

大科技信贷：一个新的信用风险管理框架

黄益平、邱晗

管理世界，2021.2

吕漫妮

2021.6.3

目录

- 引言
- 研究背景和动机
- 研究问题和结论
- 贡献
- 理论框架
- 实证检验
- 结论与潜在问题

研究背景和研究动机

- 发展普惠金融是各国政府普遍面临的挑战，其中为中小企业特别是小微企业提供融资的任务尤为艰巨
 - 由于获客难和风控难，在我国中小企业获得银行贷款的比例一度只有20%，即便过去十几年来政府采取了许多改善措施，但是“融资难”的局面并没有得到根本性改变
 - 近十年来，大科技信贷的发展为解决该问题带来了曙光，并且典型互联网银行的平均不良贷款率远低于传统商业银行，提高微观层面的风控效率，从而促进金融普惠性
- 大科技信贷建立了一个怎么样的信用风险管理框架，以解决核心痛点？它有什么样的特点？

研究背景和研究动机

- 很多银行贷款都是以不动产（主要是房地产）做抵押，在金融加速器机制下，当经济遭受负面冲击时，房价、信贷与经济之间很容易形成恶性循环，这也是许多金融危机的主要形成原因。而小微企业尤其依赖抵押贷款，所以在金融危机中往往小微企业损失最大。以大数据为依据的大科技风控模型，可以为企业尤其是小微企业提供无抵押的信用贷款，减弱了房价与信贷之间的关联。那么，房价下跌不会导致信贷收缩，也就不会导致经济加速下行
 - 从这个角度看，大科技信贷有可能增强金融与经济的稳定性
- 宏观层面上，除了增加信贷广度，大科技信贷的是否可以加强经济金融体系的稳定性？

研究问题与结论

- 大科技信贷建立了一个怎么样的信用风险管理框架，以解决核心痛点？它具有什么样的特点？
 - 大科技信贷信用风险管理框架以大科技生态系统 and 大数据风控模型为两大支柱
 - 具有模型优势和信息优势，更高维度的数据、更广泛的客户群体和更有时效性的信息有助于提升金融科技风控模型的有效性；具有群体异质性，能够改善信用历史较短用户的风险评级，从而促进普惠金融
- 宏观层面上，除了增加信贷广度，大科技信贷的是否可以加强经济金融体系的稳定性？
 - 通过对比银行信贷和大科技信贷，本文发现大科技信贷可以减弱信贷与资产价格的相关性，促进经济稳定

贡献

- 首先提出大科技信贷的信用风险管理框架并详细分析了大科技生态系统和大数据风控模型这两大工具的特点与贡献
- 从金融普惠和经济稳定性两个角度，论证了大科技信贷在宏观层面上的重要作用

大科技信贷的信用风险管理框架

- 大科技生态系统
- 大数据风控模型

大科技生态系统

微众银行——微信

网商银行——淘宝

新网银行——开放银行的平台，链接其他大科技生态平台比如滴滴打车、美团

- 大科技生态系统：基于大科技平台建立起来的包括商业、金融、社交、娱乐等活动的综合服务体系，实际上也是一个集触达用户、贷款申请、信用评价与还款管理于一体的生态闭环
- 特点：借款人的大部分活动都在平台上进行，可以降低信息不对称的程度
- 优势：
 1. **获客**：平台建立之后，可以以非常低的边际成本连接数以亿计的用户（长尾效应）
 2. **大数据**：**高维度**数据用于贷前风险评估；允许**实时**监测借款人状况，及时对风险因素做出反应
 3. **还款管理**：如期还款有益于生态系统中的信用与声誉，从而激励还款；接入央行征信加以威慑

大数据风控模型

大数据风控模型：区分于依赖传统数据与评分卡模型的传统银行风控模型

优势：

- 1. 大数据（信息优势）：**高维度数据，高体量数据；实时数据
- 2. 机器学习模型（模型优势）：**更为复杂的模型，擅长处理数量庞大的数据，更能够抓住一些复杂的非线性关系以及解释变量之间的交互作用，从而更准确的刻画用户的违约特征

