

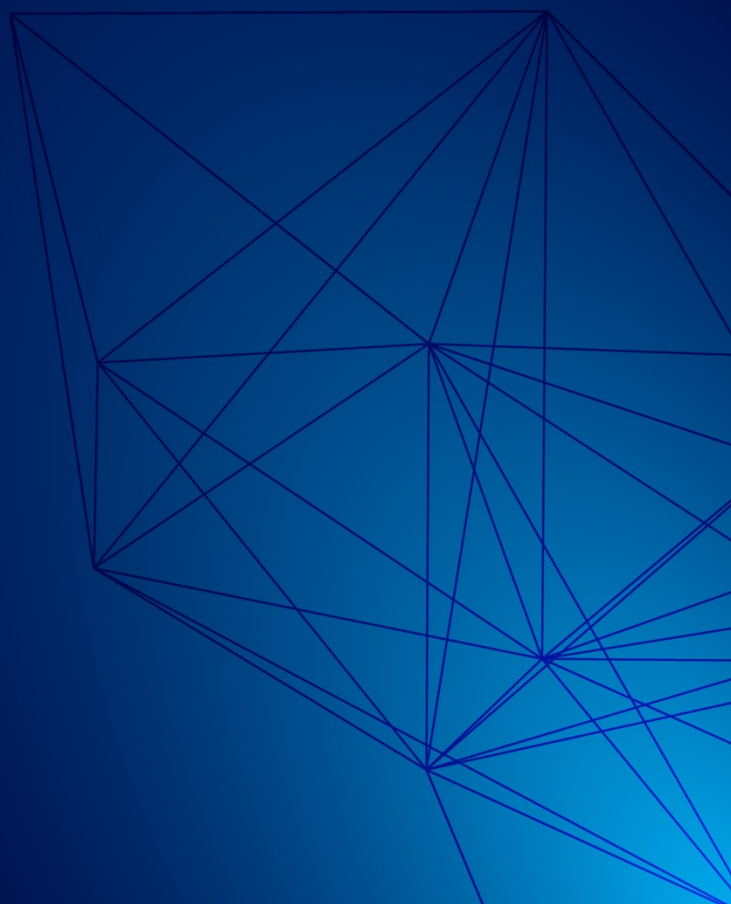


北大汇丰

PHBS FINANCIAL FRONTIER DIALOGUE

金融前沿对话

2019 年第 10 期 总第 21 期



PHBS HFRI
北京大学汇丰金融研究院

主办单位：北京大学汇丰金融研究院

院长：海闻

执行院长：巴曙松

秘书长：本力

编辑：都闻心（执行） 叶静 曹明明 鞠琳琳 唐小丽 方培豪 朱伟豪

北京大学汇丰金融研究院简介

北京大学汇丰金融研究院 (The HSBC Financial Research Institute at Peking University, 缩写 HFRI) 成立于 2008 年 12 月, 研究院接受汇丰银行慈善基金会资助, 致力于加强国内外著名高校、金融研究机构、以及知名金融学者之间的交流, 构建开放的学术、政策交流平台, 旨在提高中国金融理论与政策的研究水平, 促进学术繁荣与发展, 加强与政府决策部门的联系与合作, 为政府决策提供参考意见, 为国际金融体制改革和中国金融业的发展做出贡献。

北京大学汇丰金融研究院院长为北京大学校务委员会副主任、北京大学汇丰商学院院长海闻教授, 执行院长为中国银行业协会首席经济学家、中国宏观经济学会副会长巴曙松教授。

大数据与 AI 在证券化研究中的应用

【对话主持】

北京大学汇丰金融研究院执行院长、中国银行业协会首席经济学家、中国宏观经济学会副会长巴曙松

【特邀嘉宾】

MSCI 董事总经理及证券化产品研发主管 David Zhang

一、美国证券化产品的基本状况

美国固定收益市场中将近 1/3 是证券化产品。在 2007 年金融危机之前，证券化产品是全球最大的金融资产，大于美国国债市场。目前，因国债市场上涨较多，证券化产品市场成为第二大金融资产。

