看哪些贷款，而非贷款数量：对已部署风险模型进行精确水平不变、任意时刻有效的监测

Xin Xu
xuxin@cmu.edu 卡内基梅隆大学 美国宾夕法尼亚州匹兹堡

arXiv:2611.28342v1 [cs.LG] 20 Nov 2026

摘要

再融资浪潮可能改变抵押贷款提前还款的数量，却不改变得分与提前还款之间的关系。在逻辑模型中，以月度数量为条件，可从描述哪些贷款发生提前还款的似然中消去截距。我们的固定起点斜率乘积用于检验校准斜率的下界。当阈值为 20 时，在原假设下曾经越过该阈值的概率至多为 5%。重启水平和用于追踪偏离截距容差的情况。置信序列用于估计恒定斜率，重启交集规则则用于检测斜率变化。连续混合避免了离散先验的非网格宽度下限。在 Freddie Mac 贷款数据上，固定斜率重采样在开发集和留出集上分别重现了维护模型月度 AUC 方差的 78% 和 70%（95% 区间分别为 57–100% 和 48–100%）。从留出集的新起点开始，重启水平和在两条路径上都越过了阈值，而固定起点斜率乘积在两条路径上都低于阈值。重启交集规则在 2018年4月对冻结模型发出向下变化信号，并在 2019年8月的一次补充运行中对维护模型发出向上变化信号。

CCS 概念

• **计算数学** → **概率推断问题**； • **应用计算**；

关键词

模型监控、模型风险管理、任意时刻有效推断、E 值、置信序列、区分度、校准斜率、抵押贷款提前还款

1 引言

当抵押贷款利率下降时，更多借款人会再融资，投资组合的提前还款速度可能快于模型的预测。监测团队必须据此判断：更新预测水平是否足够，还是模型得分与提前还款之间的关联已经发生变化。这一区别对于银行业所要求的模型审查至关重要 [3, 11, 28]。

将模型对 i 在 t月的评分对数优势比记为 $\theta_{it} = \operatorname{logit} \hat{} p_{it}$。Cox [9, 21] 的逻辑重新校准模型给出的提前还款概率为

$$\mathbb{P}(Y_{it} = 1) = \sigma(\alpha_t + \beta_t \theta_{it}), \tag{1}$$

其中， α_t 是校准截距， β_t 是校准斜率 [32]。斜率衡量每单位所述对数几率的关联程度。可达到的曲线下面积还取决于截距和得分分布（附录 A），因此仅凭曲线下面积无法单独确定该关联的变化。将对数几率乘以正数 c 会使斜率除以 c，并保留排序 [29]。

给定过去情况、当前贷款簿和当月截距，我们将贷款结果建模为相互独立。给定观测到的数量，哪些贷款提前还款的似然只取决于斜率。我们使用三个监测器。固定起点斜率乘积将每月条件似然比相乘，以检验斜率的下界。阈值为 20 时，在原假设下其曾经越过该阈值的概率至多为 1/20。重启水平和 跟踪模型所述尺度上偏离截距容差的情况。重启交集规则对从连续月份起始的斜率置信序列取交集，并在这些序列没有共同斜率时发出信号。较新的绑定集合给出变化方向。

该应用基于 Freddie Mac 面板数据中的维护模型和冻结模型，这些数据覆盖 1999–2025 年。在固定斜率下对观测账簿和计数进行重采样，可以重现维护模型在开发集中的月度AUC方差的 78%，以及留出集中的 70%。从新的留出集起点开始，两条路径都跨过水平阈值，两个固定起点斜率乘积也都保持在阈值以下。重启交集规则在冻结路径上发出向下变化信号，并在补充维护运行中发出向上变化信号。在模拟中，连续混合避免了离散斜率先验的非网格宽度下限。在 Freddie Mac 面板数据上，网格实现与连续集合实现对五个开发集和冻结留出集变点日期的结果一致。

2 相关工作

任意时刻有效监测。 E 值和置信序列[18, 24, 36] 已被用于监测预测风险、校准、公平性和预测表现 [7, 8, 23],，包括存在干预的设置 [13]。Arnold 等人 [1] 使用概率积分变换（PIT）对概率预测校准进行序贯检验。Farran 的 PITMonitor [12] 在 PIT 上使用混合 E 过程，并估计变点。Wang 等人 [37] 使用e过程对风险预测进行序贯拒绝，Henzi 等人 [16] 则在信用申请中检验二元概率校准。我们的监测器针对指定对数优势尺度上的校准斜率。其截距不变性源于在合并似然之前以月度计数为条件（表3）。广义线性模型的似然比置信序列 [

20] 覆盖一个常数联合参数。

条件 E 值与变点检测。 每月因子是一个条件式（COND）E 变量 [15]。Hao 和 Grünwald 研究固定样本的 E 功效。这里的条件化分别在每个月内进行，因此条件均值界限可跨月组合。Turner 和 Grünwald [30] 使用反向信息投影，为 2×2 效应量构造置信序列。我们的条件似然使用连续评分，并纳入了

二元评分的优势比模型作为特例。我们采用序贯变化检测中的重启构造 [26, 27]。

预测模型验证。 病例组合可能改变区分度指标 [22, 31, 35]。基于模型的一致性 C^{M} [33, 34] 和基于模型的 ROC 曲线 [25] 为此类变化提供了参照。Wang [38] 使用反事实比较，评估柯尔莫哥洛夫–斯米尔诺夫统计量的下降有多少源于抽样变异或组合变化。监测实践还会使用累积准确率剖面、曲线下面积、基尼系数、稳定性指数和校准检验 [2, 4, 28]。我们将其与 ECB 统计量、基于模型的一致性以及月度联合逻辑回归再校准进行比较。我们还将斜率与其设定时的截距值一同拟合，以衡量水平偏移对该拟合的影响。

3 条件约化

固定一个月 t及其监测账簿 $\{\theta_i\}_{i=1}^{n}$；得分为样本外得分（生成 $\hat{}\,p$ 的模型仅使用该月训练截止点之前的数据进行训练）。令 $Y_i \sim \mathrm{Bern}(\sigma(\alpha+\beta\theta_i))$ 在该月内相互独立，$S=\{i:\ Y_i=1\}$，$M=|S|$，以及 $T=\sum_{i\in S}\theta_i$。

Lemma 1 (classical [6, 9, 10, 17]). $\mathbb{P}_{\alpha,\beta}(S)=\mathbb{P}_{\alpha,\beta}(M)\cdot\mathbb{P}_{\beta}(S\mid M)$ *with*

$$\mathbb{P}_{\beta}(S\mid M{=}m)=\frac{e^{\beta T}}{Z_m(\beta)},\quad Z_m(\beta)=\sum_{|A|=m}\exp\Big(\beta\sum_{i\in A}\theta_i\Big),\qquad(2)$$

条件伯努利分布 *[5, 6]*：在 β中以充分统计量 T为基础、且不含 α的单参数指数族。计数因子是在 α中以充分统计量 M和对数归一化常数 $\sum_i$对数$(1+e^{\alpha+\beta\theta_i})$构成的指数族。

对 M进行条件化会消去 α，因为它的似然因子对于所有大小为 M的事件集都相同。对于每个斜率值，对数空间初等对称递推会在 $O(nm_{\max})$时间内，计算从计数值 $m_{\max}$以内所有计数对应的 $\log Z_m(\beta)$。

4 斜率的任意时刻有效推断

固定错误水平 $0<\alpha<1$。令 F_t 包含截至第 t 个月的结果，以及第 1、……、$t+1$个月结果出现之前可获得的信息，包括投产数据和下一本账簿。第t个月的账簿和 β_t是 F_{t-1}-可测的。在给定 F_{t-1}和当月截距 α_t的条件下，账簿中的结果 Y_{it}是相互独立的 Bernoulli 变量，其概率为 $\sigma(\alpha_t+\beta_t\theta_{it})$。截距可能未被观测。第 t 个月的条件似然比为

$$E_t(\beta_1,\beta_0)=e^{(\beta_1-\beta_0)T_t}\,Z_{M_t}(\beta_0)/Z_{M_t}(\beta_1).$$

命题1（复合原假设下的 e过程）。在时间零固定 β_0 。令 E_t 为一个可预测的 $\beta_1<\beta_0$ 因子，或为此类因子的概率混合，其权重在 t月之前确定。在 H_0：$\beta_t\ge\beta_0$对所有 t、$\mathbb{E}[E_t\mid\mathcal{F}_{t-1}$、$M_t]\le 1$均成立。因此$R_T=\prod_{t\le T}E_t$是一个非负超鞅，且具有 $R_0=1$ 和 $\mathbb{P}(\sup_T R_T\ge 1/\alpha)\le\alpha$。

证明见附录 C。

命题 2（置信序列）。对于斜率域上的固定概率测度 Π，定义 $A_T(\beta)=\sum_{t\le T:\ M_t\ge 5}\{\beta T_t-\log Z_{M_t}(\beta)\}$ 且

$$\log\mathcal{M}_T(\beta_0)=\log\int e^{A_T(b)}\,\Pi(db)-A_T(\beta_0).$$

若 $\beta_t\equiv\beta^{\star}$ 属于斜率域，则过程 $\mathcal{M}_T$（$\beta^{\star}$）是均值为一的非负鞅，且集合 计算机科学$_T=\{\beta_0:\ \mathcal{M}_T(\beta_0)<1/\alpha\}$ 满足 $\mathbb{P}(\exists T:\ \beta^{\star}\notin$ 计算机科学$_T)\le\alpha$。

证明见附录 C。**R**

起点、i交集 i在ru l

e. 在第 t个月，我们将从命题 2 得到、并在各个 $r\le t$处起始的精确集合 CS$_{r,\,t}$相交，所用月份为r至 t。报警时间 τ是首次出现空交集的月份。在月度条件模型下，如果定义域内的斜率恒定，精确集合规则的平均误报时间至少为 $1/\alpha$，或在 $\alpha=0.05$ [26]时为 20 个月。当最大的下端点超过最小的上端点时（在接受边界处相等的情况除外），交集为空。如果约束集合的起始点各不相同，较新的集合位于较旧集合上方时，我们将警报记为向上；若位于其下方，则记为向下。运行长度界的推导见附录 C。

我们建议对 [0.20, 1.60]采用连续均匀先验。在似然评估之间，A_T的凹性使其下界由其弦给出。对指数化的弦积分，可得混合分子的下界，进而得到一个外部置信集。每次查看时都可以细化积分网格，以控制数值差距。抵押贷款面板序列及其交集使用同一区间上的等权 29 节点先验（附录 B）。

重启水平和。计数因子在固定 β 后，为 α 提供似然，其对数归一化常数为 $\log C(\alpha,\ \beta)=\sum_i\log(1+e^{\alpha+\beta\theta_i})$。在斜率固定且截距在其窗口内保持不变的工作模型下，水平序列是有效的。对于水平轴，我们使用 $\beta=1$，即模型所注明的对数优势尺度，因此 $\alpha=0$ 对应于所注明的概率。为监测容差 $|\alpha_t|\le 0.10$，我们针对超出任一边界的备选值，以相应边界为基准，计算月度计数似然比的平均值，然后对所有起始月份的这些因子的乘积求和。重启水平和在 20 时发出信号。

