

有效预警上市公司违规的递延所得税 异动指标和人工智能模型

郦金梁，吴谣，雷曜，黄燕婷

金融研究，2020.

叶鑫 2021/07/01

研究背景

- 我国 A 股上市公司违规情况较严重。在2000-2017年间，1641家 A 股公司有违规记录，占样本内上市公司总数的47.79%，平均年度违规率为17%，相较而言，美国市场违规率较低（<0.5%）；
- 在违规现象普遍存在的同时，监管处罚相对滞后，惩罚力度低。68.4%的 A 股公司违规事件未在当年被判定，平均滞后时间为2.7年；
- 中国上市制度、ST制度等对盈利指标的偏重可能加强了上市公司滥用盈余管理的动因，使得 A 股公司存在普遍的违规现象，其中**与公司利润相关且存在管理弹性的递延所得税信息**成为当前监管双方博弈的焦点。

研究动机

- 递延所得税数据作为违规预警指标具有多方面优势，税务数据相对真实、可查，递延所得税指标因而有效拓展了预警违规行为的信息集：
 - 税务数据是反映公司财务运行情况的硬信号，而披露数据是经公司财务粉饰后的软信号；
 - 在长期过度盈余管理之后，当常规财务粉饰方法无法延续时，将已经缴纳的所得税认定为资产成为管理层最后的手段。
- 综上所述，本文创新性地提出递延所得税异动指标在预警上市公司违规方面的重要作用，并从理论和实证两方面揭示了违规公司滥用递延所得税数据的动机。

相关文献

- 首先，如Baucus (1994) 和Povel (2007) 所述，一个公司迫于财务困境的压力可能操纵财务数据，递延所得税的异动可以反映公司面临的财务困境程度；
- 其次，Mills (2001), Phillips (2003)、Dhaliwal (2004) 所述，递延所得税的异常变动度量了单个公司的财务操纵程度，进而反映了公司财务数据的可靠程度；
- 从计量方面考虑，Guay (1996)、Bernard和Skinner (1996) 指出之前学者研究盈余管理及财务操纵时所用指标存在计量误差及系统偏差，而Phillips (2003) 的研究则表明了递延所得税指标的有效性；
- 最后，递延所得税与盈利直接挂钩，影响资本市场对公司的估值、金融市场的信息有效性和资源配置效率。研究递延所得税对警惕公司财务操纵相关的违规及促进金融市场的有效性意义重大。

研究设计

- 首先对上市公司违规机制进行分析，从文献角度分析了递延所得税信息对于上市公司违规行为的判别和预警作用；
- 其次，基于Logistic模型初步分析递延所得税指标对上市公司违规行为的样本内预测能力，并进行了一系列稳健性检验；
- 然后，通过构建二元Probit模型控制内生性和部分可观察性，进一步分析递延所得税指标与公司违规倾向和稽查率的关系；
- 最后构建决策树模型对违规事件实现样本外预警，对违规公司进行精准判别。

研究结论

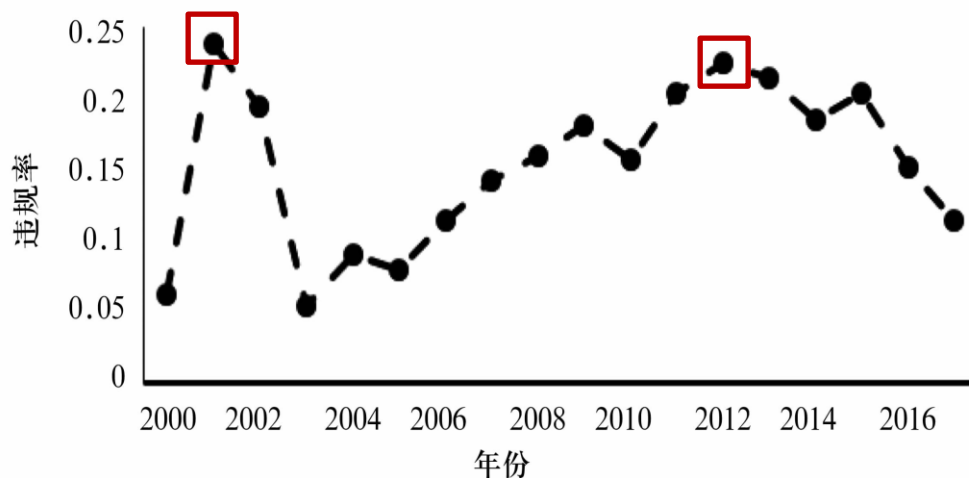
- 研究发现递延所得税异动与上市公司的违规倾向显著正相关，与监管机构的违规稽查概率显著负相关，在控制内生性之后所得研究结论依然稳健；
- 研究结论还表明本文构建的决策树模型可用于实现精准预警，实现风险规避，帮助投资者筛选出高信度未违规公司，其中识别为未违规的公司在下一年度的资产回报率和股价回报率都显著高于被判定为违规的公司。

