



上市公司股权质押风险特征识别与测度

杨柳勇 岑盛楠 贾少卿

摘要：我国经历了两次大规模股权质押风险爆发，事前风控的缺位是重要原因。本文以股权质押事件作为研究样本，以“触及平仓线或发生冻结”作为标签，使用特征工程构造质押需求、治理结构、公司财务、二级市场等多维度特征，使用机器学习中的 XGBoost 算法对质押风险进行预测和分析，关注大股东股权质押风险的事前防控。实证结果表明，XGBoost 算法训练得到的股权质押风险识别模型具有良好的预测性能，在测试集中召回率最高可达 74.44%。此外，特征重要性分析表明，在股东性质方面，境内非国有大股东的股权质押风险更高。在治理结构方面，大股东持股比例越高，股权质押风险越高，而其他大股东持股比例的提升能抑制这种风险。在公司财务方面，股权质押风险较高的上市公司具有更高的成本费用率、更高的负债率和更差的现金流情况。在二级市场方面，股权质押风险较高的上市公司具有较低的市销率和流动性水平。

关键词：股权质押 大股东 特征工程 机器学习

一、引言

我国的股权质押市场在近几年引发了诸多关注，虽然目前的市场规模相较于上一轮高峰期已有了比较显著的收缩，风险正在逐渐缓和，但股权质押市场规模仍然巨大，股权质押风险仍是当前中国资本市场的主要风险之一。2023 年 10 月底召开的中央金融工作会议强调，要全面加强金融监管，有效防范化解金融风险；对风险早识别、早预警、早暴露、早处置，健全具有硬约束的金融风险早期纠正机制。

股权质押风险的爆发，除了股市暴跌这一原因，事前风控的缺位也是重要原因。

股权质押业务缺乏独立性和专业性，准入审核不够严谨和审慎，金融机构经常将其作为与上市公司控股股东建立联系的首个业务，但股权质押的潜在风险也使得该业务成为一把“双刃剑”。例如某证券公司在 2019 年接连踩雷四个股权质押业务，累计计提的信用减值准备达到 1.77 亿元之高，导致其净利润增速从前三季度增长 113.8% 立刻转为负值 -11.5%。这类事件在股权质押市场爆发式增长之后屡见不鲜，增强业务专业性和审慎性已经成为金融机构股权质押业务的当务之急。

如何做好事前防控，防范大规模的股权质押风险再次发生？如何有效甄别股

杨柳勇，浙江大学经济学院教授、博士生导师；岑盛楠，宁银理财有限责任公司；贾少卿，浙江大学经济学院。



股权质押违约可能较高的上市公司及其大股东？这样的上市公司及其大股东在质押需求、治理结构、公司财务、二级市场等方面有什么样的特征？这些问题值得进一步探讨。

当前对股权质押的研究主要集中于大股东股权质押行为的经济后果，而对股权质押风险本身缺乏量化和归因。有少数文献涉及股权质押风险，但重点也在于对大股东规避控制权转移风险的动机进行理论或案例分析。本文尝试量化股权质押风险，将每个股权质押事件标记为高风险或低风险，并使用 XGBoost 算法研究高风险的股权质押上市公司及其大股东的特征，以提高爆雷风险的判别能力，从风险视角丰富股权质押的研究成果，对于二级市场投资者进行投资决策，金融机构等质权方开展股权质押业务，以及监管机构进一步规范市场、防范风险提供参考。

二、文献综述

国内研究主要关注股权质押可能造成的经济后果，并实证研究了大股东股权质押对公司价值、公司绩效、现金持有、盈余管理、股价崩盘风险、关联方资金占用等各方面的影响。

将股权质押变量作为被解释变量的研究，则主要关注大股东股权质押的动机，以及上市公司各方面因素对大股东股权质押行为的影响。在动机方面，相较于把股权质押获得的资金投向被质押的上市公司，大股东更倾向于将资金用于自身或第三方（张陶勇和陈焰华，2014）。对于大股