此外，云计算和人工智能则可以大大提升处理的速度与规模，从而显著降低信贷的运营成本

大数据风控模型 – 数据

贷款数据：中国一家头部金融科技公司2017年6月至9月的抽样逐笔贷款数据

业务逻辑：用户在提交贷款申请时该公司会在用户的同意下收集相关风控信息，根据这个信息计算出一个相应的信用分，并依据这个信用分决定是否批准用户的贷款申请

变量：

- 1. 传统信息变量（36个）：**包括用户的收入、年龄、性别、不同时间段下银行信用卡信息（违约情况、使用频率、额度、使用历史）。这些传统数据都是**银行可以观察到的信息，同时也被广泛运用于银行风控之中**
- 2. 非传统信息变量（99个）：**具体包括用户在网络平台上的消费数据，用户的社交数据以及用户的一些其他行为信息

大数据风控模型 – 模型设定

模型1：基于传统信息的逻辑回归模型（评分卡模型的代表）

模型2：基于所有信息的逻辑回归模型（评分卡模型的代表）

模型3：基于传统信息的随机森林模型（机器学习模型的代表）

模型4：基于所有信息的随机森林模型（机器学习模型的代表）

预测变量：用户持续30天没有按时还款（金融科技公司的**逾期**定义）

拆分数据：随机选取，7:3

大数据风控模型 – 评估

AUC: ROC曲线下的面积

- 模型1: 基于传统信息的逻辑回归模型
- 模型2: 基于所有信息的逻辑回归模型
- 模型3: 基于传统信息的随机森林模型
- 模型4: 基于所有信息的随机森林模型

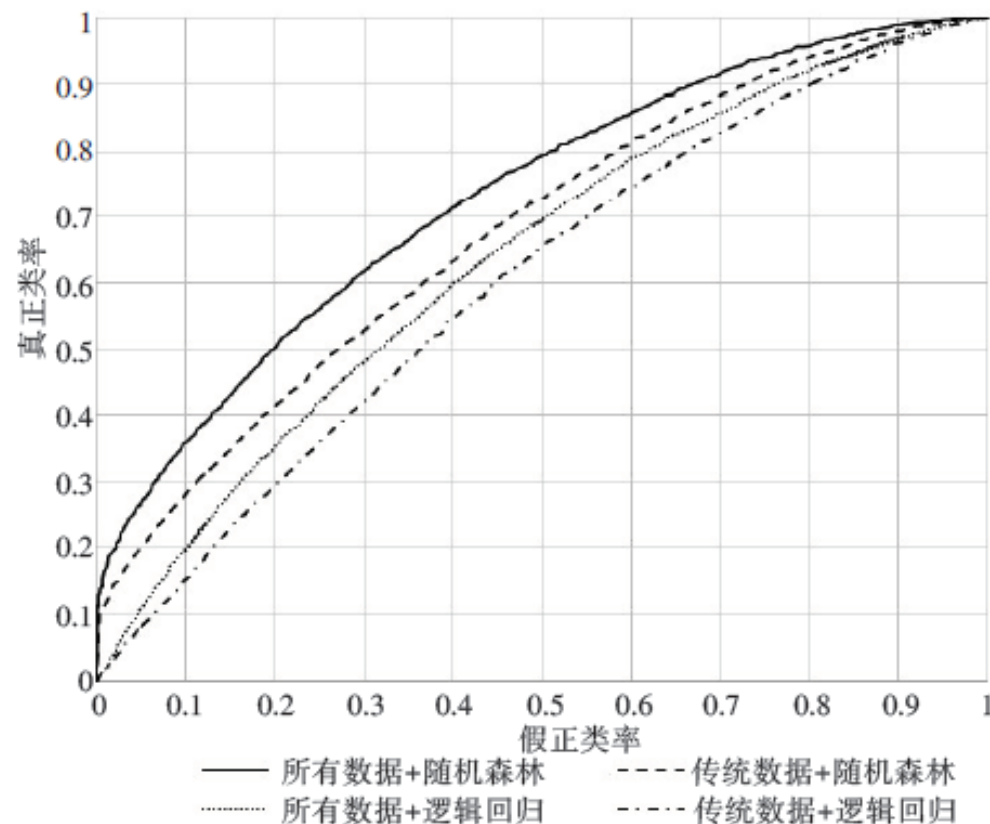


表1 4类模型的AUC比较

模型类别	AUC	95%置信区间
模型1	0.5950	[0.5891, 0.6009]
模型2	0.6361	[0.6303, 0.6418]
模型3	0.6759	[0.6696, 0.6821]
模型4	0.7286	[0.7231, 0.7340]

