GOKI WORKPAPER

TECHNICAL NOTE · TN-2026-0915

勾稽关系审计模型说明

结构 · 输入输出 · 如何用在真实财报上

一份给底稿的说明书。合成样本上训练的三头网络做什么、不做什么； 港股蓝筹年报走哪条路径、为什么未分配利润和固定资产会留下缺口。

项目
GOKI 财务报表勾稽关系审计

版本
1.0 · 2026-09-15

对象
合成 1,000 发行人 + 恒生蓝筹 FY2025 实报

密级
工作底稿 · 非鉴证意见

实现
OCaml / OCANNL（cc 后端）+ 浏览器副本

单位
金额一律折成「万元」后再入模

1. 这份说明回答什么

模型不是「读完年报判断造假」。它做三件更窄的事：

1. 把两年科目轧成十条勾稽残差；
2. 在合成样本上，用小网络把「注入的真错误」从舍入和重分类里分出来；
3. 在真实年报上，只跑规则，并把恒等拆成「硬」和「口径」，缺口要能被附注吃掉。

> 合成样本上训好的 MLP 不能直接给港股年报打舞弊分。HKFRS 比简化恒等多了 OCI、回购、处置、在建、汇兑。模型会把这些当成「真错误」。实报走规则路径，不走分类头。

2. 全链路

两年科目
利润表 / 资产负债 / 现金流量 38 栏 × {t−1, t}

→

十条勾稽
8 条恒等 + 2 条分析性 残差、相对残差、同比

→

38 维特征
10×3 + 8 行业 one-hot 训练集 z-score

HEAD A · 分类
MLP 38→64→32→1
输出 p(真错误)。只在合成样本上训练与打分。

HEAD B · 异常
自编码器 38→16→8
只在干净样本上训。重构误差高 = 勾稽形态未见过。

HEAD C · 对照
现金回归 10→32→1
用结构比率预测现金/资产，残差对照账上现金。

OCaml 侧用 OCANNL 的 `let%op` 写出同一张图，cc 后端编成 `goki_mlp_forward`；浏览器里是逐元素 FMA 副本，seed=3 可对黄金文件。

3. 输入：科目字典 YearBooks

一份发行人 = 行业、规模、两年 YearBooks。金额单位：万元。真实港股先把百万/亿元折成万元，再塞进同一套栏位。

表	栏位	说明
利润表	`revenue, cogs, gp, opex, da, ebit, interest, pretax, tax, ni, dividends`	毛利、EBIT 为派生。ni 取归母或公司口径，映射时写进 caveats。
资产负债表	`cash, ar, inv, ppe, otherCa, otherNca, ap, stDebt, taxPay, ltDebt, otherL, shareCap, re`	other* 是塞剩下的塞子，用来强制 R01 闭合。
现金流量	`cfo, cfi, cff, netCf, capex, taxPaid`	netCf 应对 Δcash。未披露经营现金流时，实报路径用间接法回填，并标记。

R01 恒等：`cash+ar+inv+ppe+otherCa+otherNca = ap+stDebt+taxPay+ltDebt+otherL+shareCap+re`。映射时把未单列的资产/负债/权益推进 other* 或 re，硬恒等必须轧平。

4. 十条勾稽 → 38 维特征

每条规则产出残差 residual、尺度 scale、相对残差 rel = residual / scale。两年样本里同比 yoy 取当年 rel。

代码	名称	种类	简化公式
R01	资产负债恒等	硬	资产 − (负债+权益) = 0
R02	货币资金滚存	口径	Δcash − netCf = 0
R03	未分配利润滚存	口径	ΔRE − (NI − 分红) = 0
R04	净利润桥	硬	pretax − tax − NI = 0
R05	毛利勾稽	硬	revenue − cogs − gp = 0
R06	间接法现金流	口径	CFO − (NI+DA−ΔAR−ΔInv+ΔAP) = 0
R07	固定资产滚存	口径	PPE − (期初 + capex − DA) = 0
R08	应交税费滚存	口径	taxPay − (期初 + tax − taxPaid) = 0
R09	应收／收入	分析	AR/rev 本年 − 上年
R10	存货／成本	分析	Inv/cogs 本年 − 上年

```
特征 x ∈ R^38
对规则 i = 0..9：
  x[3i]     = sign(residual) · log(1 + |residual|)     // 金额尺度
  x[3i+1]   = rel                                       // 相对残差
  x[3i+2]   = yoy                                       // 两年样本 = rel
对行业 k = 0..7：
  x[30+k]   = 1  若 issuer.industry = k，否则 0
训练折内 z-score：x̂ = (x − μ_train) / (σ_train + 1e-6)
```