5 固定斜率超额方差诊断

定义 1. 对于月度监测统计量 S_t，即任何基于当月账簿和事件集且观测方差为正的函数，固定斜率超额方差占比 为

$$\rho(S)=1-\mathrm{Var}_{H_0}[S]\,/\mathrm{Var}_{\mathrm{obs}}[S],$$

其中，Var_{H_0} 是跨月方差：将每个月的事件集替换为从(2) 中抽取的样本（固定 β），同时保持真实账簿和真实计数不变。

重采样在各个月份之间独立进行精确条件抽样。仅计数统计量和仅账簿统计量在每次抽样中都保持不变，并且当其观测方差为正时，$\rho=0$。原始对比量可能为负，展示的占比则截断为不小于零。一阶差分版本 ρ_{Δ} 衡量相邻月份之间的超额变异。超额变异可能反映斜率变异、依赖性或其他偏离基准的情况。区间采用对观测月份进行移动区块自助法计算，区块长度为 12。

6 实验设计

面板数据。房地美单户贷款层面数据集 [14] 包含 26 个批次（1999–2024 年），每个批次 50,000 笔贷款，共计 130 万笔贷款和截至 2025 年的 6,930 万贷款月数。截至 2015 年 12 月的面板数据包含 3,560 万贷款月数和 534,724 笔提前还款（每月 1.50%）。宏观协变量按预测日期合并，其中抵押贷款利率和失业率滞后一个月，FHFA房价滞后两个月，并采用州级插值。

模型与监测账簿。 我们使用每年1月基于滚动十年窗口重新训练的 LightGBM [19]，（maintained），以及样条逻辑回归广义线性模型。冻结变体保留指定年份拟合的评分函数。所有预测均为样本外预测。每个月度监测账簿包含20,000笔贷款。斜率分析至少需要五笔提前还款，且在计算对数几率前，会将概率截断至$[10^{-6}, 1-10^{-6}]$。监测器使用当前账簿的得分和结果。

窗口与阈值。 该设计使用2001年1月至2003年12月进行校准，使用2004年1月至2015年12月进行开发，将绝对容差设为 $\delta = 0.10$ 斜率单位，并在分析确定之前留出2016年1月至2025年8月的数据。LightGBM校准估计值为$\hat{\beta}_{\mathrm{ref}} = 0.759$（SE 0.007），其95% Wald区间的下端点为 $\beta^{\mathrm{lo}}{}_{\mathrm{ref}} = 0.746$，报警边界为 $\beta^{\mathrm{lo}}{}_{\mathrm{ref}} - \delta = 0.646$。相应的广义线性模型斜率和报警边界估计值分别为0.828和0.714。我们采用共同边界0.646，并在校准之后，于每个开发或留出窗口中第一个获得得分的月份开始计算固定起点斜率乘积。超过20是主要的单侧警报。预注册的单侧分析使用了重启斜率和（表10）。窗口、容差和假设均在下载贷款层面数据之前完成注册。

代码和数据可用性。 Freddie Mac SFLLD 样本依据其许可协议公开提供。宏观经济输入来自 FRED 和 FHFA。代码、配置和输出将随发布版本一同公开。

7 结果

7.1 固定斜率基准下的仪表盘变异

图1展示了两条开发路径上的月度仪表盘。表1比较了每个统计量的观测变异，以及其在开发期和按时间顺序排列的留出集中的变异。

实现AUC和ECB统计量。固定斜率基准在维持路径上重现了78%的实现AUC方差（95%区间：57–100%）。一阶差分的超额占比经截断后为0%（95%区间：0–25%）。在斜率固定值分别为1和0.875时，基准分别重现维持模型AUC方差的90%和83%；斜率固定值为0.875时，每月重新抽样四次。在frozen2003上，水平的超额占比为40%（0–57%），一阶差分为15%。ECB统计量在维持路径上的超额占比为13%（0–32%），在frozen2003上为0%（0–18%，经截断）。其Hanley–McNeil标准误随事件数变化，因此，该统计量既通过分母，也通过当前AUC依赖事件数。

在留出集上，基准重现了维持路径AUC方差的70%（95%区间：48–100%）。斜率固定值为1和0.875时，相应占比分别为86%和67%。在frozen2015上，原始控制方差是观测AUC方差的274%。将负的超额占比截断为零后，表1中显示的数值为0%。

校准斜率与一致性。 条件式和联合月度斜率拟合在报告精度下具有相同的超额份额，在 maintained 上为 62%，在 frozen2003 上为 76%。基于模型的一致性差距分别给出 60% 和 78%。在 maintained 路径上，预付款者中的平均得分 T/M 的截断超额份额为 0%。

固定截距拟合与观测数据中估计的截距的相关系数为 -0.963，与斜率为 1 的基准的相关系数为 -0.962。其月度标准差分别为 0.082 和 0.086。对于条件拟合，相应的相关系数为 $+0.347$ 和 -0.141。图2展示了即使斜率固定，截距约束也可能产生表观斜率变异。在 144 个开发期月份中，维护模型的条件斜率与滞后一期的 30 年期抵押贷款利率的皮尔逊相关系数为 -0.364。相应的斜率为 1 的基准相关系数为 $+0.023$。

7.2 开发窗口的两个维度

图4绘制了斜率为一时的累积截距与累积斜率之间的关系。当水平超出容差范围时，团队会重新校准截距。若有证据表明斜率已低于其边界，则将该模型转交审核其评分与结果之间的关系。

在0.646的共同斜率边界处，固定起点斜率乘积在2007年5月跨越样条广义线性模型中的20，并在其他五条开发路径上保持低于20（图5）。重启水平和在全部六条路径上均跨越（表12）。

重启交集规则在 maintained 上于2005年1月发出向上信号，在 frozen2003 上于2005年3月发出向下信号，在广义线性模型上于2006年3月发出向下信号，并在 frozen2009 上于2011年8月发出向下信号。maintained 的警报与监测开始后的首次1月重新拟合同时发生。其仅包含1月的集合始于0.765，高于始于2004年8月的集合上端点0.729。在相同的144个月度样本上，恒定斜率为0.759时，maintained 的零假设报警比例为6.15–6.25%，frozen2003 为6.45–6.54%，广义线性模型为6.74%。表13列出了所有样本序列和斜率固定值。

图3显示，maintained、frozen2003和广义线性模型的终端斜率分别为0.875、0.710和0.788。frozen2003序列拒绝斜率为0.746的恒定斜率模型，对数混合证据为11.5。相对于其自身0.714的校准边界，广义线性模型全窗口拟合得到的0.788处于容差范围内。

7.3 按时间顺序留出结果

重启水平和从2016年1月开始，在两条路径上均于2018年1月首次超过20。在maintained路径上，116个月中有69个月超过20；在frozen2015路径上则有70个月超过20。固定起点斜率乘积在两条路径上均保持低于阈值（表12）。

作为描述性背景，留出集结束时的累积斜率拟合值为：maintained（2004年1月至2025年8月）为0.907，区间为 $[0.890, 0.923]$；frozen2015（2016年1月至2025年8月）为0.760，区间为 $[0.740, 0.778]$。

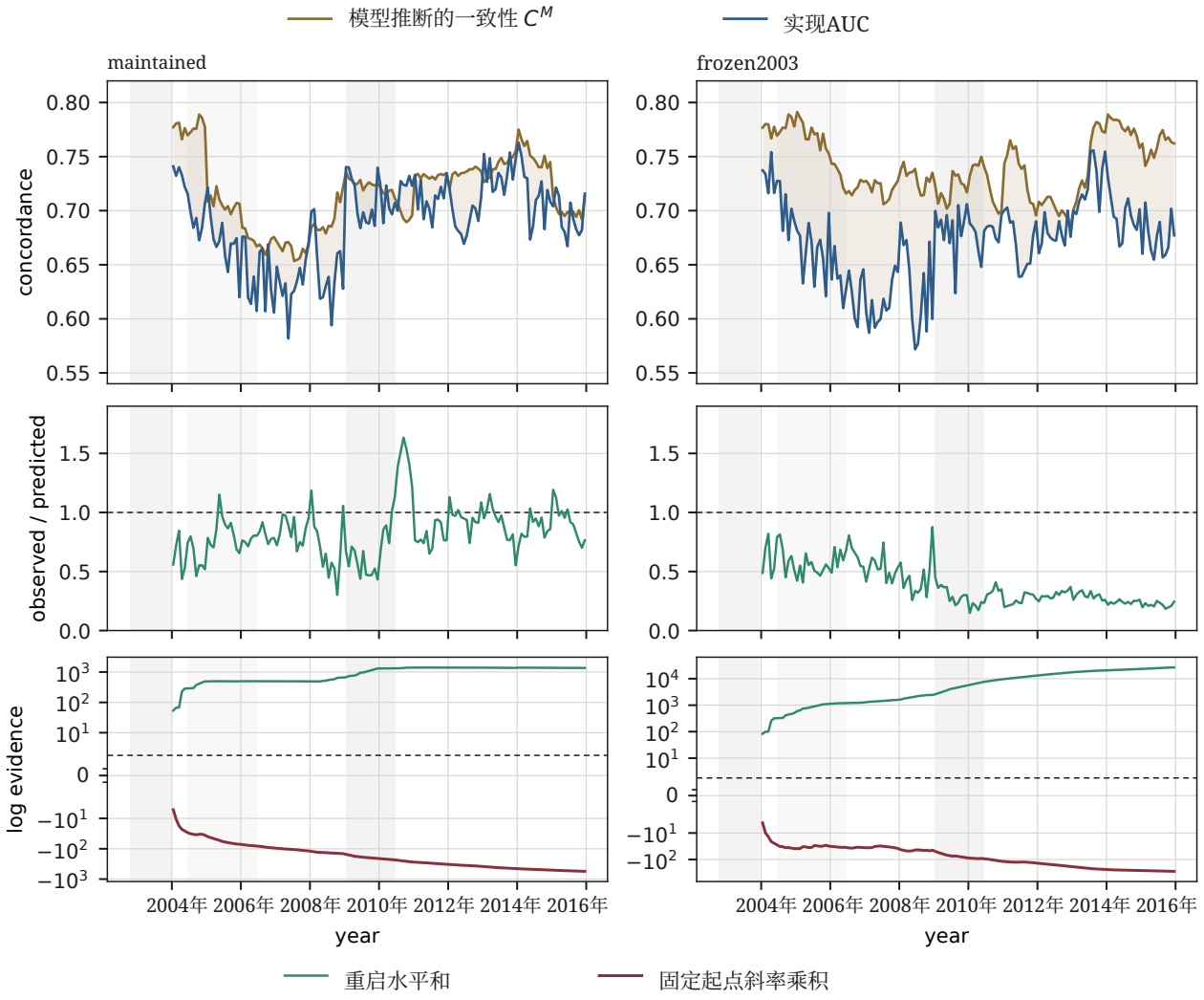


图1：月度AUC、实际与预期提前还款比及检测器证据，2004–2015年（左侧为维护模型，右侧为frozen2003）。阴影标示利率上升期和再融资期。顶部面板将实现AUC与基于模型的一致性 C^M 进行比较。底部面板显示重启水平和与固定起点斜率乘积，两者的阈值均为20。固定起点斜率乘积在两条路径上均低于阈值。