研究贡献

- 第一，本文揭示了 A 股公司大面积违规而稽查过程冗长迟缓这一重要现象，在梳理违规机制的基础上提出了有效的解决方案；
- 第二，本文指出了所得税数据在稽查违规中可以发挥的预警作用，揭示了税务数据在反映公司真实财务情况中的硬信号作用；
- 第三，本文展示了机器学习在资本市场治理中可发挥的重大作用。

研究数据

- 违规事件数据来源为CSMAR数据库，包含2000-2017年间上市公司被证监会、财政部、证交所等相关单位公开处罚的信息，A股上市公司财务数据来源于Wind数据库；
- 我们对数据做了1%极值处理，并删除了递延所得税资产指标缺失的样本；
- 经过数据处理，研究样本包含3434个公司，共24530条记录，其中有违规记录的公司达1641个，占公司总数的47.79%。



实证分析——基于Logistic模型的违规预测

- 首先设计预警违规指标——递延所得税资产异常变动指标abDTA:

$$abDTA_{i,t} = |DTA_{i,t} - DTAm_{i,t}| / |DTAm_{i,t}|$$

- $DTAm_{i,t}$ 为公司 i 在 t 年对应的前三年递延所得税资产的均值, $abDTA_{i,t}$ 用标准化比值度量了递延所得税资产异动程度。

- Logistic回归模型设计如下:

$$odss_{i,t} = \alpha_0 + \alpha_1 abDTA_{i,t} + \alpha_x X_{i,t} + \epsilon_{i,t}$$

$$p(il_{i,t} = 1 | abDTA_{i,t}, X_{i,t}) = \frac{1}{1 + \exp(-odss_{i,t})}$$

- 其中, $\epsilon_{i,t}$ 服从Logistic分布, $X_{i,t}$ 指代控制变量, $il_{i,t}$ 为发生违规事件的二元变量。
- 公司 i 于 t 年度在给定信息 ($abDTA_{i,t}, X_{i,t}$) 下发生违规的概率 $p(il_{i,t} = 1 | abDTA_{i,t}, X_{i,t})$ 和不发生违规的概率 $p(il_{i,t} = 0 | abDTA_{i,t}, X_{i,t})$ 的比值为优势比 (OR值), $OR = \exp(-odss_{i,t})$ 。

实证分析——基于Logistic模型的违规预测

控制变量 $X_{i,t}$ 说明：

- 第一类为财务变量。如应计利润率（Acta）、毛利率指数（Gmi）、资产质量指数（Aqi）、财务杠杆指数（Levi）等；
- 第二类为公司治理变量。如两权分离度、股权集中度等变量；
- 第三类为市场变量。包含股票波动率、股票换手率、市场平均违规率、行业平均违规率等变量；
- 第四类为公司特征及公司性质等变量。如高管薪酬、公司性质、Z-score(代理公司财务困境)、HHI(测量行业竞争程度)，HHI值越小，竞争越激烈。
- 此外，加入年份虚拟变量（i.year）和行业虚拟变量（i.industry）作为控制变量。

表2 变量的描述性统计

变量	样本数	均值	最小值	p25	中值	p75	最大值	标准差
$il_{i,t}$	24530	0.174	0	0	0	0	1	0.379
$abDTA_{i,t}$	22927	0.305	0	0.075	0.215	0.427	1.994	0.316
$Dist_{i,t}$	24817	0.218	0	0	0	0	1	0.413
$HHI_{i,t}$	24817	0.055	0.008	0.009	0.011	0.071	0.459	0.084
$ROA_{i,t}$	24817	0.040	-0.278	0.014	0.038	0.067	0.195	0.056

实证分析—基于Logistic模型的违规预测

- 在回归m1-m4中，abDTA的系数均在1%水平显著，这表明，递延所得税资产的异常变动值越大，相应公司当期违规概率越高；
- 在回归m1中，回归结果中abDTA对应的优势比为1.369，表明当abDTA增加1时，当期存在违规事件的优势比提升36.9%；
- 在完全回归模型m4中，加入了财务、公司治理和市场变量等控制变量，abDTA对应的优势比系数为1.25，且在1%水平下显著。