东而言，股权质押除了满足自身的流动性需求或实现资金的再投资收益，还可能是为了支持或掏空上市公司（艾大力和王斌，2012；吴静，2016）。大部分文献指出大股东股权质押对于上市公司的影响是负面的，股权质押导致现金流权与控制权分离程度增加，致使大股东的激励效应变弱、侵占效应变强，投资效率降低，股价波动加剧（郝项超和梁琪，2009；黄星刚等，2022；顾海峰和张晶，2023）。

在影响因素方面，由于实际研究中公司股东方面的数据难以获得，现有文献更多是基于被质押上市公司的数据，研究了公司股价、股权性质、控股比例等因素对大股东股权质押的影响。就公司股价而言，大股东会进行市值管理、盈余管理等活动以稳定公司股价、获得更多融资（李旻和郑国坚，2015；谢德仁和廖珂，2018）；大股东还会进行择时，在股票市场估值偏高、信贷市场宽松（徐寿福等，2016）或投资者情绪高涨（黄宏斌等，2018）时更可能进行股权质押。就股权性质而言，非国有大股东因面临着更为严苛的贷款条件，更愿意使用股权质押进行融资（王斌等，2013；林艳等，2018；周珏廷和李善民，2023）。就持股比例而言，股权质押比例与大股东的控制权转移风险正相关，持股比例高的股东会削弱谋取私利的动机（Claessens et al., 2002），且通常具有更强的融资能力和更广的融资方式。同时，管理层的自信程度会显著增加股权质押规模和风险（祁怀锦等，2022）。

已有研究大多使用线性回归模型研究



单一变量与股权质押行为的因果关系，但大股东股权质押行为通常受到多种因素的影响，仅使用单一变量难以达到良好的预测效果。此外，这些研究在实证中多使用质押率、质押次数或虚拟变量“是否股权质押”，缺乏对股权质押事件的事前风险评估，以及事后实际违约的研究。本文从大股东质押动机和上市公司信息两个角度出发，选择了质押需求、治理结构、公司财务、二级市场等方面因素作为特征，以“触及平仓线或发生冻结”作为标签，尝试对股权质押风险领域提供以预测风险而非探究因果为目标的研究视角和研究工具。

三、模型构建

1. XGBoost 算法

XGBoost 是目前最受欢迎的大规模并行提升树模型 (Boosting Tree) 的工具之一。提升 (Boosting) 则是集成学习中的一个重要方法，通过将多个弱分类器组装成一个强分类器提高模型的性能。CART (Classification and Regression Tree) 是其常用的弱分类器，这是一种典型的二叉决策树，其中的回归树使用平方误差最小化准则，递归构建二叉决策树。通过求解最优优化目标函数，获得最优切分变量 j 与切分点 s ；用选定的对 (j, s) 划分区域，并继续对两个子区域求解上述目标函数，直至满足停止条件。其所求解最优化问题的目标函数可表示为：

$$\min_{j, s} \left[\min_{j, s} \sum_{x_i \in R_1(j, s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j, s)} (y_i - c_2)^2 \right] \quad (1)$$

式中： R_1 、 R_2 是被划分后子区域； c_1 、 c_2 是

R_1 、 R_2 内部使得平方损失误差达到最小值的输出值。

Boosting Tree 是采用加法模型与前向分步算法，以 CART 回归树为基分类器的提升方法：

$$f_M(x) = \sum_{m=1}^M h_m(x) \quad (2)$$

式中： $f_M(x)$ 为 M 个弱分类器相加得到的强分类器； $h_m(x)$ 为第 m 个弱分类器。

具体而言，Boosting Tree 将所有基分类器的结果相加得到预测值，然后使用下一个基分类器去拟合预测值与真实值之间的误差。当使用平方损失函数时，可以将函数改写为残差形式：

$$L(y, f_m(x)) = L(y, f_{m-1}(x) + h_m(x)) = (y - f_{m-1}(x) - h_m(x))^2 = (r - h_m(x))^2 \quad (3)$$

