真实值，$\bar{y}$表示真实值的平均值。R^2 越接近 1 表示模型的拟合能力越好，而 RMSE、MAE 和 MAPE 则值越小表示模型拟合误差越小。各评价指标的具体含义见附录Ⅱ。

三、数　　据

本文采用我国某县级市的医疗保险参保和报销数据，该数据包含了该县级市新型农村合作医疗（以下简称“新农合”）参保居民的人口统计学特征和医保报销信息。样本包含 2011—2014 年间参加新农合的居民，各年样本量分别为 277 647，276 391，273 563

和 270 255。人口统计学信息包括年龄、性别、婚姻状况、职业、受教育水平和所属村居，医保报销信息包括疾病诊断和各类医疗费用（住院、门诊和慢病费用）。

表 1 展示了 2011—2013 年参保居民的人口统计学特征与病种分布。以 2012 年为例，参保居民以中青年为主，男性占比略少于女性，所患疾病主要集中于呼吸系统疾病。附录Ⅲ图Ⅲ1 中展示了 2012 年该县级市参保居民的人口金字塔，与我国整体人口年龄结构相符。

表 1 2011—2013 年人口统计学特征与病种分布 单位：%

		2011 年	2012 年	2013 年
年龄				
0—19 岁		13.3	13.2	13.0
20—40 岁		26.6	25.0	24.1
41—64 岁		45.7	46.5	46.4
65 及以上		14.4	15.3	16.4
性别				
男性		48.7	48.4	48.4
女性		51.3	51.6	51.6
病种				
A00—B99	某些传染病和寄生虫病	1.37	1.74	2.58
C00—D48	肿瘤	0.38	0.48	0.54
D50—D89	血液和造血器官疾病以及某些涉及免疫机能的异常	0.33	0.33	0.36
E00—E90	内分泌、营养和代谢疾病	1.02	1.23	1.52
F00—F99	精神和行为障碍	0.17	0.17	0.19
G00—G99	神经系统疾病	0.51	0.75	1.32
H00—H59	眼和附器疾病	0.76	1.26	1.70
H60—H95	耳和乳突疾病	0.60	0.76	1.18
I00—I99	循环系统疾病	3.16	4.70	6.03
J00—J99	呼吸系统疾病	37.7	43.9	45.6
K00—K93	消化系统疾病	18.5	22.8	26.0
L00—L99	皮肤和皮下组织疾病	3.38	6.25	9.42
M00—M99	肌肉骨骼系统和结缔组织疾病	8.16	14.6	18.9
N00—N99	泌尿生殖系统疾病	6.36	7.44	8.27
O00—O99	妊娠、分娩和产褥期	0.23	0.25	0.28
P00—P96	起源于围生期的某些疾病	0.00	0.00	0.05
Q00—Q99	先天性畸形、变形和染色体异常	0.03	0.04	0.08
R00—R99	症状、体征和异常的临床和化验结果	8.98	12.8	15.9
S00—T98	损伤、中毒和外因的某些其他结果	4.19	6.50	9.63
V01—Y98	发病和死亡的外因	1.05	1.68	2.15
Z00—Z99	影响健康状况和接触健康服务的因素	0.79	1.36	1.77

注：一个居民可能患多种疾病，因此病种比例加总不等于 1。

表 2 展示了 2012—2014 年个人年度医疗费用的描述性统计。样本中有大量医疗费用为零的居民。医疗费用呈现典型的左偏和厚尾特征：存在大量低费用患者，也存在部分异常高费用患者。附录Ⅲ表Ⅲ1 展示了分类别医疗费用描述性统计，分布与总医疗费用相似。图 1 展示了 2012—2014 年个人年度非零医疗费用分布，也验证了左偏和厚尾特点。附录Ⅲ图Ⅲ2 和图Ⅲ3 分别展示了 2012 年分年龄性别和分所患疾病类型的平均个人年度医疗费用，可见医疗费用在年龄、性别和病种方面的异质性。综上所述，本文所采取的样本中的医疗费用分布符合典型的医疗费用分布特征。

表 2　2012—2014 年个人年度医疗费用描述性统计

	2012 年	2013 年	2014 年
零值比例（%）	36.30	31.08	28.87
非零费用（元）			
均值	806	937	981
25%分位	44	59	63
50%分位	120	159	160
75%分位	343	399	369
95%分位	3 652	4 023	4 391
99%分位	12 431	13 684	14 853
最大值	186 817	434 542	220 380

