

外审评议与作者回复

第一份意见

审稿意见 (1.1):

论文既然是识别股权质押高风险的事前影响因素,那么在选取和构建特征变量的过程中是否考虑到了“事前”这一重要特征?文章并未对此进行明确说明。如此,一些因素可能并不是股权质押高风险的影响因素,而是高风险的表现。比如,较低的市销率一般可能是由于较低的股价所致,较低的股价也是股权质押触及平仓线的条件,那么如何解释市销率对股权质押高风险的影响呢?因此,在本文的研究问题中,明确区分特征因素的时间界限尤为重要。

修改说明:非常感谢您的宝贵意见,我们在“数据来源与样本选取”该章节中介绍了“检验股权质押前的合同特征、治理结构、公司财务以及二级市场是否影响股权质押期间风险爆发的可能性”。因此我们选取和构建特征变量的过程中已考虑了“事前”这一重要特征。基于您的建议,我们对“事前”这个特征进行了更加详细的说明,明确区分特征因素的时间界限。

审稿意见 (1.2):

文章选取的特征因素极多,但并未很好地考虑特征之间的共性以及相互解释程度。这就会导致部分类别的因素进入模型太多,影响了最终结果,甚至一些因素在特征度量并无本质区别,那么这些因素之间的重叠是否合理?比如,表4中关于流动性、盈利能力、费用率等方面的度量均有多个指标,这些指标为什么可以同时放入?为什么度量同一特征维度的指标有的解释力强,有的解释力弱?文章需要对这些处理进行详细解释,才能有很好的说服力。

修改说明:已有开源代码测试表明,包含相关变量会影响 XGBoost 模型的特征重要性,但是模型的性能在存在多重共线性的情况下仍然相当稳定和稳健,因此含有多个相关变量对模型本身判定高/低风险没有负面影响。基于此,我们选取了较多的指标,利用 XGBoost 模型识别并利用尽可能多的信息,提高模型的预测精度。

审稿意见 (1.3):

仅仅采用“触及平仓线或发生冻结”作为标签来划分股权质押高风险可能不够科学,至少仅采用这一个标准尚难以提供足够的说服力。比如,某一公司的股权质押接近平仓线是否可以被认定是低风险呢?对此,建议作者采用质押期间最低股价构造多个不同的判定标准进行更多的稳健性检验,以增强文章研究结论的可信度。

修改说明:十分感谢您的意见和建议。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,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。这种高风险的划分方式,剔除了因股价波动导致的**被动风险事件**,只保留了因大股东自身资金实力不足导致的**主动风险事件**,具有特殊性。

股权质押的警戒线和平仓线,通常由金融机构与股东协商决定,市场上通用的警戒/平仓线标准分别是 160/140% 和 150/130%。当质押股票的市场价值与质押融资金额的比值低于警戒线时,金融机构通知股东增加质押股票或补充保证金。当质押股票的市场价值与质押融资金额的比值低于平仓线时,金融机构要求股东在第二天收盘前增加质押股票或补充保证金,直到履约保障水平高于警戒线,若股东未履约,金融机构可以进行平仓处理。股票价格波动较大时,容易发生被动触及警戒线事件,若股东具有足够的股票或资金储备,则可及时恢复履约保障水平回到警戒线以上;只有当股东自身资金实力不足,无法及时补充质物,才会进一步触及平仓线甚至发生冻结。本文所用标签“触及平仓线或发生冻结”较好区分了上述被动风险事件和主动风险事件。

审稿意见 (1.4):

文章布局上,不用花费太多篇幅介绍 XGBoost 的方法,相反需要对特征工程筛选出的特征变量进行一些描述,至少应该有一些描述性统计,让读者知晓进入机器学习过程的特征变量大致情况。

修改说明:非常感谢您的意见,我们补充了入模特征变量的描述性统计,并适当修改了文章布局。

审稿意见 (1.5):

对已有文献的回顾不到位。文献综述过于蜻蜓点水，且对近年来的相关研究并未涉及，比如无一篇近五年股权质押的相关研究文献。

修改说明: 非常感谢您的意见，我们对文献综述做了进一步的完善，并补充了近五年股权质押的相关研究文献。

审稿意见 (1.6):

一些字、词、句的使用规范和表述问题需要仔细斟酌。

修改说明: 基于您的建议，我们对全文进行了文字上的审查和修改，以求字词句的使用更加规范，表述更加严谨。

第二份意见

审稿意见 (2.1):

