arXiv:2603.07813v1 [econ.EM] 8 Mar 2026

美国经济衰退预测的风险状态转换

RahulBillakanti Minchul Shin*Wayzata*
High School 费城联邦储备银行*

2025年12月

摘要

我们提出一种简单的预测变量二值化方法——“风险状态”变换，作为衰退预测模型中使用连续型标准化变量这一惯常做法的替代方案。通过依据从训练数据中估计得出的阈值规则，将预测变量转换为反映异常疲弱状态的指标，我们展示了这种方法捕捉罕见事件（如美国经济衰退）的离散特征的能力。利用包含大量美国月度宏观经济与金融数据的面板数据，我们发现，二值化预测变量能够持续提升样本外预测表现，往往使线性模型足以与灵活的机器学习方法相媲美；此外，这种提升在经济衰退开始前后尤为显著。

关键词：衰退预测、机器学习、特征工程、风险状态转换、二值化预测变量、扩散指数
*JEL*代码：C25、C53、E32、E37

*感谢 Todd Clark、Frank Diebold、Domenico Giannone、Laura Liu、Benjamin Malin、Massimiliano Marcellino、Kenwin Maung 和 Jonathan Wright 提供宝贵意见。本文所表达的观点仅代表作者本人，并不一定反映费城联邦储备银行或美国联邦储备系统的观点。电子邮箱：rahulsai.billakanti11@gmail.com、minchul.shin@phil.frb.org。

1 引言

准确、及时地预测美国经济衰退仍是宏观经济学面临的一项核心挑战，直接关系到政策制定者、投资者和家庭。近期研究主要沿着两条路径推进：（i）通过变量选择、筛选或对大型数据集进行基于因子的聚合，识别具有信息价值的预测变量（Estrella and Mishkin，1998年；Stock and Watson，1993年；Chen 等人，2011年）；（ii）利用灵活的非参数方法和机器学习方法，揭示可能有助于提升衰退预测能力的非线性特征（Qi，2001年；Ng，2014年；Vrontos 等人，2021年）。

尽管取得了这些进展，这一领域文献的一个共同特征是对预测变量的处理方式。大多数研究无论是聚焦于变量选择、结合大量预测变量，还是采用非线性和非参数方法，通常都依赖原始数据输入。除平稳化和标准化等标准程序外，这些研究很少对变量进行额外的变换。根据本文引用的论文以及对《国际预测学杂志》相关文章的针对性考察，我们发现，现有关于经济衰退临近预测和衰退预测的研究通常避免对预测变量进行非线性变换。[1] 相反，这些方法直接将原始变量输入预测模型，将任何非线性留给参数或非参数建模框架本身来捕捉。

在本文中，我们采取了不同的视角。我们提出，可以在估计任何模型之前，直接将一种关键的非线性嵌入预测变量本身。为此，我们采用了所谓的“风险暴露”变换：在将预测变量纳入其他方面均属标准的预测模型之前，先对其进行简单的二值化。具体而言，当某个序列相对于其

[1]我们回顾了截至2025年8月发表在《国际预测学杂志》上的所有基于模型的衰退预测论文。当然，如果将范围从经济衰退二元预测扩大到产出增长或通货膨胀等连续型目标，已有研究采用了预测变量的非线性变换（例如，Goulet Coulombe 等人，2021年a；Daniele 等人，2025年）。此外，在经济衰退预测之外的一些应用中，逻辑回归也明确采用了预测变量的非线性变换。例如，在一篇近期论文中，Liu and Wang（2025年）分析了具有重尾协变量的二元结果，并表明其半参数尾部目标函数在渐近意义上等价于：使用极端预测变量的对数对尾部观测值进行逻辑回归。更详细的分析请参见附录D。

历史分布进入异常疲弱状态时，将该预测变量转换为取值为 1 的指示变量。虽然这种变换看似会舍弃有价值的信息，但我们认为，它非常适合预测罕见事件，例如美国经济衰退；在这类事件中，相关信号往往在于各项指标是否进入异常不利的区间。

利用大规模的美国月度宏观经济与金融数据面板，我们发现，在样本外经济衰退预测中，二值化预测变量始终优于标准的连续预测变量。这种改进在不同预测期、模型类别和聚合方法下均稳健。在存在线性关系的情形下，采用二值化输入的逻辑回归比其连续对应变量具有更强的区分能力。在非线性情形下，应用二值化后，其他基于树的方法几乎没有带来额外收益，这表明相关非线性很大程度上已由变换本身捕捉。我们还发现，通过主成分等方法组合二值化变量，能够提升预测性能。

本文通过表明重新审视预测变量的表示方式能够带来显著的预测增益，为稀有事件预测的研究与实践作出了贡献（例如，Lahiri 和 Yang，2013年）。我们提出一种简单、依赖数据的“风险暴露”变换，将预测变量转化为反映异常疲弱状态的指标。这种变换在不同预测期以及线性和非线性模型中，始终能够提升预测表现。由于该方法易于实施且计算成本低，因此可为经济衰退或违约等罕见事件的学术研究与实际预测提供一种实用基准。我们的提议还与现有文献的若干分支密切相关，包括预警规则、扩散指数、经济周期定年方法和基于树的方法；我们将在第2.3节回顾这些文献。