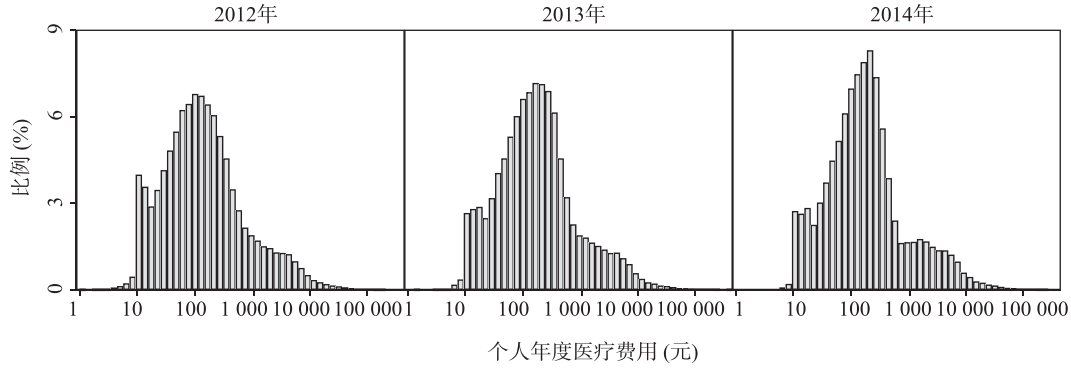

图 1　2012—2014 年个人年度非零医疗费用分布

注：本图横轴采用了对数坐标。为了取对数，此处个人年度医疗费用为原始个人年度医疗费用＋1。

四、预 测 结 果

（一）主要分析结果

我们基于 2012 年个人特征变量预测其在 2013 年的医疗费用，这一过程用于估计模型，再根据所估计的参数，基于 2013 年个人特征变量对其在 2014 年的医疗费用进行预测，并计算各项跨期样本外评价指标结果，展示于表 3。[①] 括号中展示了机器学习模型相

① 我们也对各模型的样本内预测能力进行了评价，结果见附录Ⅰ表Ⅰ2。结果显示大部分机器学习模型的样本内预测表现不差于线性回归模型，梯度提升模型在 R^2 和 RMSE 上表现最优，能使 R^2 提升 36.4%。

较于线性回归模型的提升比例。相较于线性回归模型，大部分机器学习模型的跨期样本外预测表现没有更差，梯度提升模型和前馈神经网络模型的表现最优。从 R^2 和 RMSE 来看，梯度提升模型优于其他所有模型。梯度提升模型能使样本外 R^2 提升 9.52%，达到 0.023。从 MAE 和 MAPE 来看，前馈神经网络模型优于其他所有模型。

表 3 各模型样本外预测评价结果

	R^2	RMSE	MAE	MAPE
	(1)	(2)	(3)	(4)
线性回归	0.021	3 563	980	137
拉索回归	0.020	3 563	980	139
	(−4.76%)	(0.00%)	(0.00%)	(−1.46%)
岭回归	0.021	3 563	980	137
	(0.00%)	(0.00%)	(0.00%)	(0.00%)
弹性网络	0.021	3 563	980	138
	(0.00%)	(0.00%)	(0.00%)	(−0.73%)
正交匹配追踪	0.020	3 563	981	138
	(−4.76%)	(0.00%)	(−0.10%)	(−0.73%)
随机森林	0.021	3 562	969	148
	(0.00%)	(0.03%)	(1.12%)	(−8.03%)
梯度提升	**0.023**	**3 559**	973	138
	(9.52%)	**(0.11%)**	(0.71%)	(−0.73%)
前馈神经网络	0.022	3 560	**963**	**134**
	(4.76%)	(0.08%)	**(1.73%)**	**(2.19%)**

注：括号中的数字表示该模型的该指标相对于线性回归模型的提升比例。加粗数字表示该模型在该项指标上表现最优。

由此可见，机器学习模型在预测跨期年度医疗费用上比线性回归模型更有优势。这与我们的预期相符，机器学习模型能够通过正则化缓解过拟合问题，还能更好地捕获变量间的非线性关系和交互影响。第一类模型比如拉索回归，其本质上是想通过控制模型过拟合的程度换取样本外更好的表现。而第二类模型与第三类模型除了考虑了过拟合问题，在自变量对因变量影响上还捕捉了自变量间的交叉影响，且更容易发现数据中的非线性关系。在现实当中，年龄、性别和病种存在交叉影响，比如年龄大的个体更容易患高血压，且这些因素对医疗费用的影响往往是非线性的。因此，梯度提升模型与前馈神经网络模型相对于其他机器学习模型的样本外预测表现要更好。总的来说，机器学习模型相较于线性回归模型能够更好地拟合医疗费用的分布，同时具有更强的泛化能力。

（二）模型预测的异质性：按费用类别分组

由于基本医保针对普通门诊、住院和门诊慢病具有不同的报销政策，因此我们将分别讨论模型对住院、门诊和慢病费用的预测能力。表 4 汇总了不同费用类别下各模型的样本外预测评价结果。(1)—(4) 列展示了门诊费用的预测评价结果。门诊费用的预测