- 引入非传统数据有助于提升模型的有效性
- 相比于传统的线形模型，随机森林模型有助于提高模型预测能力

大数据风控模型 – 评估

前十名中有7个变量都是传统信用卡信息（被广泛运用于银行风控之中），有3个是非传统的信息变量（来自于大科技信贷生态系统）

表2 随机森林模型下的变量重要性

排序	变量名	重要性分数
1	用户的社交活跃度	0.02747
2	贷款违约的次数	0.02133
3	3个月内信用卡使用的频率	0.02104
4	网络平台消费的额度	0.02058
5	信用卡的坏账率	0.02042
6	Pos 机消费次数	0.01926
7	信用卡的利率	0.01905
8	最近一次网络平台的消费时间	0.01881
9	年龄	0.01808
10	信用卡的额度	0.01744

注：重要性分数是根据随机森林中 Gini 不纯度做的计算。
数值越大说明变量越重要。

➤ 传统信息非常重要，但引入消费、社交等数据可以提升模型的风控效率

大数据风控模型 – 样本量的影响

- 当样本量不断增加，两个模型的预测能力都会增加，但随机森林模型的增益效果更为明显
- 随机森林模型刻画的关系更为复杂，可以做的分类更多，所以样本量的增加对于其预测效果的增加更有帮助