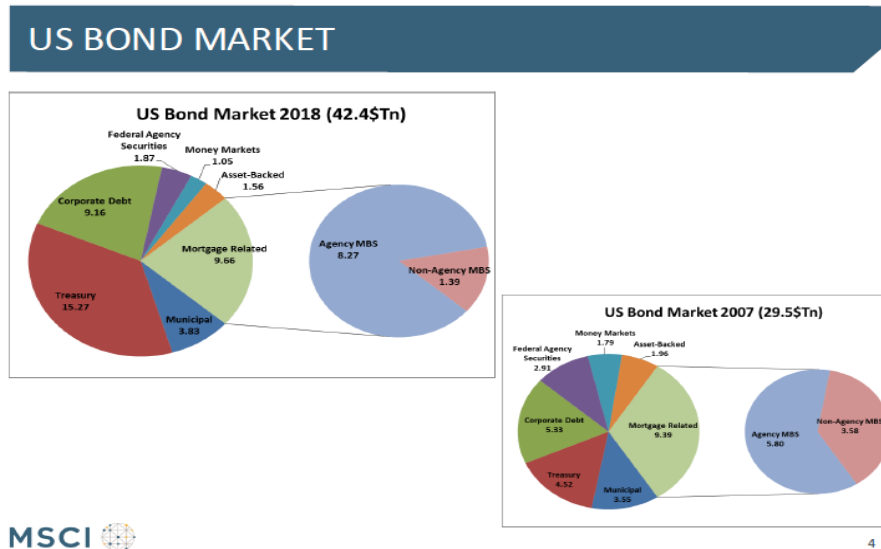


图 1

上图显示了 2007 年与 2018 年美国证券化市场分类的变化。在

2007 年之前，非“两房”持有的房地产贷款证券化市场很大，但现在已基本消失。而目前，车贷市场和信用卡贷款市场反而比金融危机前更大。现在证券化市场已恢复到金融危机之前的水准，但是在分类上尚有区别。美国早在 1930 年代就开始做证券化市场，发展至今已相对成熟，在金融危机前是其最大的金融资产类别。例如，美国的消费者信贷中 60% 的资金来自证券化市场；工业方面的信贷市场资金将近 50% 通过证券化市场而来。相比而言，欧洲证券化市场规模较小，在其整体金融资产规模中占比大概在 10% 左右。当前中国的证券化市场发展也很快。从 2013-2014 年开始，中国证券化市场突飞猛进，目前已超过欧洲，成为全球第二大证券化市场，如果按照目前的趋势发展，十年之内有可能超越美国。

在美国，证券化除了在提供融资方面起到重要作用，宏观经济政策也与证券化息息相关。例如，2009 年美联储启动量化宽松，其中有将近一半资金，约 1.8 万亿美元购买了两房的证券化资产。可见，证券化在美国有很大市场，且不光关系到经济、消费，还关系到货币政策、金融政策和税收等。

二、为什么要将大数据和 AI 用于证券化

从定价到防范风险再到模型方面，证券化具有以下几方面特点：

一是数据量大。美国证券化市场的数据量是巨大的。即使是在公共市场里面最小的数据群，也有将近两个 G。因为在美国有将近一亿个政府资助证券化项目，已有 20 余年历史，每个月都有数据沉淀，

积累下来数据量巨大。而延伸到银行贷款方面，数据比公共数据还大一千倍。

二是风险因子很多。从 30 到 300 到上万都有可能，证券化处理过程是金融机构给个人贷款，资金市场再进行打包，最早贷款的数据量比做证券化的公共数据大很多倍。

三是风险因子非线性，风险因子相互作用，因此模型、避险、定价、数据都不均匀。例如 2003 年到 2007 年信用环境较松，这种情况下的数据和现在信贷数据较紧时的数据，不能够一视同仁。这是风险、定价、模型困难的地方之一，还有技术方面的挑战。这些挑战通过大数据和人工智能更容易解决，这是大数据和人工智能较合适的切入点，如利用大数据评估资产违约、坏账等方面风险，利用人工智能做证券化提早付款模型。

