

采用机器学习方法检测美国上市 公司的会计欺诈行为

安宝, * 宾克, † 李斌, † Y. j u l i a y u



§ 和 JieZhang*

收到2015年10月7日; 接受2019年10月1日

摘要

我们使用机器学习方法开发了一种最先进的欺诈预测模型。我们论证了领域知识与机器学习方法相结合在模型构建中的价值。我们根据现有的会计理论选择我们的模型输入,但我们不同于以往的会计研究,使用原始会计数字,而不是财务比率。

* 上海交通大学安泰经济管理学院;

†新加坡国立大学NUS商学院会计系;

‡武汉大学财经管理学院; §弗吉尼亚大学商业学院; †南洋理工大学计算机工程学院。

克里斯蒂安·卢兹接受。我们谨感谢一位匿名评论员马克·塞奇尼、罗佐和研讨会与会者出席新加坡三尼会计研究会议、中大(深圳)会计和金融智能信息检索研讨会和科大的有益评论。这项研究的一部分是由新加坡教育部提供的二级赠款(编号。莫伊2012-t21-045)。杨宝承认来自自然科学基金委赠款的财政支持(编号。和上海浦江方案(编号: 7160116。16pjc045)。柯斌承认来自MOE启动赠款的财政支持(编号。r-521-000-032-133)。李斌感谢国家自然科学基金(71971164, 91646206)的财政支持)。本文件的网上附录可在 用于我们的最佳模型RUSBoost的代码和数据可在Github存储库中获得: <http://research.chicagobooth.edu/arc/journal-of-accounting-research/online-supplements>. <https://github.com/Jar欺诈/欺诈检测>。

我们采用最强大的机器学习方法之一，集成学习，而不是常用的Logistic回归方法。为了评估欺诈预测模型的性能，我们引入了一种新的性能评估度量，它通常用于排序问题，更适合于欺诈预测任务。从一组相同的理论驱动的理论驱动原始会计数字开始，我们表明，我们的新欺诈预测模型在很大程度上优于两个基准模型：Dechow等人。基于财务比率的Logistic回归模型，以及Cecchini等人。支持向量-机器模型与财务内核，映射原始会计数字到更广泛的一组比率。

JEL代码：C53；M41

关键词：欺诈预测；机器学习；集成学习

1. 引言

会计欺诈是一个世界性的问题。如果不及时发现和预防，它可能对欺诈公司的利益相关者（例如安然公司和世界公司）以及许多非欺诈公司的利益相关者（Gleason、Jenkins和Johnson[2008]、Goldman、Peyer和Stefanescu[2012]、Hung、Wong和Zhang[2015]）造成重大伤害）。不幸的是，会计欺诈很难被发现。此外，即使它被发现，通常也已经造成了严重的损害（Dyck、Morse和Zingales[2010]）。因此，有效和有效的公司会计欺诈检测方法将为监管机构、审计师和投资者提供重大价值。

本研究的目的是利用美国公开交易的现成财务报表数据，从样本中开发一种新的会计欺诈预测模型。公司。跟随Cecchini等人。[2010]和Dechow等人。[2011]，我们使用SEC的会计和审计强制执行发布（AAERs）中披露的检测到的重大会计错报作为我们的会计欺诈样本。虽然会计欺诈有有用的非财务预测因素（例如，高管的个人行为），但我们只使用现成的财务数据，原因有两个。首先，基于公开的财务数据的欺诈预测模型可以低成本应用于任何公开交易的公司。第二，现有会计文献中的大多数欺诈预测模型也依赖于公开的财务数据（例如，Green和Choi[1997]、Summers和Sweeney[1998]、Beneish[1999]、Cecchini等人。[2010]，Dechow等人。[2011]）。通过将预测器仅限于财务数据，我们的欺诈预测模型的性能可以与这些现有模型的性能进行比较。

关于会计欺诈的决定因素（例如Entwistle和Lindsay[1994]、Beasley[1996]、Dechow、Sloan和Sweeney[1996]、Beneish[1997, 1999]、Summers和Sweeney[1998]、Efendi、Srivastava和Swanson[2007]、Brazel、Jones和

Zimbelman[2009], Dechow等人。 [2011], Schrand和Zechman[2012]), 但大多数研究的主要目的是解释样本中的欺诈, 并经常强调因果推断。我们的目标是不同的: 我们希望开发一个模型, 可以准确地预测会计欺诈的样本(即预测问题)。什穆埃利[2010]表明, 因果推断和预测的问题, 虽然是相关的, 但从根本上是不同的。具体来说, 因果推理建模的目的是最小化模型错规范所产生的偏差, 以获得最准确的底层理论表示。相反, 预测建模的目标寻求最小化样本外预测误差, 即使用样本估计模型参数所产生的偏差和估计方差的组合。¹ 但是 因果关系 推理是现有社会科学研究的的主流, Kleinberg等人 al。 [2015] 表演 那个 在那里 都是很多 有意思 预测 问题 在现存的商业和经济中被忽视了 文献。

我们使用现有文献中的两种欺诈预测模型作为基准。第一种是基于比率的Logistic回归, 通常用于会计文献(例如, Beneish[1997, 1999]、Summers和Sweeney[1998]、Dechow等人。 [2011])。这种模型通常使用财务比率作为预测因素; 这些比率通常由人类专家根据理论(例如, 犯罪学文献中的动机-能力-机会框架)确定)。在这些模型中, Dechow等人的模型。 [2011]一般被认为是会计文献中最全面的欺诈预测模型。因此, 我们采用类似的Logistic回归模型作为我们的第一个基准模型(称为Dechow等人。模型)。第二个基准模型是Cecchini等人开发的欺诈预测模型。 [2010]基于更先进的机器学习方法(以下简称Cecchini等人。模型)。Cecchini等人没有单独使用人类专家确定的财务比率。 [2010]开发了一个基于支持向量机(SVM)的新的欺诈预测模型, 该模型具有一个财务核, 该核将原始财务数据映射到同一年内更广泛的比率集合中, 并将比率在不同年份的变化映射到更广泛的比率集合中。Cecchini等人。 [2010]发现, 具有金融核的SVM在会计方面优于传统的欺诈预测模型, 包括Dechow等人。模型。²

我们提出的欺诈预测模型在两个关键方面不同于这两个基准模型。首先, 我们使用集成学习(一种最先进的机器学习范式)来预测欺诈。大多数先前的会计欺诈预测研究都使用Logistic回归(见Dechow等人。 [2011]进行复核)。尽管集成学习已经成功

¹ 有关因果关系之间的差异的更详细讨论, 请参阅在线附录推理和预测。

² 值得注意的是, 我们的Dechow等人的性能结果。模型和Cecchini等人。模型与Dechow等人的模型没有直接可比性。 [2011]和Cecchini等人。 [2010]由于一些关键的研究设计差异, 在第3节中解释。

应用于许多其他领域（见周[2012]的评论），我们的研究是第一次将该方法应用于具有严重阶级不平衡问题（即欺诈的罕见）的会计环境）。Fernandez-Delgado等人。[2014]表明，在所有数据设置中都没有普遍最好的模型；因此，在我们的特殊设置中，集成学习方法是否能优于传统的欺诈预测方法是一个经验问题。

第二，我们提出的模型使用原始财务数据项，直接取自财务报表，作为欺诈预测因素。由于原始财务数据项是会计系统最基本的组成部分，因此有兴趣探讨它们是否可以直接用于欺诈预测。在此之前，尚不清楚基于原始财务数据的欺诈预测模型是否能优于基于人类专家识别的财务比率的欺诈预测模型。一方面，基于财务比率的欺诈预测模型可能更强大，因为人类专家确定的比率往往基于理论，这些理论提供了对公司经理何时参与欺诈的动机的尖锐预测。由于基于原始财务数据的欺诈预测模型没有直接与理论联系在一起，它们可能不那么强大。另一方面，关于会计欺诈驱动因素的现有理论可能是不完整的，因为根据定义，会计欺诈是秘密进行的，而且设计起来很难发现。因此，将原始会计数据转换为基于潜在不完整行为理论的有限数量的财务比率可能意味着失去有用的预测信息。相反，利用原始财务数据的欺诈预测模型可能更强大，因为它们不会对原始数据强加任何事前结构，而是让它们“为自己说话”。此外，随着计算机科学中机器学习方法的迅速进步，基于原始数据的欺诈预测模型可以采取更灵活和复杂的功能形式。因此，这种欺诈预测模型可能能够从原始数据中提取更有用的信息。由于这些相互冲突的权衡，我们认为这是一个经验问题，我们提出的基于原始数据的集成学习模型是否能优于基于财务比率的两个基准模型。

为了比较不同欺诈预测模型的样本外性能，我们采用了两种不同的性能评估指标。首先，我们遵循Larcker和Zakolyukina的[2012]，将接收机工作特性(ROC)曲线(AUC)下的区域作为性能评价指标。随机选择的欺诈观察将被分类器排名高于随机选择的非欺诈观察的概率(Fawcett[2006])。随机猜测的AUC为0.50。因此，任何合理的欺诈预测模型都必须具有高于0.50的AUC。

其次，我们介绍了一种用于排序问题的替代性能评估度量，称为在位置 k (NDCG@ k)处的归一化折扣累积增益)。因为频率不高

其中会计欺诈被SEC的AAER识别，甚至是表现最好的欺诈预测模型(例如Cecchini等人。[2010])将导致大量假阳性，远远超过测试期间的真阳性数。显然，鉴于可用于打击这种欺诈的资源有限，监管机构或公司监察员调查所有预测的欺诈案件是不切实际的(安永[2010])。即使人们希望调查所有关于欺诈的预测，直接和间接成本也将是巨大的，而收益则很小(因为预测的欺诈观察大多是假阳性)。当然，监管机构和其他监督机构会寻求调查预测欺诈可能性最小的观测数据。因此，我们还使用NDCG@k评估了不同欺诈预测模型的样本外性能。直观地说，NDCG@k评估了欺诈预测模型的能力，通过在预测欺诈概率最高的测试年度中选择最高的k个观测值来识别实际欺诈。在我们的研究中，我们选择了一个k，它等于观测值的1。我们选择了1%的截止期，因为通常不到1%的公司在一年的内是欺诈，根据SEC的AAER。NDCG@k的值在0到1.0之间有界，较高的值表示较好的模型性能。

为了使所有欺诈预测模型的绩效评估处于平等的地位，我们要求所有模型都从一组共同的原始财务数据开始。虽然一家公司的三份财务报表包含数百份现成的原始财务数据，但这两个基准模型只使用了来自14个专家确定的财务比率的总共28个原始财务数据项目。因此，我们使用这28个原始财务数据项作为所有欺诈预测模型的绩效评估的起点。德乔等人。模型使用来自28个原始数据项的14个财务比率，而Cecchini等人。模型使用一个财务内核，将28个原始数据项映射到一组更广泛的比率中。我们的许多会计欺诈案件跨越连续多年(这类案件称为连续欺诈)。正如我们在第3.3节中所解释的，如果串行欺诈的情况跨越训练和测试周期，它们可以人为地夸大诸如集成学习等更灵活的机器学习模型的性能。为了解决这一问题，我们将培训期间的所有欺诈观察重新编码为非欺诈，如果连续欺诈案件跨越培训和测试期间。在第7.3节中，我们显示了直接证据，表明不纠正串行欺诈问题将是显著的

提高集成学习模型的性能。

我们的样本涵盖了1991-2008年期间所有公开上市的公司。我们的样本从1991年开始，因为有一个重大的变化

U. 美国公司的欺诈行为以及SEC从那时开始执行的性质。我们的样本在2008年结束，因为监管机构从2009年左右开始减少了会计欺诈的执行，增加了2008年后许多会计欺诈案件仍未被发现的可能性(见Rakoff[2014]和第3节)；此外，SEC需要时间来确定和起诉这些案件

涉嫌欺诈案件，特别是2008年后期间，用于会计欺诈调查的公共资源减少。然而，如果我们选择较短的1991-2005年期间或较长的1991-2011年或1991-2014年期间，我们的推论在质量上是相似的。

我们首先报告两个基准模型的样本外性能结果。以AUC作为性能评价指标，我们发现两个基准模型的样本外性能显著优于随机猜测的性能。德乔等人的平均AUC为0.672。Cecchini等人的模型和0.626。模型。我们没有发现塞奇尼等人的证据。模型优于Dechow等人。模型表明，采用更先进的机器学习方法本身并不一定转化为更好的性能。如果我们使用NDCG@k作为我们的性能评估度量，推论在质量上是相似的。具体来说，Dechow等人的NDCG@k的平均值为0.028。Cecchini等人的模型和0.020。模型。

接下来，我们将研究将原始数据与灵活和强大的机器学习模型集成学习相结合是否能导致比两个基准模型更好的预测性能。我们找到两个关键结果。首先，我们没有发现明确的证据表明，基于28个原始财务数据项的Logistic回归模型比两个基准模型产生更好的预测性能。基于原始数据的Logistic回归模型的平均AUC为0.690，高于两个基准模型的平均AUC。然而，基于28个原始数据项的Logistic模型的平均NDCG@k仅为0.006，低于两个基准模型的平均NDCG@k。其次，我们发现基于相同的28个原始财务数据项的集成学习模型比两个基准模型产生更好的预测性能。集成学习模型的平均AUC和平均NDCG@k分别为0.725和0.049，相对于较好的基准模型的性能，Dechow等人的性能分别增加了7.9%和75。模型。这些性能差异在经济上也是显著的：使用NDCG@k方法(其中k1%)，我们的最佳模型，集成学习模型，在2003-08年的测试期间总共发现了16个欺诈案例，而Dechow等人的可比数字是9个。Cecchini等人的模型和7。模型。这些结果表明，直接使用原始财务数据，再加上更灵活和更强大的机器学习方法，可以产生更好的欺诈预测。

我们还研究了基于14个财务比率的集成学习模型，或者14个财务比率和28个原始数据项的组合，是否能够产生更好的欺诈预测性能。我们的结果没有证据表明上述两种替代集成学习模型的性能优于单独使用28个原始数据项的集成学习模型。这些结果进一步证明了将原始财务数据与强大的集成学习方法相结合的预测效益。

一个有趣的问题是，是否有可能通过添加现有商业数据库中现成的额外原始财务数据项来提高我们最佳欺诈预测模型集成学习的性能。为了找出答案，我们在集成学习模型中包括了来自Compustat的266个额外的原始财务数据项。使用AUC或NDCG@k作为性能评估度量，我们没有发现任何证据表明包括原始财务数据项的长“洗衣列表”有助于提高集成学习模型的样本外性能。虽然我们不能排除更多的数据将产生更好的预测模型的可能性，但这一初步证据表明，在没有任何理论指导的情况下，将大量原始数据项倾倒入集成学习模型中并不一定转化为更高的样本外预测性能。

我们的研究加入了一个小但不断增长的会计文献，使用财务报表数据来预测会计欺诈的样本。这一小文献中的代表性研究包括Cecchini等人。[2010]和Perols等人。[2017]。我们强调了Cecchini等人之间的差异。以及我们上面的研究。Perols等人。[2017]和我们的研究有相似之处，也有不同之处。像我们一样，佩罗尔斯等人。[2017]使用不同的统计方法来应对欺诈预测的独特挑战，例如检测到的欺诈的罕见性、串行欺诈和欺诈预测因子的丰富。然而，Perols等人。[2017]对会计欺诈采用了非常严格的定义，并将重点放在大型公司上，分析师紧随其后。因此，在1998-2005年期间，他们的最后样本只包含51家欺诈公司。更重要的是，我们的研究在几个关键方面不同于这些现有的研究。首先，我们介绍了一种最先进和强大的机器学习方法，集成学习。我们的结果表明，集成学习，如果使用得当，比Logistic回归和SVM更强大，用于欺诈预测。

第二，我们是第一项评估使用原始财务数据而不是从原始财务数据中得出的比率用于欺诈预测的有用性的研究。我们的实证研究提供了初步证据，证明通过仔细选择—借助理论指导—一组原始财务数据，然后将这些数据与强大的机器学习方法耦合，可以产生更强大的欺诈预测模型。我们的结果也提出了令人兴奋的可能性，我们可以进一步改进欺诈预测，通过使用额外的现成的原始财务数据指导新的理论。