表 4　分费用类别的各模型样本外预测评价结果

	门诊费用				住院费用				慢病费用			
	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
线性回归	0.203	142	103	27	0.016	3 551	948	456	0.264	102	15	7
拉索回归	0.203	142	103	27	0.016	3，551	948	456	0.247	103	14	6
	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(−6.44%)	(−0.98%)	(6.67%)	(14.29%)
岭回归	0.203	142	103	27	0.016	3 551	948	456	0.264	102	15	7
	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)
弹性网络	0.203	142	103	27	0.016	3 551	949	457	0.247	103	14	6
	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(−0.11%)	(−0.22%)	(−6.44%)	(−0.98%)	(6.67%)	(14.29%)
正交匹配追踪	0.200	142	103	27	0.016	3 551	948	455	0.264	102	15	7
	(−1.48%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.22%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)
随机森林	0.179	144	104	28	0.017	3 550	**942**	**449**	0.279	101	13	6
	(−11.82%)	(−1.41%)	(−0.97%)	(−3.70%)	(6.25%)	(0.03%)	**(0.63%)**	**(1.54%)**	(5.68%)	(0.98%)	(13.33%)	(14.29%)
梯度提升	**0.209**	142	102	27	**0.018**	**3，548**	943	450	**0.288**	**100**	13	**5**
	(2.96%)	(0.00%)	(0.97%)	(0.00%)	**(12.5%)**	**(0.08%)**	(0.53%)	(1.32%)	**(9.09%)**	**(1.96%)**	(13.33%)	**(28.57%)**
前馈神经网络	0.204	142	102	27	0.017	3 549	991	505	0.282	101	13	6
	(0.49%)	(0.00%)	(0.97%)	(0.00%)	(6.25%)	(0.06%)	(−4.54%)	(−10.75%)	(6.82%)	(0.98%)	(13.33%)	(14.29%)

注：括号中的数字表示该模型的该指标相对于线性回归模型的提升比例。加粗数字表示该模型在该项指标上表现最优。

力度相较于医疗总费用有较大提升，线性回归的 R^2 达到了 0.203。这与直观相符，相比于住院往往针对急重症，门诊费用更易预测。模型间比较显示，梯度提升模型和前馈神经网络模型是唯二能在所有评价指标上带来提升的模型。梯度提升模型预测下的 R^2 达到 0.209，相比于线性回归模型上升 2.96%。(5)—(8) 列展示了住院费用的预测评价结果。梯度提升模型的样本外 R^2 相较于线性回归模型上升 12.5%，达到 0.018，同时 RMSE 也有所提升。(9)—(12) 展示了慢病费用的预测评价结果。慢病费用预测的 R^2 在三类费用中最高，采用线性回归模型时达到 0.264。慢病患者的医疗服务需求较稳定，且一般在固定医疗机构常规就诊，因此更易预测。梯度提升模型在 R^2、RMSE、MAE 和 MAPE 上均表现最优，其中 R^2 达到 0.288，相较于线性回归模型提升 9.09%。对三类医疗费用进行预测的 R^2 大小与以往文献所揭示的 R^2 大小与医疗费用类别的关系一致 (Van de Ven and Ellis，2000)。

整体而言，对于不同类别医疗费用，大部分机器学习模型表现不弱于线性回归模型，其中梯度提升模型表现最优，这为将机器学习模型应用于医保支付方式改革提供了决策依据和技术支持。无论对于哪一类费用进行预测，梯度提升模型在各项评价指标上的表现都优于线性回归模型，且在 R^2 上的提升高于其他机器学习模型。另外，相较于住院，门诊和慢病费用可预测性更高，这为经风险调整的按人头付费改革路径提供了一定依据。

（三）稳健性检验

上述分析中，出于模型适用性和公平性考虑，我们仅采用前一年的性别、年龄和病种作为预测变量预测医疗费用，使用 2012—2014 年样本进行预测和评价。为了展示结果的稳健性，我们从三个方面进行扩展。

首先，我们在预测变量中加入个人的婚姻状况、职业、受教育水平和所属村居信息进行预测。[①] 个人的婚姻状况、职业和受教育水平信息存在部分缺失，可能对预测精度有所影响。[②] 参保人所属村居信息能一定程度上反映参保人之间的社会经济状况差异，且无缺失。因此我们分别在前一年的年龄、性别和病种的基础上新增婚姻状况、职业和受教育水平变量，所属村居变量，婚姻状况、职业、受教育水平和所属村居变量，进行跨期样本外预测，评价结果展示于附录Ⅰ表Ⅰ3 至表Ⅰ5。与前文结果一致，大部分机器学习模型表现不弱于线性回归模型，表现最好的梯度提升模型相较于线性回归模型在 R^2 上分别提升了 15.0%、9.52%和 20.0%。