重启交集规则在2018年4月对frozen2015发出向下信号。在2016年1月启动的补充维护运行中，该规则于2019年8月发出向上信号。仅八月集合起始值为0.932，高于2017年4月启动的集合的0.916上端点。在这116个月度样本中，斜率为0.759时，maintained和frozen2015的零假设报警比例分别为4.10%和1.56%。斜率一时，两者分别为3.32%和1.86–1.95%（表13）。网格和连续集合实现对开发集和冻结留出集的五个报警日期结果一致（表7）。

7.4 同时覆盖率与区间宽度

在表1中的两条开发路径上，条件斜率与月度联合重新校准的差异最大为0.0014。

表2的面板A比较了每种设置下500条配对路径上基于相同累积条件似然的区间，贷款组合固定为4,000笔贷款，且斜率恒定。两种无界分配在第 t次查看时分别分配 $\alpha/[t(t+1)]$ 和 $6\alpha/(\pi^2 t^2)$。条件式 Chernoff 界给出每次查看时双侧对数似然下降量为 $\log(2/\alpha_t)$，对分配误差求和可控制在任意一次查看时漏掉真实斜率的概率。邦费罗尼法对36次查看中的每次分配 $\alpha/36$。附录表格还给出了渐近威尔克斯阈值下的覆盖率。面板B给出了斜率一和36次查看时的月度重新拟合基准。

在 36 次查看的比较中，置信序列和 Chernoff–Bonferroni 区间的宽度相近。在查看 240 次、每月预期事件数为 150 且斜率为 0.875 时，连续先验

表1：在 $\beta = 0.759$ 下对探索性仪表板进行比较，每月在观测到的账簿和计数上重新抽样 64 次。开发面板数据中每条路径有 144 个月，留出集有 116 个月。在开发阶段，在 maintained AUC 估计值处， ρ 的蒙特卡洛标准误约为一个百分点。区间为 95% 移动区块自助法区间，显示的占比下限截为零。 Δ 给出一阶差分的占比。S/N $= \sqrt{\max(\mathrm{Var}_{\mathrm{obs}}/\mathrm{Var}_{H_0} - 1, 0)}$。

开发							
	maintained				frozen2003			
统计量	ρ [95%]		Δ	S/N	ρ [95%]		Δ	S/N
实现AUC	22 [0,43]		0	0.53	40 [0,57]		15	0.82
ECB z（曲线下面积与开发）	13 [0,32]		0	0.39	0 [0,18]		9	0.00
差距 C^M – 曲线下面积	60 [42,70]		13	1.23	78 [70,83]		37	1.90
$\hat{\beta}$，联合再校准	62 [47,70]		26	1.27	76 [66,81]		34	1.77
$\hat{\beta}$，条件式（我们的方法）	62 [47,70]		26	1.27	76 [66,81]		34	1.77
$\hat{\beta}$，截距固定（消融）	0 [0,31]		2	0.00	2 [0,15]		0	0.15
预付款者的平均得分 T/M	0 [0,23]		9	0.00	17 [0,41]		2	0.46
观测值/期望值； C^M； PSI	$\rho \equiv 0$按构造							
留出集								
	maintained				frozen2015			
统计量	ρ [95%]		Δ	信噪比	ρ [95%]		Δ	信噪比
实现AUC	30 [0,52]		16	0.65	0 [0,0]		20	0.00
ECB z	0 [0,20]		16	0.00	0 [0,0]		12	0.00
差距 C^M – 曲线下面积	43 [28,55]		14	0.87	84 [65,87]		29	2.30
$\hat{\beta}$，联合再校准	53 [36,62]		0	1.05	84 [69,86]		29	2.30
$\hat{\beta}$，条件式（我们的方法）	53 [36,62]		0	1.05	84 [69,86]		29	2.30
$\hat{\beta}$，截距固定（消融）	8 [0,41]		2	0.29	0 [0,14]		3	0.00
预付款者的平均得分 T/M	0 [0,0]		0	0.00	0 [0,0]		0	0.00
观测值/期望值； C^M； PSI	$\rho \equiv 0$按定义							

结果显示，中位宽度为 0.0439，而 29 节点先验分布为 0.0580，两种无界支出规则分别为 0.0605 和 0.0615。它们各自的每次观察覆盖率分别为 98.4%、98.6%、99.4% 和 99.6%。在节点斜率为 0.70 时，离散混合的宽度为 0.0396，连续混合的宽度为 0.0430。在观察次数为 240 的所有设置中，两种混合方法的每次观察覆盖率均为 97.0–99.6%。两种支出规则覆盖了 98.8–100.0% 的路径，而连续序列的宽度比望远镜式区间窄 27.4–34.3%。

对于固定离散先验分布，节点之间的真实值可能比每个受支持斜率都有更高的期望对数似然。随着信息不断累积，最佳受支持节点设定的似然水平便会低于峰值，从而留下正宽度。在期望似然压力测试中，斜率为 0.875 时，经过 240,000 次等效外观后，29 节点宽度为 0.04995。采用精细连续均匀先验时，宽度为 0.00166，因为积分包含了任意接近峰值的斜率。

附录E报告了批次级重新校准，以及按账簿规模模拟的监测时长。

8 稳健性分析

重启斜率和。 表13中的每个截点都是回溯性的，根据该运行中已实现账簿和计数进行校准。在这些截点处，frozen2003与广义线性模型相交，而frozen2006保持在其截点以下。附录E和F分别报告合成比较结果和实际账簿校准结果。

曲率。 在每条路径上，按剖面似然确定的斜率乘积都保持在20以下，包括共同边界0.646处的广义线性模型。它在二次条件模型下检验固定得分处的局部斜率。八条路径上的中位绝对曲率范围为0.025至0.085。附录 E.2给出了模型适adequacy 检查、曲率模拟，以及对共同冲击和组级冲击的敏感性分析。

9 结论

以月度事件计数为条件，验证团队无需估计校准截距的路径，即可检验斜率容差。在抵押贷款面板数据中，固定斜率基准在开发集和留出集分别再现了 78% 和 70% 的维护 AUC 方差。在条件逻辑回归零假设下，起点固定斜率乘积为 20 时，曾经越过阈值的概率上限为 5%。在留出集中，重启水平和在两条路径上均越过阈值，而起点固定斜率乘积在两条路径上均低于阈值。重启交集规则在冻结路径上发出了向下变化信号，并在补充维护运行中发出了向上变化信号。

参考文献

[1] Sebastian Arnold、Alexander Henzi 和 Johanna F. Ziegel。2023 年。《预测校准的序贯有效检验》。《应用统计年鉴》17，3（2023），1909–1935。doi:10.1214/22-AOAS1697[2] 巴塞尔银行监管委员会。2005 年。《内部评级体系验证研究》。技术报告，第14号工作论文。国际清算银行。[3] 美国联邦储备系统理事会、美国货币监理署和联邦存款保险公司。2026 年。《修订后的指导意见》

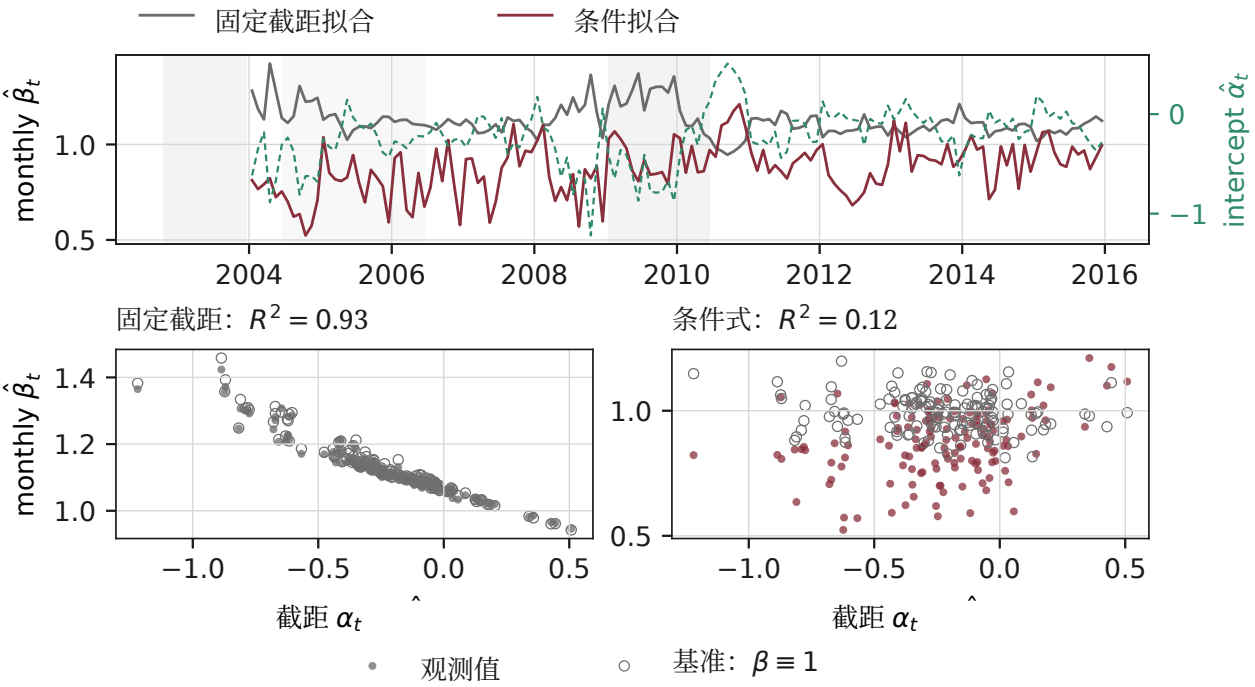


图 2：维持路径上的固定截距拟合和条件斜率拟合。上方面板显示月度拟合和估计截距。散点图比较观测拟合（实心点）与在 $\beta = 1$ 处重新抽样的结果（空心圆）。固定截距拟合将 α 固定在其校准值 0.274。

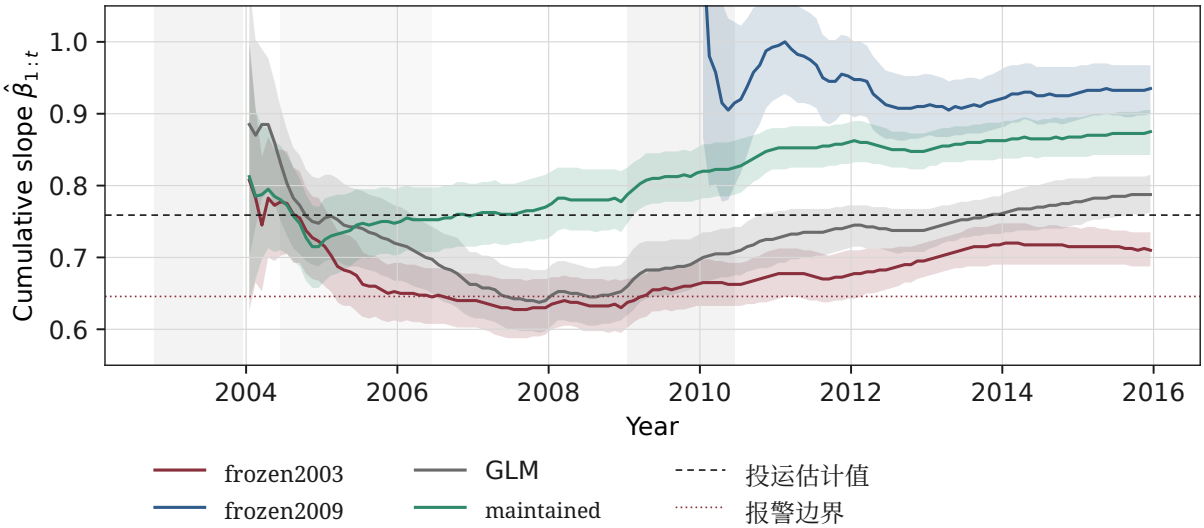


图 3：2004–2015 年累积条件斜率拟合和混合似然带，并标示投运估计值和报警边界。命题 2 中的覆盖率假设斜率恒定。

关于模型风险管理（*SR 26-2*）。技术报告。美联储。[4] Simone Casellina、Gaetano Chionsini、Raphael M. Kopp 和 Maroua Riabi。2026 年。《使用监管数据对违约概率模型进行系统性回溯测试》。技术报告，工作人员论文系列第24号。欧洲银行管理局。