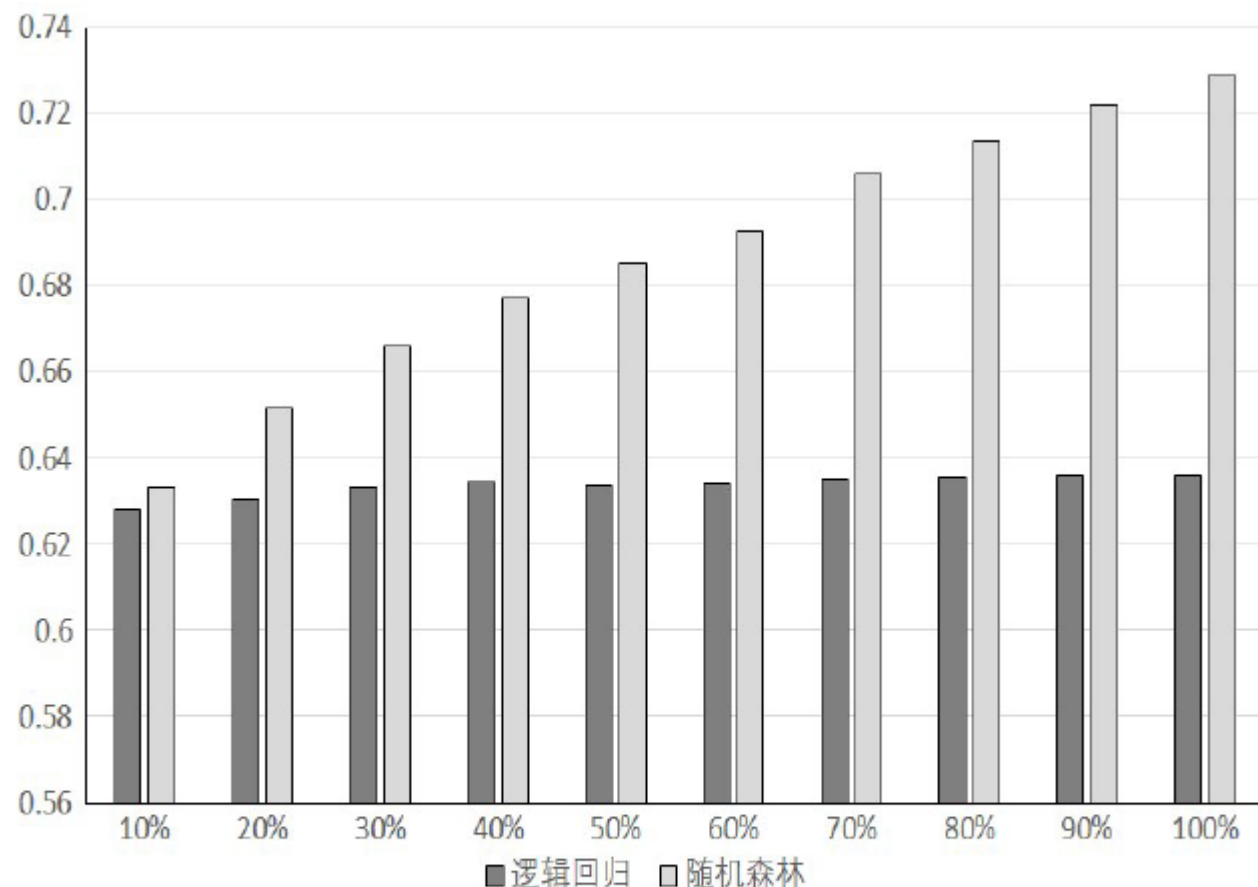


图3 不同样本量下大数据风控模型与传统模型比较

注：纵坐标为模型在测试集上的AUC值，横坐标为实际训练集样本占原训练集样本的百分比（例如10%代表使用了10%的原训练集样本训练模型）。

大数据风控模型 – 信息时效性的影响

- 信息的时效性对传统模型和随机森林模型都有帮助，但对随机森林模型增益更大
- 随机森林对于训练样本的挖掘深度要高于传统模型，挖掘出的规律更为复杂，使用时效性较高的信息，可以保证模型从训练样本中挖掘出来的复杂规律在预测样本中也适用；反之，复杂规律并不适用，随机森林模型可能出现过拟合，导致模型外推能力下降