第三，我们向欺诈预测文献介绍了一种新的绩效评价度量NDCG@k。与常用的绩效评价度量AUC相比，该度量已在以前的文献中使用(如Larcker和Zakolyukina[2012]、Perols等人。[2017])，NDCG@k对监管机构和其他监测员更有用。这是因为监管机构和其他监督员往往面临严重的资源限制；因此，他们只能调查少数指控的欺诈案件。因为NDCG@k通过选择来衡量模型的预测性能

最高的K公司预测欺诈概率最高，它为监管机构和其他监测员提供了一个简单的决策规则，以确定最可疑的公司进行调查。

本研究的结果也对正在进行的会计研究具有重要意义，该研究比较了文本数据和定量数据在预测会计欺诈方面的有用性（例如Larcker和Zakolyukina[2012]）。在这一研究领域中使用的一个典型的基准模型是Dechow等人。[2011]。我们的结果提高了这一行文本挖掘研究的门槛，因为我们表明了常用的Dechow等人。基于比率的Logistic回归模型显著低估了财务数据在欺诈预测中的价值。

2. 集合学习导论

汇编学习是机器学习的一种主要范式，最近在许多实词应用中取得了显著的成功（见周[2012]第17-19页，综述了集成学习方法的应用）。与通常产生单一估计量的传统机器学习方法（例如SVM方法）不同，集成学习方法将一组基本估计量（例如决策树）的预测结合起来，以提高泛化能力和鲁棒性。以前的研究（周[2012]）表明，组合通常优于任何单基估计器。然而，由于类不平衡问题的可能性，传统的集成学习方法通常需要与平衡训练数据的类分布的抽样技术相结合，要么在少数类中添加示例（过采样），要么从大多数类中删除示例（欠采样；刘和周[2013]）。在本研究中，我们采用了一种称为RUSBoost的集成学习变体（Seiffert等人。[2010]）。RUS Boost寻求利用高效欠采样技术（刘、吴和周[2009]）和目前最有影响力的集成算法Ada Boost（Freund和Schapire[1997]）。加拉等人的评论。[2012]发现RUSBoost表现出最好的性能，而且由于其简单性，计算效率也更高（Seiffert等人。[2010]）。Khoshgoftaar、VanHulse和Napolitano[2011]进一步证实了RUSBoost的优越性。

在接下来的段落中，我们首先介绍了AdaBoost算法，然后描述了在我们的RUSBoost的经验实现中如何将其与欠采样技术相结合。阿达Boost算法由于其坚实的理论基础、较强的预测能力和简单性，是最重要的集成学习方法之一（Wu等人。[2008]）。它的基本思想是在重复加权样本上“训练”一系列弱分类器（即只比随机猜测略好的模型，例如小决策树。具体来说，在每次迭代中，错误分类观测的权重将增加，而正确分类观测的权重将减少。这样，每次迭代中的弱分类器将被迫集中精力

在以前的迭代中很难预测的观测结果。最后，通过取所有弱分类器的加权平均值，可以产生一个强分类器，其中权重是基于训练样本中弱分类器的分类错误率。分类错误率较低的弱分类器将获得更高的权重。

RUSBoost是Ada Boost的一个变体，它利用随机欠采样(RUS)来解决类不平衡学习问题(Seiffert等人。[2010])。它的工作方式与Ada Boost大致相同，只是在每次迭代中执行RUS以解决欺诈和非欺诈公司的不平衡。具体而言，在每次迭代中训练弱分类器时，RUS算法使用训练期内欺诈公司的完整样本和同一训练期内随机生成的非欺诈公司子样本。³ 鲁斯博斯特估计数需要选择的比率之间的号码的欠采样多数上课意见(即。，非欺诈)还有号码的少数群体上课意见(即。，欺诈)。在这个报纸，我们建筑我们的鲁斯博斯特模型通过设置这个比率在1:1。就这样是，我们很简单样本的一样号码的欺诈意见还有非-欺诈性意见。

3. 样本和数据

3.1 样本期

我们的样本涵盖了所有公开上市的美国。1991-2008年期间的公司。我们在1991年开始抽样，因为美国公司的欺诈行为以及从那时起SEC执行的性质发生了重大变化。在他们对SEC执法计划的历史和演变的回顾中，Atkins和Bondi[2008]表明，SEC执法计划的目的是在1990年代从补救转向惩罚性。1990年以前，证券交易委员会的法定目的是为受害投资者提供补救救济，并制止今后的侵权行为。但在20世纪80年代中后期，国会通过了一系列法律，扩大了SEC的权力，并为其提供了新的惩罚权力。由于这些法律，证交会有权寻求或实施更多的惩罚行动，例如：(1)有能力对可能违反联邦证券法的个人和实体寻求民事罚款；(2)有权禁止公共公司的董事和官员在违反联邦反欺诈规定的情况下以这些身份任职；(3)有权发布行政禁令、临时限制令和命令，将非法所得利润退还违反联邦证券法的人(Atkins和Bondi[2008])。

³ 我们将随机数发生器的种子固定为零，以确保我们报告的结果的可复制性。这是计算机科学中常用的再现实验结果的方法。详见<http://www.mathworks.com/help/matlab/math/generate-random-numbers-that-are-repeatable.html>.

此外，股票期权作为高管薪酬的一种形式在20世纪90年代大幅上升，其他类似的绩效薪酬计划也大幅上升，如限制性股票和与绩效挂钩的奖金计划(Murphy[1999]、埃里克森、Hanlon和Maydew[2006])。因此，我们观察到，在1990年代出版的AAER中，内幕交易可能是会计欺诈的动机，比1980年代(Beasley、Carcelo、Hermanson[1999]和Beasley等人)更频繁地被引用。[2010])。⁴

最后，Beasley等人对1980年代和1990年代列入AAERs的会计欺诈案例进行了分析。(分别为[1999]和[2010])揭示了欺诈性质随时间的微妙变化。在这两个时期，用于欺诈性误报财务报表信息的两种最常见的技术都涉及虚报收入和资产。然而，在1990年代，通过少报支出/负债而进行的错报成为一种更常用的欺诈手段(从18%的案件增加到31)。

我们在2008年结束了样本，因为监管机构对会计欺诈的执法发生了明显的变化，这大约与2008年的金融危机相吻合。美国纽约南区地区法官Jed Rakoff[2014]在处理上市公司财务欺诈方面经验最丰富，他总结了该政权的转变如下。首先，联邦调查局在2001年之前有1000多名特工被指派调查金融欺诈，在911恐怖袭击后，将其中许多特工转移到反恐工作。第二，为了转移对未能发现2008年发现的马多夫欺诈的批评，证交会将其重点从会计欺诈调查转移到金融危机后出现的其他庞兹利克计划。⁵三，部门的正义制造a决定在2009去传播的调查财务欺诈案件其中许多u。律师的办公室，很多其中之一已经很少或者不以前的经验在调查还有起诉老练财务欺诈。在的一样时间，的u。律师的办公室的最伟大的专门知识在这些种类的案件，的南方地区的纽约刚刚开始起诉内幕交易引起的案件从的Raj拉贾拉特南录像带那个吸收a巨大的数额的该办公室证券欺诈股的注意。由于会计欺诈案件在公共执法方面出现了这些重大下降在周围的2008财务危机，法官拉科夫是的担心那个很多

⁴ 然而，现有学术文献中关于管理股权补偿对会计欺诈的影响的证据参差不齐(例如，Erickson等人。[2006]，约翰逊，瑞安和田[2009]，阿姆斯特朗，贾戈林泽和拉克尔[2010])。

⁵ SEC执法部门联席主管Andrew Ceresney也提出了类似意见在一次公开演讲中说，“在金融危机之后，SEC非常关注金融危机案例-涉及CDO、人民币交易、庞氏骗局和其他导致投资者大量损失的交易。因此，我们用于会计欺诈的资源较少。在此期间，我们进行的会计欺诈调查较少“(Ceresney[2013])。

在金融危机前后发生的会计欺诈案件将永远不会被发现。⁶

我们所有的欺诈预测模型都需要一个训练期和一个测试期。为了保证模型训练的可靠性，我们要求训练期超过10年。此外，我们要求在上一个培训年度的财务结果公告和一个测试年度的结果公告之间有24个月的差距。我们这样做是因为Dyck、Morse和Zingales[2010]发现，最初披露欺诈行为平均需要大约24个月。⁷因此，我们使用我们的样本2003-08的最后六年作为测试期。例如，培训期间是1991-2001年测试年2003年和1991-2003年测试年2005年。

3.2 欺诈样本

我们的会计欺诈样本来自SEC的AAERs，由加州大学伯克利分校财务报告和管理中心(CFRM)提供)。以往的欺诈研究通常使用四个流行的数据库：CFRM数据库、政府问责局(GAO)的盈余重述数据库、审计分析数据库(AA)的盈余重述数据库和斯坦福证券类诉讼信息中心(SCAC)的证券类诉讼数据库，这些数据库是自1995年《私人证券诉讼改革法案》通过以来提交的。我们选择CFRM数据库进行实证分析有两个关键原因。首先，我们的研究问题要求准确识别所有会计欺诈案件，以及Karpoff等人的调查结果。[2017]表明，CFRM数据库是最好的。具体来说，Karpoff等人。[2017]全面调查上述四个数据库的优缺点。他们的发现表明，没有一个单一的数据库占主导地位，数据库的选择取决于研究人员的具体研究问题。例如，CFRM数据库将是一个糟糕的选择，研究人员进行事件研究，股票市场的反应，最初发现的会计欺诈。然而，Karpoff等人。[2017]发现，如果想确定一个全面的欺诈案件清单，则CFRM排名第一（见表8）。

其次，我们希望将我们提出的模型的性能与来自Dechow等人的两个基准模型的性能进行比较。[2011]和Cecchini等人。[2010]。因为这些基准模型

⁶ 由于欺诈调查通常需要数年时间才能完成，2009年之前发生的许多会计欺诈的调查也可能受到这一制度转变的负面影响。因此，我们还使用较短的测试周期2003-05进行健壮性检查。

⁷ 尽管培训期和测试期之间有24个月的差距，但我们希望承认，与先前的研究一样，本研究中考虑的所有欺诈预测模型可能不会由监管机构或其他监测人员实时直接实施。这是因为在模型构建时，我们培训期间的一些欺诈观察可能不会被公开。

使用AAER数据,我们使用相同的数据源是有意义的。然而,我们承认CFRM数据库受到其自身潜在的选择偏差(Kedia和Rajgopal[2011], deHaan等人。[2015]),读者在解释我们的结果时应该记住的一个限制。⁸

我们在2017年3月获得的CFRM数据库的版本,从1982年5月17日到2016年9月30日。由于CFRM从那时起就没有更新其数据库,我们从SEC网站上收集了截至2018年12月31日的额外欺诈观察(AAER4012)。(https://www.sec.gov/divisions/强制/强制)表1显示了我们的样本中欺诈在1979-2014财政年度的分布情况。《会计准则》涵盖了早在1971年就发生的会计欺诈事件,但在1979年以前只有13家欺诈公司。因此,我们忽略了1979年以前的年份。我们列出了截至2014年的欺诈观察,因为SEC需要几年时间才能完成对涉嫌欺诈案件的调查(Karpoff等人。[2017])。

如表1所示,在1979-2014年期间,共发现了1,171个欺诈公司年,但发现的欺诈频率很低,通常不到每年所有公司的1。发现的会计欺诈的罕见性突出了欺诈预测的持续挑战。此外,观察到的欺诈频率在2003-2014年测试期间几乎单调下降。例如,2003-2005年欺诈的平均频率为0.94%,2006-2008年为0.51%,2009-2011年为0.46%,0.21

2012-2014。如果假设欺诈的真实发生率在2003-2014年期间保持不变,⁹的下降频率的欺诈在桌子1建议一个增加在未被发现欺诈结束了时间,一致的与的自2008年金融以来,监管机构对会计欺诈的执法下降危机已讨论过以上。另一个可能的理由是那个它带走更长的时间的监管机构去完成很多会计欺诈调查因为的危机后时期分配给会计欺诈调查的资源减少。因为我们从极少数被发现的欺诈开始意见每个人一年,的在场的a意义重大号码的在测试年度中隐藏的欺诈观察可能完全改变推论在

⁸ 可以将四个欺诈数据库结合起来,建立一个单一的会计欺诈综合数据库。然而,正如Dechow等人所指出的。[2011]和Karpoff等人。[2017],其他三个数据库在识别欺诈案件方面存在大量计量误差,因此建立这样一个全面的数据库将是耗时的,超出了本研究的范围。

⁹ 在这方面,值得注意的是,证交会的海法司共同司长的发言执法部门的安德鲁·塞雷斯尼(Andrew Ceresney)在一次公开演讲中[2013]:“但我对我们是否经历了财务报告中实际欺诈的下降感到怀疑,这可能是我们提交的调查和案件数量所表明的。这可能是因为我们没有像安然和世通这样的大规模会计欺诈。但我很难相信,仅仅由于治理变革、认证和其他萨班斯-奥克斯利创新等改革,我们已经从根本上减少了会计欺诈的案例。操纵财务报表的动机仍然存在,这样做的方法仍然可用。我们有额外的控制,但控制并不总是能有效地发现欺诈。”

泰布莱1
1979年至2014年按年份分列的欺诈公司分布情况

年份	公司总数	欺诈公司数目	百分比
1979	3, 782	4	0. 11
1980	4, 010	10	0. 25
1981	4, 501	12	0. 27
1982	4, 718	19	0. 40
1983	5, 056	14	0. 28
1984	5, 100	16	0. 31
1985	5, 087	10	0. 20
1986	5, 234	21	0. 40
1987	5, 406	16	0. 30
1988	5, 129	19	0. 37
1989	4, 977	23	0. 46
1990	4, 893	18	0. 37
1991	4, 981	28	0. 56
1992	5, 209	28	0. 54
1993	5, 644	31	0. 55
1994	5, 966	24	0. 40
1995	6, 561	22	0. 34
1996	7, 095	34	0. 48
1997	7, 138	45	0. 63
1998	7, 042	56	0. 80
1999	7, 230	79	1. 09
2000	7, 153	92	1. 29
2001	6, 780	87	1. 28
2002	6, 475	81	1. 25
2003	6, 285	69	1. 10
2004	6, 218	60	0. 96
2005	6, 119	47	0. 77
2006	6, 130	35	0. 57
2007	6, 073	30	0. 49
2008	5, 817	27	0. 46
2009	5, 618	30	0. 53
2010	5, 585	26	0. 47
2011	5, 583	22	0. 39
2012	5, 814	21	0. 36
2013	5, 831	11	0. 19
2014	5, 786	4	0. 07
共计	206, 026	1, 171	0. 57

表1显示了截至2018年12月31日的AAERs与COMPUSTAT1979-2014年基本数据合并后的年度欺诈百分比。

样本外测试。 因此，我们使用2003-2008年作为我们的主要测试样本，尽管我们也使用以下替代测试样本显示结果：2003-2005年、2003-2011年和2003-2014年。

3.3 连环诈骗

会计欺诈可能跨越多个连续的报告期，造成所谓的“连续欺诈”局面。在我们的样本中，披露的会计欺诈案件持续时间的平均值、中位数和第90百分位数分别为两年、两年和四年，表明这是

欺诈案件通常跨越多个连续报告期。如果欺诈性报告的实例跨越培训和测试期间,这种串行欺诈可能夸大集成学习方法的性能。这是因为集成学习比Logistic回归模型更灵活和更强大,因此可能比欺诈公司年更适合欺诈公司。¹⁰ 因此,提高了性能 的 的 合奏 学习 方法 可能 结果 从 的事实 那个 培训和测试样本都包含相同的欺诈公司;集合 学习 模型 可能 不是 表演 作为 好吧 什么时候 的 样本 包含不同的 公司。去交易 与 这个 关切, 我们 休息 向上 那些 案件 的 连环诈骗 那个 跨度 两者兼而有之 的 培训 还有 测试 期间。因为 我们 已经 a 小数目的 欺诈性的 坚定的岁月 相对的 去 的 号码 的 非欺诈公司年 在 任何 测试 一年, 我们 重新编码 全部 的 欺诈性的 多年 在 的 对于那些跨越培训和培训的连续欺诈案件, 培训期为零 测试 期间。但是 这个 方法 帮助 我们 避免 的 问题 相关的 与 串行 欺诈, 它 可能 也是 介绍 测量 错误 进入 培训数据。