三、通过大数据看美国高房贷坏账情况

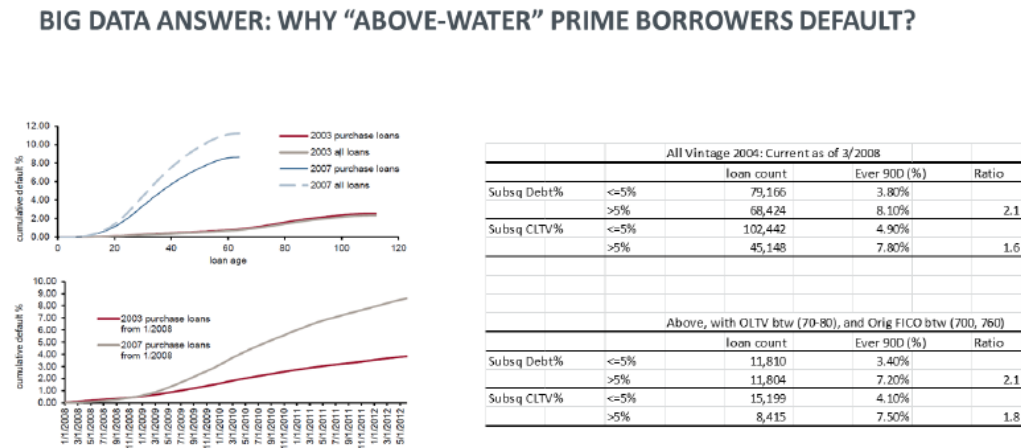


图 2

2008 年金融危机之前的房贷危机，使得美国房价下跌，产生大规模坏账。2007 年的坏账率高达百分之十几，其中有个很奇怪的现象是，将近 1/3 的房贷坏账对应的房价没有损失，如 2003 年买房，价格 100 万，贷款 80 万，到 2006 年，房价涨到 200 万，再到 2007 年跌回 100 万，虽然借款人没有亏，但坏账率仍然很高，剔除提前付款的因素，坏账率要高达 8%。这个问题引起学术界与监管的许多猜想。如，是否是购房时未贷款，而是之后用已购房产再贷款的人群坏账率较高。类似的问题我们借助大数据研究发现，2003-2005 年房价上涨一倍，房主贷款人做了很多二级房贷、消费贷款等，造成借贷压力较大，导致坏账。

资产证券化产品由于是打包出售，投资者无法获得太多关于资产、信贷安排、风险缓释安排等方面的信息和数据，我们希望通过大数据的方法把公共数据和私人数据进行连接。这样一来，对于发行债券或者发行股票，投资者就能看到更多数据。在实际推进中需要纳入三大个人征信机构等方面数据，存在数据量巨大等方面挑战。

四、为什么要建立代理 MBS 的机器学习模型

提早付款率是非常复杂的现象，且非线性，有大量的非系统性风险因子。近几年，特别是阿尔法打败了李世石后，我们觉得软件功能、硬件功能都提高了很多，所以允许我们尝试用人工智能的手段来解决模型的问题。机器学习模型在图像识别、自然语言处理、欺诈检测等领域有着广泛的应用前景。MBS 人工建模有一定的工匠性，建模时间

长，且同样的数据做出的模型会有不同，有很多个人主观因素，这会产生金融模型准确性等方面的问题，可信程度易受到质疑，因此需要通过人工智能进行一定改进。

提早付账建模困难，主要是因为以下原因。第一，数据量大。量在 2-2000G 之间。在 20 年间，美国已有将近一亿个贷款，每月都有数据沉淀。第二，多种风险因素。例如贷款规模与个人经济收入等有关，与可节省资金也呈非线性关系，与贷款目的也有关（如房贷是为了购房还是为了再融资，风险水平不同），多风险因素增加了建模的难度。一般建模有经验的需要一年时间，如果通过人工智能参与，会有革命性贡献。首先，人工智能建模非常精确；其次，建模速度极快，人工需要一年完成的建模，它仅需 3 小时。这是金融技术的颠覆性革新，后续 3-5 年有可能实现标准化推广。