建议题目改为“上市公司股权质押风险特征识别与测度——基于 XGBoost 方法”。现实中股权质押不仅有上市公司，还有大量非上市企业，两者对应的逻辑完全不同，不能混淆。

修改说明: 非常感谢您的建议，我们已将题目改为“上市公司股权质押风险特征识别与测度——基于 XGBoost 方法”。

审稿意见 (2.2):

写作太过于技术化，对于《金融市场研究》的读者而言很不友好。这个论文主要在讲股权质押风险的特别识别逻辑与测度，因此要多讲经济方面的逻辑，结果的经济含义，而不是大量描述统计的术语，让读者能够更好地理解论文的核心内容与贡献。

修改说明: 十分感谢您的意见和建议。本文主要是利用一种新的方法（机器学习中的 XGBoost 算法）来预测股权质押风险，主要关注点在于预测准确度是否有效提升。因此，本文主要注重算法的介绍和讲解、模型的构建、指标的选取，以及预测结果相较于传统方法的提升率。

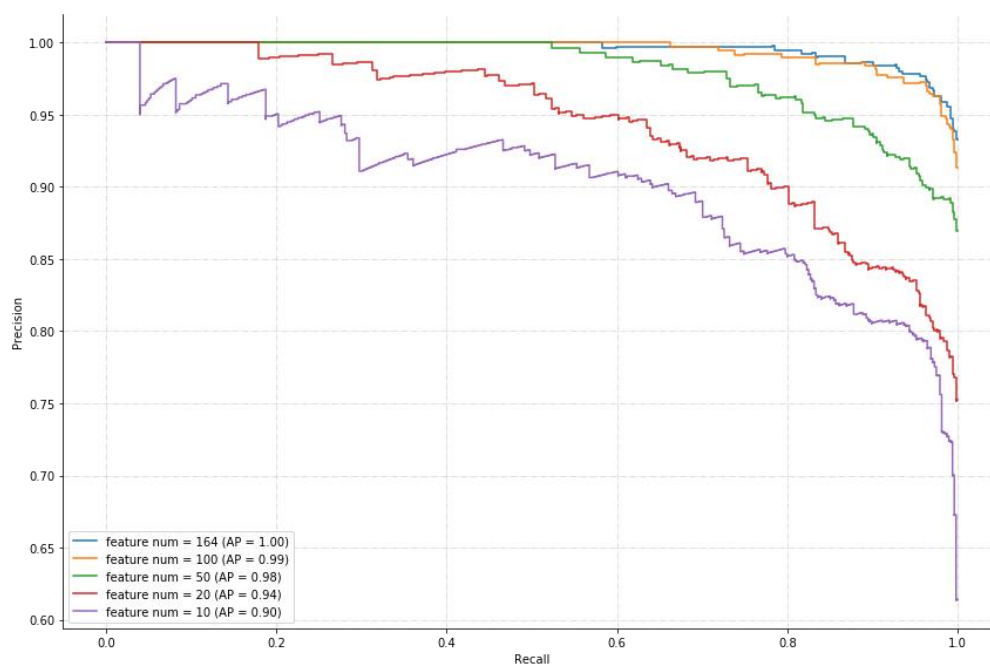
上市公司股权质押风险特征识别与测度（附录）

附表 1 特征定义和描述性统计

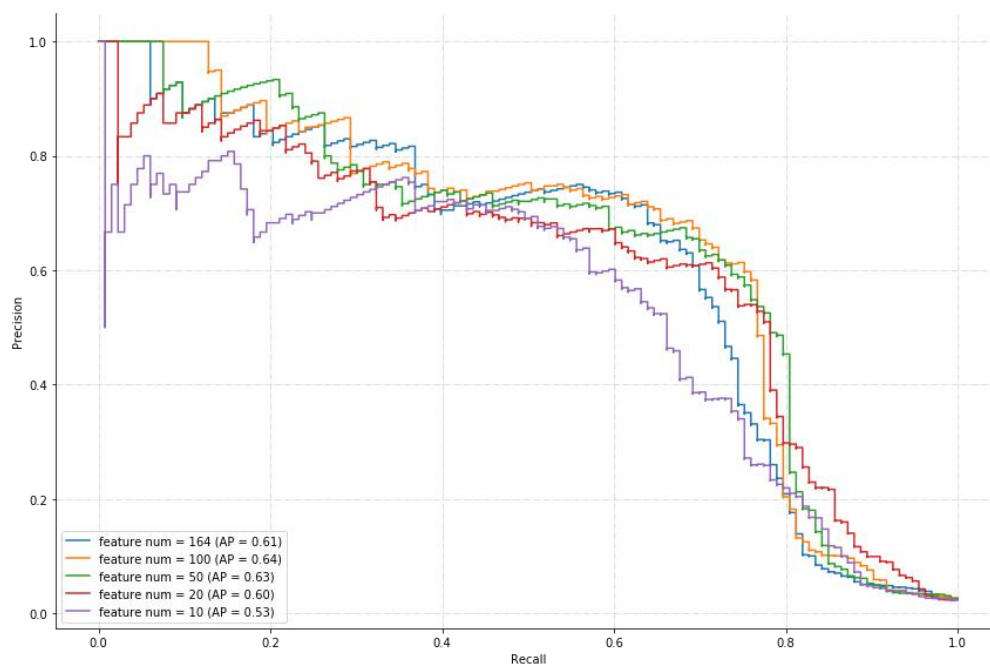
XGBoost 重要性	特征名称	特征意义	均值	标准差	25%分 位	中位数	75%分 位
1	Ret_21_diffcnt	股息率月度趋势	7.871	4.840	5	9	11
2	PledgorTypeCode_3	出质方性质编码：境内非国有	0.468	0.499	0	0	1
3	F052101C	成本费用利润率 TTM	0.159	0.246	0.047	0.111	0.209
4	LargestHolderRate	第一大股东持股比例（%）	32.315	13.796	21.430	30.320	41.590
5	PS_126_indcnt	市销率半年度行业相对水平	37.106	52.506	0	0	91
6	TopTenHoldersRate	前十大股东持股比例（%）	60.464	12.974	51.930	61.590	70.090
7	F031201A	留存收益资产比	0.169	0.114	0.098	0.159	0.232
8	F031501A	流动负债比率	0.822	0.165	0.734	0.867	0.955