3.4 原始的财务数据

原始财务数据项的列表是基于Cecchini等人选择的。[2010]和Dechow等人。[2011]。我们最初的原始财务数据项目列表是在Cecchini等人之后选择的。[2010]。在审查了一份全面的学术论文清单(包括萨默斯和斯威尼[1998]、仁慈[1999]、德乔等人。[2011]、Green和Choi[1997])、Cecchini等人。[2010年,表3]确定了先前欺诈预测研究中用于构造回归变量的40个原始财务数据项目的初步清单。Cecchini等人。保留了23个原始财务数据项的最终列表,在强制要求没有任何原始变量丢失超过25%的值之后。按照相同的样本选择程序,我们在1991-2008年的样本期间获得了24个原始财务数据项目的列表。表2显示了Cecchini等人最初40个原始财务数据项目的列表。[2010]在第1栏和我们的24个原始财务数据项目清单在第2栏。

我们还基于Dechow等人确定了原始财务数据项目的初始列表。[2011]。因为我们研究的目的是利用现成的财务报表中的数据来预测会计欺诈,所以我们不会简单地复制Dechow等人。表7中的模型。具体来说,我们从Dechow等人确定的候选回归变量的初始更大列表开始。[2011年,表3]。Dechow等人。[2011年,表3]提出了四种欺诈决定因素:(1)“应计质量相关变量”:九个变量,可根据资产负债表和损益表等年度财务报表中的数字计算;(2)“业绩变量”:衡量一家公司各种财务业绩的五个变量

¹⁰ 我们感谢匿名裁判提出这一点。

泰伯·埃2

在复制Dechow等人中选择的变量列表。 [2011]和Cecchini等人。 [2010]

	(1)	(2)	(3)
	40原始数据 塞奇尼的东西 等人。 [2010]	24原始数据 我们的东西 塞奇尼等人 的复制。 [2010]	11使用的财 务比率 在基础上 Dechow等 人。 [2011] 模型
资产负债表项目			
现金和短期投资	是的	是的	是的
应收款，合计	是的	是的	是的
应收款，预计存疑	是的	-	-
存货，总计	是的	是的	是的
短期投资，合计	是的	是的	是的
流动资产，合计	是的	-	是的
物业，厂房及设备，合计	是的	是的	是的
投资和预付款，其他	是的	是的	是的
资产，合计	是的	是的	是的
应付账款，贸易	-	-	是的
流动负债中的债务，合计	是的	是的	是的
应缴纳的所得税	是的	是的	是的
租金承诺至少第一年	是的	-	-
流动负债，合计	是的	是的	是的
长期负债，合计	是的	是的	是的
租金承诺至少第二年	是的	-	-
租金承诺至少第三年	是的	-	-
租金承诺至少第四年	是的	-	-
租金承诺最低为第5年	是的	-	-
负债，合计	是的	是的	是的
共同/普通公平，共计	是的	是的	是的
优先/优先股（资本），共计	是的	是的	是的
留存收益	是的	是的	-
损益表项目			
销售/营业额（净额）	是的	是的	是的
销售商品的成本	是的	是的	是的
折旧和摊销	是的	是的	-
折旧费用（附表六）	是的	-	-
销售、一般和行政费用	是的	-	-
利息及相关费用，合计	是的	是的	-
利息及相关收入，合计	是的	-	-
所得税，合计	是的	是的	-
所得税，递延	是的	-	-
特殊项目前的收入	是的	是的	是的
净收入（亏损）	是的	是的	-
现金流量表项目			
长期债务发行	-	-	是的
出售普通股和优先股	-	-	是的
融资活动净现金流量	是的	-	-
市值项目			
价格接近，年度，财政	是的	是的	是的

(续)

我也是 2-继续			
	(1)	(2)	(3)
	40 来 自 Cecchini的 原始数据项 等 人 。 [2010]	24原始数据 在我们的复 制 塞 奇 尼 等 。 [2010]	11使用的财 务比率 在基本 Dechow等人 中。 [2011] 模型
价格接近，年度，日历	是的	-	-
普通股未缴	是的	是的	是的
其他披露事项			
雇员	是的	-	-
订单积压	是的	-	-
养老金计划，预计长期利率	是的	-	-
计划资产的回报			

第1栏列出了Cecchini等人选择的初始40个原始数据项。 [2010]和第2栏列出了在删除缺失值超过25%的变量后，我们最终分析中保留的24个原始数据项。 第3栏显示了用于计算Dechow等人11个财务比率的23个原始数据项目。 [2011]。 我们将第2和第3栏中的原始数据项组合起来，以获得总共28个唯一的原始数据项作为我们的欺诈预测模型的模型输入。

维度：（3）两个“非财务变量”和四个“表外变量”，可以使用年度报告披露来计算；（4）八个“与市场相关的激励”变量，可以使用年度报告披露或股票价格数据来计算，或者两者兼而有之。 我们包括“应计质量相关变量”类别下的所有变量，但四种可自由支配的应计措施除外(即“修正的琼斯可自由支配应计项目”、“业绩匹配的可自由支配应计项目”、“DD残差的平均调整绝对值”和“研究的DD残差”)，因为我们希望使用可以很容易地从财务报表中计算的变量。¹¹ 我们包括 全部 五 变量 下面的 类别 的 “业绩 变量“除外 为了 “推迟 税 费用” 因为 的 变量 “收入 税收， 延迟 “需要计算这个比率有超过25%的价值缺失在我们的样本期(特别是，Dechow等人。 不包括此变量 在他们身上 随后 回归 分析； 见 他们的 表 7 还有 9)。 我们排除所有非财务变量和表外变量。 我们在“市场相关激励”类别下只保留“实际发行”和“账面到市场”，因为这两个变量的原始财务数据在Compustat中很容易获得。 因此，我们复制了Dechow等人。 欺诈预测模型(称为基本Dechow等人。 模型)包含来自Dechow等人的11个财务比率。 [2011年，表3]。

表2的最后一列显示了计算Dechow等人得出的11个财务比率所需的23个原始财务数据项目。 [2011]。 虽然我们复制Cecchini等人的原始财务数据项目之间有很大的重叠。 [2010]在第2栏和

¹¹ 值得注意的是，Dechow等人。 [2011]在随后的回归模型中也没有包括这四种可自由支配的应计措施。

Dechow等人。[2011]在第3栏中，有几个关键差异。¹² 首先，Dechow等人。第3栏中的原始变量列表包含四个原始财务数据项目（即“流动资产，总额”；“应付账款，贸易”；“长期债务发行”；“出售普通股和优先股”），这些数据被排除在Cecchini等人之外。第2栏中的原始数据列表。这四个原始财务数据项目用于构建基本Dechow等人使用的四个财务比率（“现金保证金的变化”、“实际发行”、“RSST应计”和“WC应计。模型。第二，Dechow等人。第3栏中的原始变量列表不包括Cecchini等人中包含的五个原始财务数据项。第2栏中的清单：“留存收益”；“折旧和摊销”；“利息和相关费用，总额”；“所得税，总额”；“净收入（损失）”这些原始财务数据项目被用来构建三个财务比率：效益[1999]中的“折旧指数”，以及萨默斯和斯威尼[1998]中的“留存收益比总资产”和“EBIT。最后，Cecchini等人。包括“净收入（损失）”作为正常化因素，因为Cecchini等人。使用比率和同比比率的变化。

由于两个原始财务数据项列表的差异，我们将两个列表中的原始数据项组合起来，以获得我们所有后续欺诈预测模型的28个原始财务数据项的最终样本。特别是我们的Dechow等人。模型包含14个财务比率（来自基本Dechow等人的11个比率。模型加上Cecchini等人的三个额外比率，见上文）。¹³ 我们的Cecchini等人。模型使用28个原始财务数据构造财务内核项目。¹⁴

4. 业绩评价

评估分类模型的样本外性能的一种常见方法是执行n倍交叉验证 (Efron 和 Tibshirani [1994]、Witten 和 Frank [2005]、Hastie、Tibshirani 和 Friedman [2009])。由于我们的欺诈数据在性质上是跨期的，执行标准的n倍交叉验证是不合适的。因此，正如第3节所指出的，我们使用样本期的最后几年（即2003–08年）作为测试期，所有前几年作为培训期。

4.1 样本外绩效评估度量1：AUC

由于我们的欺诈预测任务可以被转换为二进制分类问题（欺诈与非欺诈），因此我们可以使用分类问题的评估度量来度量欺诈预测性能。

¹² Dechow等人。[2011]使用“财产、工厂和设备、网络”，而Cecchini等人。[2010]使用“财产、工厂和设备，毛额。我们不认为这是一个显著的差异，因此将这两个项目视为等同的。

¹³ 基本Dechow等人的性能。基于11个财务比率的模型是略低于Dechow等人的性能。基于14个财务比率的模型(见在线附录表A1)。

¹⁴ 就像塞奇尼等人。模型需要28个原始数据项的滞后值，实际上使用比Dechow等人更多的信息。模型和我们提出的模型。

一个标准的分类性能度量是准确性，定义为

$$\frac{TP + TN}{TP + FP + FN + TN}$$

其中TP（真阳性）是欺诈公司的数量-

被正确归类为欺诈的年份；FN（假阴性）是数字被错误归类为非欺诈的欺诈公司年；TN（真阴性）是被正确归类为非欺诈的非欺诈公司年数；FP（假阳性）是被错误归类为欺诈的非欺诈公司年数。不幸的是，由于我们的欺诈与非欺诈数据的不平衡性质，这个标准分类性能度量在我们的场景中是不合适的（请记住，我们的样本中的欺诈百分比每年小于1）。例如，在我们的样本中，将所有公司年归类为非欺诈的Naïve策略将导致基于标准分类性能度量的准确率优于99。然而，这种看似高性能的欺诈预测模型在我们的场景中没有什么价值，因为我们的主要目的是准确地检测尽可能多的欺诈公司年，而不会对太多的非欺诈公司年进行错误分类。也就是说，我们既关心真正的阴性率（即下文定义的特异性），也关心真正的阳性率（即下文定义的敏感性）。

为了正确地衡量欺诈预测模型的性能，可以使用平衡精度(BAC)作为替代性能评估度量(He和Ma[2013])。BAC被定义为欺诈性观测中欺诈预测精度的平均值，以及非欺诈性观测中的非欺诈预测精度。具体来说， $BAC = \frac{1}{2} \times (敏感性 + 特异性)$ ，其中敏感性=

$\frac{TP}{TP + FN}$ 还有特殊性 = $\frac{TN}{TN + FP}$ 。拉克尔和扎科柳基纳[2012]注意到两个重要问题BAC作为绩效评价指标的局限性。首先，基于给定分类器的特定预测欺诈概率阈值构建BAC，阈值通常由分类器自动确定，使BAC最大化。换句话说，通过设置不同的分类器阈值，可以获得不同的BAC值。在不了解错误定性假阳性的成本与错误定性假阴性的成本之间的关系的情况下，人们无法确定用于对欺诈和非欺诈进行分类的最佳预测欺诈概率阈值。第二，灵敏度等度量对样本中正负实例的相对频率非常敏感（即数据不平衡）。

为了避免这两个限制，我们遵循Larcker和Zakolyukina的[2012]，使用AUC作为我们的样本外性能评估度量。ROC曲线是对分类器性能的二维描述，它将真实阳性率（即敏感性）和假阳性率（即1特异性）结合在一个图(Fawcett[2006]中)。上面定义的BAC只代表ROC曲线中的一个点。

通过计算AUC，可以将欺诈预测模型的性能降低到单个标量。因为AUC是单位正方形面积的一部分，所以它的值总是在0到1.0之间。

由于随机猜测产生 (0, 0) 和 (1, 1) 之间的对角线, 其面积为 0.5, 所以没有现实分类器的 AUC 应小于 0.5。正如 Fawcett [2006] 中所讨论的, AUC 相当于随机选择的正实例 (即真实欺诈) 被分类器排名高于随机选择的负实例 (即非欺诈) 的概率。

4.2 样本外性能评估度量 2: ndcg@k

欺诈预测任务也可以被认为是一个排名问题。具体来说, 我们可以将样本外绩效评估限制在只有少数 (即, k 如下所定义) 的公司年, 预测欺诈概率最高。在这种情况下, 欺诈预测模型的性能可以通过以下对排序问题的性能评估度量来衡量: NDCG@k. NDCG@k 是一种广泛使用的评估排名算法的度量, 如 Web 搜索引擎算法和推荐算法 (Jarvelin 和 Keka˘la˘inen [2002]), 并在理论上被证明是有效的 (Wang 等人。 [2013])。

形式上, 在 k (DCG@k) 位置的贴现累积增益)

定义如下: $DCG@k = \sum_{i=1}^k (2^{rel_i} - 1) / \log_2 (i + 1)$, 在哪里 i

如果 i 等于 1 排名表中的观察是一个真正的欺诈, 还有 0 个其他-

明智的。值 k 表示在预测欺诈概率最高的测试期内公司年的 k 个数 (称为“排名表”)。在我们随后的实证分析中, 我们选择 k , 使排名列表中的公司年数占测试年度所有公司年数的 1%。我们选择了 1% 的截止值, 因为 SEC 的 AAERs 惩罚的会计欺诈的平均频率通常不到所有公司的 1%。

DCG@k 基于两个关键假设: (1) 真正的欺诈行为得分较高 (i. e., $(2^{rel_i} - 1)$ 比非欺诈观察 (I. e., $(2^{rel_i} - 1)0$); (2) 真实的欺诈观察得分较高如果它是排名更高在的排名名单。就这样是, a 更高排名观察 (即, 与 a 更小 i) 威尔是加权更多高度通过的位置折扣, 即分母 $\log_2 (i + 1)$ 。

由理想的 DCG@k 归一化, 即 $NDCG@k = \frac{DCG@k}{\text{理想的 DCG@k}}$ 其中, 理想的 DCG@k 是 DCG@k 值, 当所有真实的欺诈实例都列在排名列表的顶部时。因此, NDCG@k 的值在 0 到 1.0 之间有界, 较高的值表示模型的更好的排序业绩。

为了更直观地说明我们的第二个性能度量 NDCG@k 的好处, 我们还报告了以下两个性能度量, 仅使用测试年度预测欺诈概率最高的前 1% 的公司: (I) 敏感性¹⁰ 其中 TP 是包含在预测欺诈最高的 1% 公司中的欺诈公司的数量 概率在 a 测试一年, 还有 fn 是 $\frac{fn}{n_{fn}}$ 的号码的欺诈性的被错误归类为非欺诈公司的公司在底部 99 的

在测试年的观察。 TP和FN的总和是测试年度的欺诈公司总数。 （二）精确性¹⁹ 其中TP是上面定义的，FP是错误分类为的非欺诈公司的数量 欺诈性的公司在 的上面 1% 的公司与 的最高 预测 测试中的欺诈概率一年。 TP和FP的总和是企业总数 在 的上面 1% 的公司与 的最高 预测 欺诈 给定测试中的概率一年。

相对于第一个绩效评估指标，AUC，NDCG@k避免了调查大量预测欺诈案件的重大直接和间接成本，其中大多数可能是由于严重的数据不平衡而产生的假阳性。 因为被SEC的AAER惩罚的会计欺诈的平均频率小于1%，即使是表现最好的欺诈预测模型(例如Cecchini等人。 [2010])将识别大量假阳性。 例如，Cecchini等人的表7。 [2010]报告说，他们的支持向量机与金融核正确地分类了80%的欺诈观察和90.6%的非欺诈观察在样本外测试期间，最好的竞争模型考虑在他们的研究。 然而，应用Cecchini等人。模型到我们的测试期2003-08将导致太多的假阳性。 具体来说，在Cecchini等人分析的30,883个公司年中，只有237个发生了欺诈。 2003-2008年试验期间的模型。 Cecchini等人。 然而，该方法将错误标记 2881 (1 90.6%)