五、机器模型、人工模型运用情况举例

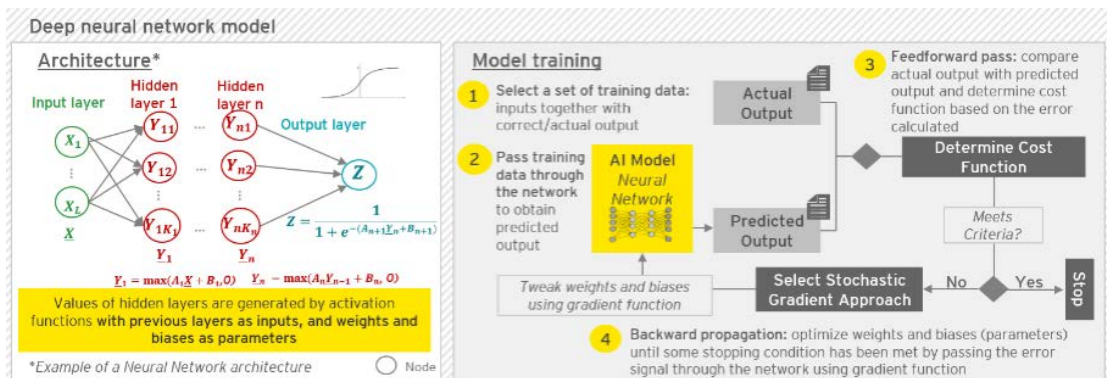


图 3

左边的图表示的是前置神经网络，X、Y 分别表示风险因子，Z 表示你想要预测的数值，这是比较经典的神经元模型。基于“两房”贷

款数据做提前偿付率模型的拟合度验证，结果如下，见图 4。

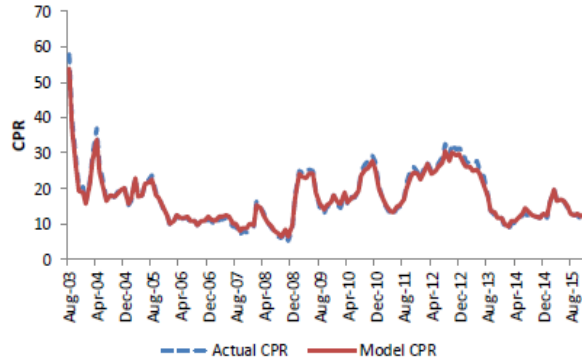


图 4

图 4 是模型的测算结果，蓝线是“两房”贷款提前偿付率实际情况，与模型测算数据基本吻合。模型建立都是基于预期风险因素，但如果出现非预期风险因素，例如飓风等非预期风险，模型的预测结果与实际情况会有出入。因此，在这一方面不能够完全预测市场。

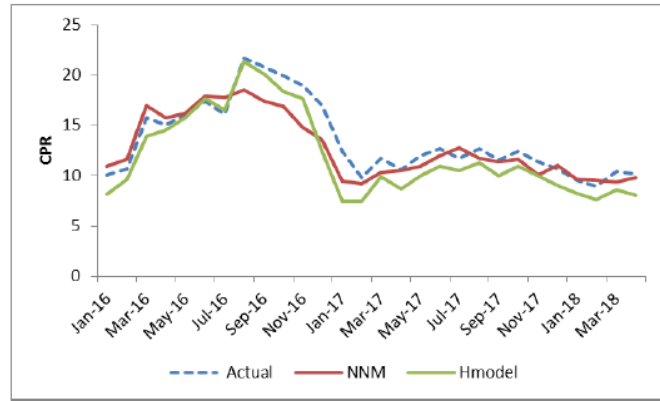


图 5

图 5 的模型中有一段不能对现实做到准确分析，因为在这段时间里有特别的（非预期性）风险因子在起作用。如果可以对历史数据模拟得较好，对预测、分析应当会有很大的帮助。