式中： r 为当前模型拟合数据的残差。

XGBoost 的损失函数展开后如下所示：

$$\overline{L}^{(t)} = \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T \quad (4)$$

式中： λ 为所有叶子节点的权值的 L2 正则之和的系数； T 是叶子节点的个数； γ 是该项的系数。

XGBoost 算法在进行分类或预测的同时，能够计算出每个特征的重要性，本文选用增益 (Gain) 作为重要性的度量，特征的增益表示了该特征被用于划分时损失函数的平均减少值。特征 X_ℓ 在单棵树 T 中的重要性通过下式计算：

$$w_\ell^2(T) = \sum_{t=1}^{J-1} \hat{\tau}_t^2 \quad (5)$$

式中： $J-1$ 为树的非叶子节点数量； X_ℓ 是和每个节点 t 相关的特征； $\hat{\tau}_t^2$ 是节点 t 分裂



后平方损失的减少数值。

则特征 X_i 的全局重要性通过其在单棵树中的重要性的平均值来衡量：

$$w_{\ell}^2 = \frac{1}{M} \sum_{t=1}^M w_{\ell}^2(T_m) \quad (6)$$

式中： M 为树的数量。

2. 模型评价准则

在机器学习中，使用最普遍的评价指标是分类准确率 (Accuracy)，该指标的含义直观，且在通常情况下有效。然而对于类别不平衡样本，它并不能很好地反映模型性能。尽管高风险类样本数量少，但正是本文需要识别和研究的对象，因此在评价模型的预测能力时，本文更关注模型正确预测高风险类样本的能力。在类别不平衡问题中，通常使用真正例 TP、真负例 TN、假正例 FP、假负例 FN 的样本数量计算得到的精确率 (Precision)、召回率 (Recall) 等评价指标。二分类下的混淆矩阵如表 1 所示。

表 1 混淆矩阵

正样本		预测值	
		负样本	
实际值	正样本	TP	FN
	负样本	FP	TN

基于混淆矩阵，可以计算精确率、召回率、F1-score 和 AUC-PRC (Area Under Curve of Precision-Recall Curve) 评价模型。其中，F1-score 是精确率和召回率的调和平均值；AUC-PRC 是精确率-召回率曲线 (PR curve) 下的面积。上述评价准则不会被样本量的分布所影响，因此被认为是无

偏的，可在类别不平衡样本中使用，其计算方式如下所示：

$$\begin{aligned} \text{Recall} &= \frac{TP}{TP+FN} \\ \text{Precision} &= \frac{TP}{TP+FP} \\ \text{F1-score} &= 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (7) \end{aligned}$$

四、实证研究与分析

1. 数据来源与样本选取

本文以 2013—2020 年沪深两市上市公司的股权质押事件作为研究对象，检验股权质押前的合同特征、治理结构、公司财务及二级市场是否影响股权质押期间风险爆发的可能性。模型选取的特征均为质押事件发生前的变量，所使用数据来自 CSMAR 数据库。本文使用“是否触及平仓线或发生冻结”作为股权质押事件样本的风险度量，即模型中的标签，如果该股权质押触及平仓线或发生冻结，则将其识别为高风险。为收集每个大股东股权质押事件的标签，获取上交所和深交所披露的带有“平仓线”“冻结”等字样的上市公司信息，进行交叉验证和数据补充。

本文的数据频率统一为“事件发生”，以某上市公司某股东或者某控制人为事件主体，记录其某次股权质押事件。例如 A 公司的股东甲于 2019 年 5 月 30 日新增质押，2019 年 10 月 12 日追加质押物，2020 年 4 月 2 日全部解押，则该样本为有效样本，初始日期为 2019 年 5 月 30 日，完结日期为 2020 年 4 月 2 日，若在此期间发生触及平仓线或发生冻结，则将标签设置为



1, 代表高风险, 否则为 0。

2. 特征工程

特征工程是指使用领域知识, 利用数据挖掘技术, 从未经处理的原始数据中提取出更有效特征以提升模型效果的过程, 通常包括特征提取、特征衍生、特征编码、特征筛选四个部分。