(30,883,237) 非欺诈性观察为欺诈-严重高估了测试期内实际欺诈案件的数量。 显然，调查所有预测的欺诈事件是不切实际的。 即使希望调查所有预测的欺诈观察，这样做的直接和间接成本也会高得令人望而却步。 同时，好处也是有限的，因为大多数预测的欺诈性观察都是假阳性。 通过将调查限制在不超过给定数量k的公司年，在测试期间预测欺诈概率最高，NDCG@k避免了这一问题。

5. 基准模型的样本外性能

在我们讨论基准模型的结果之前，我们希望提醒读者，除非另有明确说明，我们所有的绩效评估都是在纠正跨越培训和测试期间的串行欺诈案件所产生的潜在偏见之后进行的（见第3.3节）。 除了Perols[2011]和Perols等。 [2017]，没有现有的欺诈预测研究考虑连续欺诈问题。 因此，我们的基准模型的绩效评估结果不能与现有已发表的论文进行直接比较。

5.1 德乔等人。 模型

表3报告了Dechow等人的样本外绩效评估结果。 2003-2008年测试期模型，使用评估

塔布莱E3
2003-08测试期间的样本外性能评估指标

		2003-2008年测试期间的性能指标			
输入变量	方法	公制1	公制2		
		奥克	NDCG@k	敏感性	精确
14个财务比率	1) 日志	0.672 (0.167)	0.028 (0.479)	3.99%	2.63%
28项原始财务数据	2) svm-fk	0.626 (0.012)	0.020 (0.171)	2.53%	1.92%
	3) 日志	0.690 (0.211)	0.006 (0.041)	0.73%	0.85%
	4) 鲁斯博斯特	0.725	0.049	4.88%	4.48%

此表显示欺诈预测模型的性能比较，使用2003-08年测试期间平均的以下性能指标：

(1) 接收机工作特性(ROC)曲线(AUC)下的面积)。AUC是接收机工作特性(ROC)曲线下的区域，它将一个图中的真阳性率(即灵敏度)和假阳性率(即1特异性)结合起来。ROC曲线是可视化和选择分类器的标准技术。

(2) 归一化贴现累积收益在位置k(NDCG@k)，其中k是前1%公司的数量在考试年。定义DCG@k如下： $DCG@k = \frac{1}{2^{\sum_{i=1}^k \log_2(1 + \frac{1}{r_i})}}$ 。NDCG@k是

由任意DCG@k1、DCG@k2、...、DCG@kK的算术平均数。其中DCG@k是归一化贴现累积收益在位置k。

(3) 敏感性的值，通过分类前1%的公司与最高的预测欺诈概率在一个测试一年作为欺诈。具体来说，敏感性 $\frac{TP}{TP+FN}$ ，其中TP是测试年度前1%预测欺诈中包含的真实欺诈数量，FN是测试年度中99%的观测值中被错误归类为非欺诈的真实欺诈数量。TP和FN的总和是一个测试年的真实欺诈总数。

(4) 通过将在测试年预测欺诈概率最高的前1%公司归类为欺诈的准确性值。具体来说，精确 $\frac{TP}{TP+FP}$ ，其中TP是测试年度前1%预测欺诈中包含的真实欺诈数量，FP是测试年度前1%预测欺诈中错误归类为欺诈的虚假欺诈数量。TP和FP的总和是在一个测试年度中排名前1%的公司的数量。

(5) 括号中的p值是基于我们的RUSBoost模型与RUSBoost模型的双尾t检验。另一个模型。

AUC和NDCG@k的度量。表3还报告了我们的集成学习模型与括号中的其他模型的成对t检验。然而，由于我们对这种t-测试（低功率）只有六个观察结果，我们在下面的推论中不依赖它们，相反，我们将重点放在不同欺诈检测模型之间的两个性能评估差异的大小上。

与Dechow等人一致。[2011]，三个测试年的平均AUC为0.672，远高于0.50，随机猜测的AUC截止值。平均NDCG@k为0.028。灵敏度的平均值（在表3中定义）为3.99%，这意味着人口中所有真实欺诈案件中的3.99%被捕获在预测欺诈概率最高的观测结果的前1%中。同样，精度的平均值（在表3中定义）为2.63%，这意味着只有2.63%的预测欺诈概率最高的前1%的观测是实际的欺诈案件，远远高于2003-2008年全部测试样本的无条件平均欺诈频率0.73。

5.2 塞奇尼等人。模型

Cecchini等人。[2010]开发了一种创新的支持向量机方法(以下简称支持向量机,具有金融核[SVM-FK]),该方法基于金融核,将原始金融数据映射到预定义比率列表中。Cecchini等人。列表比会计欺诈预测文献中通常使用的比率列表更广泛。¹⁵ Cecchini等人。[2010年,表7]表明,它们的SVM-FK显著优于几个代表 欺诈 预测 模型 在 会计, 包括 那个 的 Dechow等人。[2011]。

在本节中,我们复制了Cecchini等人。使用我们的样本数据的SVM-FK方法。我们的复制改进了Cecchini等人。[2010]避免两种前瞻性偏见。首先,为了解决类不平衡问题,SVM-FK通过调整模型参数C,采用成本敏感的SVM¹¹ c^{-1} (即欺诈和非欺诈的错误分类成本比率)。什么时候 搜索 为了 的 最优 参数 c^{-1} : c^{-1} 去 最大化 的 AUC、Cecchini等人的价值。[2010]直接使用该搜索执行搜索测试样本而不是保持验证样本。因此,Cecchini等人。实施程序受到前瞻性偏见的影响(他们承认;Cecchini等人。[2010年,第1156页])。我们回避这种限制通过在保持验证中执行网格搜索山姆-平民百姓。尤其是,我们 火车 的 svm-fk 模型 使用 1991-1999 还有 验证模型使用2000-2001年的测试年份2003-2008年。我们用两个 多年而不是一年的验证,因为低频率的欺诈在一个典型的一年。¹⁶后确定最优参数 C^{-1} c^{-1} (20在我们的样本中),我们训练模型并相应地测试其性能。尤其是,我们 使用 的 培训 期间 1991-2001 为了 测试 一年 2003, 1991-2002年为2004年测试年等等。

其次,我们与Cecchini等人不同。[2010],我们使用所有的公司-在测试期内进行样本外性能评估的年份。Cecchini等人。[2010]只在获得一组欺诈公司年和所有匹配的非欺诈公司年后,才执行模型培训、模型验证和样本外模型评估。由于SVM-FK模型对于大型数据集的训练和验证非常耗时,因此使用较小的匹配样本是合适的

¹⁵ 具体来说,支持向量机-FK方法将公司连续两年(分别称为第一年和第二年)的原始财务数据映射为以下六种预定义的比率,表示内部比率和同比比率的变化:()