5. 三个头：结构与输入输出

5.1 分类头 MLP

```
h1 = ReLU(W1 x̂ + b1)     W1 ∈ R^{64×38}
h2 = ReLU(W2 h1 + b2)     W2 ∈ R^{32×64}
z  = W3 · h2 + b3         W3 ∈ R^{32}
p  = σ(z)                 p(真错误) ∈ (0,1)
```

项	值
参数量	64×38 + 64 + 32×64 + 32 + 32 + 1 = 4,609
初始化	中心均匀 U[−0.25, 0.25)，OCANNL `fixed_state_for_init = 3`
损失	BCE-with-logits：max(z,0) − z y + log(1+exp(−|z|))
正则	权重 L2，系数 1e-4
标签	y = 1 当且仅当 inject = true_error；clean / rounding / reclass 均为 0
解释	显著性 |∂z/∂x̂_j · x̂_j|，取前 6 个特征名

5.2 自编码器

38 → 16 → 8（瓶颈）→ 16 → 38，ReLU 隐层，线性重建。只在 y=0 的 batch 上做 SGD。输出 aeErr = 平均绝对重构误差。用来抓「勾稽形态不像干净样本」，不是分类概率。

5.3 现金回归

输入 10 个结构比率（均除以总资产）：AR、存货、PPE、应付、有息债、收入、净利润、CFO、未分配利润、所得税。隐层 32，线性输出预测现金/资产。cashResidual = 预测 − 账上。合成样本里现金被掏空时，这条会先亮。

6. 训练协议（只对合成样本）

项	值
样本量	1,000 发行人，确定种子 seed = 3
注入	clean / 舍入 / 重分类 / 真错误（现金空洞、资产负债缺口、RE 断裂、虚增收入、PPE 虚增、税断裂）
切分	80 / 10 / 10，切分种子 42；scaler 只在训练集拟合
优化	14 epoch，batch 32，学习率从 0.12 线性降到 0.01
黄金文件	参数校验和、train/val BCE 尾、test AUC 写入 goki_mlp.expected

> 合成测试折上，MLP 要把真错误与舍入/重分类分开。AUC 接近 1 只说明在自己生成的分布上可分，不说明对 HKFRS 年报有鉴别力。

7. 合成样本上如何应用

队列在合成演示里仍按 pError 降序，这只说明训练分布内部可分。审核顺序不要用 p。留出一族训练时没见过的注入后，MLP 的排序差于随机，最大恒等相对残差仍能排。p 是实验项，不进门禁，也不做排序层。判定和排序基线都用连续相对残差。

```
maxRel = max_{i=0..7} |rel_i|
例外  若 p ≥ 0.50  或  maxRel ≥ 5%
复核  若 p ≥ 0.20  或  maxRel ≥ 1%
通过  其余
```

合成底稿仍会印出 p、十条残差和上面这组旧分诊带。那是实验输出。能不能发、先看哪一份，都不靠 p。

8. 真实财报上如何应用（港股蓝筹）

真实年报不进入 MLP。路径如下。

1. 取数。年报合并口径，利润表、资产负债、现金流量、权益变动、PPE 附注。币种保留（港元 / 人民币 / 美元），期间标签写成 FY2025。
2. 折单位。百万×100、亿元×10,000，全部进万元。
3. 映射。收入、成本、税、净利润、现金、应收、存货、PPE、应付、有息债、股本、CFO/CFI/CFF、资本开支、折旧、分红，能对上的对上。对不上的推进 otherCa / otherNca / otherL；除股本外的权益推进 re，以便 R01 闭合。
4. 缺失经营现金流。若年报未给出可用的 CFO，用简化间接法回填 NI+DA−ΔAR−ΔInv+ΔAP，并在 caveats 标明「回填，非披露数」。有披露的（腾讯、移动、中海油、美团）用原文。
5. 缺失净增加额。netCf 缺则取 Δcash，避免 R02 被自己的取数空洞打爆。
6. 跑十条规则。得到 residual / rel。不标准化，不进网络。
7. 分诊只看硬恒等。

```
maxHard = max(|R01|, |R04|, |R05|)     // 映射诚实则应 ≈ 0
maxSoft = max(|R02|, |R03|, |R06|, |R07|, |R08|)
例外   若 maxHard ≥ 1%     → 取数或串户，先查映射
复核   若 maxSoft ≥ 1%     → 写底稿解释口径
通过   其余
```

> 不要把合成 MLP 的 pError 印在港股名单上。OCI 和回购会把 p 顶到「例外」，那是公式短，不是舞弊信号。

9. 硬恒等与口径残差

硬 · 必须闭合

- R01 资产负债
- R04 利润总额 − 税 − 净利润
- R05 收入 − 成本 − 毛利

这三条在同一张表、同一口径下是算术。非零只可能是漏列、串户、取数层级不一致。

口径 · 必须能解释

- R02 现金：汇兑、受限资金
- R03 未分配利润：OCI、回购、储备、NCI
- R06 间接法：其它营运资本、减值、股份支付
- R07 固定资产：处置、在建、重估、使用权、折耗
- R08 应交税：递延、预缴、税务管辖