特征衍生指在原始特征的基础上进行变换以获取新的特征。使用衍生特征的主要原因包括: 数据具有时间属性; 数据具有无序性和突发性等。在二级市场相关特征中, 考虑到日度个股数据的时间属性和不稳定性, 运用特征衍生构造出具有解释意义且更加稳定的特征。本文参考了 23 种业界常用的基于时间序列的特征聚合方法, 对二级市场日度数据进行特征聚合, 以反映这些数据的趋势、波动等。本文共获得特征 1 880 个, 其中质押特征 9 个, 治理结构 6 个, 财务指标 125 个, 二级市场 15 个, 二级市场聚合特征 1 725 个。

本文对分类特征进行了特征编码。分类特征用于表示类别或标记, 取值有限且无序。本文使用 One-hot 编码对预处理后的分类特征进行处理, 将其表示为二进制变量。

在特征筛选过程中, 本文依次使用 XGBoost 重要性、缺失率、IV 值和相关性对特征进行筛选。首先, 使用 XGBoost 重要性筛选出前 300 个重要变量。其次, 对于缺失率小于 10% 的特征进行随机森林

插补法; 对于缺失率在 10% 到 70% 的特征使用 0 值填补或将缺失单独作为一类处理; 对于缺失率高达 70% 的特征做删除处理, 共删除特征 4 个。再次, 剔除了 IV 值 (Information Value) 小于 0.1, 即信息预测能力较弱的特征 67 个。最后, 删除了相关系数大于 0.7 的特征对中 IV 值较小的特征共 65 个。经过上述四轮筛选, 最后剩余特征数量为 164 个。

在剔除出质方不为大股东、质押前一日工作日公司状态不为正常交易、特征值缺失 50 个以上的样本后, 本文共获得 33 161 条样本数据, 其中标记为 1 的高风险股权质押事件 584 条, 占全部样本的 1.76%。本文的样本数据存在明显的“类别不平衡”现象, 即标记为 1 的样本数量远小于标记为 0 的样本数量。该现象虽然与股权质押的实际场景产生较好印证, 但给模型选择和训练带来了较大困难。本文所采用的 XGBoost 算法作为 GBDT 的改进, 在处理不平衡数据上具有天然优势。在此基础上, 本文引入过采样思想, 通过赋予少数类更高权重, 改变了损失函数, 进而改变损失函数的一阶和二阶导数。

附表 1^① 报告了最终进入模型的 20 个特征的描述性统计, 包括 1 个质押特征、2 个治理结构特征、13 个财务指标特征、4 个二级市场聚合特征。

3. 模型构建

机器学习的最终目的是降低泛化误

① 附表 1~3、附图 1~4 见增强出版, 中国知网—《金融市场研究》。



差，即希望训练后的模型部署到其他未经训练的数据集时依然能取得好的预测结果。为此，机器学习通常将数据分割成训练集和测试集，本文随机抽取数据集的 80% 作为训练集，余下 20% 作为测试集。同样的模型在不同的超参数设置下，性能可能会有很大不同。如果模型过于复杂，可能导致过拟合，即模型在训练集表现优异但泛化能力很差；如果模型过于简单，则可能导致欠拟合，即模型无法有效利用特征所蕴含的信息，在训练集和测试集中均表现一般。本文使用网格搜索 (Grid Search) 进行超参数调试，并在调参时采用交叉验证 (Cross Validation) 思想以防止过拟合。

本文通过特征筛选剔除了超过 90% 的原始特征，但合理的入模变量个数还需要考虑变量关联度、样本量相对大小等因素。当变量间关联度较高，样本量相对较小时，入模变量个数过多可能导致过拟合问题。因此本文在模型训练过程中，将入模特征数量也作为另一个超参数进行了调试。使用 XGBoost 特征重要性进行排序后，分别选取前 10、20、50、100 个特征入模进行训练，获取训练集和测试集上的评分，并与全部特征入模的预测结果进行了比较，结果如附表 2、附图 1 和附图 2 所示。