其次，我们使用前三年的病种信息替换前一年的病种信息进行预测。由使用前三年的病种信息的跨期样本外预测评价结果见附录Ⅰ表Ⅰ6。与前文分析结果一致，大部分机器学习模型表现不弱于线性回归模型。梯度提升模型在 R^2、RMSE 和 MAE 上表现最好，其中 R^2 提升 16.7%，达到 0.021。但与前文不同的是，使用前三年病种信息的各模型的预测能力有所下降。这可能是由于大部分疾病的发生是短期的，前三年病种信息相较于前一年的病种信息增加了无关因素扰动，弱化了模型预测能力。

① 本节所使用的新增预测变量的描述性统计见附录Ⅲ表Ⅲ2 和表Ⅲ3。

② 在预测变量中加入个人的婚姻状况、职业和受教育水平信息后，样本流失率分别为 9.88%和 10.89%。

最后，我们更换样本时间区间进行预测。使用 2011—2013 年的样本进行跨期样本外预测的评价结果见附录Ⅰ表Ⅰ7。与主要分析结果一致，大部分机器学习模型表现不弱于线性回归模型，其中梯度提升模型在 R^2、RMSE 和 MAE 上表现最好，在 R^2 上提升了 20.0%，达到 0.024。因此，机器学习模型在费用预测方面的优势具备普遍意义。

从主要分析结果和稳健性检验可见，现有模型在不同评价指标上表现不同。梯度提升模型在 R^2 和 RMSE 上一直表现最优，而随机森林模型和前馈神经网络模型在 MAE 和 MAPE 上在部分情况下表现最优。我们进一步对随机森林模型、梯度提升模型和前馈神经网络模型进行两两和三者一起的模型平均（model averaging），以综合不同模型的相对优势，跨期样本外预测评价结果见附录Ⅰ表Ⅰ8。结果显示，采用模型平均方法可平衡不同评价指标上的提升幅度。以三者模型平均为例，其在 R^2 和 RMSE 上相对于线性回归模型的提升比例与梯度提升模型一致，高于随机森林模型和前馈神经网络模型，同时在 MAE 和 MAPE 上相对于线性回归模型的提升比例高于梯度提升模型。因此，在未来不同预测场景应用中，应根据预测目标选择机器学习模型。在风险调整领域，R^2 是用来衡量不同费用预测模型预测能力的最常用方法（Jones，2000）。因此，在下文我们采用在 R^2 上表现最好且最稳健的梯度提升模型进行应用。

五、政策模拟

在医保支付方式改革的推进过程中，医保基金运行面临巨大的成本上行压力。同时，医保基金收入占医疗机构收入的比重也不断增大，医疗机构收入不确定性也随之增加。如何降低医疗机构面临的财务风险，进而减少医疗机构对政策的抵触，是改革面临的一个重要问题。本部分设计了一套经过风险调整的按人头付费体系，分别以线性回归模型和在前述分析中表现最好的梯度提升模型的预测结果作为确定医保支付的依据，从而展示更稳健和精细的医疗费用预测模型如何降低医疗机构面临的财务风险，进而减少医保支付方式改革阻力。

作为示例，下文的模拟分析仅以新农合对乡镇卫生院的支付为例，但同样的原理可应用于其他基本医疗保险和其他医疗机构。当前县域医疗体系面临的主要问题是基层医疗机构就诊不足，大量患者涌入高级别医疗机构就诊。更全面的分析应基于医共体模式，即由县级医疗机构带领下辖基层医疗机构，作为利益共同体为患者提供服务。然而这样的模式在市级医疗体系下更有可操作性。市区下辖多个区县，每个区县成立一个或者多个医共体进行支付。由于数据的限制，我们的研究仅涵盖一个县的数据，因此只能进行乡镇卫生院层面的分析。

我们设计了两种医保支付方式。第一种方式是按人头付费。参与新农合的居民只能在其所在乡镇的乡镇卫生院就诊，乡镇卫生院为其所在乡镇所有参与新农合的居民提供医疗服务，作为补偿，乡镇卫生院接受新农合的支付。该支付以通过医疗费用风险调整后的按人头付费的方式进行。具体而言，新农合根据乡镇卫生院服务的参保居民数量和每个参保居民的人头费对乡镇卫生院进行支付。人头费根据个体的医疗费用风险进行调整，风险调整以通过模型预测的居民医疗费用值为依据。基本原则是，对于覆盖了高费用风险的居民的乡镇卫生院进行较高的支付，对于覆盖了低费用风险的居民的乡镇卫生

院进行较低的支付，以平衡由于覆盖不同健康水平的居民的差异带来的乡镇卫生院承担的治疗成本的差异。费用风险评分的计算公式如下。根据线性回归模型或梯度提升模型预测得到居民 i 的当期医疗费用，分别记为 S_i^1 和 S_i^2。个体的预测费用除以总体预测费用均值 $\overline{S^k}$，得到医疗费用风险评分 r_i^k：

$$r_i^k = \frac{S_i^k}{\overline{S^k}},\ k = 1,\ 2, \tag{5}$$

其中，风险评分 r_i^k 代表了居民 i 在所有参保居民中医疗费用风险的相对水平，$k=1$ 表示以线性回归模型预测结果为依据，$k=2$ 表示以梯度提升模型预测结果为依据。

