

# 基于语境的财务信息解读

## ——学海拾珠系列之二百二十二

报告日期: 2025-02-05

分析师: 骆昱杉

执业证书号: S0010522110001

邮箱: luoyushan@hazq.com

分析师: 严佳炜

执业证书号: S0010520070001

邮箱: yanjw@hazq.com

### 相关报告

- 《跟踪误差的构成成分、中期交易与基金业绩——学海拾珠系列之二百二十一》
- 《基于混合转移分布的投资组合优化方法——学海拾珠系列之二百二十》
- 《模糊性会引发处置效应吗？——学海拾珠系列之二百一十九》
- 《国际主动型基金的持仓拥挤与业绩影响——学海拾珠系列之二百一十八》
- 《回撤 Beta 与投资组合优化——学海拾珠系列之二百一十七》
- 《国际股票市场中的因子动量与价格动量——学海拾珠系列之二百一十六》
- 《基金中的策略背离、竞争与资金流动——学海拾珠系列之二百一十五》

### 主要观点:

本篇是学海拾珠系列第二百二十二篇，文章采用 **BERT** 编码捕捉文本多维信息，通过 **ANN** 模型完成文本与数字信息的深度交互，实现数值信息的情境化解读。结果表明，情境信息对于解释数字输入具有重要意义，融入叙述背景能大幅提升数字数据的价值。

#### ● 量化财务数字与文本背景交互的挑战与解决方案

在量化“财务数字与文本信息背景交互”时面临两大挑战：一是编码文本中的详细语义信息，二是构建模型以捕捉数字和文本间丰富且高度非线性的相互作用。

#### ● 模型方法：结合 BERT 与 ANN 预测未来收益

本研究从 MD&A 报告中提取收益相关句子，并使用 **BERT** 编码文本信息。构建 **ANN** 全连接模型，将会计信息（营收、经营现金流，按资产规模调整）与上下文信息作为输入，预测未来收益等目标变量。

**基准模型：**（1）数字输入模型；（2）BERT 编码后的文本输入模型；（3）部分连接的模型：将文本和数字输入合并，情境信息不参与数字输入的处理（反之亦然）。

#### ● 会计信息与情境互动：提升预期准确性，强化数字解读

研究表明，会计信息与其情境间的相互作用有助于形成更准确的预期，叙述性披露通过与数字数据的相互作用促进了会计数字的解读。情境性在宏观经济不确定性增加时更为重要，且分析师在预测决策中考虑了叙述情境与数字披露的相互作用，提高了预测准确性。模型通过情境衡量收益持续性异质性，成功捕捉公司-年度特征，在解释样本外收益动态方面信息价值更高。

#### ● 文献来源

核心内容摘自 ALEX G. KIM, VALERI V. NIKOLAEV 于 2024 年 10 月 31 日在 Journal of Accounting Research 上的文章《Context-Based Interpretation of Financial Information》。

#### ● 风险提示

文献结论基于历史数据与海外文献进行总结；不构成任何投资建议。

# 正文目录

1 引言.....	4
2 方法和研究设计.....	8
2.1 编码叙事信息.....	8
2.2 数值信息情境化解读的建模.....	9
3 数据与方法论.....	9
3.1 模型设计.....	11
3.2 模型训练与超参数选择.....	12
4 模型结果.....	13
4.1 利用叙事背景预测收益.....	13
4.2 利用叙述性语境预测现金流.....	16
4.3 叙述性情境与股票回报.....	17
4.4 情境性的影响因素是什么?.....	20
4.5 分析师对收益情境信息的运用.....	22
5 应用: 情境化的收益持续性.....	23
5.1 方法论.....	23
5.2 收益持续性的异质性.....	24
5.3 验证基于情境的收益持续性.....	25
5.3.1 度量的表面效度.....	25
5.3.2 基于公司特征的收益持续性是否可以推断?.....	25
5.3.3 样本外的持续性是否具有信息性?.....	26
6 讨论.....	27
7 结论.....	28
风险提示: .....	29

## 图表目录

图表 1 文章框架	4
图表 2 样本选择过程	10
图表 3 分析中使用的变量的描述性统计	10
图表 4 全连接模型	11
图表 5 部分连接模型	12
图表 6 模型按年份以及在整体样本中的表现	14
图表 7 情境信息在收益预测中的相关性：时间趋势	15
图表 8 情境信息在收益预测中的相关性：预测准确性	15
图表 9 使用模型来预测未来的经营性现金流的表现	16
图表 10 情境信息与未来经营性现金流：时间趋势	17
图表 11 情境信息与未来经营性现金流：预测准确性	17
图表 12 长期（年度）回报的结果	18
图表 13 60 日回报的结果	19
图表 14 情境信息与年度股票回报：时间趋势	20
图表 15 情境信息与年度股票回报：预测准确性	20
图表 16 情境信息与 60 日回报：时间趋势	20
图表 17 情境信息与 60 日回报：预测准确性	20
图表 18 情境性的横截面和时间序列变化	21
图表 19 不同模型在解释分析师收益预测调整方面的平均解释能力改进	23
图表 20 基于公司-年份情境的持续性系数的年度横截面平均值	24
图表 21 收益持续性的异质性	25
图表 22 基于情境的持续性能否预测实际的收益动态	27

# 1 引言

图表 1 文章框架



资料来源：华安证券研究所整理

企业通过多种沟通方式（包括数字、文字、图像、语音、视频和非言语暗示）向利益相关者披露信息。随着计算技术的最新发展，多模态数据在决策制定中变得越来越重要（例如，Zhu [2019]）。然而，多模态披露之间相互补充的程度尚待实证研究来揭示。我们研究了数字和叙述性披露之间互补性（相互作用）的经济重要性。我们之所以关注叙述性披露，是因为它们“提供了分析财务信息所需的背景”。具体来说，证券交易委员会明确要求管理层讨论与分析（MD&A）部分讨论非量化因素，这些因素有助于解释所报告的数字，从而使投资者能够对未来现金流形成更准确的预期。在本研究中，我们引入了一种基于深度学习的方法来衡量会计数字的情境性，即会计数字对由数字和文本信息相互作用产生的背景的依赖程度。

举个例子，假设两家不同的公司公布了相同的数字业绩。尽管报告的数字相同，但由于与业务战略、竞争、产品需求、运营摩擦、未来计划等相关的叙述性评论（统称为叙述性背景）不同，对这些数字的解释可能会有很大差异。尽管直接的结论是数字和文本信息会相互作用，但这种相互作用往往复杂且高度非线性，因此在实证研究中难以捉摸。这些互补性在塑造投资者对公司未来的信念方面有多重要，以及何时背景最为关键？

会计准则制定者正在积极考虑如何加强背景信息的披露，以补充数字信息。例

如，美国证券交易委员会（SEC）的评论信经常指出公司在经济业绩驱动因素方面的讨论不够充分（安永[2018]）。此外，国际会计准则理事会（IASB）目前正在修订管理层评论（类似于管理层讨论与分析 MD&A）的要求，因为公司有时“未能关注对实体前景重要的事情”等（IASB [2021]）。评估新的披露标准是否提高了背景信息的信息量是一个重大挑战。为了应对这一挑战，需要量化叙述性披露所增加的价值，这涉及测量非结构化、高度多维的文本数据和数字信息之间的非线性相互作用。我们所引入的方法使我们能够轻松应对这一问题。

先前的一系列研究表明，文本信息，特别是 MD&A，在公司的未来前景方面具有额外的信息量（例如，Tasker [1998]、Hutton, Miller, and Skinner [2003]、Baginski, Hassell, and Kimbrough [2004]、Li [2010]、Muslu 等[2015]、Mayew, Sethuraman, and Venkatachalam [2015]、Bochkay and Levine [2019]）。然而，在解释公司未来业绩方面，这种额外的信息量充其量只是中等（我们将这些结果称为直接信息价值）。除了直接信息价值外，叙述性披露还因其为数字提供了背景（与其相互作用）而具有价值。背景化在塑造对未来现金流预期方面的经济重要性可能微不足道，也可能非常巨大。先前的一些研究表明，文本属性（例如可读性）与数字披露之间存在统计上显著的相互作用（Li [2008]、Li, Lundholm, and Minnis [2013]）。然而，这种（浅层的）相互作用加上相对粗略的文本信息编码方式，导致解释能力的增量微不足道，这对叙述性背景在数字解释中的重要性造成了负面影响。

在量化背景在数字数据解释中的作用时，我们面临两个主要挑战。首先，我们需要系统地编码文本披露中包含的详细语义信息（特征）。其次，我们需要一个解释过程模型，在该模型中，叙述性和数字信息能够以丰富且高度非线性的方式相互作用。我们转向深度学习和语言建模来解决这些挑战。为了应对第一个挑战，我们使用来自 Transformer 的双向编码器表示（BERT）在高维空间中编码围绕收益披露的叙述性信息，通过人工神经网络（ANN）对解释过程进行建模。ANN 是一个数学函数，它依赖于一个由大量参数化基本单元（即“神经元”）构成的大型网络来模拟复杂关系。神经元在特定条件下被激活，并携带少量信息。ANN 的强大之处在于它能够学习神经元之间的深层相互作用，以产生有用的信号。类似的相互作用也发生在人类大脑的认知信息处理过程中，大脑会激活大量神经元，并让它们沿着突触以复杂的方式相互作用。

大多数投资者决策中的关键步骤是对公司未来前景形成信念。我们设计了一个人工神经网络（ANN），该网络能够在推断公司未来前景时，根据对会计数字（收益和现金流）的叙述性讨论来对其进行情境化处理。为了量化情境性或情境在解释数字数据中的重要性，我们的主要方法创新之处在于允许或限制叙述性信息（文本神经元）影响数字输入派生出的神经元的激活程度，反之亦然。换句话说，我们允许（或限制）文本输入来塑造对会计数字的解读。情境性与数字神经元之间相互作用的预测能力，决定了叙述性披露在解释会计数字中的价值，即情境性。

为了提取与披露收益相关的叙述性情境，我们创建了一个综合数据集，其中包含了 1995 年至 2021 年间提交的大多数可读取的机器版管理层讨论与分析（MD&A）报告。对于每份 MD&A，我们提取包含收益相关术语（如净收入、利润和收益）的句子。然后，我们提取这些句子每一侧的相邻句子，以更全面地捕捉收益的情境。我们将所有提取的句子按其在文档中出现的顺序汇总，并通过 BERT（Bidirectional Encoder Representations from Transformers，双向编码器表示）进行编码，转换为文本向量的形式。

我们首先分析了两个**基准模型**：（1）**数字输入模型**是一个仅使用会计变量（当前收益和经营现金流，按资产规模调整）以及每个变量的四个年度滞后值的 ANN；（2）**文本输入模型**是一个使用 BERT 向量编码叙述性情境作为输入的 ANN。数字和文本输入模型具有相似的架构。我们训练这些模型，通过利用数字或叙述性信息（或两者）来预测未来收益、现金流或股票价格变动的变化。特别是，我们使用  $t$  年的收益和现金流（及其滞后值），并将其与披露的情境相结合，以**预测  $t+1$  年的目标变量：未来收益、现金流或股票价格（回报）的方向性变化**。我们对模型进行了样本外评估。

我们的基准模型表明，当仅使用会计信息时，预测收益方向性变化的准确率约为 56.10%。而文本信息单独使用时，平均准确率略低，为 55.19%。为了评估叙述性情境所提供的额外信息，我们将文本和数字输入合并为一个**部分连接**的模型。在这个模型中，**情境信息不参与数字输入的处理（反之亦然），因此不影响对数字信息的解读**。这两种类型的输入通过各自的神经元层进行传播（如基准模型），并**直接映射到预测变量**。当我们把这个模型应用于数据时，我们发现**叙述性情境具有增量信息性，并且与之前的研究一致，其直接信息价值适中**。特别是，该模型在预测收益方向性变化方面的联合准确率提高到了 58.57%，相对于数字输入模型，准确率提高了 2.48 个百分点。其他目标变量的结果也类似。

相比之下，我们的**全连接模型允许通过文本对数字输入进行情境化的交互**。我们发现，**对数字的情境化处理显著提高了对未来前景的预测能力，无论是在统计上还是经济上**。在收益方向性变化的情况下，该模型的准确率达到 64.86%，相对于部分连接模型（58.57%）和数字输入模型（56.10%）的准确率，相对提高了 11%（16%）。