$$\phi_{你} = \frac{u_{1,i}}{u_{1,i}} \cdot \frac{u_{2,i}}{u_{2,i}} \cdot \frac{u_{3,i}}{u_{3,i}} \cdot \frac{u_{4,i}}{u_{4,i}} \cdot \frac{u_{5,i}}{u_{5,i}} \cdot \frac{u_{6,i}}{u_{6,i}} \cdot \frac{u_{7,i}}{u_{7,i}} \cdot \frac{u_{8,i}}{u_{8,i}} \cdot \frac{u_{9,i}}{u_{9,i}} \cdot \frac{u_{10,i}}{u_{10,i}} \cdot \frac{u_{11,i}}{u_{11,i}} \cdot \frac{u_{12,i}}{u_{12,i}} \cdot \frac{u_{13,i}}{u_{13,i}} \cdot \frac{u_{14,i}}{u_{14,i}} \cdot \frac{u_{15,i}}{u_{15,i}} \cdot \frac{u_{16,i}}{u_{16,i}} \cdot \frac{u_{17,i}}{u_{17,i}} \cdot \frac{u_{18,i}}{u_{18,i}} \cdot \frac{u_{19,i}}{u_{19,i}} \cdot \frac{u_{20,i}}{u_{20,i}} \cdot \frac{u_{21,i}}{u_{21,i}} \cdot \frac{u_{22,i}}{u_{22,i}} \cdot \frac{u_{23,i}}{u_{23,i}} \cdot \frac{u_{24,i}}{u_{24,i}} \cdot \frac{u_{25,i}}{u_{25,i}} \cdot \frac{u_{26,i}}{u_{26,i}} \cdot \frac{u_{27,i}}{u_{27,i}} \cdot \frac{u_{28,i}}{u_{28,i}} \cdot \frac{u_{29,i}}{u_{29,i}} \cdot \frac{u_{30,i}}{u_{30,i}} \cdot \frac{u_{31,i}}{u_{31,i}} \cdot \frac{u_{32,i}}{u_{32,i}} \cdot \frac{u_{33,i}}{u_{33,i}} \cdot \frac{u_{34,i}}{u_{34,i}} \cdot \frac{u_{35,i}}{u_{35,i}} \cdot \frac{u_{36,i}}{u_{36,i}} \cdot \frac{u_{37,i}}{u_{37,i}} \cdot \frac{u_{38,i}}{u_{38,i}} \cdot \frac{u_{39,i}}{u_{39,i}} \cdot \frac{u_{40,i}}{u_{40,i}} \cdot \frac{u_{41,i}}{u_{41,i}} \cdot \frac{u_{42,i}}{u_{42,i}} \cdot \frac{u_{43,i}}{u_{43,i}} \cdot \frac{u_{44,i}}{u_{44,i}} \cdot \frac{u_{45,i}}{u_{45,i}} \cdot \frac{u_{46,i}}{u_{46,i}} \cdot \frac{u_{47,i}}{u_{47,i}} \cdot \frac{u_{48,i}}{u_{48,i}} \cdot \frac{u_{49,i}}{u_{49,i}} \cdot \frac{u_{50,i}}{u_{50,i}} \cdot \frac{u_{51,i}}{u_{51,i}} \cdot \frac{u_{52,i}}{u_{52,i}} \cdot \frac{u_{53,i}}{u_{53,i}} \cdot \frac{u_{54,i}}{u_{54,i}} \cdot \frac{u_{55,i}}{u_{55,i}} \cdot \frac{u_{56,i}}{u_{56,i}} \cdot \frac{u_{57,i}}{u_{57,i}} \cdot \frac{u_{58,i}}{u_{58,i}} \cdot \frac{u_{59,i}}{u_{59,i}} \cdot \frac{u_{60,i}}{u_{60,i}} \cdot \frac{u_{61,i}}{u_{61,i}} \cdot \frac{u_{62,i}}{u_{62,i}} \cdot \frac{u_{63,i}}{u_{63,i}} \cdot \frac{u_{64,i}}{u_{64,i}} \cdot \frac{u_{65,i}}{u_{65,i}} \cdot \frac{u_{66,i}}{u_{66,i}} \cdot \frac{u_{67,i}}{u_{67,i}} \cdot \frac{u_{68,i}}{u_{68,i}} \cdot \frac{u_{69,i}}{u_{69,i}} \cdot \frac{u_{70,i}}{u_{70,i}} \cdot \frac{u_{71,i}}{u_{71,i}} \cdot \frac{u_{72,i}}{u_{72,i}} \cdot \frac{u_{73,i}}{u_{73,i}} \cdot \frac{u_{74,i}}{u_{74,i}} \cdot \frac{u_{75,i}}{u_{75,i}} \cdot \frac{u_{76,i}}{u_{76,i}} \cdot \frac{u_{77,i}}{u_{77,i}} \cdot \frac{u_{78,i}}{u_{78,i}} \cdot \frac{u_{79,i}}{u_{79,i}} \cdot \frac{u_{80,i}}{u_{80,i}} \cdot \frac{u_{81,i}}{u_{81,i}} \cdot \frac{u_{82,i}}{u_{82,i}} \cdot \frac{u_{83,i}}{u_{83,i}} \cdot \frac{u_{84,i}}{u_{84,i}} \cdot \frac{u_{85,i}}{u_{85,i}} \cdot \frac{u_{86,i}}{u_{86,i}} \cdot \frac{u_{87,i}}{u_{87,i}} \cdot \frac{u_{88,i}}{u_{88,i}} \cdot \frac{u_{89,i}}{u_{89,i}} \cdot \frac{u_{90,i}}{u_{90,i}} \cdot \frac{u_{91,i}}{u_{91,i}} \cdot \frac{u_{92,i}}{u_{92,i}} \cdot \frac{u_{93,i}}{u_{93,i}} \cdot \frac{u_{94,i}}{u_{94,i}} \cdot \frac{u_{95,i}}{u_{95,i}} \cdot \frac{u_{96,i}}{u_{96,i}} \cdot \frac{u_{97,i}}{u_{97,i}} \cdot \frac{u_{98,i}}{u_{98,i}} \cdot \frac{u_{99,i}}{u_{99,i}} \cdot \frac{u_{100,i}}{u_{100,i}} \cdot \frac{u_{101,i}}{u_{101,i}} \cdot \frac{u_{102,i}}{u_{102,i}} \cdot \frac{u_{103,i}}{u_{103,i}} \cdot \frac{u_{104,i}}{u_{104,i}} \cdot \frac{u_{105,i}}{u_{105,i}} \cdot \frac{u_{106,i}}{u_{106,i}} \cdot \frac{u_{107,i}}{u_{107,i}} \cdot \frac{u_{108,i}}{u_{108,i}} \cdot \frac{u_{109,i}}{u_{109,i}} \cdot \frac{u_{110,i}}{u_{110,i}} \cdot \frac{u_{111,i}}{u_{111,i}} \cdot \frac{u_{112,i}}{u_{112,i}} \cdot \frac{u_{113,i}}{u_{113,i}} \cdot \frac{u_{114,i}}{u_{114,i}} \cdot \frac{u_{115,i}}{u_{115,i}} \cdot \frac{u_{116,i}}{u_{116,i}} \cdot \frac{u_{117,i}}{u_{117,i}} \cdot \frac{u_{118,i}}{u_{118,i}} \cdot \frac{u_{119,i}}{u_{119,i}} \cdot \frac{u_{120,i}}{u_{120,i}} \cdot \frac{u_{121,i}}{u_{121,i}} \cdot \frac{u_{122,i}}{u_{122,i}} \cdot \frac{u_{123,i}}{u_{123,i}} \cdot \frac{u_{124,i}}{u_{124,i}} \cdot \frac{u_{125,i}}{u_{125,i}} \cdot \frac{u_{126,i}}{u_{126,i}} \cdot \frac{u_{127,i}}{u_{127,i}} \cdot \frac{u_{128,i}}{u_{128,i}} \cdot \frac{u_{129,i}}{u_{129,i}} \cdot \frac{u_{130,i}}{u_{130,i}} \cdot \frac{u_{131,i}}{u_{131,i}} \cdot \frac{u_{132,i}}{u_{132,i}} \cdot \frac{u_{133,i}}{u_{133,i}} \cdot \frac{u_{134,i}}{u_{134,i}} \cdot \frac{u_{135,i}}{u_{135,i}} \cdot \frac{u_{136,i}}{u_{136,i}} \cdot \frac{u_{137,i}}{u_{137,i}} \cdot \frac{u_{138,i}}{u_{138,i}} \cdot \frac{u_{139,i}}{u_{139,i}} \cdot \frac{u_{140,i}}{u_{140,i}} \cdot \frac{u_{141,i}}{u_{141,i}} \cdot \frac{u_{142,i}}{u_{142,i}} \cdot \frac{u_{143,i}}{u_{143,i}} \cdot \frac{u_{144,i}}{u_{144,i}} \cdot \frac{u_{145,i}}{u_{145,i}} \cdot \frac{u_{146,i}}{u_{146,i}} \cdot \frac{u_{147,i}}{u_{147,i}} \cdot \frac{u_{148,i}}{u_{148,i}} \cdot \frac{u_{149,i}}{u_{149,i}} \cdot \frac{u_{150,i}}{u_{150,i}} \cdot \frac{u_{151,i}}{u_{151,i}} \cdot \frac{u_{152,i}}{u_{152,i}} \cdot \frac{u_{153,i}}{u_{153,i}} \cdot \frac{u_{154,i}}{u_{154,i}} \cdot \frac{u_{155,i}}{u_{155,i}} \cdot \frac{u_{156,i}}{u_{156,i}} \cdot \frac{u_{157,i}}{u_{157,i}} \cdot \frac{u_{158,i}}{u_{158,i}} \cdot \frac{u_{159,i}}{u_{159,i}} \cdot \frac{u_{160,i}}{u_{160,i}} \cdot \frac{u_{161,i}}{u_{161,i}} \cdot \frac{u_{162,i}}{u_{162,i}} \cdot \frac{u_{163,i}}{u_{163,i}} \cdot \frac{u_{164,i}}{u_{164,i}} \cdot \frac{u_{165,i}}{u_{165,i}} \cdot \frac{u_{166,i}}{u_{166,i}} \cdot \frac{u_{167,i}}{u_{167,i}} \cdot \frac{u_{168,i}}{u_{168,i}} \cdot \frac{u_{169,i}}{u_{169,i}} \cdot \frac{u_{170,i}}{u_{170,i}} \cdot \frac{u_{171,i}}{u_{171,i}} \cdot \frac{u_{172,i}}{u_{172,i}} \cdot \frac{u_{173,i}}{u_{173,i}} \cdot \frac{u_{174,i}}{u_{174,i}} \cdot \frac{u_{175,i}}{u_{175,i}} \cdot \frac{u_{176,i}}{u_{176,i}} \cdot \frac{u_{177,i}}{u_{177,i}} \cdot \frac{u_{178,i}}{u_{178,i}} \cdot \frac{u_{179,i}}{u_{179,i}} \cdot \frac{u_{180,i}}{u_{180,i}} \cdot \frac{u_{181,i}}{u_{181,i}} \cdot \frac{u_{182,i}}{u_{182,i}} \cdot \frac{u_{183,i}}{u_{183,i}} \cdot \frac{u_{184,i}}{u_{184,i}} \cdot \frac{u_{185,i}}{u_{185,i}} \cdot \frac{u_{186,i}}{u_{186,i}} \cdot \frac{u_{187,i}}{u_{187,i}} \cdot \frac{u_{188,i}}{u_{188,i}} \cdot \frac{u_{189,i}}{u_{189,i}} \cdot \frac{u_{190,i}}{u_{190,i}} \cdot \frac{u_{191,i}}{u_{191,i}} \cdot \frac{u_{192,i}}{u_{192,i}} \cdot \frac{u_{193,i}}{u_{193,i}} \cdot \frac{u_{194,i}}{u_{194,i}} \cdot \frac{u_{195,i}}{u_{195,i}} \cdot \frac{u_{196,i}}{u_{196,i}} \cdot \frac{u_{197,i}}{u_{197,i}} \cdot \frac{u_{198,i}}{u_{198,i}} \cdot \frac{u_{199,i}}{u_{199,i}} \cdot \frac{u_{200,i}}{u_{200,i}} \cdot \frac{u_{201,i}}{u_{201,i}} \cdot \frac{u_{202,i}}{u_{202,i}} \cdot \frac{u_{203,i}}{u_{203,i}} \cdot \frac{u_{204,i}}{u_{204,i}} \cdot \frac{u_{205,i}}{u_{205,i}} \cdot \frac{u_{206,i}}{u_{206,i}} \cdot \frac{u_{207,i}}{u_{207,i}} \cdot \frac{u_{208,i}}{u_{208,i}} \cdot \frac{u_{209,i}}{u_{209,i}} \cdot \frac{u_{210,i}}{u_{210,i}} \cdot \frac{u_{211,i}}{u_{211,i}} \cdot \frac{u_{212,i}}{u_{212,i}} \cdot \frac{u_{213,i}}{u_{213,i}} \cdot \frac{u_{214,i}}{u_{214,i}} \cdot \frac{u_{215,i}}{u_{215,i}} \cdot \frac{u_{216,i}}{u_{216,i}} \cdot \frac{u_{217,i}}{u_{217,i}} \cdot \frac{u_{218,i}}{u_{218,i}} \cdot \frac{u_{219,i}}{u_{219,i}} \cdot \frac{u_{220,i}}{u_{220,i}} \cdot \frac{u_{221,i}}{u_{221,i}} \cdot \frac{u_{222,i}}{u_{222,i}} \cdot \frac{u_{223,i}}{u_{223,i}} \cdot \frac{u_{224,i}}{u_{224,i}} \cdot \frac{u_{225,i}}{u_{225,i}} \cdot \frac{u_{226,i}}{u_{226,i}} \cdot \frac{u_{227,i}}{u_{227,i}} \cdot \frac{u_{228,i}}{u_{228,i}} \cdot \frac{u_{229,i}}{u_{229,i}} \cdot \frac{u_{230,i}}{u_{230,i}} \cdot \frac{u_{231,i}}{u_{231,i}} \cdot \frac{u_{232,i}}{u_{232,i}} \cdot \frac{u_{233,i}}{u_{233,i}} \cdot \frac{u_{234,i}}{u_{234,i}} \cdot \frac{u_{235,i}}{u_{235,i}} \cdot \frac{u_{236,i}}{u_{236,i}} \cdot \frac{u_{237,i}}{u_{237,i}} \cdot \frac{u_{238,i}}{u_{238,i}} \cdot \frac{u_{239,i}}{u_{239,i}} \cdot \frac{u_{240,i}}{u_{240,i}} \cdot \frac{u_{241,i}}{u_{241,i}} \cdot \frac{u_{242,i}}{u_{242,i}} \cdot \frac{u_{243,i}}{u_{243,i}} \cdot \frac{u_{244,i}}{u_{244,i}} \cdot \frac{u_{245,i}}{u_{245,i}} \cdot \frac{u_{246,i}}{u_{246,i}} \cdot \frac{u_{247,i}}{u_{247,i}} \cdot \frac{u_{248,i}}{u_{248,i}} \cdot \frac{u_{249,i}}{u_{249,i}} \cdot \frac{u_{250,i}}{u_{250,i}} \cdot \frac{u_{251,i}}{u_{251,i}} \cdot \frac{u_{252,i}}{u_{252,i}} \cdot \frac{u_{253,i}}{u_{253,i}} \cdot \frac{u_{254,i}}{u_{254,i}} \cdot \frac{u_{255,i}}{u_{255,i}} \cdot \frac{u_{256,i}}{u_{256,i}} \cdot \frac{u_{257,i}}{u_{257,i}} \cdot \frac{u_{258,i}}{u_{258,i}} \cdot \frac{u_{259,i}}{u_{259,i}} \cdot \frac{u_{260,i}}{u_{260,i}} \cdot \frac{u_{261,i}}{u_{261,i}} \cdot \frac{u_{262,i}}{u_{262,i}} \cdot \frac{u_{263,i}}{u_{263,i}} \cdot \frac{u_{264,i}}{u_{264,i}} \cdot \frac{u_{265,i}}{u_{265,i}} \cdot \frac{u_{266,i}}{u_{266,i}} \cdot \frac{u_{267,i}}{u_{267,i}} \cdot \frac{u_{268,i}}{u_{268,i}} \cdot \frac{u_{269,i}}{u_{269,i}} \cdot \frac{u_{270,i}}{u_{270,i}} \cdot \frac{u_{271,i}}{u_{271,i}} \cdot \frac{u_{272,i}}{u_{272,i}} \cdot \frac{u_{273,i}}{u_{273,i}} \cdot \frac{u_{274,i}}{u_{274,i}} \cdot \frac{u_{275,i}}{u_{275,i}} \cdot \frac{u_{276,i}}{u_{276,i}} \cdot \frac{u_{277,i}}{u_{277,i}} \cdot \frac{u_{278,i}}{u_{278,i}} \cdot \frac{u_{279,i}}{u_{279,i}} \cdot \frac{u_{280,i}}{u_{280,i}} \cdot \frac{u_{281,i}}{u_{281,i}} \cdot \frac{u_{282,i}}{u_{282,i}} \cdot \frac{u_{283,i}}{u_{283,i}} \cdot \frac{u_{284,i}}{u_{284,i}} \cdot \frac{u_{285,i}}{u_{285,i}} \cdot \frac{u_{286,i}}{u_{286,i}} \cdot \frac{u_{287,i}}{u_{287,i}} \cdot \frac{u_{288,i}}{u_{288,i}} \cdot \frac{u_{289,i}}{u_{289,i}} \cdot \frac{u_{290,i}}{u_{290,i}} \cdot \frac{u_{291,i}}{u_{291,i}} \cdot \frac{u_{292,i}}{u_{292,i}} \cdot \frac{u_{293,i}}{u_{293,i}} \cdot \frac{u_{294,i}}{u_{294,i}} \cdot \frac{u_{295,i}}{u_{295,i}} \cdot \frac{u_{296,i}}{u_{296,i}} \cdot \frac{u_{297,i}}{u_{297,i}} \cdot \frac{u_{298,i}}{u_{298,i}} \cdot \frac{u_{299,i}}{u_{299,i}} \cdot \frac{u_{300,i}}{u_{300,i}} \cdot \frac{u_{301,i}}{u_{301,i}} \cdot \frac{u_{302,i}}{u_{302,i}} \cdot \frac{u_{303,i}}{u_{303,i}} \cdot \frac{u_{304,i}}{u_{304,i}} \cdot \frac{u_{305,i}}{u_{305,i}} \cdot \frac{u_{306,i}}{u_{306,i}} \cdot \frac{u_{307,i}}{u_{307,i}} \cdot \frac{u_{308,i}}{u_{308,i}} \cdot \frac{u_{309,i}}{u_{309,i}} \cdot \frac{u_{310,i}}{u_{310,i}} \cdot \frac{u_{311,i}}{u_{311,i}} \cdot \frac{u_{312,i}}{u_{312,i}} \cdot \frac{u_{313,i}}{u_{313,i}} \cdot \frac{u_{314,i}}{u_{314,i}} \cdot \frac{u_{315,i}}{u_{315,i}} \cdot \frac{u_{316,i}}{u_{316,i}} \cdot \frac{u_{317,i}}{u_{317,i}} \cdot \frac{u_{318,i}}{u_{318,i}} \cdot \frac{u_{319,i}}{u_{319,i}} \cdot \frac{u_{320,i}}{u_{320,i}} \cdot \frac{u_{321,i}}{u_{321,i}} \cdot \frac{u_{322,i}}{u_{322,i}} \cdot \frac{u_{323,i}}{u_{323,i}} \cdot \frac{u_{324,i}}{u_{324,i}} \cdot \frac{u_{325,i}}{u_{325,i}} \cdot \frac{u_{326,i}}{u_{326,i}} \cdot \frac{u_{327,i}}{u_{327,i}} \cdot \frac{u_{328,i}}{u_{328,i}} \cdot \frac{u_{329,i}}{u_{329,i}} \cdot \frac{u_{330,i}}{u_{330,i}} \cdot \frac{u_{331,i}}{u_{331,i}} \cdot \frac{u_{332,i}}{u_{332,i}} \cdot \frac{u_{333,i}}{u_{333,i}} \cdot \frac{u_{334,i}}{u_{334,i}} \cdot \frac{u_{335,i}}{u_{335,i}} \cdot \frac{u_{336,i}}{u_{336,i}} \cdot \frac{u_{337,i}}{u_{337,i}} \cdot \frac{u_{338,i}}{u_{338,i}} \cdot \frac{u_{339,i}}{u_{339,i}} \cdot \frac{u_{340,i}}{u_{340,i}} \cdot \frac{u_{341,i}}{u_{341,i}} \cdot \frac{u_{342,i}}{u_{342,i}} \cdot \frac{u_{343,i}}{u_{343,i}} \cdot \frac{u_{344,i}}{u_{344,i}} \cdot \frac{u_{345,i}}{u_{345,i}} \cdot \frac{u_{346,i}}{u_{346,i}} \cdot \frac{u_{347,i}}{u_{347,i}} \cdot \frac{u_{348,i}}{u_{348,i}} \cdot \frac{u_{349,i}}{u_{349,i}} \cdot \frac{u_{350,i}}{u_{350,i}} \cdot \frac{u_{351,i}}{u_{351,i}} \cdot \frac{u_{352,i}}{u_{352,i}} \cdot \frac{u_{353,i}}{u_{353,i}} \cdot \frac{u_{354,i}}{u_{354,i}} \cdot \frac{u_{355,i}}{u_{355,i}} \cdot \frac{u_{356,i}}{u_{356,i}} \cdot \frac{u_{357,i}}{u_{357,i}} \cdot \frac{u_{358,i}}{u_{358,i}} \cdot \frac{u_{359,i}}{u_{359,i}} \cdot \frac{u_{360,i}}{u_{360,i}} \cdot \frac{u_{361,i}}{u_{361,i}} \cdot \frac{u_{362,i}}{u_{362,i}} \cdot \frac{u_{363,i}}{u_{363,i}} \cdot \frac{u_{364,i}}{u_{364,i}} \cdot \frac{u_{365,i}}{u_{365,i}} \cdot \frac{u_{366,i}}{u_{366,i}} \cdot \frac{u_{367,i}}{u_{367,i}} \cdot \frac{u_{368,i}}{u_{368,i}} \cdot \frac{u_{369,i}}{u_{369,i}} \cdot \frac{u_{370,i}}{u_{370,i}} \cdot \frac{u_{371,i}}{u_{371,i}} \cdot \frac{u_{372,i}}{u_{372,i}} \cdot \frac{u_{373,i}}{u_{373,i}} \cdot \frac{u_{374,i}}{u_{374,i}} \cdot \frac{u_{375,i}}{u_{375,i}} \cdot \frac{u_{376,i}}{u_{376,i}} \cdot \frac{u_{377,i}}{u_{377,i}} \cdot \frac{u_{378,i}}{u_{378,i}} \cdot \frac{u_{379,i}}{u_{379,i}} \cdot \frac{u_{380,i}}{u_{380,i}} \cdot \frac{u_{381,i}}{u_{381,i}} \cdot \frac{u_{382,i}}{u_{382,i}} \cdot \frac{u_{383,i}}{u_{383,i}} \cdot \frac{u_{384,i}}{u_{384,i}} \cdot \frac{u_{385,i}}{u_{385,i}} \cdot \frac{u_{386,i}}{u_{386,i}} \cdot \frac{u_{387,i}}{u_{387,i}} \cdot \frac{u_{388,i}}{u_{388,i}} \cdot \frac{u_{389,i}}{u_{389,i}} \cdot \frac{u_{390,i}}{u_{390,i}} \cdot \frac{u_{391,i}}{u_{391,i}} \cdot \frac{u_{392,i}}{u_{392,i}} \cdot \frac{u_{393,i}}{u_{393,i}} \cdot \frac{u_{394,i}}{u_{394,i}} \cdot \frac{u_{395,i}}{u_{395,i}} \cdot \frac{u_{396,i}}{u_{396,i}} \cdot \frac{u_{397,i}}{u_{397,i}} \cdot \frac{u_{398,i}}{u_{398,i}} \cdot \frac{u_{399,i}}{u_{399,i}} \cdot \frac{u_{400,i}}{u_{400,i}} \cdot \frac{u_{401,i}}{u_{401,i}} \cdot \frac{u_{402,i}}{u_{402,i}} \cdot \frac{u_{403,i}}{u_{403,i}} \cdot \frac{u_{404,i}}{u_{404,i}} \cdot \frac{u_{405,i}}{u_{405,i}} \cdot \frac{u_{406,i}}{u_{406,i}} \cdot \frac{u_{407,i}}{u_{407,i}} \cdot \frac{u_{408,i}}{u_{408,i}} \cdot \frac{u_{409,i}}{u_{409,i}} \cdot \frac{u_{410,i}}{u_{410,i}} \cdot \frac{u_{411,i}}{u_{411,i}} \cdot \frac{u_{412,i}}{u_{412,i}} \cdot \frac{u_{413,i}}{u_{413,i}} \cdot \frac{u_{414,i}}{u_{414,i}} \cdot \frac{u_{415,i}}{u_{415,i}} \cdot \frac{u_{416,i}}{u_{416,i}} \cdot \frac{u_{417,i}}{u_{417,i}} \cdot \frac{u_{418,i}}{u_{418,i}} \cdot \frac{u_{419,i}}{u_{419,i}} \cdot \frac{u_{420,i}}{u_{420,i}} \cdot \frac{u_{421,i}}{u_{421,i}} \cdot \frac{u_{422,i}}{u_{422,i}} \cdot \frac{u_{423,i}}{u_{423,i}} \cdot \frac{u_{424,i}}{u_{424,i}} \cdot \frac{u_{425,i}}{u_{425,i}} \cdot \frac{u_{426,i}}{u_{426,i}} \cdot \frac{u_{427,i}}{u_{427,i}} \cdot \frac{u_{428,i}}{u_{428,i}} \cdot \frac{u_{429,i}}{u_{429,i}} \cdot \frac{u_{430,i}}{u_{430,i}} \cdot \frac{u_{431,i}}{u_{431,i}} \cdot \frac{u_{432,i}}{u_{432,i}} \cdot \frac{u_{433,i}}{u_{433,i}} \cdot \frac{u_{434,i}}{u_{434,i}} \cdot \frac{u_{435,i}}{u_{435,i}} \cdot \frac{u_{436,i}}{u_{436,i}} \cdot \frac{u_{437,i}}{u_{437,i}} \cdot \frac{u_{438,i}}{u_{438,i}} \cdot \frac{u_{439,i}}{u_{439,i}} \cdot \frac{u_{440,i}}{u_{440,i}} \cdot \frac{u_{441,i}}{u_{441,i}} \cdot \frac{u_{442,i}}{u_{442,i}} \cdot \frac{u_{443,i}}{u_{443,i}} \cdot \frac{u_{444,i}}{u_{444,i}} \cdot \frac{u_{445,i}}{u_{445,i}} \cdot \frac{u_{446,i}}{u_{446,i}} \cdot \frac{u_{447,i}}{u_{447,i}} \cdot \frac{u_{448,i}}{u_{448,i}} \cdot \frac{u_{449,i}}{u_{449,i}} \cdot \frac{u_{450,i}}{u_{450,i}} \cdot \frac{u_{451,i}}{u_{451,i}} \cdot \frac{u_{452,i}}{u_{452,i}} \cdot \frac{u_{453,i}}{u_{453,i}} \cdot \frac{u_{454,i}}{u_{454,i}} \cdot \frac{u_{455,i}}{u_{455,i}} \cdot \frac{u_{456,i}}{u_{456,i}} \cdot \frac{u_{457,i}}{u_{457,i}} \cdot \frac{u_{458,i}}{u_{458,i}} \cdot \frac{u_{459,i}}{u_{459,i}} \cdot \frac{u_{460,i}}{u_{460,i}} \cdot \frac{u_{461,i}}{u_{461,i}} \cdot \frac{u_{462,i}}{u_{462,i}} \cdot \frac{u_{463,i}}{u_{463,i}} \cdot \frac{u_{464,i}}{u_{464,i}} \cdot \frac{u_{465,i}}{u_{465,i}} \cdot \frac{u_{466,i}}{u_{466,i}} \cdot \frac{u_{467,i}}{u_{467,i}} \cdot \frac{u_{468,i}}{u_{468,i}} \cdot \frac{u_{469,i}}{u_{469,i}} \cdot \frac{u_{470,i}}{u_{470,i}} \cdot \frac{u_{471,i}}{u_{471,i}} \cdot \frac{u_{472,i}}{u_{472,i}} \cdot \frac{u_{473,i}}{u_{473,i}} \cdot \frac{u_{474,i}}{u_{474,i}} \cdot \frac{u_{475,i}}{u_{475,i}} \cdot \frac{u_{476,i}}{u_{476,i}} \cdot \frac{u_{477,i}}{u_{477,i}} \cdot \frac{u_{478,i}}{u_{478,i}} \cdot \frac{u_{479,i}}{u_{479,i}} \cdot \frac{u_{480,i}}{u_{480,i}} \cdot \frac{u_{481,i}}{u_{481,i}} \cdot \frac{u_{482,i}}{u_{482,i}} \cdot \frac{u_{483,i}}{u_{483,i}} \cdot \frac{u_{484,i}}{u_{484,i}} \cdot \frac{u_{485,i}}{u_{485,i}} \cdot \frac{u_{486,i}}{u_{486,i}} \cdot \frac{u_{487,i}}{u_{487,i}} \cdot \frac{u_{488,i}}{u_{488,i}} \cdot \frac{u_{489,i}}{u_{489,i}} \cdot \frac{u_{490,i}}{u_{490,i}} \cdot \frac{u_{491,i}}{u_{491,i}} \cdot \frac{u_{492,i}}{u_{492,i}} \cdot \frac{u_{493,i}}{u_{493,i}} \cdot \frac{u_{494,i}}{u_{494,i}} \cdot \frac{u_{495,i}}{u_{495,i}} \cdot \frac{u_{496,i}}{u_{496,i}} \cdot \frac{u_{497,i}}{u_{497,i}} \cdot \frac{u_{498,i}}{u_{498,i}} \cdot \frac{u_{499,i}}{u_{499,i}} \cdot \frac{u_{500,i}}{u_{500,i}} \cdot \frac{u_{501,i}}{u_{501,i}} \cdot \frac{u_{502,i}}{u_{502,i}} \cdot \frac{u_{503,i}}{u_{503,i}} \cdot \frac{u_{504,i}}{u_{504,i}} \cdot \frac{u_{505,i}}{u_{505,i}} \cdot \frac{u_{506,i}}{u_{506,i}} \cdot \frac{u_{507,i}}{u_{507,i}} \cdot \frac{u_{508,i}}{u_{508,i}} \cdot \frac{u_{509,i}}{u_{509,i}} \cdot \frac{u_{510,i}}{u_{510,i}} \cdot \frac{u_{511,i}}{u_{511,i}} \cdot \frac{u_{512,i}}{u_{512,i}} \cdot \frac{u_{513,i}}{u_{513,i}} \cdot \frac{u_{514,i}}{u_{514,i}} \cdot \frac{u_{515,i}}{u_{515,i}} \cdot \frac{u_{516,i}}{u_{516,i}} \cdot \frac{u_{517,i}}{u_{517,i}} \cdot \frac{u_{518,i}}{u_{518,i}} \cdot \frac{u_{519,i}}{u_{519,i}} \cdot \frac{u_{520,i}}{u_{520,i}} \cdot \frac{u_{521,i}}{u_{521,i}} \cdot \frac{u_{522,i}}{u_{522,i}} \cdot \frac{u_{523,i}}{u_{523,i}} \cdot \frac{u_{524,i}}{u_{524,i}} \cdot \frac{u_{525,i}}{u_{525,i}} \cdot \frac{u_{526,i}}{u_{526,i}} \cdot \frac{u_{527,i}}{u_{527,i}} \cdot \frac{u_{528,i}}{u_{528,i}} \cdot \frac{u_{529,i}}{u_{529,i}} \cdot \frac{u_{530,i}}{u_{530,i}} \cdot \frac{u_{531,i}}{u_{531,i}} \cdot \frac{u_{532,i}}{u_{532,i}} \cdot \frac{u_{533,i}}{u_{533,i}} \cdot \frac{u_{534,i}}{u_{534,i}} \cdot \frac{u_{535,i}}{u_{535,i}} \cdot \frac{u_{536,i}}{u_{536,i}} \cdot \frac{u_{537,i}}{u_{537,i}} \cdot \frac{u_{538,i}}{u_{538,i}} \cdot \frac{u_{539,i}}{u_{539,i}} \cdot \frac{u_{540,i}}{u_{540,i}} \cdot \frac{u_{541,i}}{u_{541,i}} \cdot \frac{u_{542,i}}{u_{542,i}} \cdot \frac{u_{543,i}}{u_{543,i}} \cdot \frac{u_{544,i}}{u_{544,i}} \cdot \frac{u_{545,i}}{u_{545,i}} \cdot \frac{u_{546,i}}{u_{546,i}} \cdot \frac{u_{547,i}}{u_{547,i}} \cdot \frac{u_{548,i}}{u_{548,i}} \cdot \frac{u_{549,i}}{u_{549,i}} \cdot \frac{u_{550,i}}{u_{550,i}} \cdot \frac{u_{551,i}}{u_{551,i}} \cdot \frac{u_{552,i}}{u_{552,i}} \cdot \frac{u_{553,i}}{u_{553,i}} \cdot \frac{u_{554,i}}{u_{554,i}} \cdot \frac{u_{555,i}}{u_{555,i}} \cdot \frac{u_{556,i}}{u_{556,i}} \cdot \frac{u_{557,i}}{u_{557,i}} \cdot \frac{u_{558,i}}{u_{558,i}} \cdot \frac{u_{559,i}}{u_{559,i}} \cdot \frac{u_{560,i}}{u_{560,i}} \cdot \frac{u_{561,i}}{u_{561,i}} \cdot \frac{u_{562,i}}{u_{562,i}} \cdot \frac{u_{563,i}}{u_{563,i}} \cdot \frac{u_{564,i}}{u_{564,i}} \cdot \frac{u_{565,i}}{u_{565,i}} \cdot \frac{u_{566,i}}{u_{566,i}} \cdot \frac{u_{567,i}}{u_{567,i}} \cdot \frac{u_{568,i}}{u_{568,i}} \cdot \frac{u_{569,i}}{u_{569,i}} \cdot \frac{u_{570,i}}{u_{570,i}} \cdot \frac{u_{571,i}}{u_{571,i}} \cdot \frac{u_{572,i}}{u_{572,i}} \cdot \frac{u_{573,i}}{u_{573,i}} \cdot \frac{u_{574,i}}{u_{574,i}} \cdot \frac{u_{575,i}}{u_{575,i}} \cdot \frac{u_{576,i}}{u_{576,i}} \cdot \frac{u_{577,i}}{u_{577,i}} \cdot \frac{u_{578,i}}{u_{578,i}} \cdot \frac{u_{579,i}}{u_{579,i}} \cdot \frac{u_{580,i}}{u_{580,i}} \cdot \frac{u_{581,i}}{u_{581,i}} \cdot \frac{u_{582,i}}{u_{582,i}} \cdot \frac{u_{583,i}}{u_{583,i}} \cdot \frac{u_{584,i}}{u_{584,i}} \cdot \frac{u_{585,i}}{u_{585,i}}$$