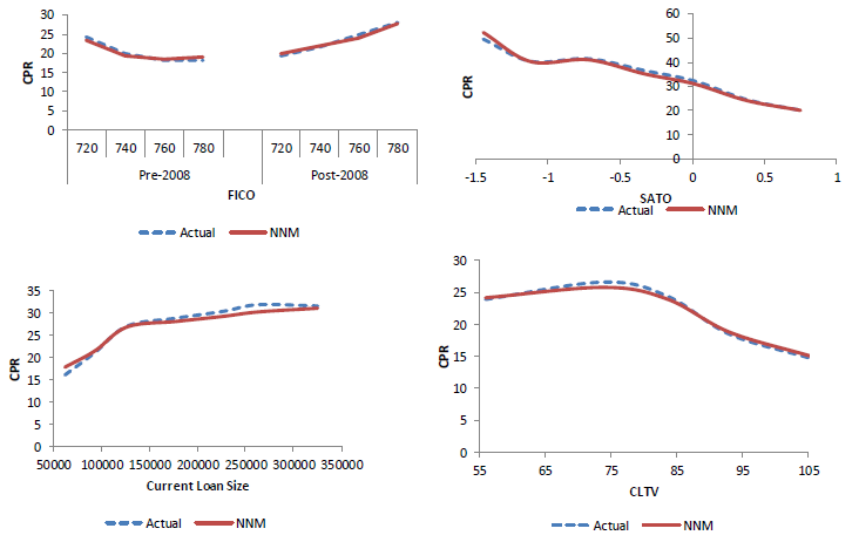


图 6

模型中有 30 余个风险因子，如何验证每一个风险因子都拟合的很好呢？我们分别验证模型在单个风险因子不同数据区间上的拟合情况。左上角验证风险因子 FICO 的拟合情况，FICO 是美国的信用分数，由图可见，在不同的 FICO 分数区间，模型数据与现实数据的拟合度均较高。右上角是 SATO 数据，左下角是贷款数目，对模型影响很大，右下角是贷款和房子总额的比率，如果是 80%，则表示首付款为 20%，在这些风险因子的不同数据区间，模型拟合的都很好。

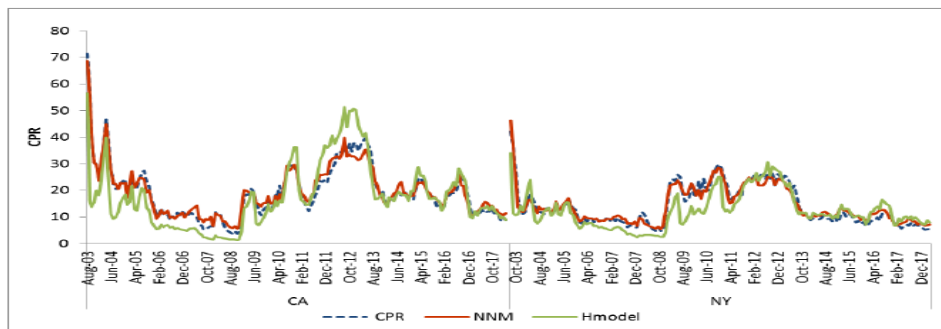


图 7

图 7 是基于加州与纽约房地产贷款数据的 AI 模型与人工模型的

拟合情况。左边是加州，右边是纽约，两者主要区别是纽约有房地产重新贷款税，如果想要重新寻求低利率贷款，政府会征收 70-100bps 的税收。由图可见加州的上涨比纽约快很多，虚线表示实际情况，红线是 AI 模型测算数据，绿线是人工模型测算数据，可见 AI 模型拟合得更好。

之前的拟合度验证都是基于单一风险因子不同数据区间的验证，我们还将不同风险因子整合进行拟合度验证，具体做法是根据预测将数据从高到低排列，再将其分为不同组，将各个组的实际数据与 AI 模型测算数据做对比，验证发现，AI 模型较人工模型拟合度更高。

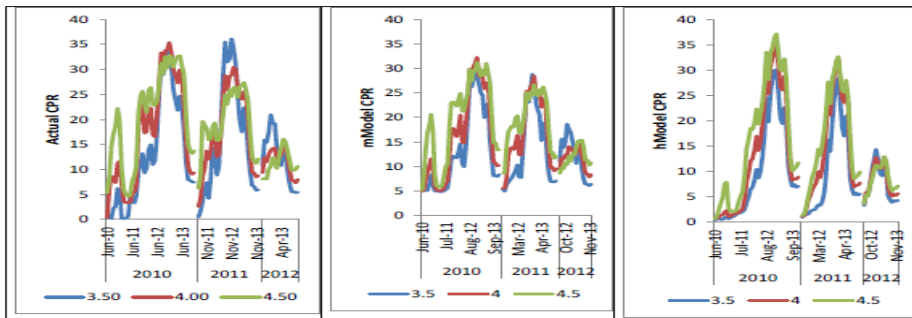


图 8

在“两房”贷款中有一个名词叫“媒介效应”。假如前者房贷利息是 5%，后者房贷利息 4.5% 相比，前者提早付款率高于前者，因为前者重新贷款节省资金更多。但如果利率连续下降，如连续下降 5 年，则后者的提早付账率反而要高，这就被称为“媒介效应”。之前提到提早付账预测模型一般有 30 到 100 个风险因子，风险因子间互相作用，上图就是 5 个风险因子互相作用，人工做模型比较困难。左一是