第二种医保支付方式由按人头付费与按服务付费相结合。与第一种方式类似，对每个居民的支付是经医疗费用风险调整后的按人头付费，而不同之处在于，第二种方式允许居民选择多个乡镇卫生院就诊，居民的人头费根据居民的医疗服务使用情况分配给其就诊的乡镇卫生院。因此，就诊人次更多和医疗服务价格更高的乡镇卫生院将获得更高的支付额。两种支付方式的具体计算方式见附录Ⅳ。

根据以上两种以按人头付费为基础设计的体系，新农合支付不依赖于每次就诊的实际支出，故都有助于纠正医疗服务提供者过度使用医疗服务的动机。为了降低医疗机构收入不确定性，我们引入政府干预来分担乡镇卫生院财务风险。我们考虑了两种政府分担方式。第一种是单边风险走廊（one-sided risk corridor），政府与乡镇卫生院共担亏损。当乡镇卫生院亏缺率小于 10％时，政府不做任何补贴。当乡镇卫生院亏缺率在 10％到 15％之间时，乡镇卫生院承担 10％的亏损，余下亏损由政府与乡镇卫生院平摊。当乡镇卫生院亏缺率大于 15％时，乡镇卫生院只承担 12.5％的亏损，余下亏损由政府承担。政府不分担乡镇卫生院盈利。第二种是双边风险走廊（two-sided risk corridor），政府不仅与乡镇卫生院分担亏损，也与乡镇卫生院分担盈利。亏损分担方式与单边风险走廊相同，同时政府以类似方式分担乡镇卫生院的盈利，故乡镇卫生院最多获得 12.5％盈利。两种分担方式的具体计算方式见附录Ⅳ。

接下来我们计算不同医保支付方式与政府分担方式组合下，以线性回归模型和梯度提升模型为确定新农合支付额依据时乡镇卫生院的财务风险。乡镇卫生院的成本为目前现实中实行的按服务付费制度下发生的医疗费用，收入为新的医保支付方式下新农合支付额与政府补贴（抽成）之和。我们假定新农合预算不变，即新农合给乡镇卫生院的支付总额不变，新的医保支付方式只改变基金在乡镇卫生院间的分配。各乡镇卫生院收入对成本的比率（收入成本比率）为乡镇卫生院财务风险指标，这是实践中常用于评估机构财务风险的指标。当收入成本比率超过 1 时，表明医疗机构的收入超过成本，医保基金支出压力较大。当收入成本比率低于 1 时，表明医疗机构的收入低于成本，医疗机构存在运营风险。故收入成本比率越接近 1，表明医疗机构财务风险越低，医保支付方式改革推进所面临的阻力越小。乡镇卫生院的收入成本比率与 1 的差值的绝对值的总和（$\sum |1-$（收入成本比率）$|$）为评估整体财务风险的指标，它体现所有乡镇卫生院的成本与支付额之间的差异，可作为评估改革可行性的关键指标。

表 5 展示了在第一种支付方式下，以线性回归模型预测结果为支付依据的结果。第（1）列展示了各乡镇卫生院的医疗成本，第（2）、（3）列展示了在无风险走廊的模式

下乡镇卫生院收到的支付总额和收入成本比率。最后一行汇报了乡镇卫生院的整体财务风险，在无风险走廊模式下为 1.097。在医疗成本不变的情况下，该县所辖的乡镇卫生院改革以后的收入与成本的平均差异为 7.84％（1.097/14）。第（4）、(5）列汇报了单边风险走廊下乡镇卫生院的收入和收入成本比率。该模式下政府不参与盈利分担，收入和收入成本比率的水平与无风险走廊下的水平一致，整体财务风险也一致。第（6）、(7）列汇报了双边风险走廊下乡镇卫生院的收入和收入成本比率。由于政府参与分担盈利，第十个乡镇卫生院的盈利由 30％降为 12.5％，整体财务风险降为 0.922。在第一种支付方式下以梯度提升模型预测结果为支付依据的结果和在第二种支付方式下以线性回归模型/梯度提升模型预测结果为支付依据的结果见附录Ⅰ表Ⅰ9 至表Ⅰ11。