表3 递延所得税异动指标对违规事件的影响分析

解释变量	m1		m2		m3		m4	
	il	优势比	il	优势比	il	优势比	il	优势比
递延所得税异动指标	0.314 **	1.369 **	0.218 **	1.243 **	0.368 **	1.445 **	0.225 **	1.253 **
(<i>abDTA</i>)	(0.068)	(0.093)	(0.079)	(0.098)	(0.078)	(0.113)	(0.091)	(0.114)
财务变量	—		控制		—		控制	
市场变量	—		—		—		控制	
行业和年份	控制		控制		控制		控制	
观察数	13,158		11,063		10,342		8,651	

实证分析—稳健性检验

- 我国在2007年前后变更了会计准则，递延所得税的计算方式有所改变，单独分析2007年后的数据(m5)；
- 为解决样本选择性偏差问题，我们采用倾向性评分匹配(PSM)方法，为违规公司组构造出一组最优匹配的未违规组。并按照表3中m4列的变量设置，对PSM样本进行回归(m6)；
- 上述Logistic回归分析控制了固定效应，公司性质变量是否加入应不影响结果。但出于稳健考虑，我们加入高管薪酬、行业违规等公司性质控制变量进行回归(m7)。

表 4 样本选择和额外控制变量

解释变量	m5		m6		m7	
	il	优势比	il	优势比	il	优势比
递延所得税异动指标	0.232 *	1.261 *	0.175 *	1.192 *	0.266 *	1.304 *
(<i>abDTA</i>)	(0.097)	(0.122)	(0.074)	(0.088)	(0.123)	(0.160)
额外控制变量	—		—		控制	
财务变量、公司治理变量、市场变量	控制		控制		控制	
行业和年份	控制		控制		控制	

实证分析—稳健性检验

- 单独考察财务操纵违规事件，标注为财务违规（88%），在全样本中对财务违规虚拟变量进行回归分析（m8）。
- 然后把违规行为分为信息披露违规（73.66%）、经营违规（63.56%）、领导人违规（23.39%）三类，这三种违规分别回归后的结果对应m9-m11列：
 - 一个合理解释是，递延所得税异动可以用于挖掘公司粉饰财务数据的踪迹，从而可以预警财务操纵和信息披露违规。经营违规相关的变量可能已经在模型的控制变量中体现，而领导人违规和递延所得税关联较小。

表 5 违规分类回归

解释变量	m8		m9		m10		m11	
	il	优势比	il	优势比	il	优势比	il	优势比
递延所得税异动指标 (<i>abDTA</i>)	0.248 **	1.282 **	0.270 **	1.310 **	0.102	1.107	0.117	1.124
	(0.085)	(0.109)	(0.084)	(0.110)	(0.087)	(0.096)	(0.125)	(0.141)
财务变量、公司治理变量、市场变量	控制		控制		控制		控制	
行业和年份	控制		控制		控制		控制	

实证分析—基于二元Probit模型的违规稽查分析

- 由于监管机构并不能及时稽查公告全部违规行为，用Logistic模型估计出的参数会存在偏差。所以本文基于二元Probit模型考察公司递延所得税异动指标与违规倾向和稽查概率的动态关系，并分析内生性问题：

$$Detect_{i,t}^* = \beta_D X_{D,i,t} + u_{i,t}, \quad Fraud_{i,t}^* = \beta_F X_{F,i,t} + v_{i,t}$$

- 潜变量 $Detect_{i,t}^*$ 表示公司 i 在年度 t 的违规行为被稽查公告的可能性； $Fraud_{i,t}^*$ 表示违规倾向； $u_{i,t}$ 、 $v_{i,t}$ 为均值为0且服从二元正态分布的残差项，相关系数为 ρ 。
- 定义哑变量 $Fraud_{i,t}$ 与 $Detect_{i,t}$ ： $Fraud_{i,t}^* > 0$ 时， $Fraud_{i,t} = 1$ ，否则 $Fraud_{i,t} = 0$ ， $Detect_{i,t}$ 类似。
- 实际可观测的数据为 $Fraud_{i,t}$ 与 $Detect_{i,t}$ 的乘积 $il_{i,t}$ 。 $il_{i,t} = 1$ 表示公司 i 在 t 年度违规且被稽查； $il_{i,t} = 0$ 表示公司 i 在 t 年度未违规或者违规但未被稽查：

$$P(il_{i,t} = 1) = P(Fraud_{i,t} * Detect_{i,t} = 1) = P(Detect_{i,t} = 1 | Fraud_{i,t} = 1) * P(Fraud_{i,t} = 1) = \Phi(\beta_D X_{D,i,t}, \beta_F X_{F,i,t}, \rho)$$