培训期间的欺诈和非欺诈。然而，在测试年度中只使用匹配的欺诈和非欺诈公司年来评估SVM-FK模型的样本外性能是有问题的，因为这样做可能会引起前瞻性偏差，从而使实时实现成为问题。具体而言，由于最初披露会计欺诈(Dyck、Morse和Zingales[2010])平均需要两年时间，相关决策者(例如监管机构或投资者)在预测某一行行业年度的公司财务报表是否在测试年中欺诈时不知道。因此，决策者不能在测试年度将欺诈公司与非欺诈公司相匹配。因此，一种更合适的样本外性能评估方法是在测试期间使用整个厂商年的总体来评估SVM-FK样本外性能。为此，我们复制了Cecchini等人。[2010]使用一个匹配的欺诈和非欺诈公司年样本进行培训和验证，但在评估SVM-FK模型的样本外性能时，使用测试期2003-08年期间公司年的全部人口。

这些区别对于评估SVM-FK模型的样本外性能评估似乎是至关重要的。具体而言，未经测试的结果显示，在2003-2008年测试期间，匹配的欺诈和非欺诈公司的样本仅占欺诈和非欺诈公司人口的22.61%(6,984/30,883)。在2003-2008年测试期的全部样本中，30,883次观测中，有237次(0.77%)是真实的欺诈事件。相比之下，在2003-2008年测试期间，根据行业和年份进行的欺诈和非欺诈匹配样本中，6,984项观察中有237项(3.39%)是实际欺诈案件。在线附录的表A1显示，使用2003-2008年测试期欺诈和非欺诈观测的匹配样本，SVM-FK模型的平均AUC为0.673。值得注意的是，当使用2003-2008年公司年的全部人口时，这一数字显著下降。具体而言，如表3所示，在修正了两个前瞻性偏差后，平均AUC仅为0.626，甚至低于Dechow等人的平均AUC。模型。以NDCG@k为替代评价准则，我们发现NDCG@k的平均值用于我们复制Cecchini等人。支持向量机-FK方法仅为0.020，也低于Dechow等人的平均NDCG@k。模型。对于SVM-FK模型预测的2003-08年欺诈公司年的前1%，灵敏度和精度的平均值分别为2.53%和1.92。以AUC或NDCG@k作为绩效评价指标，我们得出结论，总的来说，Cecchini等人的绩效。模型比Dechow等人的性能弱。模型。

6. 我们提出的模型的样本外性能

接下来，我们将研究是否有可能通过使用原始财务数据和更多的数据来提高欺诈预测的性能

强大的机器学习方法，集成学习。为了更清楚地看到将原始数据与集成学习相结合的价值，我们将这一评估分为三个步骤。首先，我们通过继续使用Logistic回归来检验预测性能，但将模型输入从14个财务比率改为28个原始财务数据项。其次，我们使用28个原始数据项和集成学习方法来检验我们提出的模型的欺诈预测性能。第三，我们使用集成学习方法，但将模型输入从28个原始数据项改为14个财务比率或14个财务比率和28个原始数据项的组合来检验欺诈预测性能。

6.1 使用原始财务数据和Logit预测欺诈

表3显示了基于28个原始财务数据项的Logistic回归模型的预测性能。为了最小化不同原始数据项之间的尺度差异的影响，我们将每个公司年度观测的输入向量（即原始数据项列表）标准化，使归一化向量具有单位长度，即 $x^r = \frac{x}{\|x\|}$ 在哪里部门是按要素划分的。例如，(1, 2)是标准化的到 $(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}})$ 。¹⁷ 基于Logistic回归模型的平均AUC

这28个原始财务数据项目为0.690，高于Dechow等人的平均AUC。模型和Cecchini等人。模型。然而，基于28个原始数据项的Logistic回归的NDCG@k的平均值仅为0.006，低于Dechow等人的NDCG@k的平均值。模型和Cecchini等人。模型。总的来说，在不采用更先进的机器学习方法的情况下单独使用原始数据是否能提高预测性能方面，结果好坏参半。

6.2 使用原始数据和集成学习预测欺诈

将自己限制在相同的28个原始财务数据项上，接下来我们将研究是否可以通过使用更先进的数据挖掘方法集成学习来提高样本外欺诈预测性能。¹⁸ 虽然每种算法都有一些参数需要微调，但值得注意的是，AdaBoost(以及RUSBoost等变体)被认为是最好的开箱即用分类算法之一，因为它在不微调参数的情况下表现良好。RUSBoost有两个主要参数需要调整：决策树的数量和树的复杂性。我们通过调整这些参数 表演

¹⁷ 虽然在会计文献中不常见，但这种缩放方法在机器学习文献（汉、坎伯和裴[2006]）中得到了广泛的应用）。还请参阅对特征缩放的直观介绍（即在我们的情况下，原始数据项）。http://en.wikipedia.org/wiki/Feature_scaling

¹⁸ 在表A1的在线附录中，我们还尝试了基于SVM的方法
28个原始财务数据项，没有发现任何证据表明它优于下面讨论的集成学习方法。

与SVM-FK相同的坚持验证。具体来说，我们使用1991-1999年对RUSBoost模型进行了训练，并使用2000-2001年对模型进行了验证。我们将树的数量设置为3000棵，因为我们发现RUSBoost模型的性能在3000棵“生长”树之后趋于稳定。从理论上讲，有几个参数可以用来控制树木的复杂性，例如树木的深度或树叶上所需的最小样本数(即“minleaf”)。我们选择调整参数“minleaf”而不是“树的深度”，因为Matlab中决策树的实际实现不支持参数“树的深度”。我们将参数“minleaf”设置为5，这是在2000-2001年验证期最大限度地提高AUC和NDCG@k性能指标的范围内(1200)的最佳值。

除了上述两个重要参数外，我们还考虑了两个额外的参数：“学习率”和“与最小比率”。“学习率”缩小了每个基本模型(即我们的决策树)的贡献)。“学习率”的值在0到1之间，我们将其设置为0.1，因为这种选择通常使集成模型能够收敛到更好的解。另一个参数“比率到最小”是RUSBoost的一个特定参数，它指定执行RUS时非欺诈案件数量与欺诈案件数量之间的比率。“最小比率”通常定为1:1。因此，我们将参数“比最小”设置为1:1。