当我们预测经营现金流或股票收益的变化方向时，这些差异同样显著。例如，在 60 天收益的情况下，将情境信息与会计变量相结合，可以使准确性提高约 2.5 个百分点，而考虑情境交互则能进一步带来 6 个百分点的准确性提升。在所有情况下，与先前的研究一致，叙述性信息本身即包含有用信息。然而，**叙述性信息的主要价值并非来自其直接的增量信息性，而是来自与数字的交互，即允许情境化**。

我们的分析揭示了数字披露情境性的积极时间趋势。在解释年度收益时，这一趋势的增长最为显著：二十年间，其占比从大约 1% 增长到大约 10%。因此，叙述性情境在解释会计数字时扮演着越来越重要的角色。这些模式与 MD&A（管理层讨论与分析）讨论质量随时间提高的趋势相一致。

我们的证据表明，来自叙述性情境的交互有助于解释股票市场如何处理会计信息。为了进一步探究人类决策者在做决策时是否考虑情境因素，我们研究了情境是否有助于解释由 MD&A 披露引发的分析师预测修订。事实上，我们发现叙述性披露与数字披露之间的交互有助于解释人类决策者如何解读会计信息。

我们还研究了**在不同经济情境下，叙述性情境的重要性如何变化**。我们发现，当数字数据可靠性较低时，情境信息变得更为重要，例如当过去业绩由于高收益波动性或极端应计项目而无法预示未来结果时，或当估值相对具有挑战性时。在宏观经济不确定性时期，如 COVID-19 疫情期间或 2008 年金融危机期间，情境信息也更具价值，因为这些时期公司的文本披露对于解读报告业绩尤其有用。**结果表明，当经济条件使得数字数据更难解读时，叙述性情境的价值最大**。

最后，为了进一步证明我们发现的实践相关性，我们利用**情境交互的价值来量化收益持续性**——收益质量的主要指标（例如，Dichev 等人[2013]）。我们遵循 Farrell、Liang 和 Misra[2020,2021]中的非参数方法，根据伴随收益披露的叙述性计

论，构建了一个公司年度收益持续性指标。具体来说，我们将收益对其滞后值进行回归，并将斜率系数建模为一个深度神经网络，该网络将情境信息浓缩为一个单一数字。与我们先前的发现一致，**收益情境识别出了收益持续性中的巨大异质性，这有助于更准确地预测未来收益动态。**一种替代的、无情境的衡量公司层面持续性的方法依赖于滚动窗口回归，但这种方法被我们的基于情境的衡量方法所主导。

我们的研究在四个方面为文献做出了贡献。首先，我们引入了一种新方法来量化文本披露的情境性，即文本披露与伴随的叙述性情境交互后，数字披露预测价值的变化。文本和数字信息之间的交互对会计至关重要，因为财务报表中的数字不应孤立解读，而应通过管理层评论来丰富其内涵。我们利用类似于人类信息处理方式的现代技术，提供了一种新颖的方法来衡量这些互补性。我们的方法使人们能够研究情境信息更加相关的经济环境，还能比较不同披露制度（如一刀切的要求与自由裁量的情境披露）下情境的有用性。然而，这种方法的价值不仅限于文本披露，还可以很容易地扩展到量化其他沟通模式之间的互补性（即复杂交互）（见第 7 节讨论）。

其次，我们量化了数字披露与其叙述性情境之间交互的总体价值。先前的研究表明，MD&A（管理层讨论与分析）为数字数据提供了额外的信息（Bryan [1997]、Li [2010]、Mayew, Sethuraman, and Venkatachalam [2015]、Frankel, Jennings, and Lee [2016]、Bochkay and Levine [2019]），并且文本属性与会计数字之间存在交互（Li [2008]、Li, Lundholm, and Minnis [2013]）。然而，先前证据显示，文本属性和数字信息交互带来的解释（预测）能力提升在经济上并不显著。相比之下，我们的分析表明，这两种披露模式之间的交互在经济上具有重要影响，并且根据公司的经济环境而有所不同；它们还反映在投资者（和分析师）的决策中。此外，这些交互的价值始终且显著超过叙述性信息的直接（非交互）信息价值。这些结果通过显示情境披露的主要价值来源在于其与数字数据的交互，增强了我们对文本在决策制定中作用的理解。

第三，我们为会计和金融领域关于企业沟通文本分析的文献做出了贡献。先前的研究主要探索单一维度的特征，如情感（Loughran and McDonald [2011]）、可读性（Li [2008]、Bonsall et al. [2017]）、复杂性（Loughran and McDonald [2024]）和相似性（Brown and Tucker [2011]、Cohen, Malloy, and Nguyen [2020]）。在其最近的综述中，Bochkay 等人（[2023]）呼吁应用深度学习技术来进一步推动会计领域的文本分析。特别是，他们主张在需要从数据中学习未知的多维特征（文本特征工程）以及存在复杂非线性关系时使用深度学习。我们的研究完全符合这一描述，并展示了深度学习如何有助于回答以前无法解决的问题（即测量情境性和基于情境的持续性）。

最后但同样重要的是，我们引入了一种基于情境的收益持续性衡量指标。持续性是会计研究人员非常感兴趣的一个指标，它是收益质量的关键指标。与现有的持续性衡量指标不同，我们的衡量指标是针对公司年份的。它在样本外优于公司层面的无情境持续性衡量指标。我们的衡量指标扩展了一系列可以探索的研究问题。我们也是会计文献中第一个应用 Farrell、Liang 和 Misra（[2020, 2021]）的方法来学习公司层面参数异质性的研究。这种方法可以在许多其他环境中用于模拟公司之间的横截面异质性。

## 2 方法和研究设计

在本节中，我们首先讨论如何量化收益披露的周围环境信息，然后讨论一个解释过程模型，该模型允许文本为数字信息提供背景。

### 2.1 编码叙事信息

在过去的几十年里，文本分析已经从关键词搜索迅速发展 to 大型语言模型，如 BERT 和 GPT。在早期，给定文本（如段落）中信息的编码依赖于词典和词频统计（例如 Loughran 和 McDonald [2011]，Li、Lundholm 和 Minnis [2013]），主要涉及情感或不确定性等单一维度属性。在给定文本中捕捉多维信息的传统方法依赖于词袋表示法，即构成词频向量。类似地，tf-idf 表示法也依赖于词频，但会根据这些词在整个语料库中的常见程度进行调整。当测量文档间的相似性（Brown 和 Tucker [2011]，Hoberg 和 Maksimovic [2015]）或使用朴素贝叶斯分类器将文档分类到类别中时（Li [2010]），这些方法非常有用。

尽管基于频率的方法具有简洁性，但它们在捕捉和编码给定文本中包含的信息方面能力有限。其中一个原因是，为了克服“维度灾难”（Li [2010]），它们将所有单词视为彼此独立。因此，它们不考虑单词在句子中的使用方式以及不同单词之间的相互关系。为了理解信息的丢失，可以想象一下根据词袋频率重建文本的任务——这几乎是不可能完成的任务。

过去十年自然语言处理领域的一个关键发展是出现了能够“学习”以密集向量形式表示单词意义的机器算法（例如 Bochkay 等人 [2023]）。这些算法包括 Word2Vec 模型（Mikolov 等人 [2013]）或 GloVe 模型等。例如，Word2Vec 使用在大规模语料库上训练的神经网络（ANN）来学习单词向量（嵌入）。这些嵌入表现出一些吸引人的特性。一个典型的例子是，在对相应单词嵌入进行数学运算时，“king” - “man” + “woman” 大约等于 “queen”（Mikolov 等人 [2013]）。然而，这些模型还不够复杂，无法捕捉单词意义如何根据文本段落中使用的其他单词而变化，因此它们无法完全捕捉代表叙事信息的相关维度。

**BERT**——被视为自然语言处理领域的突破性进展，并且在概念上与 ChatGPT 使用的 GPT 模型相似（都依赖于 Transformer 架构）——与之前的模型不同，它根据单词使用的上下文来学习单词的语义意义（Devlin 等人 [2019]）。与之前的语言模型不同，BERT 使用自注意力机制来编码单词的意义，这取决于这些单词在何处以及如何使用（Devlin 等人 [2019]）。**BERT 在大型英语维基百科文本语料库上进行预训练，可以处理长达 512 个标记的相对较长的文本段落。**该模型输出的单词向量会根据句子中其他单词的存在而变化。此外，**BERT 还编码了整个输入序列的复合意义**，该意义由与分类（CLS）标记关联的隐藏状态（向量）捕获。CLS 标记被插入到每个输入序列（句子或句子序列）的开头。尽管它不可直接解释，但它编码了整个序列中的信息。

BERT 被广泛应用于机器翻译、文本分类和文本生成等领域。在会计文献中，BERT 最近被用于 Siano 和 Wysocki [2021] 以及 Huang、Wang 和 Yang [2023] 中的情感分析，以及 Lee 和 Zhong [2022] 中的主题分类。

BERT 能够编码讨论经营业绩的句子中传达的多维信息，这使得该模型特别适合我们的研究。与上述研究相比（并且符合我们的目标），我们提取了整个 768 维向量，该向量总结了给定文档中的信息（即分类标记的最后一个隐藏状态向量）。

## 2.2 数值信息情境化解读的建模

我们始终坚持一个假设，即关于公司报告业绩的叙述性讨论中包含了一些单独的定性信息，这些信息本身往往并不重要，但联合起来却能形成一个“马赛克”图案。基于这一理念，我们的设计使得情境信息与报告中的数字能够以揭示有用信息的方式相互作用。人工神经网络（ANNs）非常适合于这项任务，因为它们能够以复杂且高度非线性的方式使高维文本信息与数字信息实现深度交互（Bengio, Ducharme, and Vincent [2000]）。

为了捕捉情境性，即叙事语境对于数字披露解读的重要性，我们使文本输入能够影响（或不能影响）源自数字输入的神经元的激活，反之亦然。这一点在后边的图表 4 和图表 5 中得到了说明。特别是，人工神经网络输入层中的信息会向目标变量传播，经过多个隐藏层，每一层都由多个信息单元（神经元）表示。一个神经元会处理来自前一层的信息，并在满足特定条件下被激活。在这种情况下，它会将信号向前传递到下一层。这个过程简化了人类大脑中信息处理的方式，即大脑依赖于通过神经网络相互连接的多个生物神经元层。每个单独的神经元行为相对简单，但组合起来后，神经元能够执行高度复杂的任务。

我们设计该模型，使其能够启用或限制叙事语境对数字输入进行情境化（即影响源自数字输入的神经元的激活）的能力，从而使文本信息直接负责对数字信息的解读。

## 3 数据与方法论

我们首先收集了 1995 年至 2021 年间所有可机器读取的管理层讨论与分析（MD&A）报告。在我们样本期间，美国证券交易委员会（SEC）提供了 165,921 份文件目录，我们成功提取了 138,129 份 MD&A 报告（约占 83%）。由于我们的研究聚焦于收益情境，因此我们采用了 Kothari, Li, and Short [2009] 中的关键词列表来识别包含收益相关词汇的句子。此外，为了更全面地捕捉相关情境，我们还提取了所识别句子之前和之后的一个句子。我们将这些提取的句子称为收益相关句子。平均而言，一份 MD&A 部分包含 305 个句子，其中 13.94%（即 43 个句子）与收益相关。我们还尝试使用不同数量的相邻句子作为披露收益的情境，范围从零到整个 MD&A。实际上，扩大情境窗口涉及权衡：一方面，它增加了可能有助于解释收益的相关信息量；另一方面，收益讨论附近的直接叙述信息可能具有更高的信噪比。从实证角度看，我们发现使用一个相邻句子是最优的，尽管不同的情境窗口在定性上得出了相似的结果。

我们使用 Compustat 数据检索财务信息，并通过中央索引键（CIK）将其与 MD&A 报告进行匹配。图表 2 概括了我们的样本选择过程。我们的匹配样本中有 87,201 个观测值包含非缺失的收益和现金流量数据，62,287 个观测值包含股票收益数据。图表 3 报告了我们主要分析中使用的变量的描述性统计。

我们对提取的收益相关句子应用了几个过滤器，使其可机器读取。我们删除了表格、图形、特殊字符和 html 语法。我们将句子中的数字替换为掩码标记，使模型能够学习数字信息的位置。与删除数字的方法（例如，Siano and Wysocki [2021]）相比，这种方法更受欢迎。