表 5　第一种医保支付方式下以线性回归模型预测结果为支付依据的结果

乡镇	成本	无风险走廊		单边风险走廊		双边风险走廊	
		收入	收入成本比率	收入	收入成本比率	收入	收入成本比率
	(1)	(2)	(3)＝(2)/(1)	(4)	(5)＝(4)/(1)	(6)	(7)＝(6)/(1)
1	6 113 787	5 649 442	0.924	5 649 442	0.924	5 649 442	0.924
2	10 125 700	10 317 197	1.019	10 317 197	1.019	10 317 197	1.019
3	12 775 556	12 803 402	1.002	12 803 402	1.002	12 803 402	1.002
4	11 525 740	12 021 390	1.043	12 021 390	1.043	12 021 390	1.043
5	10 818 740	9 854 199	0.911	9 854 199	0.911	9 854 199	0.911
6	8 140 266	7 663 444	0.941	7 663 444	0.941	7 663 444	0.941
7	13 266 642	14 436 363	1.088	14 436 363	1.088	14 436 363	1.088
8	12 926 489	12 494 371	0.967	12 494 371	0.967	12 494 371	0.967
9	11 891 490	13 573 766	1.141	13 573 766	1.141	13 573 766	1.141
10	890 593	1 157 968	1.300	1 157 968	1.300	1 001 918	1.125
11	4 631 023	4 211 377	0.909	4 211 377	0.909	4 211 377	0.909
12	7 791 352	8 245 079	1.058	8 245 079	1.058	8 245 079	1.058
13	11 492 682	11 357 070	0.988	11 357 070	0.988	11 357 070	0.988
14	16 203 138	14 808 139	0.914	14 808 139	0.914	14 808 139	0.914
整体财务风险（$\sum$ |1－(收入成本比率)|）		1.097		1.097		0.922	

四种组合下乡镇卫生院面临的整体财务风险水平汇总于表 6。(1)—(3）列展示了第一种支付方式下的结果。可以发现，相较于线性回归模型，梯度提升模型在不同风险走廊模式下都降低了医疗机构的财务风险。在无风险走廊和单边风险走廊下，梯度提升模型下医疗机构财务风险降低了 4.19％。在双边风险走廊下，由于政府已经分担了部分医疗机构风险，梯度提升模型额外的风险降低程度低于其他模式。尽管如此，梯度提升模型依然使整体收入成本比率的差异降低了 2.60％。(4)—(6）列展示了第二种支付方式下的结果。在患者可自由选择医疗机构的模式下，乡镇卫生院面临的财务风险更高。梯度提升模型依然降低了医疗机构的财务风险，但相较于第一种支付方式更有限。这是因为，在第二种模式下医疗机构面临的风险不仅来自患者自身医疗花费的风险，也来自患者对医疗机构的选择。医疗费用预测模型精准度的提升有助于降低前者的不确定性，但

不影响后者。模拟发现，梯度提升模型在无风险走廊、单边风险走廊和双边风险走廊的模式下，分别降低了 1.87%、1.67%和 0.83%的整体财务风险。即在模型应用最受限的情况下（患者自由选择医疗机构且政府参与分担盈利或亏损），机器学习模型依然能够降低医疗机构面临的财务风险。以一个中型的乡镇卫生院为例，假设该乡镇卫生院年收入为 1 000 万元，医保超支造成亏损 100 万元。若经梯度提升模型调整医保支付额，根据支付和分担方式的不同，亏损可降低 0.83 万—4.19 万元，显著节省了财务成本。

表 6 四种组合方式下乡镇卫生院整体财务风险

	第一种支付方式			第二种支付方式		
	无风险走廊	单边风险走廊	双边风险走廊	无风险走廊	单边风险走廊	双边风险走廊
	(1)	(2)	(3)	(4)	(5)	(6)
线性回归	1.097	1.097	0.922	2.621	2.461	1.325
梯度提升	1.051	1.051	0.898	2.572	2.420	1.314
降低比例	4.19%	4.19%	2.60%	1.87%	1.67%	0.83%

注：(1) 四种组合方式为患者只能在固定乡镇卫生院就诊的第一种支付方式或患者可自由选择乡镇卫生院就诊的第二种支付方式，与以线性回归预测结果为支付依据或以梯度提升模型预测结果为支付依据的组合；(2) 不同组合方式下整体财务风险为所有乡镇卫生院的收入成本比率与 1 的差值的绝对值的总和（$\sum |1-$（收入成本比率）$|$）。

综上所述，更好的医疗费用预测模型不单对医保支付方来说更有利于其成本预测从而控费降费、有利于医疗资源的跨区域调度，还能有效降低医疗机构面临的财务风险，缓解医疗机构和医保基金的运行压力。模拟结果提示，更精确的医疗费用预测模型可有效减低医保支付方式改革推进过程中所受的阻力。

六、结 论

本文比较了传统线性回归模型与多种机器学习模型在微观医疗费用预测方面的能力，以支持机器学习模型在医保支付领域的应用。实证结果表明，大多数机器学习模型在对医疗费用进行跨期样本外预测时的性能不劣于线性回归模型，其中梯度提升模型表现最优。无论是对于总医疗费用进行预测，还是对于门诊、住院和慢病费用进行预测，机器学习模型相较于线性回归模型都存在明显优势。本文进一步发现，机器学习模型相较于线性回归模型的优势十分稳健。最后，本文构建了一套根据风险调整的按人头付费的支付体系，展示了如何将模型预测结果应用于实际医保支付。通过政策模拟，我们发现以梯度提升模型预测结果作为医疗机构支付水平的依据时，医疗机构的整体财务风险将低于以线性回归模型作为依据的整体财务风险。