[5] Sean X. Chen 和 Jun S. Liu。1997。泊松-二项分布与条件 Bernoulli 分布的统计应用。《中华统计学志》7, 4 (1997), 875–892。[6] Xiang-Hui Chen、Arthur P. Dempster 和 Jun S. Liu。1994。用于最大化熵的加权有限总体抽样。*Biometrika* 81, 3 (1994), 457–469。

表2：覆盖率和区间宽度。每个单元格给出每次查看/最终查看的覆盖率（%）、两个二项分布蒙特卡洛标准误（百分点），以及 [最终宽度中位数]。面板数据 A 在每种设置下使用 500 条路径，每月预期事件数为 150。真实值在 0.70 时位于 29节点网格上，在 0.875 时则不在网格上。连续组使用固定的 225 点网格生成的外层集合，在第 240 次查看时，对数积分差距最大为 0.647 纳特。Bonferroni/36 在第 36 次查看时结束。面板数据 B 使用单独的模拟，包含 800 条路径，斜率为 1，并在每个符合条件的查看时重新拟合。表4–6报告了所有斜率和事件率设置。

A. 累积条件似然

	36 种外观		240 种外观	
方法	0.70	0.875	0.70	0.875
CS, 29 个节点	99.6/100.0 (0.28/0.00) [0.1024]	99.0/100.0 (0.44/0.00) [0.1067]	99.6/100.0 (0.28/0.00) [0.0396]	98.6/100.0 (0.53/0.00) [0.0580]
CS, 连续均匀分布	99.6/100.0 (0.28/0.00) [0.1033]	98.8/99.8 (0.49/0.20) [0.1054]	99.4/100.0 (0.35/0.00) [0.0430]	98.4/100.0 (0.56/0.00) [0.0439]
望远镜式支出	99.6/100.0 (0.28/0.00) [0.1318]	99.6/100.0 (0.28/0.00) [0.1347]	99.6/100.0 (0.28/0.00) [0.0592]	99.4/100.0 (0.35/0.00) [0.0605]
反平方支出	99.4/100.0 (0.35/0.00) [0.1346]	99.6/100.0 (0.28/0.00) [0.1376]	99.4/100.0 (0.35/0.00) [0.0602]	99.6/100.0 (0.28/0.00) [0.0615]
切尔诺夫-邦费罗尼/36	99.8/100.0 (0.20/0.00) [0.1077]	100.0/100.0 (0.00/0.00) [0.1101]	—	—

B. 在 36 次外观下每月重新拟合，真实斜率为 1

	每月预期事件数		
方法	10	40	150
月度概况	15.88/93.88 (1.29/0.85) [1.20]	13.75/95.25 (1.22/0.75) [0.63]	14.88/96.00 (1.26/0.69) [0.34]
月度概况 邦费罗尼/36	94.25/99.62 (0.82/0.22) [1.82]	95.50/100.00 (0.73/0.00) [1.04]	94.38/99.88 (0.81/0.12) [0.57]
月度沃尔德	18.25/94.62 (1.37/0.80) [1.27]	15.00/95.25 (1.26/0.75) [0.64]	15.25/96.12 (1.27/0.68) [0.35]

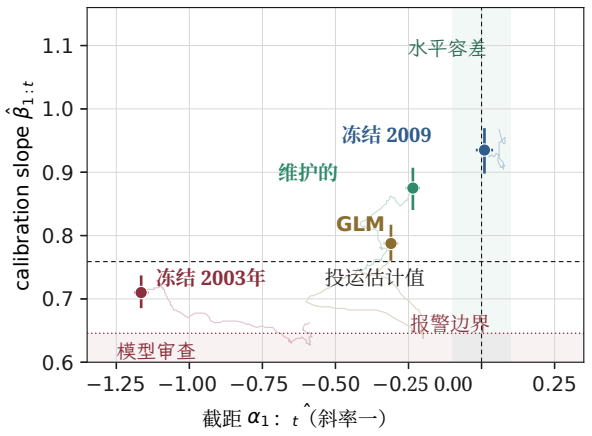


图4：斜率为一时的累积截距拟合（横轴）以及在所述对数优势尺度上的条件斜率拟合（纵轴）。实心点表示终端拟合及其边际95%带。浅色曲线表示预热期后的路径。竖直色带标示截距容差 $|\hat{\alpha}_{1:t}| \le 0.10$。水平色带位于共同斜率警报边界 0.646 以下。

[7] Yo Joong Choe 和 Aaditya Ramdas。2024。《比较序贯预测器》。*Operations Research* 72，4（2024），1368–1387。[8] Ben Chugg、Santiago Cortes-Gomez、Bryan Wilder 和 Aaditya Ramdas。2023。《通过投注审计公平性》。载于神经信息处理系统进展 *36*。[9] D. R. Cox。1958。《二元回归模型的另外两种应用》。*Biometrika* 45，3–4（1958），562–565。

[10] D. R. Cox. 1970. 二元数据分析。Methuen，伦敦。 [11] 欧洲中央银行。2019. 内部模型验证结果报告说明：信用风险*IRB*第一支柱模型。技术报告。欧洲中央银行，银行监管。 [12] Tristan Farran. 2026. 当模型不再奏效：任意时刻有效的校准监测。*arXiv*预印本 *arXiv:2603.13156* (2026)。[13] Jean Feng、Alexej Gossmann、Gene Pennello、Nicholas Petrick、Berkman Sahiner 和 Romain Pirracchio。2024。混杂医疗干预下的机器学习风险预测算法监测。载于第*27*届国际人工智能与统计学会议论文集。[14] Freddie Mac。2026。单户贷款级数据集。https://www.freddiemac.com/research/datasets/sf-loanlevel-dataset。访问日期：2026-07-22。[15] Yunda Hao 和 Peter Grünwald。2026。指数族的E值：一般情形。*Bernoulli* 32, 4 (2026), 3113–3142。[16] Alexander Henzi、Marius Puke、Timo Dimitriadis 和 Johanna Ziegel。2024。一种安全的Hosmer-Lemeshow检验。新英格兰数据科学统计学杂志 2, 2 (2024), 175–189。[17] Karim F. Hirji、Cyrus R. Mehta 和 Nitin R. Patel。1987。精确逻辑回归分布的计算。美国统计学会会刊 82, 400 (1987), 1110–1117。[18] Steven R. Howard、Aaditya Ramdas、Jon McAuliffe 和 Jasjeet Sekhon。2021。时间一致、非参数、非渐近置信序列。统计学年刊 49, 2 (2021), 1055–1080。[19] Guolin Ke、Qi Meng、Thomas Finley、Taifeng Wang、Wei Chen、Weidong Ma、Qiwei Ye 和 Tie-Yan Liu。2017。LightGBM：一种高效的梯度提升决策树。载于神经信息处理系统进展 *30*。[20] Junghyun Lee、Se-Young Yun 和 Kwang-Sung Jun。2024。广义线性模型的统一置信序列及其在多臂老虎机中的应用。载于神经信息处理系统进展 *37*。[21] Michael E. Miller、Siu L. Hui 和 William M. Tierney。1991。逻辑回归模型验证技术。医学统计学 10, 8 (1991), 1213–1226。[22] Anthony P. Morise、George A. Diamond、Robert Detrano、Marco Bobbio 和 Erdogan Gunel。1996。疾病患病率调整对逻辑预测模型准确性的影响。医学决策 16, 2 (1996), 133–142。[23] Aleksandr Podkopaev 和 Aaditya Ramdas。2022。跟踪已部署模型的风险并检测有害分布偏移。载于国际学习表征会议。