图表 2 样本选择过程

TABLE 1  
Sample Construction

Description	Number of Observations
Observations with valid earnings-related information in MD&A	138,129
Less:	
Texts that are too long to process with BERT	(6,750)
Insufficient financial information (including 4 lagged terms)	(44,179)
Total (for tables 3 and 4)	87,201
Less:	
Insufficient return data	(24,914)
Total (for tables 5 and 6)	62,287
Test set samples used in tables 3 and 4 (from 1999 to 2020)	77,609
Test set samples used in tables 5 and 6 (from 1999 to 2020)	52,321

资料来源：《Context-Based Interpretation of Financial Information》，华安证券研究所

注：表格总结了我们的样本的构建情况。样本期是从 1995 年到 2020 年。我们将样本观测值限制为在其年报的 MD&A 部分包含有效收益相关文本信息的公司。我们排除了对于自然语言处理模型而言过于冗长（超过 10,240 个词元）而无法处理的观测值。对于所有模型，我们要求观测值包含当期经营现金流、当期收益以及它们过去最多四年的滞后值。此外，对于图表 12 和图表 13 中的收益预测测试，我们需要与 CRSP 有有效链接且收益观测值非缺失。我们的测试结果基于 1999 年至 2020 年的数据。

图表 3 分析中使用的变量的描述性统计

TABLE 2  
Descriptive Statistics

Variable	N	Mean	Median	Std. Dev.
Positive_Return	62,287	0.5075	1.0000	0.4994
Positive_Earnings	87,201	0.5582	1.0000	0.4966
Positive_OCF	87,201	0.5497	1.0000	0.4975
Returns	62,287	0.1427	0.0056	0.8455
Net_Income	87,201	-0.0311	0.0263	1.0939
OCF	87,201	0.0316	0.0659	1.0622
n(Sentences)	87,201	42.5200	42.3671	42.2018
n(Words)	87,201	4,790.66	4,198.00	3,619.11

资料来源：《Context-Based Interpretation of Financial Information》，华安证券研究所

注：该表格总结了我们的样本的构建情况。样本期是从 1995 年到 2020 年。我们将样本观测值限制为在其年报的 MD&A 部分包含有效收益相关文本信息的公司。我们排除了对于自然语言处理模型而言过于冗长（超过 10,240 个词元）而无法处理的观测值。对于所有模型，我们要求观测值包含当期经营现金流、当期收益以及它们过去最多四年的滞后值。此外，对于图表 12 和图表 13 中的收益预测测试，我们需要与 CRSP 有有效链接且收益观测值非缺失。我们的测试结果基于 1999 年至 2020 年的数据。

然后，我们使用 BERT 分词器将每个 MD&A 中收集的句子通过 BERT 基础无大小写模型进行信息编码。由于单个 MD&A 中的收益相关句子的总长度通常超过 510 个词元（平均 4,790 个单词），因此我们将这些提取的句子分成几个段落。

CLS 标记的最后一个隐藏状态向量总结了输入段落的上下文信息，并且通常用于分类任务。我们通过计算多个段落向量的平均值来聚合这些向量，从而获得

MD&A 级别的上下文向量。我们使用五台 Tesla V100 32GB GPU 来处理我们样本中所有收益相关句子的整个集合。

### 3.1 模型设计

我们将解释过程建模为一个复杂的非线性函数，该函数在形成对公司未来前景的预期时，学会根据叙述性披露为会计数字提供上下文背景。我们将目标变量  $T_{t+1}$ （未来收益、现金流和回报的方向性变化）建模为会计信息向量  $A_t$  和上下文信息向量  $C_t$  在时间  $t$  时的函数  $\Lambda_t(A_t, C_t)$  的预期值  $E_t[\cdot]$ ，如公式 (1) 所示：

$$E_t(T_{t+1} | A_t, C_t) = \Lambda_t(A_t, C_t) \quad (1)$$

其中， $A_t$  是会计信息向量， $C_t$  是上下文信息向量，两者均在时间  $t$  时可得； $E_t$  是预期算子； $\Lambda_t(\cdot)$  是参数依赖于时间  $t$  的函数。

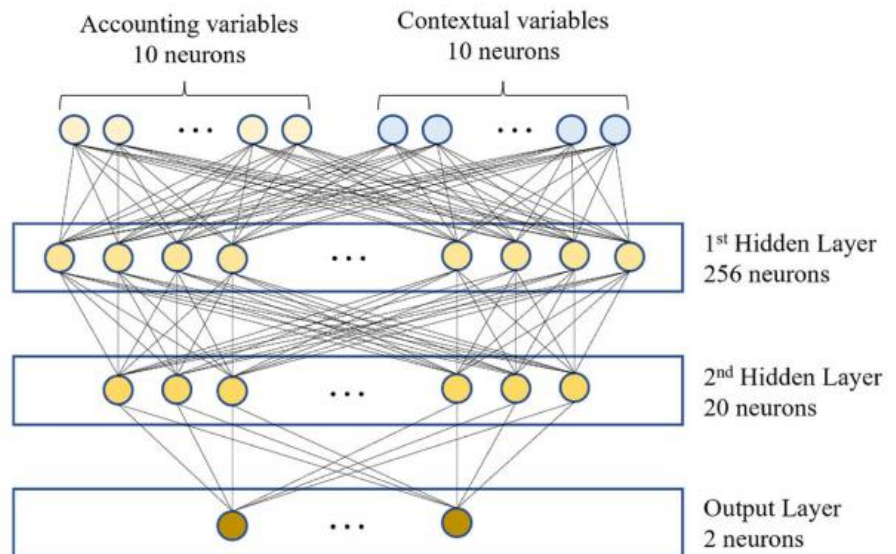
如前所述，基于上下文的财务信息解释假设由向量  $A_t$  和  $C_t$  总结的信息位之间存在高度非线性交互。这些交互是  $\Lambda_t(\cdot)$  的关键部分，并通过来自  $A_t$  和  $C_t$  输入的神经元之间的深层网络连接进行建模。允许这些交互使得上下文能够影响财务数字的解释。

我们使用当前经营现金流、当前收益及其四年年度滞后值来构建  $A_t$ （因此， $\dim(A_t)=10$ ）。对于  $C_t$ ，我们使用之前描述的平均 CLS 向量。由于上下文向量的原始维度为 768，远大于  $A_t$  的维度，我们在模型中增加了一层，学习将上下文向量减少到 10 维（我们选择 10 维以与数值输入的对称性相匹配）。

为了测试上下文在解释数字信息中的作用，我们比较了全连接和部分连接的人工神经网络（ANN）模型的性能，这将在下面讨论。我们的全连接模型如图表 4 所示。它有一个输入层、两个隐藏层和一个输出层：第一个和第二个隐藏层分别包含 256 和 20 个神经元，输出层有两个神经元（增加或减少）。我们允许所有神经元之间的完全交互，并为填充两个隐藏层的神经元使用修正线性单元（ReLU）激活函数。该模型输出目标变量（例如，收益）在后续时期减少或增加的对数概率。

图表 4 全连接模型

(a) Fully Connected Neural Network Model

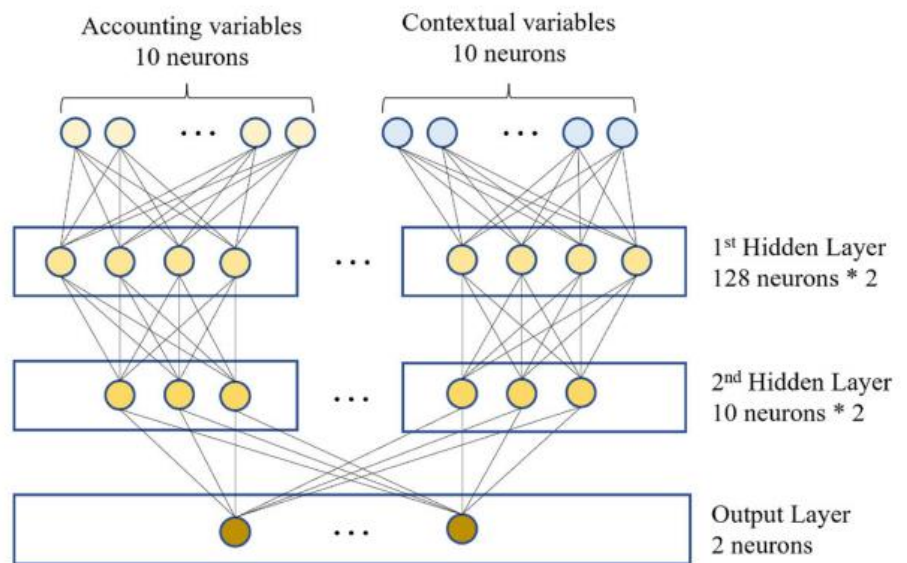


资料来源：《Context-Based Interpretation of Financial Information》，华安证券研究所

为了评估  $A_t$  和  $C_t$  之间交互的作用，我们还构建了一个**部分连接**的模型。该模型如图表 5 所示。在此模型中，各层的结构与全连接模型相同。然而，该网络没有  $A_t$  和  $C_t$  之间的交互。这意味着  $C_t$  不参与来自  $A_t$  的神经元的激活，反之亦然。因此，在第一个隐藏层中，128 个神经元分配给  $A_t$ ，其余 128 个神经元分配给  $C_t$ 。类似地，第二个隐藏层中的 10 个神经元分配给  $A_t$ ，其余分配给  $C_t$ 。第二个隐藏层有 20 个神经元连接到输出层的两个神经元。如图所示，这两个模型之间的唯一区别是来自  $A_t$  和  $C_t$  的神经元之间是否存在连接。

图表 5 部分连接模型

(b) Partially Connected Neural Network Model



资料来源：《Context-Based Interpretation of Financial Information》，华安证券研究所

全连接模型和稀疏连接模型的准确性之间的差异被称为会计信息的上下文性。高上下文性意味着当与数字信息结合时，叙述性上下文增加了更大的价值。为了评估模型的准确性，我们计算正确分类观测值的百分比。当模型的预测显示增加的可能性大于减少时，该观测值被分类为“增加”。然后，我们比较真实和预测的目标变量结果来评估准确性。

### 3.2 模型训练与超参数选择

我们训练模型以基于  $A_{t-1}$  和  $C_{t-1}$  来预测时间  $t$  的目标  $T_t$ 。我们使用前四年作为滚动训练窗口，并使用估计的模型参数创建样本外预测  $\hat{\Lambda}_{t+1|t} = \Lambda_{t-1}(A_t, C_t)$ 。我们从 1999 年到 2020 年每年重复此过程，并设定四个目标变量（收益、现金流、短期 60 天回报和长期 360 天回报）。

训练有效的机器学习模型涉及一个关键步骤，即超参数选择（Goodfellow、Bengio 和 Courville [2016]，Géron [2022]）。超参数是在训练之前选择的，用于指导学习过程。对于每个目标变量，我们执行网格搜索以优化多个超参数并最大化全连接模型的性能。具体来说，我们尝试了三种批量大小（0、64 和 128），以确定传递给优化器以计算梯度的数据量；三种丢弃率（0、0.2 和 0.4），以控制正则化

的程度；以及三种学习率（ $e-1$ 、 $e-2$  和  $e-4$ ），以控制优化过程的速度。对于我们四个目标变量中的每一个，我们估计了 27 种不同的超参数组合（总共 108 个模型），并为每个任务选择了产生全连接模型最高平均准确率的超参数集。对于收益、现金流和长期回报预测，我们的程序选择了 64 个观测值的批量大小、20%的丢弃率和  $e-4$  的学习率。对于短期回报预测，我们选择了  $e-1$  的学习率、20%的丢弃率和 64 的批量大小。我们的超参数选择过程的详细信息见在线附录表 OA-1。

对于所有实验，我们使用 Adam 优化器，并结合早停策略，基于自熵损失函数训练模型。我们从训练集中随机分配 20%的观测值作为验证集，以评估每个训练周期后损失函数的改进情况。如果在第  $n$  个训练周期后，连续五个训练周期内损失函数没有改进，我们将使用第  $n$  个训练周期的估计参数。我们的训练程序每年  $t$  都会重置并重新初始化参数的起始值。由于神经网络模型依赖于随机初始化，相同的训练和测试集可能会产生略有不同的输出。为了提高结果的泛化能力，并减少随机初始化带来的噪声（这可能导致找到局部最小值），对于每个训练集，我们重复估计 30 次。然后，我们报告平均准确率。