当前，我国正在进行医保支付方式改革，以提高医疗体系运行效率并控制医疗费用不合理增长。在此时代背景下，希望本研究能引起更多有关根据风险调整的医保人头费改革方式的探讨。机器学习模型所提供的强大的跨期样本外预测能力能为医疗支付改革提供理论和实践支持，助力经风险调整的按人头付费制、总额预付制等多种医保支付方式的推行和完善。另外，随着医疗大数据的可获得性大幅提升，使用机器学习模型预测医疗费用可为政府规范医保基金管理提供技术协助，在降低医保基金透支风险的同时促

进医疗资源公平惠及所有社会成员。最后，利用机器学习模型作为对医疗机构支付的依据，将减少医疗机构的收入不确定性，减少医疗资源过度使用，促进医疗资源实现更好的跨地区平衡。

但需要指出的是，首先，本文的分析是基于某县级市的早期新农合数据，由于地理位置、经济发展水平等差异，基于其他地区数据的分析得出的结论可能不同，因此，本文结论存在一定的局限性。但是，本文的主要目的是在方法论上比较传统线性回归模型和多种经典的机器学习模型，并展示更好的医疗费用跨期预测模型如何降低医疗机构所面临的财务风险。加之，文中采用的线性回归模型和机器学习模型均具有广泛的适用性，后续研究可采用近年的城镇职工医保或城乡居民医保数据，利用该支付框架探索适合本地区的支付方式。其次，本文构建的预测模型仅采用有限的个人客观特征变量，从而限制了模型的预测潜力。这是由于如若包含个人经济状况相关变量或医疗机构特征等与个人选择相关的变量作为最终风险调整依据可能会影响医保支付的公平性。最后，疾病类型与医疗费用高度相关，虽两类模型都控制了疾病信息，但都只把病种分成了 21 大类。理论上如何将病种类别划分更细以供更精确的医疗费用预测，如何有针对性地为我国医保设计更精确的费用预测模型都是值得下一步深入研究的课题。

参 考 文 献

[1] Buchner, F., D. Goepffarth,, and J. Wasem, "The New Risk Adjustment Formula in Germany: Implementation and First Experiences", *Health Policy*, 2013, 109 (3), 253-262.

[2] Duncan, I., M. Loginov, and M. Ludkovski, "Testing Alternative Regression Frameworks for Predictive Modeling of Health Care Costs", *North American Actuarial Journal*, 2016, 20 (1), 65-87.

[3] Ellis, R. P., D. G. Fiebig, M. Johar, G. Jones, and E. Savage, "Explaining Health Care Expenditure Variation: Large Sample Evidence Using Linked Survey and Health Administrative Data", *Health Economics*, 2013, 22 (9), 1093-1110.

[4] 封进，"夯筑可持续的中国医保体系"，《中国经济报告》，2016 年第 11 期，第 32—36 页。

[5] Garber, A. M., and J. S. Skinner, "Is American Health Care Uniquely Inefficient", *Journal of Economic Perspectives*, 2008, 22 (4), 27-50.

[6] Iommi, M., S. Bergquist, G. Fiorentini, and F. Paolucci, "Comparing Risk Adjustment Estimation Methods Under Data Availability Constraints", *Health Economics*, 2022, 10.1002/hec.4512. Advance online publication.

[7] Jones, A. M., "Health Econometrics", In: Culyer, A. J., and J. P. Newhouse (eds.), *Handbook of Health Economics Volume* 1. Amsterdam: Elsevier, 2000.

[8] Jones, A. M., "Models for Health Care", The University of York, HEDG Working Paper 10/01, 2010.

[9] Kautter, J., G. C. Pope, M. Ingber, S. Freeman, L. Patterson, M. Cohen, and P. Keenan, "The HHS-HCC Risk Adjustment Model for Individual and Small Group Markets under the Affordable Care Act", *Medicare & Medicaid Research Review*, 2014, 4 (3), 16.

[10] 刘国恩，"中国医改谏言"，《中国经济报告》，2016 年第 3 期，第 36—38 页。

[11] Malehi, A. S., F. Pourmotahari, and K. A. Angali, "Statistical Models for the Analysis of Skewed Healthcare Cost Data: A Simulation Study", *Health Economics Review*, 2015, 5, 11.

[12] Mesike, G. C., I. A. Adeleke, and A. Ibiwoye, "Predictive Actuarial Modeling of Health Insurance Claims Costs", *International Journal of Mathematics and Computation*, 2012, 14 (1), 34-45.