- 根据上述 $P(il_{i,t} = 1)$ 和 $P(il_{i,t} = 0)$ 的推导式得到对数似然函数 $L(\beta_D, \beta_F, \rho)$ 。

实证分析—基于二元Probit模型的违规稽查分析

- 考虑模型中潜在的内生性，比方说存在被忽视的第三者因素（行业、地区）共同影响递延所得税异动和公司违规：
 - 由于同行业或同地区的经济环境与监管环境相近，单个公司的递延所得税变动与同行业或同地区其它公司的延所得税变动相关，而该公司的违规情况与同行业或同地区其它公司的递延所得税变动没有直接联系。故选取同行业以及同省 / 直辖市 / 自治区所有其它公司的abDTA平均值作为工具变量，分别设为IV_Industry和IV_Region；
 - 从总样本中筛选出同行业或同地区每年不少于50家公司的观测值，然后采用工具变量进行两阶段回归检验：
 - ◆ 在第一阶段，用IV_Industry（IV_Region）和选定的财务控制变量对abDTA做回归，然后求得回归预测值abDTA*；
 - ◆ 在第二阶段，用abDTA*和所选定控制变量，根据 $L(\beta_D, \beta_F, \rho)$ 对 $il_{i,t}$ 做二元Probit回归分析。

实证分析—基于二元Probit模型的违规稽查分析

- $abDTA^*$ 与稽查率（Detect）的相关性显著为负。说明在控制了行业或地区的递延所得税异动之后，公司的递延所得税异动指标越高，该公司违规倾向越高，被稽查通告的概率越低。
- 一个合理的解释是，由于监管机构偏重于对A股公司财务指标的监测，财务状况不良可能是公司成为稽查目标的一个表征诱因，而通过操控递延所得税可以在当期改善公司财务表现，减低被稽查的概率。
- 这个实证结果意味着，监管机构未能识别递延所得税异动导致的业绩粉饰，实际上是激励了上市公司通过操控递延所得税掩盖违规行为。

表 6 二阶段二元 Probit 回归

解释变量	A. 行业工具变量(IV_Industry)			B. 地区工具变量(IV_Region)		
	第一阶段 $abDTA$	第二阶段		第一阶段 $abDTA$	第二阶段	
		Detect	Fraud		Detect	Fraud
$abDTA^*$	—	- 1.961 ** (0.605)	2.402 ** (0.531)	—	- 1.746 ** (0.403)	1.525 ** (0.388)
工具变量	0.833 ** (0.036)	—	—	0.740 ** (0.036)	—	—
财务控制变量	控制			控制		
违规倾向变量	—	—	控制	—	—	控制
违规稽查变量	—	控制	—	—	控制	—
行业和年份	—	—	控制	—	—	控制

实证分析—基于二元Probit模型的违规稽查分析

- 在使用行业和地区工具变量的同时，进一步排除递延所得税异动和公司违规都是公司经营面临困境的表征结果的内生性：每年分别按照Z-score（数值低对应财务困境）和ROA（数值低对应业绩差）两个指标的赋值将上市公司排序，等分成两组。
- 回归结果说明在控制了财务困境状态后，递延所得税异动指标与违规倾向的关系依然存在。

表 7 分组二阶段回归

解释变量	财务状况分组				盈利能力分组			
	高 Z - score		低 Z - score		高 ROA		低 ROA	
	Detect	Fraud	Detect	Fraud	Detect	Fraud	Detect	Fraud
<i>abDTA</i> *	-2.730 **	2.048 **	-1.137	1.983 **	-2.368 **	1.535 **	-1.115	1.783 **
	(1.139)	(0.883)	(0.829)	(0.734)	(0.660)	(0.586)	(0.835)	(0.714)
违规倾向变量	控制	控制	控制	控制	控制	控制	控制	控制
行业和年份	控制	控制	控制	控制	控制	控制	控制	控制
样本数	4030	4030	4232	4232	4153	4153	4421	4421

实证分析—基于决策树模型的违规判定

- 我们使用XGboost算法构建决策树模型做违规预警的深入分析，包括对违规公司的精准判别和对违规风险规避的实证分析。
- 基于上文Logistic回归和稳健性测试，选取如下指标做模型训练：递延所得税资产均值、递延所得税变动指标（abDTA）、折旧率、应收账款周转率、销售毛利率、ROA等；
- 在数据方面，对主要变量做1%极值处理，随机取90%的观测值作为样本内数据，剩余的10%作为样本外数据，训练时采用五折交叉验证得出模型参数；
- 我们从准确率和覆盖率两个维度衡量模型质量：准确率= $TP/(TP+FP)$ ；覆盖率= $TP/(TP+FN)$ 。