## 4 模型结果

### 4.1 利用叙事背景预测收益

我们首先研究叙事背景在预测未来收益方面的重要性。该模型预测了按总资产计算的收益增减。图表 6 展示了模型按年份以及在整个样本中的表现。第 1 列报告的是数值输入模型（仅基于会计信息），而第 2 列展示的是情境输入模型（仅依赖情境信息）。这两个模型的结构和训练计划相同。第 1 列显示，**基于会计输入的模型平均准确率为 56.10%**，而第 2 列显示，**基于情境输入的模型准确率略低，为 55.19%**。

接下来的两列结合了会计和情境输入。第 3 列和第 4 列分别报告了部分连接模型和完全连接模型的准确率。请注意，**部分连接模型排除了情境输入与数值输入之间的相互作用，从而可以评估文本数据的增量信息内容。平均而言，我们观察到部分连接模型的准确率为 58.57%，比数值输入模型高出约 2.5 个百分点。这一适度提升与 Bochkay 和 Levine [2019]所记录的文本带来的准确率提升平均相符。相比之下，完全连接模型的准确率为 64.86%，比部分连接模型高出 6.29 个百分点（或约 10.7%的增长）。准确率的这一提升在统计上和经济上都具有显著性（ $t$  值为 4.66）。**此外，在我们样本中的每一年，完全连接模型的准确率都超过了其他三个模型，情境输入对准确率的提升幅度在 4.43%至 8.72%之间。为了更直观地理解这些数据，需要注意的是，相对于数值输入基准（平均准确率为 56.1%），将文本输入纳入模型后，模型性能提升了 2.47 个百分点（准确率为 58.57%）。当考虑到数值和文本数据之间的相互作用时，这一提升几乎翻了两番，达到了 64.86%的准确率。**这一结果表明，与无情境的数字相比，情境化的数字在塑造未来信念方面具有更高的价值。换句话说，数值与其相关的叙事背景之间的相互作用具有相当大的经济重要性。**

图表 6 模型按年份以及在整体样本中的表现

**TABLE 3**  
*Contextual Information and Earnings Prediction*

Year	Accounting Info Only (1)	Contextual Info Only (2)	Partially Connected (3)	Fully Connected (4)	Contextuality (4) - (3)
1999	55.94%	53.19%	58.15%	62.89%	4.74%***
2000	54.80%	45.71%	55.21%	61.12%	5.91%***
2001	54.90%	53.44%	59.27%	64.43%	5.16%***
2002	57.38%	51.19%	60.23%	65.29%	5.06%***
2003	59.60%	54.47%	61.36%	67.05%	5.69%***
2004	58.75%	59.71%	62.12%	66.55%	4.43%***
2005	59.32%	58.52%	61.16%	67.89%	6.73%***
2006	54.71%	55.40%	58.15%	64.38%	6.23%***
2007	42.90%	42.64%	46.42%	53.56%	7.14%***
2008	53.49%	53.06%	54.73%	62.36%	7.63%***
2009	58.90%	59.06%	59.23%	66.78%	7.54%***
2010	57.16%	56.97%	58.86%	64.74%	5.88%***
2011	55.62%	56.15%	58.23%	63.19%	4.97%***
2012	58.55%	58.21%	59.85%	66.37%	6.53%***
2013	58.95%	57.94%	61.05%	67.26%	6.21%***
2014	55.57%	53.93%	55.34%	62.62%	7.28%***
2015	58.16%	57.51%	59.79%	66.60%	6.82%***
2016	55.64%	58.15%	59.88%	65.33%	5.45%***
2017	58.61%	60.31%	63.72%	68.44%	4.72%***
2018	55.73%	54.59%	57.03%	65.75%	8.72%***
2019	44.28%	45.60%	48.31%	56.58%	8.27%***
2020	65.24%	68.55%	70.35%	77.60%	7.25%***
Average	56.10%	55.19%	58.57%	64.86%	6.29%***
AUC	0.5190	0.5096	0.5023	0.5889	0.0866*** (4.66)
Macro F1	0.4812	0.4759	0.4733	0.6126	0.1393*** (7.32)
Trend	0.0003 (0.15)	0.0032 (1.48)	0.0008 (0.39)	0.0017 (0.93)	0.0010** (11.52) (2.74)

资料来源：《Context-Based Interpretation of Financial Information》，华安证券研究所

注：该表格报告了四个人工神经网络（ANN）模型预测未来收益方向变化的准确性。这些模型预测的是随后财年中收益变化的方向（增加或减少）。我们以正确预测的比例来衡量预测准确性。模型（1）仅依赖会计信息（当前收益、当前经营现金流及其最多四年的滞后项）。模型（2）仅依赖情境信息。模型（3）和（4）在其初始层中同时使用会计和情境信息。模型（3）是部分连接模型，不允许会计和情境信息之间的相互作用，而模型（4）是完全连接模型，允许这种相互作用。情境性，即情境信息在解释数字信息中的重要性，是通过模型（4）和（3）之间预测准确性的差异来衡量的。每个模型都估计了 30 次，我们报告的是预测准确性的平均值。为了检验统计显著性，我们首先计算所有年份的合并标准差，并检验模型（3）和（4）之间方差的等同性。然后，我们报告了针对合并平均值方差差异进行调整后的 t 值。AUC 表示受试者工作特征曲线（真阳性率和假阳性率）下的面积，Macro-F1 是精确率和召回率的逆调和平均值。趋势表示预测准确性对线性时间趋势进行时间序列回归的斜率系数。我们报告了经自相关调整后的 t 统计量，以检验时间趋势的显著性。括号内报告的是 t 值。\*、\*\*和\*\*\*分别表示在 10%、5%和 1%水平上的统计显著性。

除了预测准确性之外，我们还使用另外两个性能指标来确认我们的结果：(1)受试者工作特征（ROC）曲线下的面积（AUC）和(2)宏观 F1 分数。这两个指标通常

用于评估二分类器，并且是 I 类（假阳性）和 II 类（假阴性）错误的函数。ROC 曲线显示了每个可接受的假阳性率下的真正阳性率。一个完美的模型在每个阈值下都将正例排在负例之前，因此其 AUC 为 1，而一个随机分类器无法区分真阳性和假阳性，其 AUC 为 0.5。F1 分数是精确率（预测为正例中有多少是正确的）和召回率（真正为正例的有多少）的逆调和平均值，当类别不平衡时表现良好。因此，这两个指标在评估分类器性能时相互补充。这两个指标都表明，全连接模型显著优于部分连接模型，这意味着数值数据具有相当大的情境性。后者与两个基准模型（第 1 列和第 2 列）的表现相当，这证实了叙事语境中的直接增量信息有限。相比之下，全连接模型的 ROC AUC 为 0.5889，比部分连接模型的 ROC AUC（0.5023）提高了 8.66%。F1 分数也得出了类似的结论，证实了全连接模型在 I 类和 II 类错误方面都表现良好。

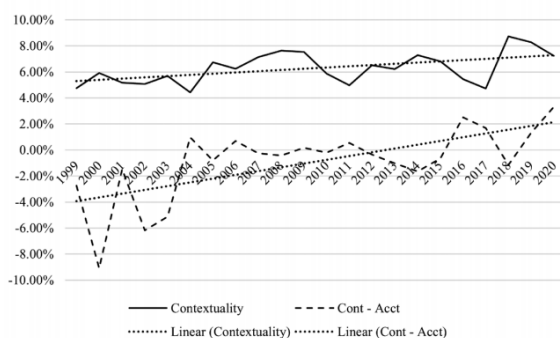
接下来，我们转向时间趋势的分析。我们关注同一张表格中报告的年度准确性。我们观察到，在样本期间，最后一列报告的情境性平均呈上升趋势。在样本早期，它围绕 5 波动，到样本结束时达到 8，增长了 60%。为了更正式地分析时间趋势，我们进行了以下回归：

$$Dep\_Var_t = \lambda_0 + \lambda_1 Time_t + \epsilon_t$$

并在图表 6 的最后一行报告了  $\lambda_1$  的估计值及其时间序列相关调整后的 t 值。第 1、3 和 4 列没有显示出统计上显著的时间趋势。然而，在我们样本期间，会计信息的情境性稳步增长。表格的最后一行表明，模型（4）和（3）之间的差异（报告在最后一列）呈现出正向且统计上显著的时间趋势（ $\lambda_1 = 0.0010$ ，t 值为 2.74）。这一结果表明，情境性每年增长 0.1 个百分点，在过去 21 年（1999 年至 2020 年）中总计增长了 2.1 个百分点。考虑到 1999 年的情境性为 4.74%，2.1 个百分点的增长在我们样本期间代表了近 50% 的增长。图表 7 证实了情境性的正向时间趋势，并进一步表明，叙述性披露的直接信息含量（模型（2）和（1）之间的差异）也呈现出正向趋势。这两个属性的增长幅度相当。

图表 7 情境信息在收益预测中的相关性：时间趋势

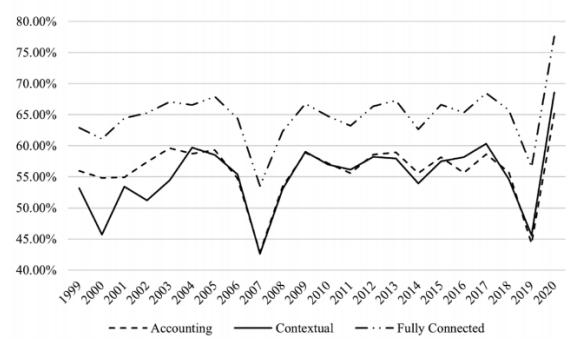
(a) Time Trend in Contextuality



资料来源：《Context-Based Interpretation of Financial Information》，华安证券研究所

图表 8 情境信息在收益预测中的相关性：预测准确性

(b) Time Trend in the Predictive Accuracy of Accounting vs. Contextual Information



资料来源：《Context-Based Interpretation of Financial Information》，华安证券研究所

综上所述，图表 6 表明，会计信息与其情境之间的相互作用有助于形成对未来收益更为准确的预期，而叙述性披露通过与数字数据的相互作用促进了会计数字的解读。我们的研究结果还表明，随着时间的推移，情境性的重要性日益增加，这与管理层讨论与分析（MD&A）披露变得更具信息量的趋势相一致。

## 4.2 利用叙述性语境预测现金流

接下来，我们使用模型来预测未来的经营性现金流。由于MD&A讨论往往侧重于收益，因此尚不清楚情境信息是否有助于预测现金流。然而，现金流对于许多经济决策而言至关重要，会计领域有大量文献专注于现金流的预测（有关该领域文献的讨论，参见Nallareddy、Sethuraman和Venkatachalam [2020]）。与之前一样，我们关注经营性现金流的变动方向（增加或减少）。

图表9展示了我们的分析结果。仅基于会计信息，模型的准确率为57.77%（第1列），高于仅依赖叙述性语境的模型所达到的54.97%的准确率（第2列）。当我们在部分连接模型中结合这两种输入（第3列）时，模型的性能为58.47%。与第1列相比，预测性的提升微乎其微（不足1个百分点）。这些结果符合我们的直觉，即在预测未来现金流时，与预测收益不同，叙述性披露的信息含量（直接）较低。然而，当我们在全连接模型中允许情境输入与会计数字相互作用（第4列）时，模型的性能相对于其他三个模型有了显著提升，达到了64.74%的准确率，这是现金流预测问题的最先进水平。与部分连接模型（58.47%的准确率）相比，准确率提高了6.27个百分点，即提升了10%。重要的是，根据表格底部报告的AUC和F1分数指标，全连接模型的优越性甚至更加明显。全连接模型和部分连接模型之间的差异在经济上和统计上都具有显著性（AUC为0.621 vs. 0.519，宏观F1分数为0.684 vs. 0.524），证实了数字和情境信息之间相互作用的统计和经济重要性。