实际情况，表示利率分别是 3.5%、4%、4.5% 的三种债券的提前还款率。提前还款率一般是 3.5%、4%、4.5% 依次递增。但实际上在 2011 年、2012 年，9-12 个月的时间蓝线在红线之上，红线在绿线之上，这是在特定情况下才会出现的“媒介效应”。中间是机器模型，右边是人的模型，可见机器模型模拟的比人的模型更精确。机器模型只需要三个小时就可以完成一次，由于用时短，还可以做很多不同模型的分析。

六、问答环节：

Q1: 将来随着技术进一步发展，在资产证券化领域，若资产本身产生的过程当中就把人工智能、区块链技术融合进来，使得资产的质量更高，这对后期的分析或将来整个资产证券化跟技术的结合有什么新的好处和进展？

A1: 证券化的产生，一个主要原因就是对于资产如果技术非常发达，任何数据都能得到，这种趋势很明显。通过数据，不光可以看到借款人的负债情况，还有他的收入数据、驾驶数据以及多维度行为数据等，就可以对信贷资产的风险进行定价，这样就不需要再做证券化。国内支付宝就有很大的数据量，计算能力很强，在有的时候就能够做到因人而异的风险预测。将来如果所有的资产就可以引进数据分析，就不用做证券化了，我想这方面也是一种可能性。

Q2: 据我所知人工智能或者说机器学习包含了各种不同的技术，包括你提到的神经网络。我想知道其他的的技术，比如基因算法或全脑

模拟有没有在证券化这个领域被应用，将来会不会被应用？

A2: 人工智能各项技术目前各公司、业界都在发掘，有很多应用，比如神经网络、短期/长期记忆技术等。人工智能近五年发展非常快，几乎日新月异。以前做人工智能需要很多训练，需要很专业的人工智能博士等。但现在人工智能软件方面进展很快，已经做了很多很成熟的 package，在软件技术分享等方面进展很大，一些非人工智能专业的人，对于人工智能可以不知道很多细节，也可以使用这些 package。

Q3: 美国有些团队正在研究如何把区块链的技术直接介入到资产证券化行业里。目前中国的资产证券化发展势头很不错，是不是可以考虑把区块链技术跟资产证券化市场更加紧密结合起来，让中国在这个领域能够走到世界的前列？

A3: 区块链与资产证券化的结合，在美国有好几家公司同时在做，这可能也是一个突破性的技术，通过区块链可以看到资产是否被篡改。但是对于中国的情况我了解不多，到目前为止，我不知道有哪家机构在做区块链与证券化结合。一个月前在北京的会议与几家银行的交流得知，他们正在考虑通过区块链技术做资产证券化，也包括大数据、人工智能等技术，这是一个创新的方向。

本文根据北京大学汇丰金融研究院执行院长巴曙松教授发起并主持的“全球市场与中国连线”第二百九十三期（2019年4月24日）内容整理而成，特邀嘉宾为 MSCI 董事总经理及证券化产品研发主管 David Zhang 先生。

David Zhang, MSCI 董事总经理及证券化产品研发主管。加入 MSCI 之前，David 担任瑞士信贷董事总经理及证券化产品建模主管逾十年，其团队建立的 MBS 模型是固定收益机构投资者最广泛使用的模型之一，多年被机构投资者评为全美最顶级房屋抵押债券早偿风险研究团队。David 还担任全美华人金融协会（TCFA）董事，全球华人地产协会（GCREC）董事，和纽约国际风险管理协会（NYPRMIA）董事等职务。David 获得普林斯顿大学博士学位。

【免责声明】

“全球市场与中国连线”为中国与全球市场间内部专业高端金融交流平台。本期报告由巴曙松教授和王志峰博士共同整理，未经嘉宾本人审阅，文中观点仅代表嘉宾个人观点，不代表任何机构的意见，也不构成投资建议。

本文版权为“全球市场与中国连线”会议秘书处所有，未经事先书面许可，任何机构和个人不得以任何形式翻版、复印、发表或引用本文的任何部分。



PHBS HFRI
北京大学汇丰金融研究院