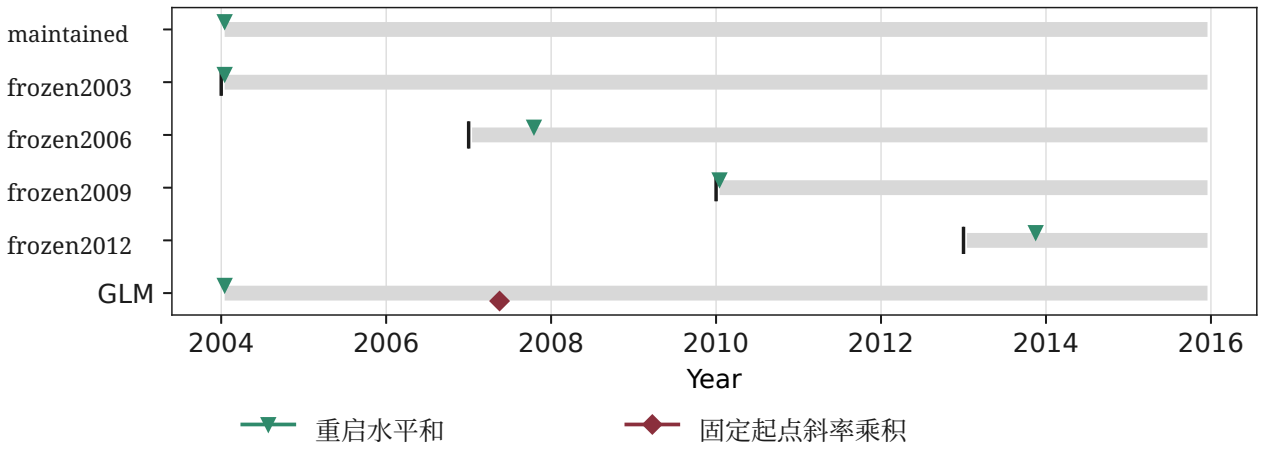


图 5：固定起点斜率乘积和重启水平和在阈值 20 时产生的首次开发预警。斜率边界为 0.646。竖线标记模型被冻结的月份。

[24] Aaditya Ramdas、Peter Grünwald、Vladimir Vovk 和 Glenn Shafer。2023。博弈论统计学与安全的随时有效推断。*Statist. Sci.* 38, 4 (2023), 576–601。[25] Mohsen Sadatsafavi、Paramita Saha-Chaudhuri 和 John Petkau。2022。基于模型的ROC曲线：考察病例组合和模型校准对ROC图的影响。医学决策 42, 4 (2022), 487–499。[26] Shubhanshu Shekhar 和 Aaditya Ramdas。2024。将序贯变化检测化归为序贯估计。载于第*41*届国际机器学习会议论文集（*PMLR*，第*235*卷）。44628–44642。[27] Jaehyeok Shin、Aaditya Ramdas 和 Alessandro Rinaldo。2024。E检测器：用于序贯变化检测的非参数框架。新英格兰数据科学统计学杂志 2, 2 (2024), 229–260。[28] Naeem Siddiqi。2017。智能信用评分：构建并实施更优的信用风险评分卡 （第2版）。Wiley。[29] Richard J. Stevens 和 Katrina K. Poppe。2020。临床预测模型验证：“校准斜率”究竟衡量什么？临床流行病学杂志 118 (2020), 93–99。[30] Rosanne J. Turner 和 Peter D. Grünwald。2023。列联表及其他情形的精确随时有效置信区间。统计与概率快报 198 (2023), 109835。[31] Wouter A. C. van Amsterdam。2024。病例组合变化下预测模型性能的因果视角。*arXiv*预印本 *arXiv:2409.01444* (2024)。[32] Ben Van Calster、David J. McLernon、Maarten van Smeden、Laure Wynants 和 Ewout W. Steyerberg。2019。校准：预测分析的阿喀琉斯之踵。*BMC*医学 17 (2019), 230。[33] David van Klaveren、Mithat Gönen、Ewout W. Steyerberg 和 Yvonne Vergouwe。2016。用于外部验证情境下风险预测模型的新一致性度量。医学统计学 35, 23 (2016), 4136–4152。[34] David van Klaveren、Ewout W. Steyerberg、Mithat Gönen 和 Yvonne Vergouwe。2019。基于校准模型的一致性：改进了对小样本患者聚类判别能力的评估。诊断与预后研究 3 (2019), 11。[35] Yvonne Vergouwe、Karel G. M. Moons 和 Ewout W. Steyerberg。2010。风险模型的外部有效性：利用基准值将病例组合效应与错误系数区分开来。美国流行病学杂志 172, 8 (2010), 971–980。[36] Jean Ville。1939。集体概念批判研究。Gauthier-Villars，巴黎。[37] Qiuqi Wang、Ruodu Wang 和 Johanna F. Ziegel。2026。E回测。管理科学 (2026 2209 00991 10 1287 2023 01659)。arXiv：。[38] 2026 doi：。/mnsc.。Yiqing Wang。。用于解释信用风险模型验证中KS性能恶化的反事实诊断框架。*arXiv*预印本 *arXiv:2604.11561* (2026)。

A 校准水平与曲线下面积的稀有事件极限

在 (1) 中，曲线下面积将一个事件的得分与一个非事件的得分进行比较，平局时计半分。这两个得分分别取自与 $f(\theta)\sigma(\alpha+\beta\theta)$ 和 $f(\theta)\{1-\sigma(\alpha+\beta\theta)\}$ 成正比的分布，其中 f 是该书的评分分布。当 $-\infty$ σ $\sim e^{\alpha}e^{\beta\theta}$ e^{α} 时，$(\alpha+\beta\theta)$。在该极限下，归一化的事件得分分布中的公共因子会相互抵消。在有限 α 下，两个类别条件分布都可能取决于截距。在 $\beta=1$ 处进行的探索性20万笔贷款计算中，将 α 从 -1.22 变为 $+0.51$ 会使可达到的曲线下面积改变0.0034。

B 数值实现

29节点先验在间隔为 0.05 的节点上赋予相等权重。$\log Z_{M_t}$ 的凸性使 A_T 为凹函数，并使 $\log \mathrm{M}_T(\beta_0)$ 为凸函数。在使其达到最大值的节点处，混合比至多为 1，因此接受区间包含一个节点。所包含节点的外侧邻点将其圈定在报告范围内。我们在 561 个点上对对数归一化常数进行插值，以供展示。直接检查表明，展示的开发期终端端点位于精确集合内。位于报告限值处的端点表示发生了截断。

固定起点斜率乘积和重启斜率和分别对低于边界 0.05、0.10、0.20、0.35 和 0.50 的偏移量处的似然比取平均。重启交集规则的网格实现会在报告网格上评估插值后的似然。对于连续集合复核，我们首先在 1,009 个节点上评估每个条件式似然，利用约束端点括区识别重启点，并通过直接评估似然来确定连续端点的范围（表7）。回顾性方向诊断对警报月份之前以及从警报月份起的累积斜率拟合取平均。在维护的开发路径上，这些平均值分别为 0.765 和 0.816。在维护的留出集重置上，这些平均值分别为 0.954 和 0.927。重置后的回溯性下降，与 2019年8月约束集合向上的分离方向不同。

对于表1，开发阶段原假设标准差的相对蒙特卡洛标准误为0.74%。将其传播到标准差平方后，原始超额份额约为1.47(1 – ρ)个百分点，或在 ρ = 22%时为1.15个百分点。

C 序贯保证的证明

命题1的证明。令 $\mathcal{H}_t = \mathrm{H}_t \vee \sigma(\alpha_t)$ 和 $\mathrm{G}_t = \mathrm{H}_t \vee \sigma(M_t)$。令 S_t 为事件集。给定 H_t，由独立性以及式 (1) 中的伯努利优势比可得，对于大小为 m 的子集 A，

$$\mathbb{P}(S_t = A \mid \mathcal{H}_t) = \frac{\exp\{m\alpha_t + \beta_t \sum_{i \in A} \theta_{it}\}}{\prod_{i} (1 + \exp\{\alpha_t + \beta_t \theta_{it}\})}.$$

将其除以所有大小为 m 的子集对应概率之和，公共截距和分母便会约去。因此，给定 G_t，对于截距的每一种取值，S_t 在 $b = \beta_t$ 处均服从式 (2) 中的分布。

固定该历史和计数。记 $\mathbb{E}_b$ 为斜率为 b 时所得有限条件族下的期望。对于固定的 $\beta_1 < \beta_0$，似然比是充分统计量的一个递减函数 $g(T_t) = E_t(\beta_1,\ \beta_0)$。将似然比对其分母分布求和，得到 $\mathbb{E}_{\beta_0}[g(T_t)] = 1$。对有限指数族求和式求导可得

$$\frac{d}{db} \mathbb{E}_b[g(T_t)] = \mathrm{Cov}_b(g(T_t), T_t) \le 0.$$

事实上，协方差的两倍等于 $\mathbb{E}_b[(g(T_t) - g(T_{t'}))(T_t - T_{t'})]$，其中 $T_{t'}$ 是一个独立副本，且被积函数非正。这也适用于退化计数，此时比值为一。当 $M_t < 5$ 时，我们也将该因子设为一；这一选择关于 G_t 可测。对于 $b = \beta_t \ge \beta_0$，可得 $\mathbb{E}[E_t \mid \mathrm{G}_t] \le 1$。对 $\beta_1 < \beta_0$ 进行可预测的概率混合，通过条件积分仍可保持该界。

首先对截距取平均，得到在 F_{t-1}，M_t 条件下的界。对计数取平均可得 $\mathbb{E}[E_t \mid \mathrm{F}_{t-1}] \le 1$。由于 R_{t-1} $t-1$ $\mathbb{E}[R_t \mid \ _{t-1}] \le$ 非负 R_{t-1} $R_0 = 1$ 且 F-可测，F 。从 开始，归纳可得可积性和非负超鞅性质。Ville不等式 [36] 给出 $\mathbb{P}(\sup_t R_t \ge 1/\alpha) \le \alpha$。□

命题 2 的证明。在 $\beta_t = \beta^\star$ 条件下，给定 G_t 的条件分布是每个比值 $E_t(u,\ \beta^\star)$ 的分母分布。由于分子和分母的支持集相同，且分子概率之和为 1，因此对于斜率域中的任意 u，其条件均值均为 1。塔式性质可得，在给定 F_{t-1} 的条件下均值为 1。因此，对于起点 r，每个乘积 $\prod_{s=r}^{t} E_s(u,\ \beta^\star)$ 都是非负鞅，并在 $r-1$ 处以 1 为初值。

在任意候选 b 处，定义重启的似然比混合

$$\mathcal{M}_{r,t}(b) = \int \exp\{A_t(u) - A_{r-1}(u) - A_t(b) + A_{r-1}(b)\}\, \Pi(du).$$

在 $b = \beta^\star$ 处，由于非负性，条件积分保持鞅性质。取 $r = 1$ 后，即得到命题中的均值为 1 的过程。对于每个起点 r，根据以 F_{r-1} 为条件的 Ville不等式，$\mathrm{M}_{r,\ t}(\beta^\star)$ 曾达到 $1/\alpha$ 的概率至多为 α。在时刻 t 排除 b，意味着 $\mathrm{M}_{r,\ t}(b) \ge 1/\alpha$，从而对每个固定起点 r 都有时间一致覆盖。□

重启交集规则的运行长度。令 Θ 为斜率域，即包含恒定真实斜率 $\beta^\star$ 的一个区间，并使用精确集合 $\mathrm{CS}_{r,t} = \{b \in \Theta : \mathrm{M}_{r,t}(b) < 1/\alpha\}$。这是 ShekharandRamdas [26]提出的前向重启构造。对于此处使用的当前集合交集，该界可直接由重启混合推出。在无变化且给定月度模型的条件下，$\sum_{r=1}^{t}$ $_{r\,t}$ $\beta^\star$ $\mathbb{E}[L_t \mid \ _{t-1}] = L_{t-1} + 1$ $L_0 = 0$ $\mathrm{M}_{,}(\,)$ 满足 F，且 。当交集为空时，某个集合排除了 $\beta^\star$，因此 $L_\tau \ge 1/\alpha$。精确集合在 Θ 中相对开，这是似然比连续性的结果。Θ 的一个可数稠密子集即可确定交集是否为空，因此 τ 是停止时刻。对 $L_t - t$ 应用有界可选停止可得

$$\mathbb{E}[\tau \wedge n] = \mathbb{E}[L_{\tau \wedge n}] \ge \alpha^{-1} \mathbb{P}(\tau \le n).$$

如果 τ 几乎必然有限，则取 $n \to \infty$，由单调收敛可得 $\mathbb{E}[\tau] \ge 1/\alpha$。如果 $\tau = \infty$ 以正概率发生，则均值为无穷大。

D 补充比较与研究结果

在每月 150 个预期事件下，使用渐近 Wilks 截断值的 Bonferroni/36，在斜率为 0.70 时，每次查看和最终查看的覆盖率分别为 99.0% 和 100.0%（蒙特卡罗标准误分别为 0.44 和 0.00 个百分点）。在斜率为 0.875 时，相应的覆盖率为 98.0% 和 99.8%（标准误分别为 0.63 和 0.20 个百分点）。表10记录了预注册标准和结果。

E 二次分析

重启斜率乘积和 [27] 为 $D_t = (1 + D_{t-1})\, E_t$，其中 $D_0 = 0$。在命题1的零假设下，$D_t - t$ 是超鞅，在阈值 20 下的平均误报时间至少为 20 个月。对于覆盖整个时域的比较，我们通过重采样校准阈值。

E.1 边界校准监测器比较

我们模拟了由 4,000 笔贷款组成、评分离散度呈周期性变化的月度样本。每次变化均从第 49 次观测开始。P1 在事件平均数不变的情况下，使斜率在 12 个月内从 1 降低。P2 对部分得分进行置换，同时根据未置换评分生成结果。N1 在 24 个月内提高截距，N2 压缩评分离散度，同时保持条件响应不变。校准针对每个账簿集合分别使用 400 条路径，以及监测器实际看到的账簿，并采用随机化并列决胜规则。在固定的账簿序列上，PSI 保持不变。其匹配后的报警率反映了随机化并列决胜规则。