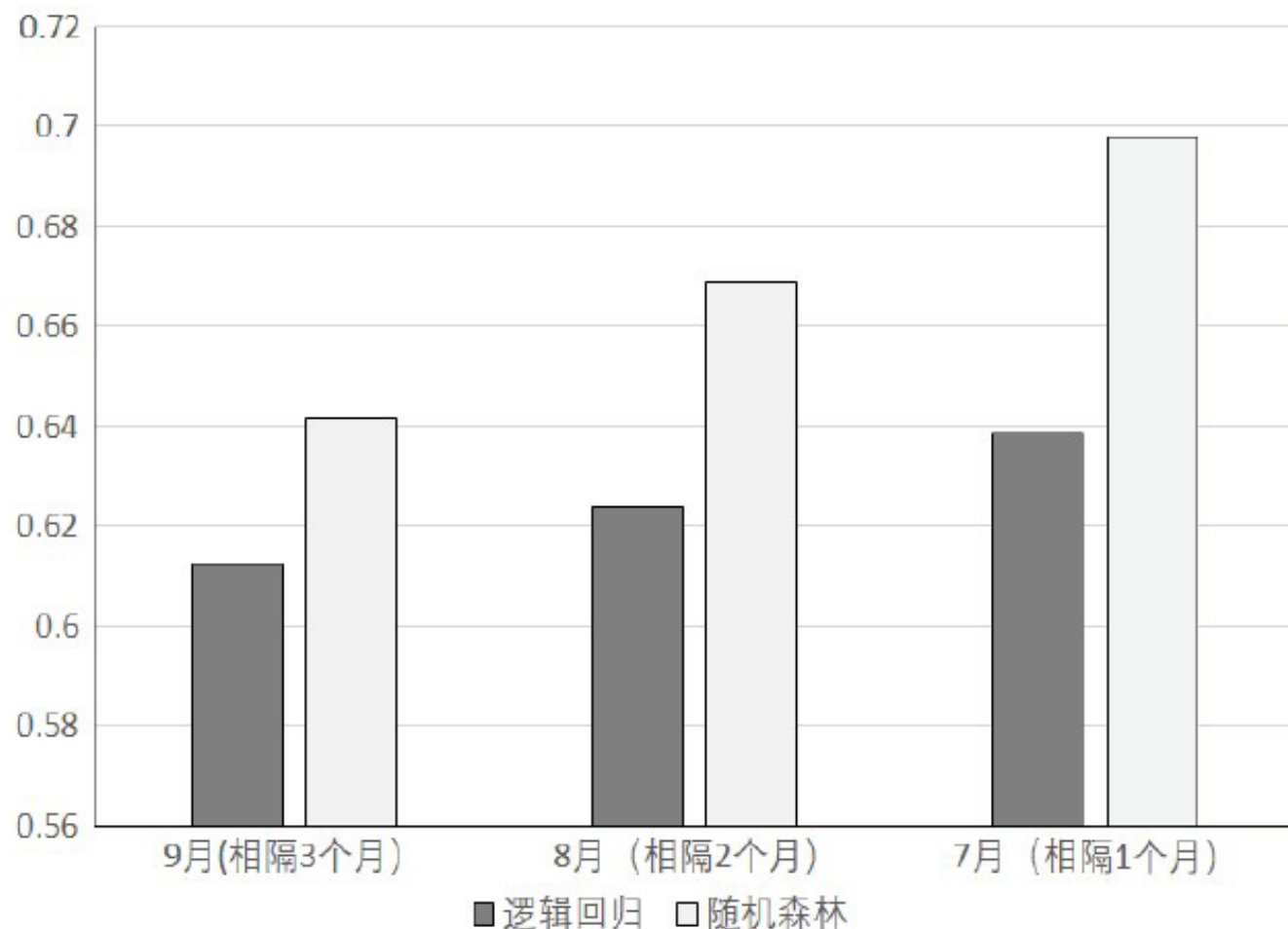
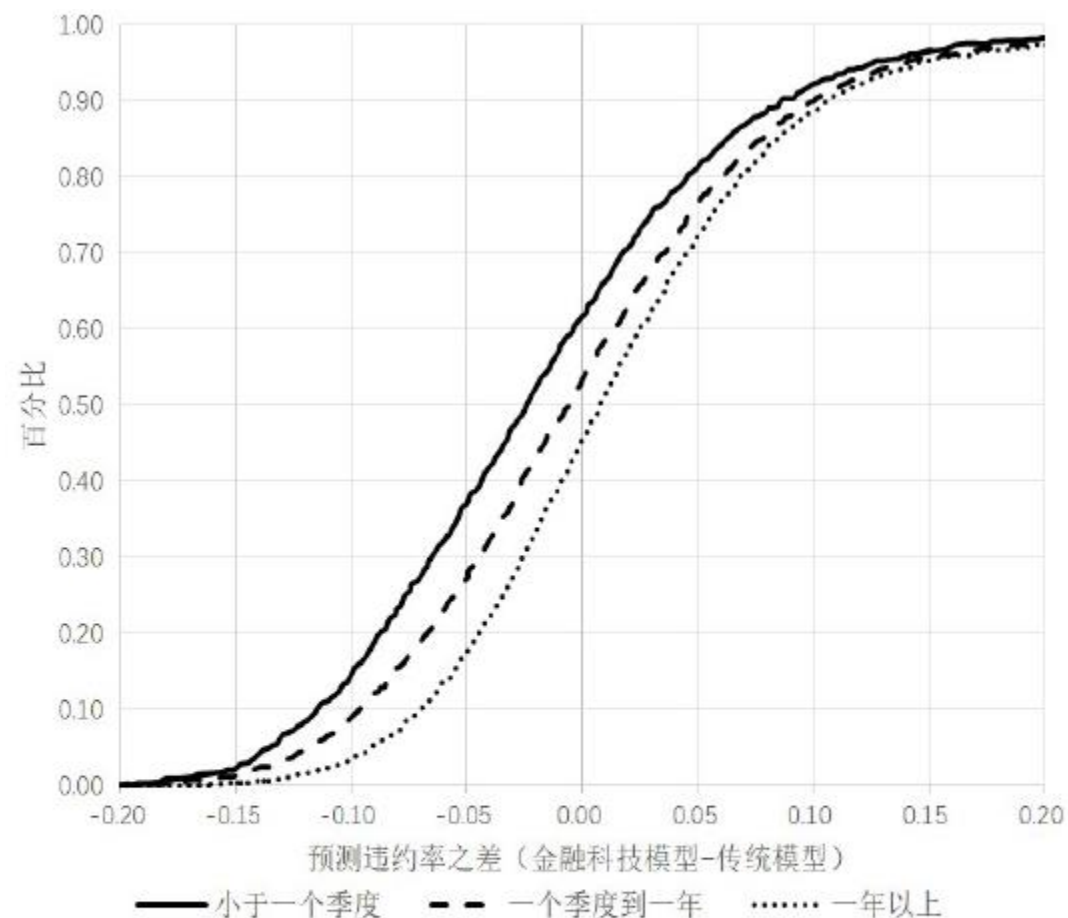