表3报告了基于2003-2008年测试期间原始财务数据的集成学习模型的样本外性能。集成学习方法的平均AUC为0.725，远大于较好的基准模型的平均AUC，Dechow等人。模型(0.672)。以NDCG@k为替代评价准则，我们发现集合学习方法的NDCG@k的平均值为0.049，大于Dechow等人的NDCG@k的平均值。模型(0.028)。对于2003-2008年测试期预测的前1%的欺诈公司，集成学习模型的灵敏度和精度平均值分别为4.88和4.48。相反，Dechow等人的相应值。模型分别为3.99%和2.63。总的来说，这些结果表明，我们提出的模型将原始财务数据与更强大的集成学习方法相结合，优于两个基准模型。

为了更好地理解2003-2008年测试期各种预测模型在差异性能方面的经济意义，我们还计算了使用NDCG@k方法确定的真实欺诈观测数，其中k1%(见在线附录表A2)。我们发现，我们最好的模型，集成学习模型，在2003-2008年的测试期间总共发现了16个欺诈案件。相比之下，Dechow等人的可比数字为9。Cecchini等人的模型和7。模型。这些结果表明，集成学习模型与两个基准模型的性能差异也具有经济意义。

泰布莱B4
样本外性能评估指标：基于不同输入变量集的集合学习方法

2003-2008年测试期间的性能指标					
输入变量	方法	公制1	公制2		
		奥克	NDCG@k	敏感性	精确
28项原始财务数据 14个财务比率	1) RUSBoost (来自表3)	0.725	0.049	4.88%	4.48%
	2) 鲁斯博斯特	0.659	0.017	2.03%	1.69%
	3) 鲁斯博斯特	0.696	0.035	3.19%	2.54%
14+财务比率 28项原始财务数据					
全部294个原始财务数据项	4) 鲁斯博斯特	0.692	0.015	1.92%	1.41%

此表显示了基于不同输入变量集的集成学习方法的性能。 作为基准，我们还根据表3中的28个原始数据项单独列出了集成学习方法的結果。 所有性能指标均在2003-2008年测试期间进行平均。 性能指标的定义见表3。

6.3 比率与原始数据使用集成学习方法

在这一部分中，我们进一步研究我们是否可以通过单独使用14个比率来提高集成学习方法的性能，或者同时使用28个原始数据项和14个比率。表4报告了这两种替代模型的样本外性能统计。 我们没有发现证据表明这两种替代的集成学习模型优于仅基于28个原始数据项的集成学习模型。这一证据与我们的猜想是一致的，即一旦我们考虑了28个原始数据项，再加上一种灵活而强大的机器学习方法，从相同的原始数据项构建的比率就不再增量地用于预测欺诈。

7. 补充分析

7.1 替代测试周期

如第3.2节所述，所观察到的欺诈频率在2003-2014年期间几乎单调下降，这种下降的一个可疑原因是存在未被发现的欺诈，特别是在金融危机后时期。 因此，我们在表3中将测试期限限制在2003-2008年为了显示主要结果的鲁棒性，我们使用以下替代测试样本在表3中复制结果：2003-2005、2003-2011和2003-2014。 由于第3.2节中所述的原因，可以合理地假定未被发现的欺诈问题随着时间的推移而增加。 因此，使用较长测试周期的性能评估结果应该不那么可靠。

结果见表5。 A-C小组分别报告2003-2005年、2003-2011年和2003-2014年测试年份的结果。 有几个关键的发现。首先，对于所有的测试周期，集成学习模型继续使用AUC或NDCG@k表现最好，这表明集成学习模型的鲁棒性。 第二，集合学习模型的性能从2003-2005年到2003-2008年、2003-2011年和2003-2014年几乎单调下降。 基于原始数据的集成学习模型的性能在2003-2005年最早的测试期是最令人印象深刻的：AUC和NDCG@k的平均值分别为0.753和0.085，而0.725和0.725的值是不同的表3为2003-2008年试验期0.049。 这些结果表明，未被发现的欺诈观测的存在可能显著降低了集成学习模型的性能。¹⁹ 第三，有薄弱的证据表明Dechow等人的表现。 基于AUC的模型(而不是NDCG@k)从2003-2005年测试期增加去2003 - 2014, 其中 是 违反直觉, 但是 的 业绩 的 的 Cecchini等人。 模型没有显示出清晰的模式 时间到了。

7.2更多的原始数据有帮助吗？

迄今为止的实证分析只使用了28个原始财务数据项目，这些数据项目是由人类专家确定的，在解释会计欺诈方面很重要。 这28个原始数据项目仅占三个主要公司财务报表可用数据项目的一小部分。 因此，我们接下来询问是否有可能通过没有任何理论指导的情况下从三个财务报表中包含更多的原始数据项来进一步提高集成学习方法的性能。 虽然对这个研究问题进行系统分析超出了本研究的范围，但我们采取了一种“蛮力”的方法，将满足相关样本选择条件的Compustat基本文件中所有可用的原始财务报表数据项都包含在集成模型中。 我们从三个财务报表中确定了总共266个额外的原始数据项目，这些数据项目适用于1991-2008年样本期间的COMPUSTAT基本年度工业格式公司。 由于集成学习方法可以处理缺失值的观测，我们不会对放大的原始数据集施加任何缺失值限制。

¹⁹ 我们试图使用表3中的集合学习模型来粗略估计2009-2014年危机后时期未被发现的欺诈率为了做到这一点，我们提出了三个假设：（1）2003-2008年测试期间报告的欺诈频率代表真实欺诈率；（2）2003-2008年和2009-2014年期间的真实欺诈率相同，这可能不是真实的（例如，Wang、Winton和Yu[2010]）；（3）如果观察到所有真实的欺诈2003-2008年和2009-2014年测试期间的集成学习模型的性能是相同的。 由于使用NDCG@k方法的集成学习模型的平均报告精度为2003-2008年测试期的4.48%（见表3）和2009-2014年测试期的0.93%（未测试），我们推断2009-2014年期间未发现欺诈的估计率为3.55%（即4.48%-0.93%）。

泰布莱5

2003-2005年、2003-2011年和2003-2014年测试年度的样本外绩效评估度量

小组A：2003-2005年测试期的样本外绩效评估指标

		2003-2005年测试期间的性能指标			
输入变量	方法	公制1	公制2		
		奥克	NDCG@k	敏感性	精确
14个财务比率	1) 日志	0.649 (0.041)	0.012 (0.153)	1.37%	1.29%
	2) svm-fk	0.637 (0.029)	0.024 (0.035)	2.28%	2.53%
	3) 日志	0.685 (0.091)	0.012 (0.054)	1.45%	1.69%
	4) 鲁斯博斯特	0.753	0.085	7.64%	7.83%

小组B：2003-2011年测试期间的样本外绩效评估指标

		性能度量在2003-2011年测试期间进行了验证			
输入变量	方法	公制1	公制2		
		奥克	NDCG@k	敏感性	精确
14个财务比率	1) 日志	0.672 (0.143)	0.024 (0.395)	3.49%	2.23%
	2) svm-fk	0.647 (0.045)	0.025 (0.364)	3.07%	1.98%
	3) 日志	0.702 (0.732)	0.012 (0.074)	1.87%	1.19%
	4) 鲁斯博斯特	0.710	0.040	4.40%	3.60%

小组C：2003-2014年测试期的样本外性能评估指标

		性能度量在2003-2014年测试期间进行了验证			
输入变量	方法	公制1	公制2		
		奥克	NDCG@k	敏感性	精确
14个财务比率	1) 日志	0.702 (0.509)	0.023 (0.616)	3.45%	1.86%
	2) svm-fk	0.628 (0.006)	0.019 (0.356)	2.30%	1.48%
	3) 日志	0.709 (0.649)	0.011 (0.125)	1.84%	1.04%
	4) 鲁斯博斯特	0.717	0.030	3.30%	2.70%

面板A-C显示欺诈预测模型的性能比较使用以下性能指标平均在测试年2003-2005，2003-2011和2003-2014。性能指标的定义见表3。括号中的p值是基于我们的RUSBoost模型与其他模型的双尾t检验。

六岁

2003-2008年测试期间的样本外性能评估指标：忽略串行欺诈问题

		2003-2008年测试期间的性能指标			
输入变量	方法	公制1	公制2		
		奥克	NDCG@k	敏感性	精确
14个财务比率	1) 日志	0. 674 (0. 011)	0. 029 (0. 035)	3. 99%	2. 63%
28项原始财务数据	2) svm-fk	0. 661 (0. 001)	0. 025 (0. 008)	2. 90%	2. 24%
	3) 日志	0. 708 (0. 005)	0. 002 (0. 006)	0. 24%	0. 28%
	4) 鲁斯博斯特	0. 801	0. 158	13. 56%	10. 74%

此表显示了基于忽略串行欺诈问题的示例的欺诈预测模型的性能比较。 所有性能指标均在2003-2008年测试期间进行平均。 性能指标的定义见表3。 括号中的p值是基于双尾t检验与我们的RUSBoost模型配对的。

表4最后一行报告结果。 对于这种更全面的集成学习模型，AUC的平均值低于基于表3中28个原始数据项的集成方法 (0. 692vs。 0. 725)。 对于这种更全面的集成学习模型，NDCG@k的平均值也比基于28个原始数据项 (0. 015vs) 的集成学习方法低得多。 0. 049)。 这些结果表明，在没有任何明确的理论指导的情况下，在我们的最佳欺诈预测模型中包含大量原始财务数据项无助于提高模型的性能。 这一发现表明，即使采用集成学习等强大的机器学习方法，理论指导在模型输入选择中仍然很重要。

7.3 连环诈骗

迄今为止报告的结果是基于一个没有串行欺诈问题的样本。 由于大多数先前的欺诈预测研究没有显示串行欺诈对欺诈预测性能的影响，在本节中，我们还复制了我们所有的欺诈预测模型，而不考虑2003-2008年同一测试期的串行欺诈问题。 结果见表6。 与我们的猜想一致，表6中集成学习模型的性能相对于表3中相同模型的性能提高最显著。 这一发现表明了欺诈预测模型中处理串行欺诈的重要性。 然而，表6所示的模型性能排名在质量上与表3相似。 因此，我们的总体推论不受串行欺诈问题的影响。

7.4 揭开集成学习方法性能背后的黑匣子

许多机器学习方法（如神经网络）的一个众所周知的缺点是这种模型的内部工作缺乏透明度。虽然单个决策树可以通过简单地可视化树结构来轻松地解释，但集成学习方法包含数百棵树，因此不能通过对单个树的视觉检查来轻松地解释。幸运的是，已经提出了一些技术来帮助揭示集成模型的重要性能驱动因素，方法是估计欺诈预测中各种特征（在本研究的背景下，原始数据项）的重要性（Tuv等人。[2009]）。在本文中，我们使用MATLAB中实现的“预测重要性”函数来估计我们集成学习模型中使用的28个原始财务数据项的重要性。²⁰尤其是，个人决定树木本质上通过选择适当的分裂点来执行特征选择。这些信息可以用来衡量每个特征的重要性：杂质的变化（衡量分裂的质量）由于分裂在一个特征上表明它的重要性（Breiman等人。[1984]）。这种重要性的概念可以通过简单地平均每个树的特征重要性来扩展到决策树组合树。

表7的A面板报告了对我们的集合学习模型在六个测试年中28个原始数据项的平均重要性的描述性统计。根据平均特征重要性将28个原始数据项从高到低排序。最重要的10个最重要的特征（或原始数据项）有助于集成学习模型的卓越性能，列出如下：(1) 普通股流通；(2) 流动资产总额；(3) 出售普通股和优先股；

(4) 不动产、厂房和设备共计；(5) 应付账款、贸易；(6) 现金和短期投资；(7) 价格接近、年度、财政；(8) 留存收益；(9) 存货，共计；(10) 普通股/普通股本，共计。值得注意的是，“普通股流通”和“价格接近、年度、财政”为欺诈预测提供了重要信息。我们怀疑“未偿普通股”在预测会计欺诈方面是有用的，仅仅是因为错误陈述的公司经常发行普通股（Dechow、Sloan和Sweeney[1995]、Dechow等人。[2011]）。同样，“价格接近，年度，财政”提供了关于欺诈可能性的宝贵信息，因为公司雇员和其他相关利益攸关方（Summers和Sweeney[1998]、Kedia和Philippon[2007]、Agrawal和Cooper[2015]）在错报期知情交易）。

为了对集成学习模型的性能进行基准测试，表7的B面板还列出了受影响的具体财务报表帐户

²⁰ 有关函数“预测器重要性”的更详细描述，请参阅
<http://www.mathworks.com/help/stats/compactclassificationensemble.predictorimportance.html>。

泰布莱7

表3中RUSBoost模型中使用的28个原始财务数据项的重要性

小组A: RUSBoost模型中28个原始数据项的重要性			
排名	28原始数据项	特征重要性的平均值 测试期2003-2008年	B组的相关帐户类别
1	普通股未缴	1.448	公司出口
2	流动资产, 合计	1.362	资产
3	出售普通股和优先股	1.324	公司出口
4	物业, 厂房及设备, 合计	1.289	资产
5	应付账款, 贸易	1.237	报酬
6	现金和短期投资	1.210	资产
7	价格接近, 年度, 财政	1.146	公司出口
8	留存收益	1.119	公司expense; rev; cogs
9	存货, 总计	1.064	inv
10	共同/普通公平, 共计	1.054	公司出口; Res
11	流动负债中的债务, 合计	1.042	利亚布
12	折旧和摊销	1.041	公司出口
13	应收款, 合计	0.982	rec
14	销售商品的成本	0.968	猪
15	资产, 合计	0.873	资产
16	长期债务发行	0.801	利亚布
17	特殊项目前的收入	0.788	公司出口
18	长期负债, 合计	0.784	利亚布
19	利息及相关费用, 合计	0.731	公司出口
20	所得税, 合计	0.712	公司出口
21	流动负债, 合计	0.704	利亚布
22	销售/营业额 (净额)	0.625	Rev
23	应缴纳的所得税	0.592	利亚布
24	投资和预付款, 其他	0.588	资产
25	负债, 合计	0.546	利亚布
26	短期投资, 合计	0.476	证券交易委员会; 资产
27	净收入 (亏损)	0.371	公司expense; rev; cogs
28	优先/优先股 (资本), 共计	0.247	公司出口
B小组: 2003-2008年测试期间被发现的会计欺诈所影响的具体账户类别			
排名	账户类别来自 Dechow等人。 [2011]	来自德乔的定义 等人。 [2011] 频率	相关的前10名原始数据项 A. 小组
1	公司出口	等于1如果错报 受影响的净收入, 因此, 股东 股权, 但不可能 分类在任何具体的 收入, 费用或 下面的股权账户 本表, 否则为0	普通股 未付; 出售 普通和首选 股票; 价格接近, 年度, 财政; 保留 收入; 普通/普通 股权, 总计

(续)

TA B L E7-继续

B小组：2003-2008年测试期间被发现的会计欺诈所影响的具体账户类别				
排名	账户类别来自 Dechow等人。 [2011]	来自德乔的定义 等人。 [2011]	频率	相关的前10名原始数据项 A. 小组
2	Rev	等于1如果错报 受影响的收入，0 否则	127	留存收益
3	资产	等于1如果错报 影响了一个资产账户 那不可能 分类在一个单独的 个人资产账户 在本表中，否则为0	63	流动资产，合计； 财产，植物和 设备，总计； 现金 和短期 投资； 存货，总计
4	rec	等于1如果错报 受影响的账户 应收款，否则为0	46	
5	inv	等于1如果错报 受影响的库存，0 否则	29	存货，总计
5	决议	等于1如果错报影响准备 金账户，否则为0。 Dechow等人。 [2011]	29	共同/普通平等， 共计
7	利亚布	等于1如果错报影响负 债，否则为0		
7	猪	等于1如果错报 影响货物成本 卖了，否则是0	26 26	应付账款，贸易 留存收益
9	报酬	等于1如果错报 受影响的账户 应付，否则为0	17	应付账款，贸易
10	MKT证券交易委员会	等于1如果错报 受影响的证券市场 证券，否则是0	0	
10	债务	等于1如果错报 受影响的津贴 坏账，否则为0	0	
共计			532	