图表9 使用模型来预测未来的经营性现金流的表现

**TABLE 4**  
*Contextual Information and Cash Flow Forecasts*

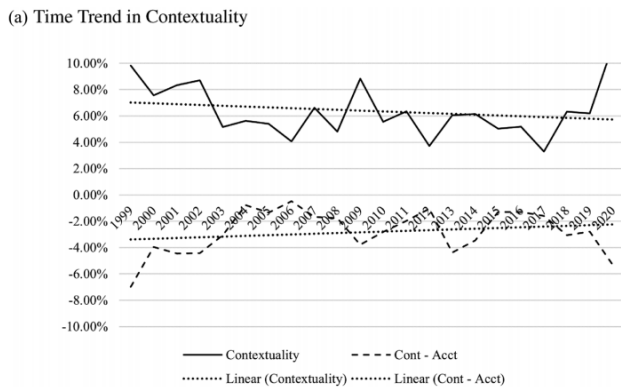
Year	Accounting Info Only (1)	Contextual Info Only (2)	Partially Connected (3)	Fully Connected (4)	Contextuality (4) - (3)
1999	58.07%	51.09%	55.01%	64.84%	9.83%***
2000	60.76%	56.81%	60.59%	68.15%	7.56%***
2001	60.91%	56.47%	59.60%	67.93%	8.33%***
2002	58.40%	53.97%	56.65%	65.37%	8.71%***
2003	58.75%	55.69%	60.14%	65.30%	5.16%***
2004	57.40%	56.62%	58.70%	64.32%	5.63%***
2005	58.31%	56.97%	59.84%	65.24%	5.40%***
2006	58.90%	58.43%	62.60%	66.65%	4.05%***
2007	53.54%	51.87%	54.59%	61.20%	6.61%***
2008	56.89%	55.13%	58.98%	63.80%	4.82%***
2009	57.43%	53.65%	56.53%	65.37%	8.83%***
2010	57.98%	55.18%	58.34%	63.91%	5.56%***
2011	56.45%	54.48%	57.91%	64.24%	6.34%***
2012	59.27%	58.21%	61.69%	64.42%	3.73%***
2013	55.60%	51.22%	55.16%	61.22%	6.06%***
2014	58.45%	54.99%	59.25%	65.41%	6.16%***
2015	59.39%	58.08%	61.69%	66.72%	5.03%***
2016	56.82%	55.52%	58.50%	63.67%	5.17%***
2017	60.11%	58.53%	63.31%	66.61%	3.30%***
2018	56.54%	53.49%	57.22%	63.55%	6.32%***
2019	57.21%	54.43%	57.45%	63.64%	6.19%***
2020	53.85%	48.50%	52.52%	62.90%	11.39%***
Average	57.77%	54.97%	58.47%	64.74%	6.27%***
AUC	0.5158	0.5093	0.5193	0.6213	0.1020***
Macro F1	0.5089	0.4888	0.5239	0.6836	0.1597***
Trend	-0.0011* (-1.89)	-0.0006 (-0.55)	-0.0004 (-0.34)	-0.0010* (-1.89)	-0.0006 (-0.69)

资料来源：《Context-Based Interpretation of Financial Information》，华安证券研究所

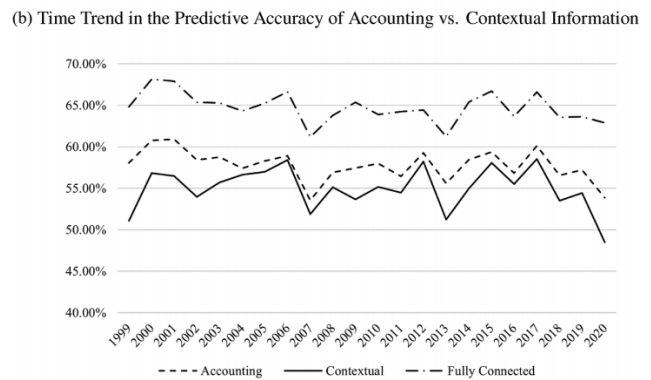
在时间趋势的分析中，我们观察到未来现金流的可预测性随时间略有下降，这由负向时间趋势所指示。第 1 列和第 4 列中的趋势在统计上显著，且似乎是由会计变量的预测能力下降所驱动。我们在第 5 列中发现的时间趋势在统计上不显著，这表明在预测现金流的情况下，文本与数字的相互作用随时间变化不显著。

综上所述，本分析证实了情境化数字数据在形成未来现金流预期时的巨大价值，从而进一步强调了情境信息对于解释数字输入的重要性。

图表 10 情境信息与未来经营性现金流：时间趋势



图表 11 情境信息与未来经营性现金流：预测准确性



资料来源：《Context-Based Interpretation of Financial Information》，华安证券研究所

资料来源：《Context-Based Interpretation of Financial Information》，华安证券研究所

### 4.3 叙述性情境与股票回报

证明情境化在预测未来收益或现金流方面很重要，并不一定意味着投资者认识到了这一点，即在解读会计数字时融入了情境。为了解决这个问题，我们研究了叙述性情境在解释股票价格变动方面的相关性，股票价格变动反映了投资者如何将收益情境纳入定价决策中。

我们采用了与之前相同的方法，但目标变量现在是累积股票回报，以捕捉信息处理情况。更具体地说，我们关注两种回报结果：(1) 在收益公告的财年（即收益赚取之后的年份）内衡量的年度买入并持有回报；(2) 从 10-K 报告提交日期起算的 60 天回报。长期窗口衡量指标旨在纳入收益公告日左右股价中释放的收益惊喜（包括任何信息泄露），并允许情境信息缓慢纳入股价（Cohen, Malloy, 和 Nguyen [2020]）。相比之下，60 天回报窗口旨在排除实际的收益惊喜，而是专注于投资者处理管理层讨论与分析（MD&A）信息的最活跃时期。在这两种情况下，模型都是用来预测回报是正还是负。

图表 12 展示了长期（年度）回报的结果。与之前一样，第 1 列和第 2 列分别关注单一输入（会计数字和情境信息），而第 3 列和第 4 列则分别列出了部分连接模型和全连接模型的结果。长期回报比收益更难预测。基于会计信息的模型平均准确率为 50.31%，与仅基于情境信息的模型平均准确率相似。当两种输入在部分连接模型中结合时（第 3 列），准确率提高了约 5.4 个百分点（准确率为 55.71%），这一点值得注意。然而，当我们允许数字输入情境化时，这一准确率进一步提高。实际上，第 4 列显示，全连接模型的准确率为 60.26%。与部分连接模型相比，这一改进幅度为 4.55 个百分点，相当于准确率提高了 8%。在评估 F1 分数和 AUC 指标时，我们也观察到了同样显著的结果。

图表 12 长期（年度）回报的结果

**TABLE 5**  
*Contextual Information and Annual Returns Forecasts*

Year	Accounting Info Only (1)	Contextual Info Only (2)	Partially Connected (3)	Fully Connected (4)	Contextuality (4) - (3)
1999	59.79%	60.25%	63.85%	65.24%	1.38%*
2000	52.34%	52.63%	58.30%	59.01%	0.72%
2001	68.82%	68.43%	72.97%	74.75%	1.78%**
2002	20.44%	20.87%	26.09%	27.76%	1.67%*
2003	58.05%	54.16%	62.14%	64.01%	1.87%**
2004	55.41%	49.74%	55.92%	60.35%	4.42%***
2005	57.50%	56.87%	62.01%	65.06%	3.05%***
2006	43.55%	41.14%	45.95%	50.72%	4.76%***
2007	40.33%	39.58%	46.92%	49.70%	2.78%**
2008	35.98%	35.80%	41.84%	46.97%	5.13%***
2009	32.61%	31.92%	36.99%	44.54%	7.55%***
2010	59.65%	59.10%	65.67%	72.28%	6.61%***
2011	44.77%	42.20%	50.91%	52.81%	1.90%*
2012	72.42%	74.92%	80.68%	84.26%	3.58%***
2013	53.57%	54.58%	60.98%	63.43%	2.45%**
2014	43.36%	43.43%	46.17%	53.24%	7.07%***
2015	62.73%	62.04%	68.36%	73.54%	5.19%***
2016	53.37%	54.55%	58.62%	65.94%	7.33%***
2017	39.63%	37.32%	48.80%	52.09%	3.28%***
2018	48.27%	53.89%	54.39%	63.86%	9.47%***
2019	44.69%	49.74%	54.70%	60.30%	5.60%***
2020	59.45%	59.05%	63.46%	75.93%	12.47%***
Average	50.31%	50.10%	55.71%	60.26%	4.55%*** (5.30)
AUC	0.5099	0.5121	0.5335	0.5982	0.0647*** (6.35)
Macro F1	0.4682	0.4893	0.5223	0.5833	0.0610*** (5.89)
Trend	-0.0002 (-0.05)	0.0014 (0.35)	0.0013 (0.33)	0.0045 (1.1)	0.0032*** (4.22)

资料来源：《Context-Based Interpretation of Financial Information》，华安证券研究所

在分析时间趋势时，我们的发现如下。尽管第 1 至 4 列并未展现出统计上显著的时间趋势，就像盈利预测（图表 6）的情况一样，但我们观察到了情境性（ $\lambda_1$  为 0.0032，t 值为 4.22）随时间增加的明确证据。这也可以从图 4 中看出。该图还揭示了叙述背景（未与其他因素相互作用）相对于报告的收益和现金流数字所传达的信息量呈现出的正趋势，即第 1 列和第 2 列之间的差异。

图表 13 展示了基于 60 天回报窗口的结果。由于管理层讨论与分析（MD&A）提交前会通过新闻稿公布收益，因此 60 天回报窗口通常排除了公告日的收益惊喜。相反，它旨在捕捉投资者在处理 MD&A 细节时的过程，因为这些信息会反映在价格中。第 1 列表明，仅会计信息就能达到 57.03% 的预测准确性。这一较高的解释能力是直观的，因为在此期间披露的信息会持续流入股票价格。在第 2 列和第 3 列中，这一数字略有增加，我们报告的预测准确性与其他使用文本输入的近期回报预测模型相似（Jiang, Kelly, and Xiu [2022]）。然而，当模型允许情境和数字输入之间的相互作用时，准确性大幅提升至 64.80%，即提高了 11%。在 ROCAUC 和 F1 分数

这两个替代性能指标中也观察到了类似的效果。这些证据证实，收益披露的情境对股票市场处理会计信息的方式具有统计上显著且经济上显著的影响。

图表 13 60 日回报的结果

**TABLE 6**  
*Contextual Information and 60-Day Stock Returns*

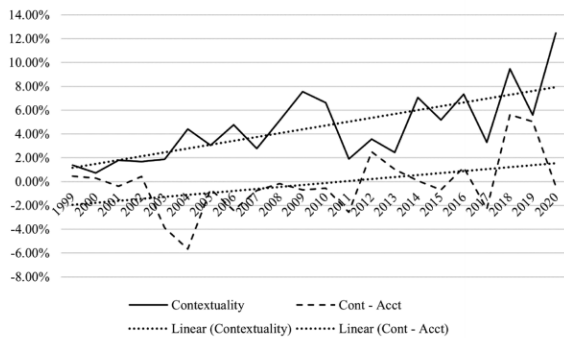
Year	Accounting Info Only (1)	Contextual Info Only (2)	Partially Connected (3)	Fully Connected (4)	Contextuality (4) - (3)
1999	53.23%	50.59%	52.75%	59.97%	7.23%***
2000	57.34%	52.49%	56.94%	62.00%	5.06%***
2001	63.80%	57.30%	61.87%	67.86%	5.98%***
2002	63.40%	68.14%	70.97%	77.64%	6.67%***
2003	53.22%	52.71%	54.19%	60.10%	5.92%***
2004	54.41%	53.58%	54.86%	59.42%	4.56%***
2005	52.72%	53.78%	54.83%	59.56%	4.73%***
2006	58.41%	60.41%	61.25%	65.09%	3.84%***
2007	54.95%	54.42%	53.55%	59.80%	6.25%***
2008	57.92%	62.98%	64.64%	70.64%	6.00%***
2009	66.84%	65.72%	66.95%	74.00%	7.04%***
2010	57.33%	58.28%	58.59%	63.98%	5.39%***
2011	53.72%	53.19%	51.50%	59.17%	7.67%***
2012	67.47%	69.23%	70.41%	75.39%	4.98%***
2013	60.66%	58.31%	56.43%	65.45%	9.03%***
2014	55.84%	55.81%	58.64%	63.16%	4.52%***
2015	67.73%	68.99%	72.77%	76.05%	3.28%***
2016	55.85%	55.17%	54.52%	60.32%	5.80%***
2017	55.20%	55.92%	56.88%	65.07%	8.19%***
2018	56.17%	51.16%	53.74%	61.24%	7.51%***
2019	32.22%	36.62%	37.68%	46.56%	8.87%***
2020	56.19%	59.14%	63.65%	73.18%	9.52%***
Average	57.03%	57.00%	58.53%	64.80%	6.27%***
AUC	0.5189	0.5335	0.5395	0.5894	0.0499***
Macro F1	0.5195	0.5339	0.5392	0.6662	0.1270***
Trend	-0.0023 (-0.77)	-0.0009 (-0.33)	-0.0014 (-0.46)	-0.0004 (-0.12)	0.0011* (1.91)