9	F061901B	资本支出与折旧摊销比	2.864	6.762	0.852	1.694	3.363
10	F031601A	经营负债比率	0.537	0.235	0.352	0.510	0.706
11	F051701C	销售费用率 TTM	0.083	0.096	0.026	0.050	0.101
12	F052001C	销售期间费用率 TTM	0.210	0.195	0.114	0.169	0.260
13	F061301C	公司现金流 TTM1	-136.469	320.699	-107.503	-46.906	-22.074
14	F053301C	营业毛利率 TTM	0.322	0.175	0.197	0.295	0.410
15	Liquidity_126_indcnt	流动性指标半年度行业相对水平	20.806	29.118	0	6	32
16	F050601C	息税前利润 TTM	5.466	20.633	0.739	1.926	4.981
17	Ret_21_maxfrac	股息率月度趋势	1.051	0.474	0.936	1.007	1.105
18	F030901A	无形资产比率	0.044	0.047	0.018	0.033	0.056
19	F041705C	总资产周转率 TTM	0.615	0.495	0.344	0.510	0.734
20	F060201C	营业收入现金含量 TTM	0.992	0.281	0.867	0.995	1.098

附表 2 基于不同入模特征数量的模型评分

入模特征 数量	训练集			测试集		
	Recall	Precision	F1 score	Recall	Precision	F1 score
164	1.0000	0.8587	0.9240	0.6165	0.7130	0.6613
100	1.0000	0.7980	0.8876	0.6917	0.6715	0.6815
50	1.0000	0.7488	0.8564	0.7293	0.6178	0.6690
20	1.0000	0.6423	0.7822	0.7444	0.5657	0.6429
10	1.0000	0.5442	0.7048	0.6767	0.4265	0.5233