面板A显示了表3中RUSBoost模型中使用的28个原始数据项的特征重要性。 报告的值是使用MATLAB中的“预测器重要性”函数计算的特征重要性估计（乘以10，000）。 小组B显示了2003-2008年期间AAERs查明的错报所影响的具体账户类别。 这些数据是Dechow等人从AAERs手工收集的。 [2011]。 在2003-2008年期间，有268个错报公司年，涉及532个特定账户类别。 注意，一个单一的错报公司年可能影响多个特定帐户。

^a虽然“rev”和“cogs”不是股票账户，但它们每年都不能获得“留存收益”。 因此，我们仍然将“rev”和“cogs”分配给“留存收益”。

AAERs在2003-2008年期间报告的错报，按帐户类别受会计欺诈影响的频率排序。然后，我们将A面板中的前10个原始会计数据项映射到B面板中的这些单独帐户。我们直接从Dechow等人那里获得了B面板中使用的数据，包括11个个人帐户类别。[2011]，谁又从AAERs收集这些数据。在最后一栏中，我们的前10名原始数据项与Dechow等人之间有很大的重叠。第二栏中排名第一的个人账户。这一证据表明，我们的集成学习方法在识别受错报影响最频繁的特定帐户方面相对有效。还值得注意的是，最常见的帐户类别受到Dechow等人的错报的影响。[2011]是“incexpse”，它是受影响账户的一个杂项类别，不能在表7B小组中归类为独特的收入、费用或权益账户。“Incexse”的相应原始数据项目是“普通股未清”、“出售普通股和优先股”、“价格接近、年度、财政”、“留存收益”和“共同/普通股本总额”；有趣的是，这些原始数据项目是欺诈的一些最重要的预测因素，如表7所示。

8. 结论

会计舞弊极难察觉。因此，会计研究的一个重要领域是开发及时发现公司会计欺诈的有效方法，从而限制与欺诈有关的损害的程度。本研究的目的是建立一个新的样本外欺诈预测模型的基础上，公开交易的美国公司在1991-2008年期间的样本。为了保持欺诈预测的跨期性质，我们使用我们的样本期的最后六年，2003-2008年，作为样本外测试期，前几年作为培训期。我们还使用三个替代测试周期（2003-2005、2003-2011和2003-2014）显示了我们的结果的鲁棒性。为了减少潜在的前瞻性偏见，我们还要求在上一次培训期间的财务结果公告和测试年度的财务结果公告之间至少有24个月的差距。我们这样做是因为披露欺诈行为平均需要24个月(Dyck、Morse和Zingales[2010])。

根据现有的研究，我们只使用现成的财务数据作为欺诈预测的输入。然而，我们在几个重要方面偏离了大多数现有的会计研究。首先，我们预测样本外欺诈，而不是解释样本中的欺诈决定因素。第二，我们利用财务报表中的原始财务数据来预测欺诈行为。相比之下，现存的研究通常使用人类专家确定的财务比率来预测欺诈。第三，我们使用集成学习，这是机器学习中最先进的范式之一，用于欺诈预测，而不是常用的Logit回归。最后，我们介绍一部小说

通常用于排序问题的方法，用于评估欺诈预测模型的性能，称为NDCG@k。

虽然财务报表中有许多原始财务数据项，但我们的实证分析仅限于Cecchini等人的28个原始数据项。[2010]和Dechow等人。[2011]为了比较我们提出的欺诈预测模型与更传统的欺诈预测方法（其欺诈预测因子也来自同一组原始财务数据）的性能）。这28个原始数据项目已由财务专家根据先前的欺诈研究确定。

我们采用两种基准欺诈预测模型。首先，我们跟随Dechow等人。[2011]使用基于来自28个原始数据项的14个财务比率的物流欺诈预测模型，称为Dechow等人。模型。我们的第二个基准模型是Cecchini等人开发的欺诈预测模型。基于SVM-FK，将28个原始财务数据项目映射到一组更广泛的财务比率和财务比率的变化。

我们发现两个基准模型的样本外性能都优于随机猜测，但Dechow等人的性能。模型优于Cecchini等人的性能。模型。更重要的是，我们发现基于28个原始财务数据项的集成学习模型直接优于两个基准模型。然而，我们没有发现证据表明，基于原始数据的简单Logistic回归模型，或基于14个财务比率的集成学习模型，或14个财务比率和28个原始数据项的组合，仅优于基于28个原始数据项的集成学习模型。这些结果表明，Dechow等人。基于人力专家确定的财务比率的模型没有充分利用原始财务数据中包含的宝贵信息。此外，我们还表明，通过构建一个更先进的机器学习模型来利用这些数据，可以从原始数据中提取更有用的预测信息。

由于这28个原始数据项只代表会计系统中可能出现的数百个原始财务数据项中的一小部分，我们还检查是否包括来自三个财务报表的所有合格的、现成的原始财务数据项，从而进一步提高了我们的集成学习模型的性能。我们没有发现这样的证据，表明仅仅添加更多的原始数据项，而没有任何补充理论是不够的。然而，我们不排除通过从数百个现成的原始财务数据项中进行更系统和理论驱动的模式输入选择来开发更好的欺诈预测模型的可能性。

我们的发现与越来越多的会计文献有关，这些文献试图收集公司文件中的文本数据，以预测欺诈或公司业绩（例如，Li[2010]、Larcker和Zakolyukina[2012]、Lo、Ramos和Rogo[2017]）。为了证明这些文本数据的有用性，常用的基准是列表

从原始财务数据中得出的定量变量。未来的研究人员可能会探讨的一个有趣的问题是，如果使用先进的数据挖掘技术更有效地提取现成的原始金融数据中的信息，那么文本数据的有用性是否继续存在。

参考资料

- 阿拉瓦尔A和T. 库珀。 “会计丑闻前的内部交易” *公司财务杂志* 34 (2015) : 169-90。
- 阿姆斯特朗, c.s.; a.d. jagolinzer; 和d.f. larcker。 “首席执行官股权激励和会计违规” *会计研究杂志* 48 (2010) : 225-71。
- 阿特金斯, p.s. 和b.j. 邦迪。 “评估使命: 对SEC执行计划的历史和演变的批判性回顾”, *Fordham公司和金融法杂志* 13 (2008) : 367-417。
- 贝斯利, M. “董事组成与财务报表欺诈关系的实证分析”, 《会计评论》 71 (1996) : 443-65。
- 贝斯利, M. S.; j. 五。 卡塞罗; 和d. r. 赫尔曼森。 “欺诈性财务报告: 1987-1997年: 美国分析。 “由特雷德韦委员会赞助组织委员会赞助, 1999年。
- Beasley, m.s.; j.v. carcelo; d.r. hermannson; and t.l. neal。 “欺诈性金融报告: 1998-2007年: 美国分析。 公共公司。 “由特雷德韦委员会赞助组织委员会赞助, 2010年。
- Beneish, M. d. 《发现违反公认会计原则的行为: 对评估极端财务绩效公司的收益管理的影响》, 《会计和公共政策杂志》 16 (1997) : 271-309。
- Beneish, M. d. “盈余操纵的检测” *金融分析师杂志* 55 (1999) : 24-36。
- 布拉泽尔, j. f.; k. l. 琼斯; 和m. f. 津贝尔曼。 “使用非财务措施评估欺诈风险” *会计研究杂志* 47 (2009) : 1135-66。
- 布雷曼, l.; j. 弗里德曼; c. j. 树类分类和回归树。 博卡拉顿, FL: CRC出版社, 1984年。
- 塞奇尼, M.; H. 艾图格; G. J. Koehler; 和P. Pathak。 “检测公共公司的管理欺诈” *管理科学* 56 (2010) : 1146-60。
- 塞雷斯尼, a. 执行司 “财务报告和会计欺诈” 联合主任在美国法律研究所继续法律教育处的演讲, 华盛顿, D. 2013年9月19日。
- 德乔, 下午; 通用电气; 拉尔森和斯隆。 “预测材料会计错误陈述” *当代会计研究* 28 (2011) : 17-82。
- 德乔, p. m.; 斯隆; 以及。 p. 斯威尼。 “检测收益管理” *会计评论* 70 (1995) : 193-226。
- 德乔, p. m.; r. g. 斯隆; 和a. p. 斯威尼。 “盈余操纵的原因和后果: 受SEC强制执行行为约束的公司分析” *当代会计研究* 13 (1996) : 1-36。
- 德汉, e.; s. 凯迪亚; k. 哦; 还有S. 拉杰戈帕尔。 “循环门和SEC的执行结果: 民事诉讼的初步证据”, 《会计和经济学杂志》 60 (2015) : 65-96。
- 戴克, a. a. morse和l. zingales。 “谁吹了公司欺诈的口哨?” *金融杂志* LXV (2010) : 2213-53。
- 埃芬迪, j.; a. 斯里瓦斯塔瓦; 和e. p. 天鹅森。 “为什么公司经理会误解财务报表? 期权补偿和其他因素的作用。” *金融经济学杂志* 85 (2007) : 667-708。
- EFRON, B. 和R. J. TIBS HIRANIAN对引导的介绍。 英国伦敦: CRC出版社, 1994年。

- 昆虫, G. 和D. 林赛。 “收入影响财务报表错报存在、原因和发现的档案研究” 当代会计研究11 (1994): 271-96。
- 埃里克森, m.; 汉隆, 和梅德。 “高管股权激励和会计欺诈之间有联系吗?” 会计研究杂志44 (2006): 113-43。
- <http://www.ey.com/Publication/vwLUAssets/Driving安永>。 “推动道德增长——新市场, 新挑战”, 第11次全球欺诈调查, 2010年。 可在道德增长新市场, 新挑战: 第11次全球欺诈调查/\$文件/EY第11次全球欺诈调查。 - - - - -
- 福塞特, t. - “ROG分析简介” 模式识别信函27 (2006): 861-74。 费尔南德斯-德尔加多, M.; E. 玉米饼。 巴罗; 和d. 阿莫林。 “我们需要几百个分类解决真实世界分类问题?” 机器学习研究杂志 15 (2014): 3133 - 81。
- 弗伦德, Y. 和R. Schapire。 “在线学习的决策论概括和应用用于促进”, 《计算机和系统科学杂志》, 55 (1997): 119-39。
- 加拉, m.; 费尔南德斯; 布林切亚; 巴宾斯; 海雷拉。 “回顾一下类不平衡问题的集合: Bagging-、Boosting-和基于混合的方法。” IEEE关于系统、人和控制论的交易, C部分-应用和评论42 (2012): 463-84。
- 葛莱森, c.; N. T. 詹金斯; 和W. B. 约翰逊。 “会计报告的传染效应” 会计评论83 (2008): 83-110。
- 高盛, E.; U. Peyer; 和I. Stefanescu。 “财务失实及其对竞争对手的影响” 财务管理41 (2012): 915-45。
- 格林, P. 和J. h. 崔。 “通过神经网络技术评估管理欺诈的风险” 审计: 实践与理论杂志16 (1997): 14-29。
- . HAN, J. W.; M. KAMBER; 和J. PEIDATA采矿: 概念和技术。 旧金山, 加利福尼亚州: 摩根·考夫曼, 2006年。
- . HASTIE, T.; R. TIBS HIRANI; 和J. H. FRIED MAN统计学习的要素。 纽约: 斯普林格, 2009年。
- 他, H. 和Y. 妈妈。 均衡学习: 基础、算法和应用。 霍博肯, 新泽西州: 威利, 2013年。
- 洪先生。 王j. 和F. 张。 “政治谎言的价值与市场信誉: 来自中国企业丑闻的证据” 当代会计研究, 32 (2015): 1641-75。
- ja `` 瑞林K. 和j. 凯卡 `` 洛杉矶 `` 伊宁。 “基于累积增益的IR技术评估” ACM信息系统交易20 (2002): 422-46。
- 约翰逊, 旧金山; 赫·瑞安; 还有Y. 田s.。 “管理激励和公司欺诈: 激励的来源重要。” 金融评论13 (2009): 115-45。
- 卡波夫, j. m.; a. 科斯特; 李德生和马丁。 “金融不当行为研究中的Proxies和数据库” 会计评论92 (2017): 129-63。
- 凯迪娅和菲利普。 “欺诈会计的经济学” 金融研究评论22 (2007): 2169-99。
- 凯迪娅, S. 拉杰戈帕尔。 “SEC的强制执行偏好会影响公司的不当行为吗?” 《会计和经济学杂志》51 (2011): 259-78。
- khoshgoftaar, t. m.; j. 范·赫尔斯; 还有纳波利塔诺。 “比较Boost和Bag-带有噪音和不平衡数据的Ging技术。” IEEE关于系统、人和控制论的交易, A部分-系统和人类4 (2011): 552-68。
- 克莱恩伯格, j.; j. 路德维希; s. mullainathan; 和z. obermeyer。 “预测政策问题。” 美国经济评论: 论文和论文集, 105 (2015): 491-95。
- 拉克尔, D. 和A. a. zakolyukina。 “在会议呼吁中探测欺骗性的讨论。” 会计研究杂志50 (2012): 495-540。
- 李, F. “企业筛选中的前瞻性声明的信息内容-Na `` ve贝叶斯机器学习方法” 会计研究杂志 48 (2010): 1049-102。

刘, x-y. ; WU, J.; and Zhou, Z.-H. "Exploratory Underampling for Class-Imbalance Learning." *IEEE Transactions on Systems, Man, and Cybernetics, Part B-Cybernetics* 39 (2009): 539-50. 刘, X. Y. 和 Z. H. 周. 《平衡学习中的整合学习方法: 基础、算法和应用》, 由 H. He 和 Y. Ma 编辑。

霍博肯, 新泽西州: 约翰威利和儿子公司, 2013年: 61-82。

罗, K; F. 拉莫斯; R. 罗戈. "收益管理和年度报告可读性。"

会计研究杂志 63 (2017): 1-25。

墨菲, K. J. 《行政报酬》劳工经济学手册 3 (1999年): 2485-563. PEROLS, J. L. "财务报表欺诈检测: 统计和机器分析

学习算法。"审计: 实践与理论杂志" 30 (2011): 19-50。

perols, j. l.; r. m. bowen; c. zimmermann; 和 b. samba. "大海捞针: 利用数据分析来改善欺诈预测" 会计评论 92 (2017): 221-45. RAKOFF, J. S. "金融危机: 为什么没有高层管理人员被起诉?"

纽约图书评论, 2014年1月9日。

施兰德, c. 和 S. l. c. 泽克曼. "高管过度自信和财务错误报告的滑铁卢", 《会计和经济杂志》 53 (2012): 311-29。

西弗特, C; T. M. Khoshgoftaar; J. 范·赫尔斯; 还有纳波利塔诺. "Rusboost: 一种杂交降低类失衡的方法。" *IEEE关于系统、人和控制论的交易, A部分-系统和人类* 40 (2010): 185-97。

什穆利, g. "解释或预测" 统计科学 25 (2010): 289-310。

萨默斯. 和 j. t. 斯威尼. "欺诈性错报财务报表与内幕交易: 实证分析" 会计评论 73 (1998): 131-46。

图夫, e.; a. borisov; g. runger; 和 k. torkkola. "特征选择与组合, 人工变量, 和冗余消除。" 机器学习研究杂志 10 (2009): 1341-66。

王, T. 是. 温顿; 和 x. 于. "公司欺诈和商业条件: IPO的证据", 《金融杂志》, 65 (2010): 2255-92。

王毅; 王立李; 他; W. 陈; 和 t-y. 刘. "NDCG的理论分析排名措施。" 在第26届学习理论年会论文集, 2013年. 维滕, 我. 和 E. 弗兰克. 数据挖掘: 实用的机器学习工具和技术. 圣

旧金山, 加利福尼亚州: 摩根·考夫曼, 2005年。

吴, x; v. 库马尔; j. r. 昆兰; j. ghosh; q. yang; h. motoda; g. j. mclachlan; a. ng; b. liu; p. 余 s.; 周志赫; 史丹巴赫; 德杰. 和 D. STEINBERG "十大算法在数据挖掘中。" *知识和信息系统* 14 (2008): 1-37。

周, z-h. 集合学习方法: 基础和算法. 博卡拉顿, FL: CRC出版社, 2012年。