带边界校准的重启斜率和在斜率为 0.646 时对 2.50% 的路径发出警报，在 0.70 时则对所有路径均不发出警报（表11）。在内部斜率 0.70 处进行校准时，该斜率下的报警率为 4.25%，边界处则为 23.25%。在行业阈值下，AUC 图表在两个零假设斜率处分别对 99.75% 和 100% 的路径发出警报，而 PSI 在两个斜率处均对所有路径发出警报。基准合成账簿上重启斜率和的对数阈值为 6.62。

在边界校准阈值下，重启斜率和在 99% 的路径上检测到向 0.55 的突变，中位延迟为 45 个月。得分 CUSUM 在 100% 的路径上检测到突变，中位延迟为 16 个月；在较大的截距偏移下，报警率为 81%。对于降至 0.85 和 0.70 的情况，重启斜率和均低于阈值，且都在容差范围内。在两种纯水平偏移和病例组合对照下，其报警率均为零。在两种纯水平偏移下，重启水平和在所有路径上均触发警报。

尖锐零假设比率所引用的是一个固定截距的AUC分布。在单独的账簿规模敏感性运行中，将截距增加1.2会使其AUC中心下移0.0060–0.0069；在这一变化下，其报警率从2,000笔贷款时的0%上升至50,000笔贷款时的100%。

表3：序贯监测方法的目标与适用范围。截距不变性指在保留监测参数的同时消除逻辑回归截距。破折号表示该特征不在方法的适用范围内。

方法	目标	截距不变	任意时刻有效	变点规则	真实金融应用
本文	校准斜率	Yes	Yes	重启交集规则	抵押贷款面板
Arnold 等人 [1]	PIT校准	—	Yes	—	—
Wang 等人 [37]	风险度量	—	Yes	—	市场数据
Farran [12] (PITMonitor)	PIT分布	—	Yes	混合报警与估计	—
Henzi 等人 [16]	二元概率校准	—	Yes	—	信用申请
Hao 和 Grünwald [15](COND)	指数族参数	是的，在此模型中	固定样本因子	—	—

表4：每月10个预期事件。每次查看覆盖率（%）（二项分布蒙特卡洛标准误，单位为百分点）[每个斜率和事件率下的最终宽度中位数]。每种设置包含500条路径。29节点网格上测试的斜率为0.70和1.00。按顺序，最后四种方法分别为望远镜式和反平方切尔诺夫支出、切尔诺夫-邦费罗尼/36以及威尔克斯-邦费罗尼/36。

斜率	CS 29	CS 均匀	望远镜式	逆平方	邦费罗尼/36	威尔克斯/36
36 次外观						
0.646	97.8(0.66) [0.3415]	97.8(0.66) [0.3404]	99.4(0.35) [0.4845]	99.4(0.35) [0.4949]	99.8(0.20) [0.3961]	98.4(0.56) [0.3320]
0.7	98.6(0.53) [0.3419]	98.6(0.53) [0.3408]	99.4(0.35) [0.4851]	99.2(0.40) [0.4955]	100.0(0.00) [0.3966]	99.4(0.35) [0.3324]
0.875	97.6(0.68) [0.3401]	97.6(0.68) [0.3390]	99.6(0.28) [0.4824]	99.4(0.35) [0.4927]	99.8(0.20) [0.3943]	98.8(0.49) [0.3305]
0.93	98.4(0.56) [0.3403]	98.2(0.59) [0.3392]	100.0(0.00) [0.4828]	99.8(0.20) [0.4931]	100.0(0.00) [0.3946]	98.6(0.53) [0.3307]
1	97.8(0.66) [0.3389]	97.8(0.66) [0.3379]	100.0(0.00) [0.4806]	100.0(0.00) [0.4909]	99.8(0.20) [0.3928]	98.4(0.56) [0.3292]
1.07	98.2(0.59) [0.3388]	98.2(0.59) [0.3377]	99.2(0.40) [0.4805]	99.0(0.44) [0.4908]	99.8(0.20) [0.3927]	98.6(0.53) [0.3291]
1.33	98.6(0.53) [0.3409]	98.6(0.53) [0.3399]	100.0(0.00) [0.4798]	100.0(0.00) [0.4885]	100.0(0.00) [0.3948]	98.8(0.49) [0.3316]
240 外观						
0.646	97.6(0.68) [0.1437]	97.6(0.68) [0.1434]	99.4(0.35) [0.2182]	99.4(0.35) [0.2218]	—	—
0.7	98.2(0.59) [0.1438]	98.2(0.59) [0.1434]	99.4(0.35) [0.2183]	99.2(0.40) [0.2220]	—	—
0.875	97.0(0.76) [0.1427]	97.0(0.76) [0.1424]	99.6(0.28) [0.2167]	99.4(0.35) [0.2203]	—	—
0.93	97.8(0.66) [0.1428]	97.4(0.71) [0.1424]	100.0(0.00) [0.2166]	99.8(0.20) [0.2202]	—	—
1	97.6(0.68) [0.1424]	97.6(0.68) [0.1422]	100.0(0.00) [0.2162]	100.0(0.00) [0.2199]	—	—
1.07	97.8(0.66) [0.1425]	97.8(0.66) [0.1421]	99.2(0.40) [0.2161]	99.0(0.44) [0.2197]	—	—
1.33	97.6(0.68) [0.1434]	97.6(0.68) [0.1429]	100.0(0.00) [0.2174]	100.0(0.00) [0.2210]	—	—

表5：每月预期事件数为40。每次查看覆盖率（%）（二项分布蒙特卡洛标准误，单位为百分点）[各斜率和事件率下的最终宽度中位数]。每种设置包含500条路径。29节点网格上测试的斜率为0.70和1.00。最后四种方法依次为望远镜式和反平方切尔诺夫支出、Chernoff Bonferroni/36，以及威尔克斯-邦费罗尼/36。

斜率	CS 29	CS 均匀	望远镜式	逆平方	邦费罗尼/36	威尔克斯/36
36 外观						
0.646	98.0(0.63) [0.1828]	98.0(0.63) [0.1823]	99.4(0.35) [0.2443]	99.4(0.35) [0.2495]	100.0(0.00) [0.1997]	98.8(0.49) [0.1674]
0.7	99.6(0.28) [0.1828]	99.6(0.28) [0.1823]	100.0(0.00) [0.2443]	100.0(0.00) [0.2495]	100.0(0.00) [0.1997]	99.2(0.40) [0.1674]
0.875	99.0(0.44) [0.1835]	99.0(0.44) [0.1830]	99.8(0.20) [0.2453]	99.8(0.20) [0.2506]	100.0(0.00) [0.2005]	98.8(0.49) [0.1681]
0.93	98.2(0.59) [0.1836]	98.2(0.59) [0.1832]	99.2(0.40) [0.2456]	99.2(0.40) [0.2509]	100.0(0.00) [0.2007]	98.2(0.59) [0.1683]
1	98.4(0.56) [0.1842]	98.4(0.56) [0.1838]	99.4(0.35) [0.2465]	99.0(0.44) [0.2518]	99.8(0.20) [0.2015]	98.4(0.56) [0.1689]
1.07	98.6(0.53) [0.1848]	98.6(0.53) [0.1844]	99.0(0.44) [0.2474]	98.8(0.49) [0.2527]	99.8(0.20) [0.2022]	99.0(0.44) [0.1695]
1.33	99.0(0.44) [0.1906]	99.0(0.44) [0.1901]	99.8(0.20) [0.2558]	99.8(0.20) [0.2612]	99.6(0.28) [0.2090]	98.0(0.63) [0.1752]
240 种外观						
0.646	97.8(0.66) [0.0741]	97.8(0.66) [0.0761]	99.4(0.35) [0.1098]	99.4(0.35) [0.1117]	—	—
0.7	99.4(0.35) [0.0737]	99.4(0.35) [0.0761]	100.0(0.00) [0.1099]	100.0(0.00) [0.1117]	—	—
0.875	99.0(0.44) [0.0814]	98.8(0.49) [0.0763]	99.8(0.20) [0.1101]	99.8(0.20) [0.1120]	—	—
0.93	98.0(0.63) [0.0806]	97.8(0.66) [0.0764]	99.2(0.40) [0.1103]	99.2(0.40) [0.1121]	—	—
1	98.2(0.59) [0.0745]	98.2(0.59) [0.0767]	99.4(0.35) [0.1107]	99.0(0.44) [0.1126]	—	—
1.07	98.4(0.56) [0.0806]	98.2(0.59) [0.0770]	99.0(0.44) [0.1113]	98.8(0.49) [0.1131]	—	—
1.33	99.0(0.44) [0.0832]	99.0(0.44) [0.0793]	99.8(0.20) [0.1148]	99.8(0.20) [0.1168]	—	—

E.2 二次校准

二次条件逻辑回归允许斜率随得分变化。对于 $x = \theta - c$，其中 $c = -4.35488$ maintained 校准中位数对数优势比，β 在 $\beta x + \gamma x^2$ 处的斜率为 c。月度似然比检验在144个维持的开发月份中的16个月以及116个留出月份中的19个月拒绝了线性形式（名义显著性水平为5%）。

表6：每月预期事件数为150。每次查看覆盖率（%）（二项分布蒙特卡洛标准误，单位为百分点）[在各斜率和事件率下的] 最终宽度中位数。每种设置包含500条路径。29节点网格上的测试斜率为0.70和1.00。按顺序，最后四种方法为望远镜式和反平方切尔诺夫支出、Chernoff Bonferroni/36以及威尔克斯-邦费罗尼/36。

斜率	CS 29	CS 均匀	望远镜式	逆平方	邦费罗尼/36	威尔克斯/36
36 个外观						
0.646	99.6(0.28) [0.1020]	99.8(0.20) [0.1029]	99.6(0.28) [0.1312]	99.6(0.28) [0.1340]	100.0(0.00) [0.1072]	99.0(0.44) [0.0899]
0.7	99.6(0.28) [0.1024]	99.6(0.28) [0.1033]	99.6(0.28) [0.1318]	99.4(0.35) [0.1346]	99.8(0.20) [0.1077]	99.0(0.44) [0.0903]
0.875	99.0(0.44) [0.1067]	98.8(0.49) [0.1054]	99.6(0.28) [0.1347]	99.6(0.28) [0.1376]	100.0(0.00) [0.1101]	98.0(0.63) [0.0923]
0.93	98.6(0.53) [0.1075]	98.8(0.49) [0.1063]	99.4(0.35) [0.1360]	99.0(0.44) [0.1389]	99.8(0.20) [0.1111]	98.2(0.59) [0.0932]
1	98.6(0.53) [0.1072]	98.6(0.53) [0.1076]	99.2(0.40) [0.1378]	99.2(0.40) [0.1407]	99.6(0.28) [0.1126]	97.6(0.68) [0.0944]
1.07	98.6(0.53) [0.1102]	98.6(0.53) [0.1092]	98.8(0.49) [0.1399]	98.8(0.49) [0.1429]	99.8(0.20) [0.1144]	98.4(0.56) [0.0959]
1.33	99.2(0.40) [0.1173]	99.2(0.40) [0.1166]	99.4(0.35) [0.1502]	99.4(0.35) [0.1534]	99.8(0.20) [0.1228]	99.0(0.44) [0.1029]
240 种外观						
0.646	99.6(0.28) [0.0402]	99.6(0.28) [0.0428]	99.6(0.28) [0.0589]	99.6(0.28) [0.0599]	—	—
0.7	99.6(0.28) [0.0396]	99.4(0.35) [0.0430]	99.6(0.28) [0.0592]	99.4(0.35) [0.0602]	—	—
0.875	98.6(0.53) [0.0580]	98.4(0.56) [0.0439]	99.4(0.35) [0.0605]	99.6(0.28) [0.0615]	—	—
0.93	98.0(0.63) [0.0562]	98.2(0.59) [0.0443]	99.4(0.35) [0.0611]	99.0(0.44) [0.0621]	—	—
1	98.4(0.56) [0.0415]	98.4(0.56) [0.0448]	99.2(0.40) [0.0619]	99.2(0.40) [0.0629]	—	—
1.07	98.4(0.56) [0.0571]	98.2(0.59) [0.0454]	98.8(0.49) [0.0629]	98.8(0.49) [0.0639]	—	—
1.33	99.2(0.40) [0.0586]	99.2(0.40) [0.0485]	99.4(0.35) [0.0675]	99.4(0.35) [0.0686]	—	—