资料来源：《Context-Based Interpretation of Financial Information》，华安证券研究所

最后，分析还证实了在解释会计数字时，情境相互作用的重要性存在时间趋势。的确，图表 13 最后一列报告的情境性展现出了正向且统计上显著的时间趋势，这也可以从图表 14 和图表 15 中看出。与之前一样，我们观察到情境信息（未相互作用）的增量信息内容呈现正向趋势，即第 2 列和第 1 列之间的差异。

图表 14 情境信息与年度股票回报：时间趋势

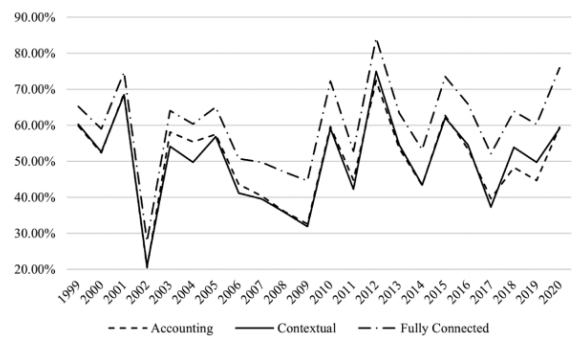
(a) Time Trend in Contextuality



资料来源：《Context-Based Interpretation of Financial Information》，华安证券研究所

图表 15 情境信息与年度股票回报：预测准确性

(b) Time Trend in the Predictive Accuracy of Accounting vs. Contextual Information

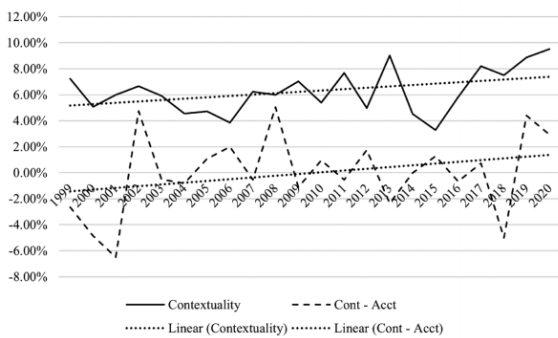


资料来源：《Context-Based Interpretation of Financial Information》，华安证券研究所

我们的结果与 Brown 和 Tucker [2011]的研究结果相反，他们报告称市场对 MD&A 披露的短期反应减弱，并得出结论认为这些信息随时间推移失去了信息性（由于内容重叠）。虽然我们没有关注短期市场反应，但我们的证据与他们的发现相协调，因为我们观察到情境信息解释 60 天股票回报的预测能力呈现负向（尽管统计上不显著）趋势。同时，我们的证据表明，报告的数字及其情境的相互作用越来越重要。

图表 16 情境信息与 60 日回报：时间趋势

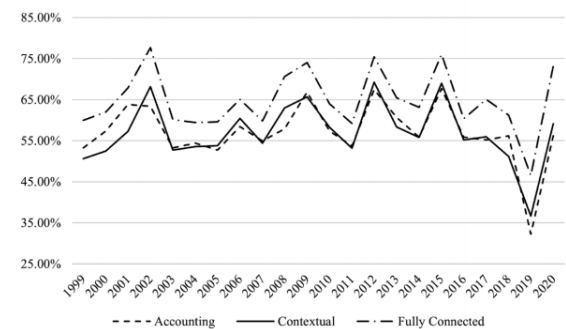
(a) Time Trend in Contextuality



资料来源：《Context-Based Interpretation of Financial Information》，华安证券研究所

图表 17 情境信息与 60 日回报：预测准确性

(b) Time Trend in the Predictive Accuracy of Accounting vs. Contextual Information



资料来源：《Context-Based Interpretation of Financial Information》，华安证券研究所

综上所述，我们始终发现情境性具有经济上的重大影响，即融入叙述背景能大幅提升数字数据的价值。特别是，在推断公司未来时，文本与数字的相互作用的重要性超过了情境信息本身的直接信息价值。

#### 4.4 情境性的影响因素是什么？

在不同的经济情境下，文本情境的重要性是否会发生变化？何时情境的影响最大？我们的方法可以通过衡量根据公司经济状况变化的情境性来回答这个问题。我们探讨了几个可能更加重视情境信息的经济情境。

首先，我们预计当基于数字信息更难评估（估值）公司时，叙述情境可能更重

要。我们预计在以下情况下会如此：(i) 公司的过往业绩可能不代表未来；(ii) 会计数字不容易转化为估值（资产更难衡量）。为了代表第一种情况，我们根据过往收益的波动性（Dichev and Tang [2009]）、异常高（前 20%）或低（后 20%）——因此预计会反转——的应计项目（Sloan [1996]）以及是否报告净亏损的指标（Hayn [1995]）对样本进行划分。为了代表第二种情况，我们根据市净率（Beaver and Ryan [2005]）和政治风险（Hassan et al. [2019]）对样本进行划分。我们采用与主要分析相同的方法，并报告每组和每个目标变量的情境性的平均值。

结果见图表 18，面板 A。括号内为不同子样本间平均情境性差异的 t 统计量。与我们的预期一致，当收益不太具有前瞻性时，情境的重要性更高。特别是，当公司收益波动性更大、应计项目更极端以及报告经营亏损时，情境通常更重要。此外，当公司更难估值或面临更多不确定性时，情境性也更高。直观地看，与公司在稳定状态下运营的情况相比，在这些情况下情境更为重要。综上所述，我们的结果表明，当公司的数字信息更难评估时，情境往往更重要。

图表 18 情境性的横截面和时间序列变化

**TABLE 7**  
*Cross-Sectional and Time-Series Analysis of Contextuality*

Panel A. Cross-sectional analysis																
Var =	Earning Volatility			Accruals			Loss			M/B			Political Uncertainty			
	High (1)	Low (2)	Diff (3)	High (4)	Low (5)	Diff (6)	=1 (7)	=0 (8)	Diff (9)	High (10)	Low (11)	Diff (12)	More (13)	Less (14)	Diff (15)	
Earnings	7.15	5.36	<b>1.79</b> (2.45)	7.00	5.56	<b>1.34</b> (1.99)	7.95	5.01	<b>2.94</b> (3.98)	6.69	5.19	<b>1.50</b> (2.13)	7.32	5.24	<b>2.08</b> (2.47)	
Cash Flow	6.85	5.81	1.04 (1.56)	6.41	6.13	0.28 (0.81)	6.98	5.46	<b>1.52</b> (2.15)	6.39	6.21	0.18 (0.52)	6.58	6.43	0.15 (0.48)	
Return_Ann	5.12	4.01	1.11 (1.60)	5.32	3.99	<b>1.33</b> (1.95)	5.71	3.45	<b>2.26</b> (2.58)	5.17	4.03	1.14 (1.65)	6.51	3.99	<b>2.52</b> (3.01)	
Return_60D	6.90	5.39	<b>1.51</b> (2.05)	6.94	5.21	<b>1.73</b> (2.36)	6.75	5.73	1.02 (1.33)	6.66	5.93	0.73 (1.02)	7.13	5.88	<b>1.25</b> (1.85)	
Panel B. Time-series analysis																
Dep Var =	Earnings (1)				Cash Flow (2)				Return_Annual (3)				Return_60Day (4)			
EPU	0.010*** (3.09)				0.001 (0.33)				0.014* (1.94)				0.014*** (3.52)			
Financial Crisis	0.015*** (5.18)				0.008 (0.47)				0.024** (2.37)				0.004 (0.91)			
COVID	0.007** (2.49)				0.053*** (14.54)				0.080*** (11.32)				0.029*** (8.95)			
N	22				22				22				22			
Adjusted R <sup>2</sup>	0.265				0.210				0.389				0.262			

(Continued)

资料来源：《Context-Based Interpretation of Financial Information》，华安证券研究所整理

注：在面板 A 中，对于每一年，我们取收益波动性、市净率（从 2002 年起）和公司层面政治不确定性的中位数。损失是一个指标，当公司报告经营损失时取值为 1，否则为 0。应计项目是营业收入与经营现金流之间的差额。我们根据每年的应计项目金额对观测值进行排序，并取前 20% 和后 20% 作为“高”应计组和“低”应计组。其余部分被归类为“中”应计组（虽然此处未直接提及“中”应计组，但根据上下文可以推断出这一分类）。基于预测结果，我们每年为每个组计算情境性，并报告平均年度值。我们使用基准模型的预测值。我们估计了 30 次模型，并报告了所有四个目标变量的平均情境性值。加粗的差异表示在 10% 水平下具有统计显著性。括号内报告的是双尾 t 检验的 t 值。在面板 B 中，我们将所有四个目标变量的年度情境性值对几个宏观经济指标进行回归。EPU（经济政策不确定性）是年化后的经济政策不确定性指标。金融危机在 2008 年和 2009 年取值为 1。COVID 在 2020 年取值为 1。标准误差按年份聚类。我们在括号内报告了经自相关调整后的 t 统计量。

其次，我们预计在宏观经济不确定性增加或经济衰退期间，数字披露将更难解读。因此，在这种情况下，情境预计将更具价值，并能增强对数字信息的解读。例如，在新冠疫情期间，许多公司提供了大量关于其供应链中断或产品需求变化的信息披露，这些信息对于推断当前业绩具有重要意义。为了理解什么因素解释了情境性的宏观经济趋势，我们将年平均情境性对三个变量进行回归：年化宏观经济政策不确定性指数（Baker, Bloom, and Davis [2016]）、金融危机指标（2008 年和 2009 年取值为 1，其他年份取值为 0）和新冠疫情指标（2020 年取值为 1，其他年份取值为 0）。

结果见图表 18，面板 B。与我们的横截面分析一致，我们发现情境性与经济政策不确定性（EPU）呈正相关。它在新冠疫情和全球金融危机指标上显示出正向且统计上显著的系数。总之，当公司面临增加的宏观经济不确定性时，情境更为重要。

总体而言，我们的分析表明，情境的价值因公司而异，并随时间变化，这取决于公司所面临的经济状况。

## 4.5 分析师对收益情境信息的运用

在本节结尾，我们更直接地探讨了人类决策者在解读会计信息时是否会考虑叙述情境与数字披露之间的相互作用。为此，我们重点关注证券分析师的预测决策。如果分析师考虑了收益披露周围的叙述情境，他们应该能够提高预测的准确性。

为了测试分析师是否将情境信息纳入其预测中，我们采用了以下设计。基于我们之前训练的模型，我们对第  $t+1$  年的收益是否会增长进行样本外预测。这是基于图表 6 中列出的完全连接和部分连接模型进行的。然后，我们评估与部分连接模型的预测相比，纳入收益情境解释的模型预测是否能更好地解释分析师对未来收益预测的调整。换句话说，我们研究了我们的模型预测的收益变化方向如何映射到发布 MD&A（管理层讨论与分析）信息后分析师预测调整的方向变化上。

我们以两种方式衡量分析师的预测调整，这两种方式在基准方面有所不同：

（1）基于分析师对第  $t+1$  年收益的共识每股收益（EPS）预测减去第  $t$  年的实际 EPS（实际 EPS 基准）；或（2）基于第  $t$  年 10-K 报告发布后发布的第  $t+1$  年共识 EPS 预测减去发布 10-K 报告前相同的预测（预测 EPS 基准）。28 分析师的共识预测基于紧随第  $t$  年 10-K 报告发布之后的 EPS 预测中位数值。我们要求至少有五个预测值来计算中位数。

我们在图表 19 中呈现了结果。为了节省空间，我们只报告了平均准确性。我们发现，完全连接模型更好地解释了分析师的预测调整。特别是，依赖于实际 EPS 基准的模型（1）规范在分析师预测调整方面的解释能力提高了 6.45 个百分点（ $t$  值 = 4.37）。这意味着当模型在插值收益时纳入收益的叙述情境时，它产生的变化与分析师所做的更为一致。这一发现表明，人类决策与基于情境的会计数字解释相一致。