图4 不同信息时效性下大数据风控模型与传统模型的比较
注：纵坐标为模型在测试集上的AUC值，横坐标为测试集的月份。

宏观影响 – 促进金融普惠

- 对于信用卡历史小于一个季度的用户，平均而言大数据风控模型给出的评价要高于传统模型（61%的用户获得了更高的评价）
- 对于信用卡历史大于1 年的用户，只有45%的用户获得了更高的评价
- 这可能是因为信用历史较短的用户，传统财务数据不足，传统方法往往会给出偏谨慎的评价，而大数据风控模型则结合了大量企业主维度的非传统信息，可以给出更为合理的评价，从而提高该类别用户的信贷准入概率



注：横坐标为大数据风控模型预测违约率与传统风控模型预测违约率的差，曲线为其累积分布函数图，图例中，“小于一个季度”表示信用卡历史小于一个季度（3个月）的用户；“一个季度到一年”表示信用卡历史一个季度到一年之间的用户；“一年以上”表示信用卡历史高于1年的用户。

宏观影响 – 提高经济稳定性

- 金融加速器机制：企业融资依赖于企业抵押品的价值，信贷市场与实体经济往往具有很强的同向性
- 抵押品可以有效缓解借贷双方信息不对称的问题，但金融加速器模型机制下，金融市场会加剧实体经济的波动，从而放大经济的不稳定性
- 小微企业贷款严重依赖抵押品