表7：网格和连续集合实现中的首次报警。有符号宽度是报警时最小上端点减去最大下端点。方向根据较新的绑定集合确定。维护的留出运行仅作补充。

Path	运行开始	Grid	连续	方向	有符号宽度
maintained	2004年1月	2005年1月	2005年1月	Up	−0.03537
frozen2003	2004年1月	2005年3月	2005年3月	Down	−0.02176
GLM	2004年1月	2006年3月	2006年3月	停用	−0.06845
frozen2009	2010年1月	2011年8月	2011年8月	Down	−0.02649
frozen2015	2016年1月	2018年4月	2018年4月	Down	−0.06156
maintained	2016年1月	—	2019-08	Up	−0.01658

表8：条件模型充分性与单侧报警。LR 拒绝次数统计名义 $p < 0.05$的月份。日期指阈值为 20 时的重启斜率和。滞后列将前 12 个月的中位数曲率视为已知。分箱斜率使用每月得分十分位数。

			LR 拒绝		首次报警			分箱斜率		
Path	个月	中位数 $\lvert\gamma\rvert$	二次样条		线性	滞后	剖面似然	指定	线性	二次
开发										
maintained	144	0.031	16	22	—	—	—	0.890	1.023	1.028
frozen2003	144	0.071	47	46	2005年3月	2005年3月	2006年10月	0.665	0.921	1.006
frozen2006	108	0.085	40	37	2011年8月	2009年8月	2011年8月	0.802	0.940	1.023
frozen2009	72	0.045	14	17	—	—	—	0.969	1.030	1.064
frozen2012	36	0.037	11	10	—	—	—	0.896	0.990	1.050
GLM	144	0.083	40	39	2005年12月	2006年3月	2007年5月	0.767	1.015	1.002
留出集										
maintained	116	0.034	19	21	—	—	—	0.951	1.012	1.017
frozen2015	116	0.025	21	17	—	2024-04	—	0.773	1.006	0.995

限制性三次样条产生了22次和21次拒绝。八条路径上的中位数 $|\gamma|$ 介于0.025至0.085之间（表8）。

在所述得分上，maintained 和 frozen2015 的留出集十分位校准斜率分别为0.951和0.773。经过样本内线性拟合后，斜率分别为1.012和1.006；经过二次拟合后，分别为1.017和0.995。

二次链接函数下的病例组合变化可能会改变线性斜率投影。在所测试的最接近非零曲率下，$\gamma = 0.10$ 以面板数据的对数优势单位计，半个标准差的评分漂移使前缀投影的每次查看覆盖率达到95.6%，而固定终点投影的每次查看覆盖率为79.2%。重启斜率和及其模拟的120次观察阈值（对数值6.58）在边界处对2.2%的路径发出警报，而线性链接函数下这一比例为4.2%。表9包含曲率更强的设置。

限制性三次样条使用月度评分第10、35、65和90百分位作为节点。在二次模拟中，以总体均值为中心对平方项进行中心化会改变截距；将评分原点平移至 c 则会使线性系数改变 $2\gamma c$，同时保持 γ 不变。前缀投影会最大化截至该次查看时的预期累积条件对数似然。在 $\gamma = 0.30$ 处，漂移相同的情况下，边界报警率为7.8%，每次查看的前缀覆盖率为88.2%（表9）。

表9：二次逻辑模拟，在每种设置下沿500条路径进行120次查看，每次查看有40个预期事件。漂移衡量平均得分以账本标准差为单位的偏移。比率以百分比表示（蒙特卡洛标准误以百分点表示）。每次查看时的覆盖率，目标为该前缀的投影或固定终点投影。

β	γ	漂移	最终投影	警报	前缀覆盖率	终端覆盖率
0.646	0	0	0.6460	4.2 (0.90)	98.8 (0.49)	98.8 (0.49)
0.646	0	−0.5	0.6460	4.2 (0.90)	98.8 (0.49)	98.8 (0.49)
0.646	0.1	0	0.7878	0.2 (0.20)	95.8 (0.90)	95.8 (0.90)
0.646	0.1	−0.5	0.7271	2.2 (0.66)	95.6 (0.92)	79.2 (1.82)
0.646	0.3	0	1.1810	0.0 (0.00)	88.6 (1.42)	88.6 (1.42)
0.646	0.3	−0.5	0.9212	7.8 (1.20)	88.2 (1.44)	0.0 (0.00)
0.700	0	0	0.7000	0.0 (0.00)	98.6 (0.53)	98.6 (0.53)
0.700	0	−0.5	0.7000	0.0 (0.00)	98.6 (0.53)	98.6 (0.53)
0.700	0.1	0	0.8519	0.0 (0.00)	94.8 (0.99)	94.8 (0.99)
0.700	0.1	−0.5	0.7915	0.6 (0.35)	94.6 (1.01)	80.6 (1.77)
0.700	0.3	0	1.2669	0.0 (0.00)	90.0 (1.34)	90.0 (1.34)
0.700	0.3	−0.5	1.0130	1.0 (0.44)	87.6 (1.47)	0.2（0.20）

表10：开发窗口和留出窗口中的预注册标准与结果。

假设	结果
H1：C^M 解释了大部分已实现AUC方差（$R^2 \geq 0.7$；终止 < 0.3）	**未满足**：在各路径中为 R^2 0.068–0.450（< 0.3 在六个路径中的五个上）。
H2：maintained 上无斜率警报；冻结状态下有斜率警报在冻结状态下	**已满足**：在开发阶段，按预注册的重启斜率和，阈值为 20：maintained 在 144 个月内未触发警报（峰值对数证据为 1.8，低于 3.0）；frozen2003 于 2005年3月触发警报。**未满足**于留出集：maintained 达到其无警报标准（峰值 −2.7），而 frozen2015 也低于阈值（峰值为 1.41）。其重启交集规则于 2018年4月发出警报。固定起点斜率乘积为在面板结果可用后被指定为主要单侧统计量。
H3(i)：在纯水平（N1）和纯病例组合（N2）阴性结果中，当水平敏感基线触发了	**在区分度方面达标** 于预注册评估中。在边界校准下，重启后的斜率和在N1或N2下均未产生警报，而重启水平和在每条N1路径上都产生警报。
H3(ii)：中位检测延迟 ≤ 是最佳方法的一半基线的	**未满足** 预注册评估。对于在边界校准下低于容差至0.55的突变，重启斜率和的中位延迟为45个月，而得分CUSUM的中位延迟为16个月。
H5(a)：稀有事件近似 $\Phi(\sigma/\sqrt{2})$，适用于评分离散度 σ 跟踪 C^M 在 ±0.02内	**部分**：相关系数为 0.49–0.92；偏差 +0.010至 +0.034。
H5(b)：利率差离散度可以解释评分离散度 ($R^2 \geq 0.6$)	**符合**维持路径（$R^2 = 0.678$）。
H6：警报月份的顶部表现明显更差——十分位预付款、12个月现金流和匹配水平下的加权平均存续期误差	**小幅增量**：在水平误差中加入斜率证据，使 R^2 最多增加 0.021，适用于前十分位预付款误差。我们没有计算现金流或加权平均存续期终点指标。
误差	

要在二次条件模型下检验 $\beta_t \geq \beta_0$，令 $p_{b,\,g}(S_t \mid M_t)$ 表示其事件集概率。我们使用

$$\frac{q_t(S_t \mid M_t)}{\sup_{b \geq \beta_0,\, g \in \mathbb{R}} p_{b,g}(S_t \mid M_t)},$$

其中，q_t 将较低斜率的备择假设与曲率设为前12个月度拟合的中位数相结合。对所有事件集而言，上确界给出了真实零假设概率的上界，因此条件期望至多为1，且固定起点乘积具有相同的越界界限。对于经验比较，我们通过数值方法近似剖面似然。无约束拟合采用 2×10^{-5} 的梯度容差，并在返回的梯度范数高于 10^{-3}时拒绝该结果。当拟合值低于边界时，我们在边界处以 10^{-9}的标量容差对曲率进行最大化。在阈值20下，将曲率作为剖面似然参数消去后的重启斜率和，在与线性校准相同的三条开发路径上越界，另有两条路径较晚越界。固定滞后曲率后，重启斜率和也在2024年4月于frozen2015上越界（表8）。

E.3 依赖性和分数尺度

所有贷款优势比上的共同乘数会进入截距。在模拟对数阈值处（重启斜率和为 2.07，重启水平和为 -0.21），重启斜率和的共同冲击报警率为 4.8–7.4%，水平的报警率为 63.8–100%。当组水平脆弱性最高达到 0.40 时（这可能改变排列分布），重启斜率和的报警率范围为 3.6% 至 6.4%。

E.4 批次水平斜率估计

以每个批次内的事件数为条件，可为该分段拟合条件斜率。在对维护模型的探索性分析中，2006–07批次的拟合斜率为0.828，而1999–2001为1.068（图6）。将分析限制在共同的2011–2015日历窗口内，二者分别为0.762和1.103。在该窗口的前五个批次中，月度提前还款率介于1.73%和1.98%之间。在整个开发窗口的六个维护模型批次中，固定截距的均值

表11：在每种设置下的400条独立评估路径上进行144次外观时，边界校准的报警率和检测率（%）。在斜率为0.646的独立路径上，每个监测器的阈值均以其观测到账簿集合为依据，将第144次外观时的报警率设定为5%。零假设列采用基础账簿序列上的恒定生成斜率。P列给出首次报警发生在起始时点或之后的路径占比，并在方括号中列出检出样本中的中位延迟（月）。N列给出总体报警率，随后是相同的起始时点后汇总结果。短横线表示没有检出。

	零假设斜率		P1终端斜率			P2 置换比例	
监测器	0.646	0.70	0.85	0.70	0.55	0.10	0.20
重启斜率和	2.50	0.00	0.00 [—]	0.00 [—]	99.00 [45]	0.00 [—]	0.00 [—]
重启水平和	8.00	6.75	5.75 [34]	8.25 [21]	7.00 [28]	2.25 [22]	2.75 [23]
Sharp-null AUC 逻辑回归	3.00	0.00	0.00 [—]	0.00 [—]	0.00 [—]	0.00 [—]	0.00 [—]
AUC 控制图	7.00	6.25	4.75 [80]	4.00 [78]	4.25 [79]	6.50 [80]	8.00 [40]
得分CUSUM	4.50	10.75	98.50 [19]	100.00 [17]	100.00 [16]	91.50 [20]	99.75 [18]
校准检验	5.00	4.75	0.75 [19]	4.50 [17.5]	14.50 [17.5]	0.75 [17]	0.50 [15]
适当损失 e-检测器	8.00	2.25	0.00 [—]	0.75 [85]	2.25 [84]	0.00 [—]	0.00 [—]
mROC 差距	3.25	0.50	0.00 [—]	0.00 [—]	31.00 [56]	0.00 [—]	7.00 [59]