根据附表 2 可知，在训练集中，所有模型的召回率 (Recall) 均为 100%，即所有实际为高风险股权质押的样本全部被模型正确分类为高风险；精确率 (Precision) 则随着入模特征数量的减少从 85.87% 下降到 54.42%。在测试集中，召回率和精

确率相较于训练集均有所下降，其中入模特征数量为 20 的模型具有最高的召回率 74.44%；而精确率同样随着入模特征数量的减少而下降，精确率在本数据集中的具体意义为所有被模型预测为高风险的样本中实际为高风险的比率。由此可知，随着入模特征数量的减少，XGBoost 算法通过牺牲精确率，即将更多的样本预测为高风险的方式，保证了高召回率，即尽量减少实际为高风险的样本被误判为低风险的情况。

根据附图 1 和附图 2 可知，训练集的 AUC-PRC 均高于 0.9，而测试集的 AUC-PRC 最差也达到 0.53。将上述指标与一个随机分类器的指标进行对比：假设有一个随机分配器，其随机预测所有样本中的 q ($0 < q < 1$) 为高风险样本，其余为低风险样本；已知实际情况下，所有样本中有 p ($p=1.76%$) 为高风险样本。由于随机分类器预测结果与实际结果独立，可计算得到 $TP=pq$, $FN=p(1-q)$, $FP=(1-p)q$, $TN=(1-p)(1-q)$ ，进而计算得到召回率为 q ，精确率为 p ，因此该随机分类器的 PR 曲线为一条平行于 x 轴、与 y 轴相交于 1.76% 的直线，AUC-PRC 为 0.0176。综合分析上述指标，可以认为所有 XGBoost 模型在测试集中仍有较好的预测性能。

4. 进一步分析：Logistic 回归

Logistic 回归是一种广义的线性回归分析模型，也是机器学习中用于分类预测的常用模型之一。本文使用 Logistic 回归进行模型预测性能的比较，并展开进一步分析。



表 2 Logistic 回归评价指标

入模特征数量	模型	训练集			测试集		
		Recall	Precision	F1 score	Recall	Precision	F1 score
20	XGBoost	1.0000	0.6423	0.7822	0.7444	0.5657	0.6429
20	Logistic	0.7207	0.0424	0.0800	0.6364	0.0350	0.0664

本文选择 XGBoost 重要性排名为前 20 的特征，首先进行最大最小值归一化处理，以加快算法的收敛速度和运算精度，然后将这些特征作为自变量进行 Logistic 回归，得到 Logistic 模型评价指标表 2。

由表 2 可知，尽管 Logistic 模型预测性能仍优于随机分类器，但显著差于 XGBoost 模型。为了保证较高的召回率，模型精确率在训练和测试集上均小于 5%。精确率较小的模型在实际应用中会将较多的低风险股权质押样本识别为高风险，如果金融机构采用该预测模型，将会导致大部分“好客户”被拒之门外。

附表 3 展示了 Logistic 回归的参数检验结果，可以发现在 XGBoost 中较为重要的特征，并不一定在 Logistic 回归模型中具有显著性，这是因为 XGBoost 能够通过特征之间的串行组合，挖掘出特征与标签之间更加复杂的非线性关系。XGBoost 所属的树模型将特征空间按照最优化信息增益的策略对各个特征依次进行划分，进而逐渐解构问题；而 Logistic 回归尝试直接获得分离样本集的边界，对于复杂问题的处理能力相较于树模型稍显薄弱。附图 3 展示了这两种模型常见的决策边界，可以发现树模型能够更精准地刻画不够平滑的决策边界。