完整式（底稿要写的，模型没写进去）

```
R03 完整：
  ΔRE = NI − 分红 + OCI − 回购 ± 储备结转 ± 少数股东 ± 股份支付

R07 完整：
  期末 PPE = 期初 + 购置 + 在建结转 − 折旧 − 处置净值 − 减值 ± 汇兑 ± 重估
             ± 使用权资产（若并入 PPE）
```

合成样本按截断式生成，所以截断式在那里是恒等。实报按完整式编报，截断式必然有缺口。缺口方向可提示附注该翻哪一页：

规则	缺口符号	先看什么
R03	负（ΔRE 小于 NI−分红）	回购、拨入法定/一般储备、OCI 亏损、外币折算
R03	正	OCI 盈余、重估、注资进储备、股份支付、投资物业公允
R07	期末低于滚存	资本开支进了在建/无形/投资物业，或处置、折耗、减值
R07	期末高于滚存	并表购置、在建结转、重估、汇兑、使用权资产

10. 工作底稿示例：00700 腾讯 FY2025

映射后硬恒等均为 0。口径残差最大在 R03。

项目	金额（万元人民币）
Δ 未分配利润（映射进 re 的权益变动）	11,840,600
净利润 − 分红	22,980,100
R03 简化缺口	−11,139,500
相对残差	−48.5%
Δ PPE	6,972,000
资本开支 − 折旧	8,748,200
R07 简化缺口	−1,776,200

> 分诊：硬恒等 0 → 不是例外。R03 / R07 有缺口 → 复核。解释：股份回购与其它综合收益进权益、资本开支部分进在建/无形，不是利润作假。写入 caveats，处置记「复核」。

同批 11 家恒生蓝筹：硬恒等全部为 0；全部进入复核。口径残差从恒生银行 2.3% 到港交所 132%（总部资本开支未进 PPE 净额）不等。残差大小排序的是「简化公式有多不完整」，不是「谁更可能舞弊」。

11. OCaml / OCANNL 编译件

权威训练图在 OCaml：`hid1 = relu (w1 * x + b1)`，再 32 维，再 logit。BCE-with-logits 与浏览器副本同一式。cc 后端把前向编成：

```
void goki_mlp_forward(const float *x /* [38] */, float *y /* [1] */);
/* packed FMA, dtype=f32, seed=3, params=4609
   输入必须是「训练集 μ/σ 标准化后的 38 维」
   输出 y[0] = sigmoid(logit) = p(真错误) */
```

黄金文件核对：参数校验和、尾部 BCE、test AUC。比特级可复现是这条链路存在的理由——底稿要能说出「同一个种子、同一份 .expected」。它不自动获得在真实年报上的外部效度。

12. 边界：模型不做什么

- 不读 PDF 年报，不抽附注文本。映射是人做的。
- 不处理多币种折算、分部、保险合同（IFRS 17）、银行存贷款的完整监管表。
- 不把「复核」当成审计意见。复核 = 需要一句能被附注支撑的解释。
- 不在实报上微调或再训练。要用真实标签，得先有「已确认错报」样本，目前没有。
- 分析性程序 R09 / R10 不是恒等，周转变化要结合行业，单独不升级为例外。

13. 把下一份年报喂进去的检查单

1. 合并利润表、资产负债、现金流量、权益变动、PPE 变动表齐了没有。
2. 单位、币种、期间写进 issuer 元数据。
3. R01 / R04 / R05 是否为零。不为零先改映射，不要改阈值。
4. R03 缺口能否用权益变动表的 OCI / 回购 / 储备行对上数量级。
5. R07 缺口能否用 PPE 附注的购置、在建、处置、折旧行对上数量级。
6. R02 是否被汇兑或受限现金解释；R06 是否被「其它营运资本」解释。

7. 解释得了 → 复核并归档 caveats。解释不了、且硬恒等也裂了 → 例外，回到取数。

8. 仍然不要把 MLP 概率写进这张底稿。

一句话

38 维勾稽特征进一个 4,609 参数的 MLP，只为在合成世界里分开「真错误」和「舍入」。真实年报用同一套规则当尺子，用硬恒等找映射错误，用口径残差逼附注。模型结构很浅；难的是科目怎么进这 38 栏。

GOKI TN-2026-0915 · 勾稽关系审计模型说明 · 工作底稿，不构成鉴证 · 金额单位万元