本文其余部分安排如下。第 2 节介绍所提出的风险转化方法及相关预测方法，并将其与现有文献联系起来。第 3 节介绍样本外预测设计和评估指标。第 4 节呈现主要实证发现。第 5 节通过比较不同模型的概率预测和变量重要性，探究这些增益的来源。最后，第 6 节通过验证这些发现是否在精简预测变量集中依然成立作结。我们还提供了其他稳健性

附录中的检验。

2 运用风险状态转换预测美国经济衰退

我们从一个用于预测美国经济衰退的标准模型开始，

$$P(y_{t+h} = 1) = f(X_t'\theta),$$

其中，y_{t+h} 是一个衰退指标：如果美国经济在时间 $t + h$处于美国国家经济研究局定义的经济衰退期，则其值为1。X_t 是一个 $N \times 1$ 预测变量向量，这些变量有助于预测美国经济衰退。函数 $f(\cdot)$将预测变量映射为经济衰退概率。对于$f(\cdot)$，一种常见的选择会得到逻辑回归或Probit回归。在我们的设定中，时间以月为单位，我们考虑$h = 3$，6，12，分别对应提前3个月、6个月和12个月的预测。

这一通用框架涵盖了文献中出现的大多数预测模型。关于在 X_t中纳入哪些内容，已有大量研究。例如，一些研究者发现，收益率曲线是经济衰退的有力预测变量（例如，Estrella and Mishkin，1996年；Wright，2006年；Rudebusch和Williams，2009年）。近年来，研究者发现，纳入大量经济指标能够提升预测性能：一种做法是在估计逻辑回归之前，对 X_t 应用降维技术；另一种做法是将正则化方法与逻辑回归结合使用。

另一项重要选择是 $f(\cdot)$ 的函数形式以及线性指标 $X_t'\theta$。一些研究者发现，随机森林或梯度提升树等非参数方法的表现可能优于逻辑回归等参数方法（Ng，2014年；D¨opke 等人，2017年；Vrontos 等人，2021年）。另一些研究者发现，长短期记忆网络和门控循环单元等深度学习神经网络展现出良好的效果（Qi，2001年；Chung，2024年）。总体而言，文献表明，与普通的逻辑回归模型相比，将非线性纳入模型可带来显著的预测增益；在普通逻辑回归模型中，X_t 是以线性形式纳入的。

2.1 风险状态转换

经济衰退之前，某些指标往往会持续走弱。与其利用某个预测变量的全部变化范围，不如关注该变量相对于自身历史表现是否处于异常疲弱状态，这可能更有参考价值。为将这一想法形式化，我们将每个平稳序列 x_{it} 转换为一个新的二元变量 z_{it}，并将其称为风险状态转换。

对于每个变量 x_{it}，风险状态转换定义为

$$z_{it} = \mathbf{1}\Big\{ s_i \bar{x}_{it}^{h_g} \le Q_{i,h_g}(\tau_g) \Big\}, \tag{1}$$

其中 $s_i \in \{+1, -1\}$ 表示该变量的周期性方向，顺周期变量取 $s_i = +1$，逆周期变量取 $s_i = -1$； $\bar{x}_{it}^{h_g}$ 是原始序列 x_{it}的 h_g个月移动平均，

$$\bar{x}_{it}^{h_g} = \frac{1}{h_g} \sum_{s=0}^{h_g-1} x_{i,t-s},$$

且$Q_{i,\ h_g}(\tau_g)$表示 $s_i \bar{x}_{it}^{h_g}$的历史分布的 τ_g分位数。

该转换将每个原始序列 x_{it} 转换为二元指标 z_{it}： 若 x_{it} 的历史移动平均低于由其历史分布的第 τ_g分位数给出的指定阈值，则该指标等于1。参数 τ_g 和 h_g 是该转换的关键调优参数。在下一节中，我们将讨论如何选择这些调优参数，并提供实践指导。在我们的实证分析中，每个变量被预先划分为顺周期或逆周期，完整列表见附录A.2。该分类主要基于经济理论，并在需要时由训练数据中的历史相关性加以支持。

核心思想是，相较于某些预测变量的完整变化范围，经济衰退与这些变量的极端值——即尾部行为——关联更为密切。通过聚焦这些尾部观测值，风险状态变换旨在提升预测性能，相较于使用