需要指出的是，Logistic 回归在解释

变量筛选和模型参数估计时假设各个变量是相互独立的。但在真实世界中，这一假设无法满足，解释变量之间必然存在一定程度的线性相关关系。这种多重共线性通常会导致估计参数的方差增大，进而导致其不稳定和不显著。附图 4 展示了 Logistic 回归中解释变量的相关系数矩阵，发现部分变量之间存在明显的线性相关，说明了多重共线性问题的客观存在。该问题也是导致 Logistic 回归预测能力不如 XGBoost 的重要原因之一。

在附表 3 中，仍有部分 XGBoost 重要性较高的特征在 Logistic 回归中是显著的，如出质方性质编码、成本费用利润率 TTM、第一大股东持股比例 (%) 等。根据参数检验结果，相较于其他大股东，境内非国有大股东的股权质押风险更高。大股东持股比例越高，股权质押风险越高，而其他大股东（前十大股东）持股比例的提升能抑制这种风险；这一现象支持了大股东股权质押的弱化激励效应和强化侵占效应，并证明其他大股东的监督能够抑制这种侵占。在公司经营财务方面，股权质押风险较高的上市公司具有更高的成本费用率、更高的负债率和更差的现金流情况。在二级市场表现方面，股权质押风险较高的上市公司具有较低的市销率和流动性水平。



五、结论

本文以股权质押事件作为研究样本,以“触及平仓线或发生冻结”作为样本的风险度量标签,关注大股东股权质押风险的事前防控,研究股权质押发生前的质押需求、治理结构、公司财务、二级市场等特征对风险的影响。利用数据处理与特征工程,通过特征提取和特征衍生获得 1 880 个特征,以覆盖更加完整和全面的信息;对行业代码等分类特征进行 One-hot 编码处理;根据特征重要性、缺失率、IV 值和相关性进行特征筛选,最终获得 164 个人模特征。针对高风险样本占比较小、数据集类别不平衡这一特点,选用 XGBoost 算法进行分类预测,同时使用 Logistic 回归进行了模型预测性能比较和参数显著性检验,获得以下结论。

①XGBoost 算法训练得到的股权质押风险识别模型具有良好的预测性能,在测试集中召回率最高可达 74.44%,而精确率随着入模特征数量的减少而下降。该算法通过牺牲精确率,即将更多的样本预测为高风险的方式,来保证高召回率,即尽量减少实际为高风险的样本被误判为低风险

的情况。

②Logistic 模型预测性能虽然优于随机分类器,但显著差于 XGBoost 模型。为了保证较高的召回率,模型精确率在训练和测试集上均小于 5%。精确率较小的模型在实际应用中会将较多的低风险股权质押样本识别为高风险,让质权方错失大量好项目。

③在 XGBoost 中较为重要的特征,并不一定在 Logistic 回归中仍具有显著性,因为 XGBoost 算法能够挖掘出特征与标签之间更加复杂的关系,且不受多重共线性的影响。

④根据 XGBoost 重要性和 Logistic 回归参数检验结果,相较于其他大股东,境内非国有大股东的股权质押风险更高。在治理结构方面,大股东持股比率越高,股权质押风险越高,而其他大股东持股比率的提升能抑制这种风险。在公司经营财务方面,股权质押风险较高的上市公司具有更高的成本费用率、更高的负债率和更差的现金流情况。在二级市场表现方面,股权质押风险较高的上市公司具有较低的市销率和流动性水平。^[N]

学术编辑: 卢超群

参考文献

- [1] 艾大力,王斌.论大股东股权质押与上市公司财务:影响机理与市场反应[J].北京工商大学学报(社会科学版),2012,27(04):72-76.
- [2] 顾海峰,张晶.控股股东股权质押是否会影响企业投资效率?——基于中国A股上市公司的证据[J].湖南大学学报(社会科学版),2023,37(03):41-53.
- [3] 郝项超,梁琪.最终控制人股权质押损害公司价值么?[J].会计研究,2009,7:57-63+96.
- [4] 黄宏斌,肖志超,刘晓丽.股权质押的时机选择及市场反应——基于投资者情绪视角的研究[J].金融论坛,2018,23(02):65-80.
- [5] 黄星刚,叶似剑,侯宝升.股权质押融资、非效率投资与股价波动性——基于“中介效应”模型的实证检验[J].金融监管研究,2022,08:78-93.