监测器	N1 +0.6	N1 +1.2	N2病例组合
重启斜率和	0.00 / 0.00 [—]	0.00 / 0.00 [—]	0.00 / 0.00 [—]
重启水平和	100.00 / 99.75 [12]	100.00 / 98.00 [7]	3.00 / 2.25 [17]
Sharp-null AUC 逻辑回归	0.00 / 0.00 [—]	0.00 / 0.00 [—]	0.00 / 0.00 [—]
AUC 控制图	6.75 / 2.75 [80]	5.25 / 0.75 [79]	16.50 / 16.50 [0]
得分CUSUM	80.75 / 80.75 [21]	81.00 / 81.00 [21]	74.25 / 74.25 [26]
校准检验	0.00 / 0.00 [—]	0.00 / 0.00 [—]	1.50 / 1.50 [15]
适当损失 e-检测器	100.00 / 100.00 [21]	100.00 / 100.00 [14]	0.75 / 0.75 [80]
mROC 差距	65.25 / 65.25 [54]	100.00 / 100.00 [19]	0.00 / 0.00 [—]

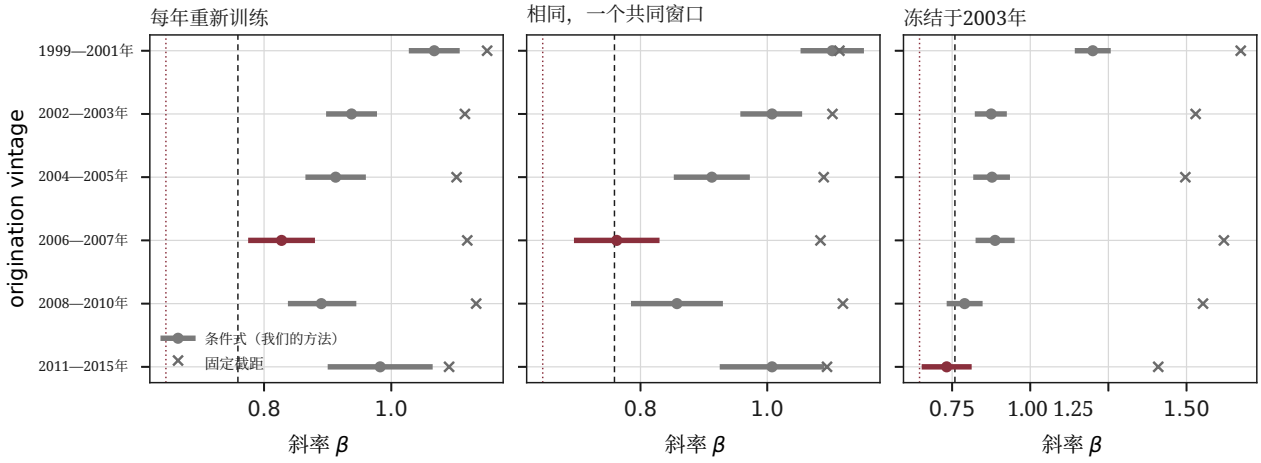


图6：探索性批次特定斜率拟合和混合似然带。每个批次使用 $\alpha/6$ 个。叉号表示固定截距拟合。中间面板使用共同的2011–2015日历窗口。

斜率拟合与月度平均提前还款率的皮尔逊相关系数为 +0.84 。

对于frozen2003，拟合结果显示，最新群组的斜率为0.732，最早群组的斜率为1.200。共享仿射逻辑校准会将所有分段斜率乘以同一个系数。因此，最新和最早的群组需要分别调整，才能使两者的拟合斜率都达到1。

E.5 监测设计（按账簿规模）

我们针对每种账簿规模和斜率模拟 200 条路径，每条路径包含 60 个月（图 7）；在每种规模下固定账簿，并选择截距，使不同斜率下的平均事件率保持不变。当斜率为 0.90 时，在检出中位时间内，整个置信带都降至 1 以下

对于账簿规模分别为 2,000、5,000、20,000 和 50,000 笔贷款的账簿，时间分别为 26、13、4 和 2 个月。斜率为 0.85 时，在三个较大的账簿上，将整个区间置于 0.90 以下所需的中位时间分别为 40、15 和 6 个月。在 2,000 笔贷款的账簿上，16% 的路径在 60 个月内达到该标准。斜率为 1 时，各账簿规模下低于 1 的相应比例为 0–1%，最终区间宽度的中位数范围为 0.158 至 0.032。

E.6 账簿规模与尺度敏感性

受监测账簿规模。 使用20,000、60,000和120,000笔贷款时，在48个月的开发期内，AUC超额占比分别为20%、15%和29%。在使用约200,000笔贷款的完整风险集的12个月子集中，该占比为11%。

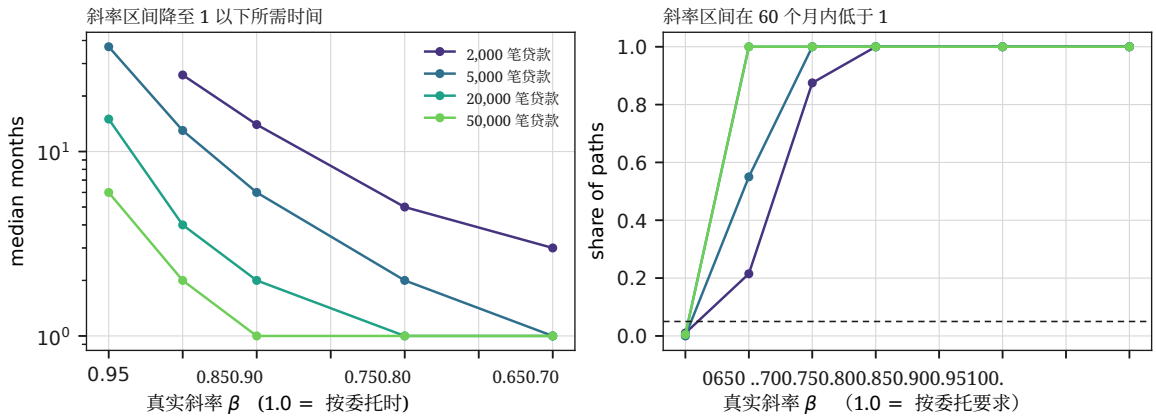


图7：模拟斜率带上端降至1以下所需的时间（左），以及在60个月内满足该标准的路径占比（右）。虚线表示5%的参考率。

保秩分数变换。将对数几率加倍或减半后，斜率拟合值分别为0.50和2.00。非线性单调变换得到的拟合值为0.64。不同变换下，实现AUC保持不变。在第19个月将对数几率加倍会在当月触发交集警报。这种敏感性在重新拟合时很重要，而冻结评分函数会保持其分数尺度。

衰减。监测规则使用$\delta = 0.10$斜率单位的绝对容差。在探索性分数噪声扫描中，对于真实绝对差距为0.15的情况，在事件强度匹配时，校准斜率与监测斜率之间估计的绝对差距从0.149缩小至0.072。在不同噪声水平和事件强度设置下，对于真实比率为0.85的情况，监测斜率与校准斜率的估计比率范围为0.845至0.860。

F 实际账本零假设校准

对于每次运行，我们保留观测到的月度分数和事件计数。在恒定斜率下，根据条件伯努利分布进行独立抽样，并使用与表1基准相同的后缀递归。在边界处，我们通过对1,024次完整重新抽样所得的最大值进行排序，校准重启斜率和。阈值为第$\lceil 0.95(1,\ 024 + 1)\rceil$个排序后的最大值，其中穿越定义为严格超过。每个零假设估计使用1,024条独立验证路径。这些实验以完整实现的账本和计数序列为条件，并描述该序列时域内的警报概率。

重启斜率和在 2005年3月的 frozen2003 数据中，以及在 2006年3月的广义线性模型中，越过了校准后的截止点。在 frozen2006 数据中，重启斜率和在 2011年8月于阈值 20 处越界。其峰值对数和为 4.83，低于校准后的日志截止点 5.46。在 144 个月运行的边界重新抽样中，重启斜率和在阈值 20 处于 49.2–57.8% 的路径上越界。在校准后的截止点处，所有运行的独立重新抽样中，警报比例为 3.8% 至 5.4%（表13）。在两条观测到的留出路径上，它均低于校准后的截止点（表12）。

为评估连续集合下的重启交集规则，我们在449个等间距的斜率值以及生成斜率处评估条件似然。若任一评估点位于所有重启集合中，则交集非空。对于其他外观，重启对数似然边际的最小值是凹函数。只有在两个相邻单元之一中，其最大值才可能超过网格中的最佳值。在每个单元内，我们用其弦给出各个边际的下界，并通过外推相邻割线给出上界。下界的重叠可以证明存在保留斜率，而上界不相交则可以证明交集为空。表13中的一个范围将模糊路径分别按无警报和有警报各计一次。这些范围框定了模拟比例。它们不同于蒙特卡洛不确定性。

表12：从全新起点开始的首次观测越界。两种斜率统计量均检验共同边界0.646。固定起点斜率乘积、重启斜率和以及重启水平和使用阈值20。校准列使用表13中的对数阈值。短横线表示无警报运行。

Path	起点	斜率			水平
		乘积	20处的和	校准后的和	20处的和
开发					
maintained	2004年1月	—	—	—	2004年1月
frozen2003	2004年1月	—	2005年3月	2005年3月	2004年1月
frozen2006	2007年1月	—	2011年8月	—	2007年10月
frozen2009	2010年1月	—	—	—	2010年1月
frozen2012	2013年1月	—	—	—	2013年11月
GLM	2004年1月	2007年5月	2005年12月	2006年3月	2004年1月
留出集					
maintained	2016年1月	—	—	—	2018年1月
frozen2015	2016年1月	—	—	—	2018年1月

表13：实际账簿和计数上的零假设概率估计（%）。重启斜率和的验证使用边界斜率。每个交叉列使用相应的恒定斜率。在恒定斜率模型下，连续集合实现的平均误报时间至少为20个月。范围反映数值夹逼结果。每个单元格使用1,024条独立验证路径。

Path	个月	重启斜率和			交点零假设斜率		
		日志截止点	在 20 时	在截止点	0.646	0.759	1.00
开发							
maintained	144	5.74	52.5	3.9	5.66	6.15–6.25	7.42–7.52
frozen2003	144	5.39	49.2	5.4	5.47	6.45–6.54	6.35
frozen2006	108	5.46	43.7	4.8	3.03	3.12	2.54
frozen2009	72	4.68	29.1	5.2	1.46	1.95	1.17
frozen2012	36	4.00	15.4	5.4	0.59	0.29	0.49
GLM	144	5.61	57.8	5.0	6.05–6.15	6.74	8.40–8.50
留出集							
maintained	116	5.76	49.3	3.9	4.20	4.10	3.32
frozen2015	116	5.21	43.7	3.8	0.98	1.56	1.86–1.95