对于依赖于预测 EPS 基准的模型（2）规范，我们同样发现由于收益的情境性，解释能力提高了 4.98%（ $t$  值 = 3.82）。有趣的是，在这两种规范下，我们都观察到随着时间的推移，时间趋势在不断增加，这表明分析师利用了随时间增加的收益情境性。这些结果共同表明，分析师在解释数字时会对其进行情境化处理。

图表 19 不同模型在解释分析师收益预测调整方面的平均解释能力改进

**TABLE 8**  
*Analysts' Use of Context Information*

Model	Benchmark	Prediction	Contextuality	Time Trend
(1)	Actual $EPS_t$	Forecasted $EPS_{t+1}$	6.45%*** (4.37)	0.0563*** (4.02)
(2)	Forecasted $EPS_{t+1}$	Forecasted $EPS_{t+1}$	4.98%** (3.82)	0.0493*** (2.86)

资料来源：《Context-Based Interpretation of Financial Information》，华安证券研究所  
注：该表格报告了完全连接模型相较于部分连接模型（情境性）在解释分析师收益预测调整方面的平均解释能力改进情况。分析是按年份进行的（未制表），随后计算了解释能力改进的平均值以及改进中的相关时间趋势。对于这项分析，我们使用了之前训练的模型（图表 6 中报告）来进行方向性收益变化的样本外预测。这是基于完全连接和部分连接模型进行的。然后，我们评估了与部分连接模型的预测相比，纳入情境解释的模型预测是否能更好地解释分析师对未来收益预测的调整。我们报告了两种规范。模型（1）预测第  $t+1$  年中位数共识每股收益（EPS）预测与实际第  $t$  年 EPS 之间差异的符号。模型（2）预测第  $t$  年 10-K 报告发布后第  $t+1$  年的中位数 EPS 预测与第  $t$  年 10-K 报告发布前测量的第  $t+1$  年共识 EPS 预测之间差异的符号。我们要求至少有五个观测值来计算中位数，对于模型（2），要求 10-K 报告发布前后的分析师群体相同。趋势基于年度估计值对线性时间趋势的时间序列回归的回归系数。我们报告了趋势系数的自相关调整  $t$  统计量。

## 5 应用：情境化的收益持续性

为了说明我们的结果的价值和适用性，我们以一个实际应用作为结论，该应用量化了基于情境的收益持续性异质性。收益持续性既受到学术界的关注，也受到决策者的重视，因为它通常被视为衡量收益质量的主要指标（Francis 等[2004]，Dichev 等[2013]）。同时，衡量持续性也颇具挑战性。以往的研究通常依赖于横截面代理变量或使用滚动窗口来估计公司层面的持续性。这两种方法都不能衡量特定公司-年份的收益持续性，而这正是我们模型得出的基于情境的持续性所具有的独特特征。为了将收益持续性的异质性建模为情境的函数，我们采用了 Farrell、Liang 和 Misra 在[2020, 2021]中最近开发的一种方法。这种方法允许将持续性参数建模为公司特征高维向量的非参数函数。该模型生成的估计量与回归持续性估计量具有类似的解释性，但它们在横截面和时间上存在差异。

### 5.1 方法论

我们采用了标准的收益持续性回归模型，但其参数被建模为文本向量的灵活函数：

$$\begin{aligned} E_{i,t+1} &= \gamma_{0it} + \gamma_{1it}E_{i,t} + \delta_{i,t+1} \\ &= \gamma_0(\text{context}_{it}) + \gamma_1(\text{context}_{it})E_{i,t} + \delta_{i,t+1}, \end{aligned} \quad (2)$$

其中， $E_{i,t}$ 表示公司  $i$  在第  $t$  年的标准化收益， $\text{context}_{it}$ 是 768 维的情境向量。我们采用了相同的神经网络架构，但现在模型预测的是回归系数作为其输出。为了实现这一点，我们从输出层中移除了标准的激活函数，而是添加了一个自定义层，

用于计算预测误差  $\delta_{i,t+1} = E_{i,t+1} - \gamma_0(context_{it}) - \gamma_1(context_{it})E_{i,t}$ 。该模型被训练以生成能够最小化均方根误差 (RMSE) 的参数。模型训练后的输出 ( $\hat{\gamma}_{1it}$ ,  $\hat{\gamma}_{2it}$ ) 是在公司-年份层面估计的基于情境的收益持续性参数。

## 5.2 收益持续性的异质性

图表 20 报告了基于公司-年份情境的持续性系数的年度横截面平均值，并将其与标准横截面普通最小二乘法 (OLS) 回归得到的系数进行了比较。第 1 列和第 2 列分别展示了 OLS 系数及其标准误差，而第 3 列和第 4 列则展示了平均基于情境的持续性及其相关的年度标准偏差。比较第 1 列 (OLS) 和第 3 列 (基于情境) 可以看出，两个模型的估计值随时间共同变动，这是符合预期的。例如，在 2000-2002 年期间，持续性一直较低。基于情境的收益持续性在所有年份的平均时间序列值为 0.776，与 OLS 平均值 0.774 几乎相同。

图表 20 基于公司-年份情境的持续性系数的年度横截面平均值

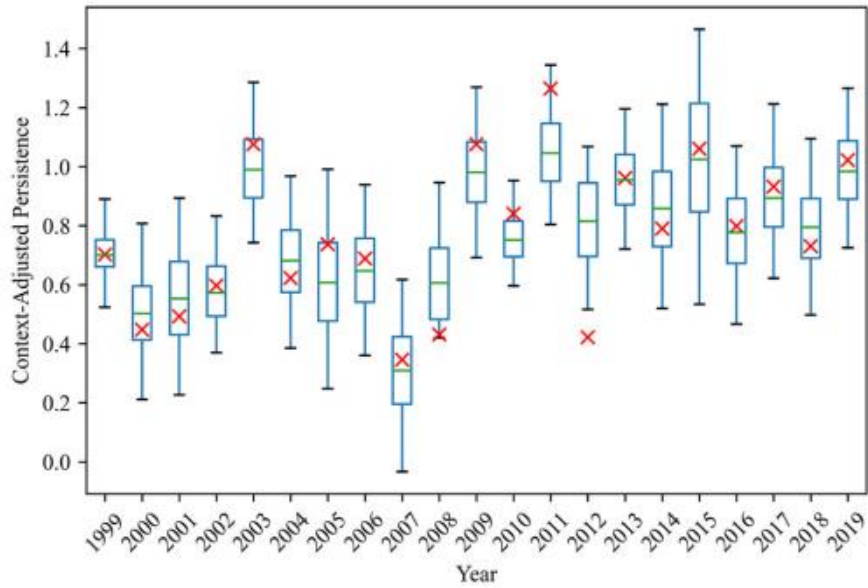
**TABLE 9**  
*Heterogeneity in Context-Based Earnings Persistence*

Year	Persistence (OLS)		Context-Based Persistence	
	Slope (1)	Standard Error (2)	Slope (3)	Standard Error (4)
1998	0.9827	0.0121	0.9875	0.1889
1999	0.7025	0.0260	0.7019	0.2432
2000	0.4473	0.0167	0.5022	0.2083
2001	0.4922	0.0118	0.5561	0.1409
2002	0.5972	0.0138	0.5816	0.1873
2003	1.0762	0.0122	0.9990	0.1944
2004	0.6223	0.0167	0.6727	0.2551
2005	0.7375	0.0113	0.6142	0.1499
2006	0.6888	0.0131	0.6501	0.1802
2007	0.3463	0.0074	0.3011	0.1966
2008	0.4311	0.0088	0.6284	0.1471
2009	1.0765	0.0079	0.9811	0.0847
2010	0.8411	0.0116	0.7586	0.1533
2011	1.2644	0.0227	1.0520	0.1752
2012	0.4216	0.0076	0.8077	0.1680
2013	0.9615	0.0126	0.9566	0.1972
2014	0.7906	0.0071	0.8569	0.2937
2015	1.0602	0.0120	1.0059	0.1613
2016	0.7987	0.0127	0.7724	0.1973
2017	0.9324	0.0175	0.9003	0.1786
2018	0.7306	0.0185	0.7911	0.1489
2019	1.0219	0.0288	0.9879	0.2201
Average	0.7738	0.0140	0.7757	0.1850
Trend	0.0129 (1.56)		0.0132*** (2.26)	

资料来源：《Context-Based Interpretation of Financial Information》，华安证券研究所

尽管与 OLS 估计值相关的标准误差非常小 (第 2 列)，但它们是误导性的，因为我们观察到基于情境的斜率系数存在更大的变异性 (第 4 列)。这种异质性在图表 21 中以图形方式进行了总结。

图表 21 收益持续性的异质性



资料来源：《Context-Based Interpretation of Financial Information》，华安证券研究所

### 5.3 验证基于情境的收益持续性

由于基于公司-年份的收益持续性度量可能对不同的实证研究有用，我们进行了几个步骤来验证它并探索其实证特性。

#### 5.3.1 度量的表面效度

作为第一步，我们检查了基于情境的持续性的表面效度。我们从基于情境的持续性系数分布的前 5% 和后 5% 中随机选择了八个文本片段。在阅读这些例子时，人们很快就会感受到相应的公司正在经历经济冲击（中断）并对未来表达不确定性。例如，公司披露了（1）对未来利润率的潜在不利影响，（2）可能需要通过降价来应对竞争压力，（3）对负面宣传的担忧，（4）由于财政年度多出一周而对业绩的提升等。这些因素是收益暂时性的直观指标。这些例子通常传达了对经营业绩的信心。总体而言，这些例子与模型根据情境区分暂时性收益与永久性收益成分的能力相一致。

#### 5.3.2 基于公司特征的收益持续性是否可以推断？

作为第二步，我们调查了我们的基于情境的收益持续性的变化是否由公司特征驱动。我们估计了以下回归：

$$Persistence_{it} = \gamma X_{it} + \delta_t + \lambda_j + \delta_{it}, \quad (3)$$

其中， $Persistence_{it} = \hat{\gamma}_1(context_{it})$  是估计的基于情境的持续性。公司特征向量  $X_{it}$  包括规模（总资产的对数）、市账比、资产回报率、杠杆率、资本密集度和亏损指标。我们还控制了基于情境的截距  $\hat{\gamma}_{0,it}$ ，以方便解释。模型包括时间固定效应  $\delta_t$  和行业固定效应  $\lambda_j$ 。我们按行业聚类标准误差，并在适用的情况下将连续变量在 1% 和 99% 的水平上进行缩尾处理。我们列出了有和无行业固定效应的模型，因为

它们与公司特征相关。两列的结果在性质上是相似的。我们发现，基于情境的持续性在统计上和经济上与公司特征没有关联。唯一的例外是公司规模，它表现出显著的负相关性。这些发现意味着基于情境的持续性的异质性超出了公司特征的范围。

### 5.3.3 样本外的持续性是否具有信息性？

基于情境的持续性所带来价值的最终检验在于它是否能预测实际的收益动态，即收益是持续存在还是均值回归，以及相比其他现有的公司-年度特定替代指标，它是否能更有效地做到这一点。具体而言，估计公司-年度收益持续性的一个常见方法是依赖公司层面的滚动窗口回归。然而，滚动窗口越长，对当前年度持续性的描述就越不准确。我们采用这种方法，使用公司  $i$  从  $t-10$  年到  $t$  年的数据来估计其在  $t$  年的收益持续性。请注意，这种方法假设持续性在 10 年窗口内是恒定的，因此是无情境的。

为了评估情境化持续性估计与无情境持续性估计中的信息，我们估计了以下样本外回归：

$$E_{i,t+1} = \beta_0 + \beta_1 E_t + \beta_2 Persistence_{it} + \beta_3 E_t \times Persistence_{it} + \gamma_1 X_{it} + \gamma_2 Persistence_{it} \times X_{it} + \delta_t + \lambda_j + \delta_{it}, \quad (4)$$

其中变量的定义与之前相同。请注意， $Persistence_{it}$  是基于  $t$  时或  $t$  时之前的信息估计的，因此模型看不到  $E_{i,t+1}$ 。当前收益与基于情境的持续性之间交互项的系数为正，表明存在样本外持续性可预测性。