- [6] 李旒,郑国坚.市值管理动机下的控股股东股权质押融资与利益侵占[J].会计研究,2015,5:42-49+94.
- [7] 林艳,魏连宾,李炜.控股股东股权质押、股权性质与公司绩效研究[J].商业研究,2018,2:50-56.
- [8] 祁怀锦,刘艳霞,曹修琴.自信程度与股权质押风险——基于控股股东兼任公司高管视角的分析[J].商业研究,2022,05:23-34.
- [9] 王斌,蔡安辉,冯洋.大股东股权质押、控制权转移风险与公司业绩[J].系统工程理论与实践,2013,33(07):1762-1773.
- [10] 吴静.控股股东股权质押等于“掏空”吗?——基于中国上市公司股权质押公告的实证分析[J].经济论坛,2016,8:65-70.
- [11] 谢德仁,廖珂.控股股东股权质押与上市公司真实盈余管理[J].会计研究,2018,8:21-27.
- [12] 徐寿福,贺学会,陈晶萍.股权质押与大股东双重择时动机[J].财经研究,2016,42(06):74-86.
- [13] 张陶勇,陈焰华.股权质押、资金投向与公司绩效——基于我国上市公司控股股东股权质押的经验数据[J].南京审计学院学报,2014,11(06):63-70.
- [14] 赵玉芳,余志勇,夏新平,等.定向增发、现金分红与利益输送——来自我国上市公司的经验证据[J].金融研究,2011,11:153-166.
- [15] 周珏廷,李善民.国有股东如何化解民营企业股权质押危机?——对国有股东“用脚投票”治理机制的实证检验[J].证券市场导报,2023,10:25-38.
- [16] Claessens S, Djankov S, Fan J P H, et al. Disentangling the incentive and entrenchment effects of large shareholdings[J]. The Journal of Finance, 2002, 57(6): 2741-2771.
- [17] Chen T, Guestrin C. Xgboost: A scalable tree boosting system[C]. In Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 2016: 785-794.
- [18] Friedman J H. Greedy function approximation: a gradient boosting machine[J]. Annals of Statistics, 2001: 1189-1232.
- [19] Huang Z, Xue Q. Re-examination of the effect of ownership structure on financial reporting: Evidence from share pledges in China[J]. China Journal of Accounting Research, 2016, 9(2): 137-152.
- [20] Lins K V. Equity ownership and firm value in emerging markets[J]. Journal of Financial and Quantitative Analysis, 2003, 38(1): 159-184.

Identifying and Quantifying Share Pledge Risks of Listed Companies

YANG Liuyong¹ CEN Shengnan² JIA Shaoqing¹

(1. School of Economics, Zhejiang University; 2. Ningyin Wealth Management LLC)

Abstract China has undergone two rounds of large-scale share pledging crises, and the absence of prior risk controls were important factors in both of these instances. This paper examines these events and looks at the demand for share pledge financing, as well as corporate governance, corporate finance practices and secondary market characteristics. It uses the XGBoost algorithm in machine learning to predict and analyze pledge risks, focusing on prior prevention and control of the share pledge risk of majority shareholders. Empirical results show that the share pledge risk classification model trained by XGBoost has a very good predictive performance, and the highest recall in the test set reaches 74.44%. An analysis of feature importance shows that, compared with other majority shareholders, share pledge risk is highest among domestic non-state shareholders. In terms of governance structure, the higher the shareholding ratio of the majority shareholders, the higher the risk of share pledges. Conversely, an increase in the shareholding ratio of other shareholders inhibits this risk. Companies with a higher share pledge risk have lower cost expense ratios, higher debt ratios, and worse cash flow. Companies with higher share pledge risk have lower price-to-sales ratios and liquidity levels.

Keywords Share Pledge, Majority Shareholder, Feature Engineering, Machine Learning

JEL Classification C45 G14 G32