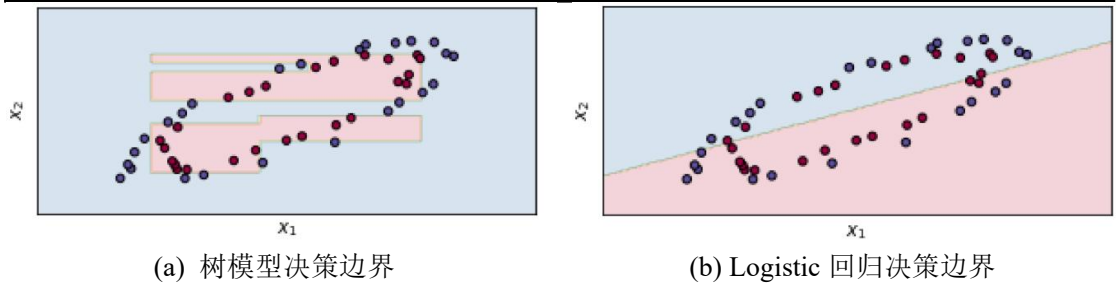
附图 1 训练集 PR 曲线



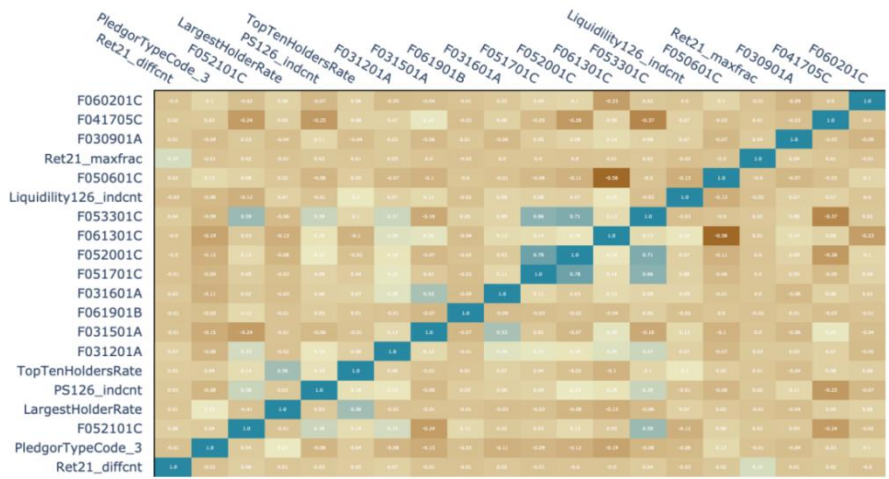
附图 2 测试集 PR 曲线

附表3 Logistic 回归结果

XGBoost 重要性	特征名称	特征意义	coef	std err	z	P> z
1	Ret_21_diffcnt	股息率月度趋势	-0.1438	0.327	-0.439	0.660
2	PledgorTypeCode_3	出质方性质编码：境内非国有	0.6864***	0.104	6.582	0.000
3	F052101C	成本费用利润率 TTM	-13.4233***	4.984	-2.693	0.007
4	LargestHolderRate	第一大股东持股比率（%）	2.8535***	0.357	7.990	0.000
5	PS_126_indcnt	市销率半年度行业相对水平	-0.4016***	0.140	-2.873	0.004
6	TopTenHoldersRate	前十大股东持股比例（%）	-1.8305***	0.332	-5.508	0.000
7	F031201A	留存收益资产比	-0.8222***	0.328	-2.507	0.012
8	F031501A	流动负债比率	1.1214***	0.293	3.829	0.000
9	F061901B	资本支出与折旧摊销比	-2.3899	4.711	-0.507	0.612
10	F031601A	经营负债比率	0.5641***	0.224	2.524	0.012
11	F051701C	销售费用率 TTM	1.589	1.380	1.152	0.249
12	F052001C	销售期间费用率 TTM	1.1041	4.809	0.230	0.818
13	F061301C	公司现金流 TTM1	-0.5488	0.447	-1.227	0.220
14	F053301C	营业毛利率 TTM	-0.1396	0.894	-0.156	0.876
15	Liquidity_126_indcnt	流动性指标半年度行业相对水平	-1.4023***	0.268	-5.237	0.000
16	F050601C	息税前利润 TTM	-2.4607	4.549	-0.541	0.589
17	Ret_21_maxfrac	股息率月度趋势	-8.2133***	3.455	-2.377	0.017
18	F030901A	无形资产比率	0.8929***	0.306	2.915	0.004
19	F041705C	总资产周转率 TTM	-0.3043	1.183	-0.257	0.797
20	F060201C	营业收入现金含量 TTM	-8.6994***	2.219	-3.921	0.000
Pseudo R ²						0.1118



附图3 树模型和 Logistic 回归的决策边界



附图4 Logistic 回归解释变量相关系数矩阵