实证分析—基于决策树模型的违规判定

- 对监管者而言，精准地发现违规公司是首要决策目标。
- 经检验，模型可以100%的准确率分别在样本内找出38%的违规事件，在样本外找出2%的违规事件。模型的准确率越高，识别比例越低，反之则相反
- 对于投资者而言，规避违规公司是首要考虑，精确判定为未违规公司的样本形成的投资标的池越大越好。这就需要模型在给定准确率下识别未违规样本的比例最大化。
- 经检验，模型能够在样本内以100%的准确率找出46%的未违规样本。在样本外以95%的准确率找出39%的未违规样本。

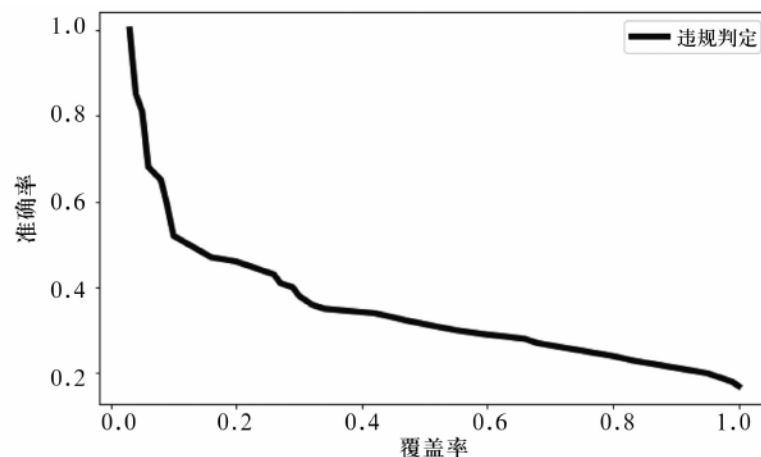


图2 决策树模型样本外“违规样本判定准确率和覆盖率曲线”

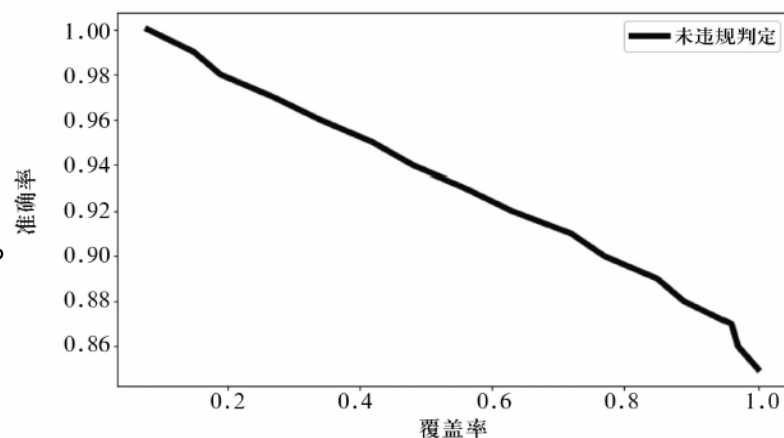


图3 决策树模型样本外“未违规样本判定准确率和覆盖率曲线”

实证分析—基于决策树模型的违规判定

- 为进一步测试决策树模型对违规预警的实践效果，我们根据决策树的违规风险规避模型来构建投资组合：根据决策树模型所得样本内对未违规的预测评估准确率1.00、覆盖率0.06和样本外为0.98、0.06对应的预测结果做组合；
- 所得结果显示，未违规组的年均ROA比违规组高2.80%，未违规组的年均股票回报率比违规组高3.69%；
- 这一比较说明，违规公司的经营业绩和投资回报率都显著差于未违规公司。本文提出的预警方法对于投资者回报和市场有效性可产生直接作用。

研究结论

- 计量分析结果显示，递延所得税异动指标与上市公司违规倾向显著相关，而与违规事件的稽查率显著负相关。
- 这一结果说明，监管机构未能充分识别递延所得税指标与公司违规和财务粉饰行为的关系，并实质上激励了上市公司通过操纵递延所得税实施财务粉饰的违规行为。
- 本文提出的递延所得税异动指标填补了预警违规事件的可选指标集。对于监管者而言，及时识别疑似违规的公司样本对于打击违规行为、维护市场健康发展具有重要意义。
- 本文构建的决策树模型可用于实现精准预警和风险规避，帮助投资者筛选出高信度未违规公司。