[13] Mihaylova, B., A. Briggs, A. O'Hagan, and S. G. Thompson, "Review of Statistical Methods for Analysing Healthcare Resources and Costs", *Health Economics*, 2011, 20 (8), 897-916.

[14] Morid, M. A., K. Kawamoto, T. Ault, J. Dorius, and S. Abdelrahman, "Supervised Learning Methods for Predicting Healthcare Costs: Systematic Literature Review and Empirical Evaluation", In AMIA Annual Symposium Proceedings (American Medical Informatics Association), 2017, 1312.

[15] Newhouse, J. P., M. Price, J. Huang, J. M. McWilliams, and J. Hsu, "Steps to Reduce Favorable Risk Selection in Medicare Advantage Largely Succeeded, Boding Well for Health Insurance Exchanges", *Health Affairs*, 2012, 31 (12), 2618-2628.

[16] Pope, G. C., J. Kautter, M. J. Ingber, S. Freeman, R. Sekar, and C. Newhart, "Final Report: Evaluation of the CMS-HCC risk adjustment model.", DC: Center for Medical and Medicaid Services, 2011.

[17] Rose, S., "A Machine Learning Framework for Plan Payment Risk Adjustment", *Health Services Research*, 2016, 51 (6), 2358-2374.

[18] Shi, J., Y. Yao,, and G. Liu, "Modeling Individual Health Care Expenditures in China: Evidence to Assist Payment Reform in Public Insurance", *Health Economics*, 2018, 27 (12), 1945-1962.

[19] Song, Z., D. G. Safran, B. E. Landon, M. B. Landrum, Y. He, R. E. Mechanic, M. P. Day, and M. E. Chernew, "The 'Alternative Quality Contract,' Based on a Global Budget, Lowered Medical Spending and Improved Quality", *Health Affairs*, 2012, 31 (8), 1885-1894.

[20] Van de Ven, W. P. M. M., and R. P. Ellis, "Risk Adjustment in Competitive Health Plan Markets", In: Culyer, A. J., and J. P. Newhouse (eds.), *Handbook of Health Economics Volume* 1. Amsterdam: Elsevier, 2000.

[21] Van Kleef, R. C., R. C. Van Vliet, and W. P. Van de Ven, "Risk Equalization in the Netherlands: An Empirical Evaluation", *Expert Review of Pharmacoeconomics & Outcomes Research*, 2013, 13 (6), 829-839.

[22] 温小霓、刘鹏、杨楠堃，"人口老龄化与经济增长下的医疗费用预测"，《中国卫生经济》，2014 年第 33 卷第 2 期，第 56—59 页。

[23] 夏涛、徐辉煌、郑建立，"基于机器学习的冠心病住院费用预测研究"，《智能计算机与应用》，2019 年第 9 卷第 5 期，第 35—39 页。

[24] Yip, W., and W. Hsiao, "Harnessing the Privatisation of China's Fragmented Health-Care Delivery", *The Lancet*, 2014, 384 (9945), 805-818.

[25] 张宁、陈浩、周亮、包竹青、高珊、赵颖旭，"基于机器学习模型的糖尿病带病人群医疗险风险保费测算"，《保险研究》，2020 年第 11 期，第 79—95 页。

[26] 赵颖旭、包竹青、高珊、周亮、刘逸圣、陈浩、张宁，"考虑老年痴呆症的医疗险住院费用预测与比较——基于机器学习模型"，《保险研究》，2020 年第 9 期，第 64—76 页。

[27] 周绿林、刘石柱、周以林、梅强，"我国医疗费用趋势预测研究"，《中国卫生经济》，2008 年第 5 期，第 16—18 页。

[28] 朱恒鹏、彭晓博，"医疗价格形成机制和医疗保险支付方式的历史演变——国际比较及对中国的启示"，《国际经济评论》，2018 年第 1 期，第 24—38＋4 页。

Modeling Individual Medical Expenditures: From Linear Regression to Machine Learning

SHI Julie
(Peking University)
WANG Xiaoqian
(Nankai University)
WANG Xi*
(Peking University)

Abstract: Using a large sample of medical data, we compare the performance of the traditional linear model and machine learning methods in predicting individuals' medical expenditures. We find that many machine learning models outperform the linear model in terms of out-of-sample predictability. We also propose a risk-adjusted payment method for the social health insurance system and simulate the payment to providers according to predictions based on the linear model and the Gradient Boosting Decision Tree (GBDT) model, the machine learning method with the best predictive performance. We find that the GBDT model could reduce the financial risk faced by healthcare providers.

Keywords: medical expenditure prediction; machine learning; payment system reform

JEL Classification: I18, C40, I13

* Corresponding Author: Wang Xi, School of Economics, Peking University, Beijing 100871, China; Tel: 86-10-62753635; E-mail: wang.x@pku.edu.cn.