我们在图表 22 的面板 A 中列出了这一分析的结果。第 1 列和第 2 列提供了基于无情境持续性（公司层面的滚动窗口估计）的估计值，而第 3 列和第 4 列则依赖于基于情境的持续性。我们发现，交互项“收益×持续性”在第 1 列中勉强显著，但在第 2 列控制持续性决定因素后失去了统计显著性。换句话说，无情境持续性在样本外解释收益动态的能力有限。相比之下，在第 3 列和第 4 列中，交互项的系数为正且统计显著。在控制公司特征后，收益与持续性的交互项（ $Earnings_{it} \times Persistence_{it}$ ）仍然显著为正。这些结果表明，基于情境的收益持续性包含了预测收益动态的有用信息，而这些信息在无情境替代指标中是不存在的。

为了促进对上述发现的经济解释，我们还采用了一种更直观的设计，即每年根据公司持续性（ $Persistence_{it}$ ）将公司分为五个等分位数。然后，我们估计当前收益对未来（样本外）收益的横截面回归。我们在面板 B 中报告了每个等分位数的回归估计值。同样，前两列（第一列和第二列）依赖于无情境估计，而后两列（第三列和第四列）依赖于基于情境的估计。对于这些估计值，我们发现实际收益持续性在五个等分位数之间呈非单调（倒 U 形）关系（预测持续性）。相比之下，基于情境的持续性（第三列和第四列）呈现出跨等分位数的向上倾斜关系。它还显示出在区分低收益持续性和高收益持续性的公司方面具有更强的能力。

最后，在面板 C 中，我们对两种方法进行了比较，这使我们能够评估与基于情境的替代指标相关的经济意义改善程度。我们计算平均绝对误差（MAE）作为实际收益与基于上述规范的预测收益之间的差异。对于每种持续性指标，我们报告乘以 100 的平均 MAE。我们发现，无情境模型的 MAE 为 5.6529，是基于情境的持续性估计 MAE 的两倍多。

图表 22 基于情境的持续性能否预测实际的收益动态

**TABLE 10**  
*Context-Based Earnings Persistence and Future Earnings Dynamics*

---

**Panel A. Predictive power of context-based earnings persistence**

Dep Var = Earnings <sub><i>it</i>+1</sub>	Context-Free		Context-Based	
	(1)	(2)	(3)	(4)
ML_Intercept <sub><i>it</i></sub>	0.5581*** (14.29)	0.5158*** (12.73)	10.3135 (1.55)	16.2398 (1.23)
Earnings <sub><i>it</i></sub>	0.5854*** (23.64)	0.5886*** (22.16)	0.7667*** (17.23)	0.7367*** (12.99)
Persistence <sub><i>it</i></sub>	18.7834*** (3.82)	13.6024*** (2.94)	-12.3251 (-1.32)	-0.2153 (-1.13)
Earnings <sub><i>it</i></sub> × Persistence <sub><i>it</i></sub>	0.2293* (1.97)	0.2193 (1.37)	0.1657*** (3.23)	0.1837*** (3.06)
Controls	No	Yes	No	Yes
Controls × Persistence <sub><i>it</i></sub>	No	Yes	No	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes
Cluster	Industry	Industry	Industry	Industry
<i>N</i>	47,706	47,706	47,706	47,706
Adjusted <i>R</i> <sup>2</sup>	70.30%	72.97%	72.55%	74.40%

---

**Panel B. Quintile analysis**

Quintile of Estimated $\beta_1$	Context-Free		Context-Based	
	With FE (1)	Without FE (2)	With FE (3)	Without FE (4)
First Quintile	0.7951*** (28.30)	0.7845*** (27.68)	0.7853*** (24.96)	0.7599*** (22.18)
Second Quintile	0.8115*** (24.11)	0.8008*** (23.71)	0.8237*** (26.95)	0.8102*** (23.87)
Third Quintile	0.9314*** (30.68)	0.9266*** (27.56)	0.9110*** (37.63)	0.9086*** (35.11)
Fourth Quintile	0.8796*** (26.55)	0.8649*** (24.16)	0.9008*** (35.95)	0.8914*** (34.16)
Fifth Quintile	0.8635*** (25.94)	0.8518*** (24.57)	0.9307*** (38.07)	0.9302*** (38.22)

---

**Panel C. Mean absolute errors**

Accuracy	Context-Free	Context-Based
MAE × 100	5.6529	2.2135

资料来源：《Context-Based Interpretation of Financial Information》，华安证券研究所

综上所述，我们的基于情境的衡量指标成功捕捉到了收益持续性的公司-年度异质性。这种异质性没有体现在公司特征中，并且在解释样本外收益动态方面具有比无情境（基于滚动窗口）估计更高的信息价值。

## 6 讨论

深度学习方法在需要大量学习特征而非手动构建特征，以及模型关系中可能存在深度非线性（交互）关系的场景中尤为重要（Bochkay 等人[2023]）。这些确实

是本研究的关键方面，使我们能够得出性质上截然不同的结论。衡量数值披露的上下文性需要建模高维语义含义，并捕捉文本数据所特有的深度非线性关系（Bengio、Ducharme 和 Vincent[2000]）。相比之下，以往的文献主要关注单维文本属性，如可读性（例如，Li[2008]）。尽管这些研究通常建立了这些交互作用的统计显著性，但它们并未表明这些交互作用是否具有经济意义上的重要解释力。

我们全面评估了这一问题。例如，我们复制了 Li[2008] 中考察的未来收益对当前收益的回归，以及当前收益与迷雾指数（可读性）的交互作用。我们发现，迷雾指数与当前收益的交互作用仅使该回归的调整  $R^2$  增加了 0.0001，即解释变异的 0.03%。与我们在图表 6 中观察到的上下文性值相比，这种增量解释力微不足道。此外，与文本数据的主要影响相比，这些交互作用在经济上的重要性较低，这与我们的发现相反。我们将此分析扩展到其他一些文本特征，如情感，并研究了替代规范。总体而言，我们得出结论，以往依赖于回归方法和浅交互作用的实证证据与文本披露在经济上具有显著上下文性的观点不一致。

## 7 结论

不同的企业沟通模式往往旨在相互补充和增强。特别是，叙述性披露通常旨在为财务报表中报告的数值信息的解释提供背景。我们研究了背景性的程度，即围绕报告的收益数字周围的叙述信息在多大程度上有助于投资者解释这些数字。我们引入了一种新的方法，通过允许或限制多维文本和数值信息之间的深度交互，来量化数值解释中背景的经济重要性。特别是，我们使用了密集连接与稀疏连接的人工神经网络（ANN）相结合，以及大型语言模型（BERT），以模拟解释过程，这一过程类似于人类大脑中多模态数据的信息处理。我们表明，将管理层讨论与分析（MD&A）中披露的收益背景纳入其中，使得财务报表使用者能够在对实体未来前景形成准确预期方面取得实质性改进。

我们的证据加深了对背景在解释财务报表中报告的数值信息方面的重要性的理解。具体而言，我们的结果表明，MD&A 披露之所以有用，不仅因为其直接的信息内容，还因为它们改变了数值信息的意义。特别是，允许数值披露与其叙述讨论之间存在深度非线性交互，可以显著改进对未来公司收益、现金流和股票表现的推断，因此在决策中具有首要重要性。此外，我们的结果表明，文本-数值交互的信息价值通常远超过文本信息的直接增值。

我们还表明，在数值信息不太可能具有信息性的情况下，收益背景性（即背景在解释收益数字中的重要性）会增加。它还呈现出积极的时间趋势，这表明 MD&A 报告对决策者的相关性随着时间的推移而增强。鉴于我们没有观察到会计数字的信息性随时间发生系统性变化，这一证据为投资者对收益公告的反应增强（Beaver、McNichols 和 Wang [2020]）提供了潜在解释。

最后，我们通过基于背景的收益持续性建模展示了研究结果的实际应用。我们表明，从收益背景中得出的收益持续性存在丰富的异质性，这难以用公司特征轻易解释。此外，基于背景的持续性在样本外预测收益动态方面很有用，并且优于特定公司年度的无背景估计。这一发现与 Ball 和 Nikolaev [2022] 的结果相呼应，即在使用收益预测现金流时，考虑公司特定的异质性至关重要。

我们的研究为未来研究开辟了多个途径。多模态通信越来越普遍，而多模态数据在机器学习（ML）中的应用也势头正盛（例如，Mayew 和 Venkatachalam [2012]，Hobson、Mayew 和 Venkatachalam [2012]，Baik 等 [2023]，De la Parra

和 Gallemore [2024])。尽管我们专注于文本-数值互补性,但基于我们引入的方法,也可以解决许多有趣的问题。潜在研究问题的示例如下。

首先,与叙述性披露为数字提供背景类似,语音线索为叙述性沟通提供背景(例如, Mayew 和 Venkatachalam [2012], Hobson、Mayew 和 Venkatachalam [2012], Baik 等[2023], De la Parra 和 Gallemore [2024])。这些背景特征是否以及在多大程度上增强了管理层传达的叙述,是一个开放性问题,而以往的方法无法有效解决这一问题。其次,言语沟通与非言语线索(如手势、面部表情和肢体语言)之间的交互在检测欺骗和虚假报告方面很有前景(例如, Peng 等[2022], Dávila 和 Guasch [2022], Jiang、Kelly 和 Xiu [2023])。心理学研究表明,检测欺骗仍然是一个极具挑战性的问题(参见 DePaulo 等[2003]的综述)。有理由预期,参与欺诈的个人在公开讲话时,其非言语动作和语音线索之间会存在深度交互。我们的方法能够轻松解决口头和视觉沟通之间的交互在检测欺骗行为中的重要性。最后,我们得出的基于背景的持续性可用于理解在衡量会计质量时背景的价值,并且可以将分析扩展到研究背景对其他会计属性(如收益操纵)的作用。总之,能够为企业沟通提供背景,使得能够提出许多新的研究问题。

文献来源:

核心内容摘自 ALEX G. KIM, VALERI V. NIKOLAEV 于 2024 年 10 月 31 日在 Journal of Accounting Research 上的文章《Context-Based Interpretation of Financial Information》。

## 风险提示:

文献结论基于历史数据与海外文献进行总结;不构成任何投资建议。

## 重要声明

### 分析师声明

本报告署名分析师具有中国证券业协会授予的证券投资咨询执业资格，以勤勉的执业态度、专业审慎的研究方法，使用合法合规的信息，独立、客观地出具本报告，本报告所采用的数据和信息均来自市场公开信息，本人对这些信息的准确性或完整性不做任何保证，也不保证所包含的信息和建议不会发生任何变更。报告中的信息和意见仅供参考。本人过去不曾与、现在不与、未来也将不会因本报告中的具体推荐意见或观点而直接或间接接收任何形式的补偿，分析结论不受任何第三方的授意或影响，特此声明。

### 免责声明

华安证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。本报告中的信息均来源于合规渠道，华安证券研究所力求准确、可靠，但对这些信息的准确性及完整性均不做任何保证。在任何情况下，本报告中的信息或表述的意见均不构成对任何人的投资建议。在任何情况下，本公司、本公司员工或者关联机构不承诺投资者一定获利，不与投资者分享投资收益，也不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。投资者务必注意，其据此做出的任何投资决策与本公司、本公司员工或者关联机构无关。华安证券及其所属关联机构可能会持有报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供投资银行服务或其他服务。

本报告仅向特定客户传送，未经华安证券研究所书面授权，本研究报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。如欲引用或转载本文内容，务必联络华安证券研究所并获得许可，并需注明出处为华安证券研究所，且不得对本文进行有悖原意的引用和删改。如未经本公司授权，私自转载或者转发本报告，所引起的一切后果及法律责任由私自转载或转发者承担。本公司并保留追究其法律责任的权利。

## 投资评级说明

以本报告发布之日起 6 个月内，证券（或行业指数）相对于同期沪深 300 指数的涨跌幅为标准，定义如下：

### 行业评级体系

- 增持—未来 6 个月的投资收益率领先沪深 300 指数 5%以上；
- 中性—未来 6 个月的投资收益率与沪深 300 指数的变动幅度相差-5%至 5%；
- 减持—未来 6 个月的投资收益率落后沪深 300 指数 5%以上；

### 公司评级体系

- 买入—未来 6-12 个月的投资收益率领先市场基准指数 15%以上；
- 增持—未来 6-12 个月的投资收益率领先市场基准指数 5%至 15%；
- 中性—未来 6-12 个月的投资收益率与市场基准指数的变动幅度相差-5%至 5%；
- 减持—未来 6-12 个月的投资收益率落后市场基准指数 5%至；
- 卖出—未来 6-12 个月的投资收益率落后市场基准指数 15%以上；
- 无评级—因无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，致使无法给出明确的投资评级。市场基准指数为沪深 300 指数。