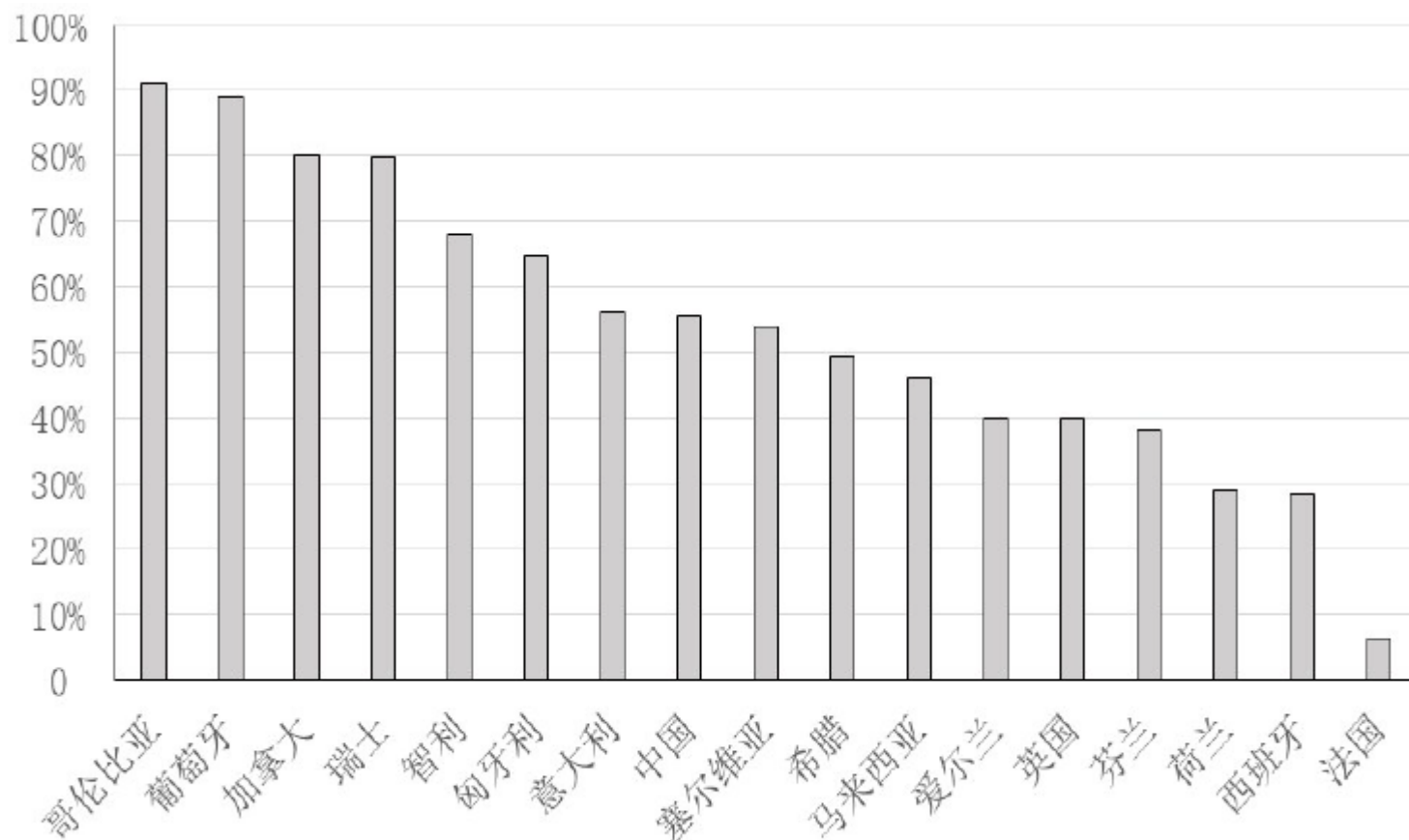


图6 各国小微企业抵押贷款占比(2015)

宏观影响 – 提高经济稳定性 – 数据和模型

- 数据：2005~2013年传统银行小微企业（调查网点层面）、2011~2017年网商银行小微企业信用贷款金额数据（地级市层面）、城市房价、人均GDP、常住人口、银行网点覆盖率（每平方公里银行网点数）
- 传统银行数据来源于北京大学县域金融机构调查（边文龙等，2017）
- 网商银行数据来源于蚂蚁研究院

$$y_{i,t} = \alpha + \beta \times \ln hp_{i,t} + \gamma \times Z_{i,t} + \delta_i + \eta_t + \varepsilon_{i,t}$$

- 被解释变量：对数总信贷额
- 解释变量：房价
- 控制变量：对数人均GDP、银行网点覆盖率、对数常住人口数

宏观影响 – 提高经济稳定性

- 相比于传统银行贷款，基于平台生态和大数据模型的科技贷款与资产价格的共振并不强烈，能够提升经济体系的稳定性
- 问题：
 1. 样本期间不一致，但是以控制时间固定效应进行缓解
 2. 大科技信贷所参考的经营流水等数据是否会带来新的加速器机制？可能减弱了“金融加速器”的作用，也可能增强“其他加速器”的作用？

表 3 资产价格与银行信贷供给

	传统银行				网商银行
贷款额度(log)	总贷款	信用贷款	抵押贷款	小微企业贷款	小微企业贷款
房地产价格(log)	0.373* (0.207)	-0.0694 (0.428)	0.908** (0.432)	0.491** (0.192)	0.218 (0.451)
控制变量	是	是	是	是	是
年度固定效应	是	是	是	是	是
城市固定效应	是	是	是	是	是
R	0.229	0.169	0.253	0.033	0.702
观测值	3086	2207	2176	2385	1469

注：因变量为总信贷额度(取对数)，自变量为房地产价格(取对数)，控制变量为人均GDP(取对数)，银行网点覆盖率(网点数/城市面积)，常住人口数(取对数)，回归控制了城市固定效应和年份固定效应，*** $p < 0.01$ ；** $p < 0.05$ ；* $p < 0.1$ 。

结论

- 本文总结了全新的大科技信贷的信用风险管理框架，主要依靠两大数字技术工具，一是大科技生态系统，二是大数据风控模型
- 本文的实证研究表明，这类新型的风控模型具有模型优势和信息优势。它不仅更加准确地预测违约率（微观风控），更重要的是可以为许多缺少银行信用历史的借款人提供融资服务，充分地体现了大科技信贷的普惠性特点。同时，由于大科技信贷不依赖抵押资产，弱化了资产价格与信贷之间的顺周期机制，有利于提高金融与经济的稳定性。

潜在问题

- 大科技信贷的风控方法解释性较低，类似于“黑箱子”，监管部门并不能很好地识别这类贷款的稳定性
- 如果企业与个人明白大科技平台及其生态系统根据他们留下的数字足迹作为信用风险管理框架的基础，他们的线上“足迹”可能会因此改变，从而带来欺诈风险
- 大科技利用用户的社交、购物、网页浏览和金融账户等数据进行建模，客观上提高了风控能力，但也可能带来数据隐私和“算法歧视”等问题
- 基于数据的贷款虽然提高了金融的普惠程度，但也可能加大一些特殊群体（例如中老年群体）与社会的差距，导致“数字鸿沟”的产生