原始序列 x_{it} 直接进行预测。此方法能否提高预测准确性，归根结底是一个经验问题；我们的结果表明，这种特定的非线性变换显著提升了对美国经济衰退的预测能力。

2.2 聚合风险状态信号

应用风险状态 变换后，预测模型变为

$$P(y_{t+h} = 1) = f(Z_t'\theta),$$

其中， $Z_t = [z_{1t}, z_{2t}, ..., z_{Nt}]'$ 汇集了经过 N 变换的二元指标。这一表述将二元预测变量直接与衰退指标联系起来。将每个 z_{it} 解读为一个独立的"衰退信号"模型的输出时，单一指标 $Z_t'\theta$ 表示通过组合这些独立信号得到的聚合信号。函数 $f(\cdot)$随后将聚合信号映射为预测的经济衰退概率。

正如我们将在实证部分所展示的，"风险暴露"变换会将大量连续型宏观经济预测变量面板数据转换为二元状态矩阵 Z_t。尽管该矩阵可能包含预测信息，但其高维性带来了重大建模挑战。在预测模型中直接使用完整的未聚合矩阵，过拟合风险很高，并可能导致参数估计值不稳定。因此，本研究将探讨的另一个核心经验问题是，如何以最佳方式聚合并建模这些信息，从而最大限度地提升样本外预测表现。为此，我们考虑了若干聚合策略。

聚合策略。 每种聚合策略都会将 Z_t 降维为表示形式 W_t，其中 $\dim(W_t) \ll \dim(Z_t)$:

> **未聚合（基准）** 作为直接基准，我们的基准模型使用完整的二值化预测变量 Z_t 矩阵，不再进行进一步聚合。

简单平均 该预测变量是未聚合二值化预测变量的横截面均值：

$$W_t = \frac{1}{n}\sum_{i=1}^{n} z_{it}$$

类似于构建扩散指数时广泛采用的“计数规则”，Moore（1961）提供了一个经典示例。

无监督聚合（主成分分析）预测变量是 Z_t的前 K 个主成分：

$$W_t = V_k' \cdot Z_t,$$

其中， V_K 是一个 $N \times K$ 矩阵，其列为 Z_t样本协方差矩阵中对应于最大 K 个特征值的特征向量。我们的研究遵循 McCracken and Ng (2016)，设定 $K = 8$。

主要预测模型。对于每种聚合方法，我们考虑两类预测模型：

逻辑回归我们将 $f(\cdot)$设定为应用于线性指标 $W_t'\theta$的逻辑链接函数。在使用完整 Z_t的基准情形下，我们采用 ℓ_2 惩罚（岭回归）来缓解高维性导致的过拟合。

梯度提升我们还考虑一种非线性基于树的模型（XGBoost）。尽管风险状态转换已引入非线性，但基于树的方法能够捕捉预测变量之间的复杂交互，从而可能进一步提升预测性能。

2.3 二值化预测变量及相关文献

将经济指标转化为基于阈值的信号这一理念，已见于多个研究领域，其中最突出的是预警系统框架。在衰退预测中，一个著名的例子是负利差规则，即将收益率曲线倒挂（利差

< 0）解读为即将到来的经济衰退的预警信号（Laurent，1988年；Lahiri and Yang，2023年b）。类似地，萨姆规则 规定，当失业率的三个月移动平均值相较于前12个月三个月移动平均值中的最低值上升0.50个百分点或以上时，即识别为经济下行期；这一规则采用了特定阈值（Sahm，2019年）。

与我们所提出的方法最接近的先例是 Keilis-Borok 等人（2000），他们将地震预测中的模式识别算法改编用于预测美国经济衰退。他们将六项经济指标转换为二元信号，并在足够数量的指标越过预先设定的阈值时发出警报。他们证明，此类二值化变换包含线性模型未能捕捉到的预测信息。我们的贡献基于这一洞见，将二值化变量的应用扩展到高维环境。我们不再依赖一小组固定指标和启发式聚合规则，而是将风险状态 变换应用于广泛的宏观经济预测变量面板，并探索多种数据驱动的聚合方法，将这些信号合成为衰退概率。

对大量二元信号取平均，也被用于商业周期分析。例如，Moore (1961) 的经典研究提出了一种由二元指标简单平均构成的扩散指数；Mathy and Zhao（2025年）最近重新探讨了这一方法，并在其研究背景下分析这些扩散指数能否预测类似大萧条的经济衰退。

更一般地说，商业周期测定中所谓的“日期—平均”方法也基于同一原理：首先将各个序列划分为收缩阶段或经济扩张阶段，随后对这些二元分类进行聚合，以确定总体时间序列（例如，参见 Burns and Mitchell，1946；Harding and Pagan，2006年；Stock and Watson，2014年；Crump 等人，2020年）。这种聚合通常通过利用不同序列中转折点的聚类来实现。这一逻辑自然地启发了我们的方法：将连续型预测变量转换为表示收缩或扩张的二元信号，并根据这些信号与 NBER 衰退指标的历史关联生成预测值。

我们的方法也与树模型方法有关，因为我们的二值化变量可以

被解释为具有数据依赖型截断值的受限决策树桩。虽然我们的结果表明，简单的二值化预测变量线性聚合通常优于标准的基于树的方法，但我们也发现，在特定情境下，基于树的模型具有优势，尤其是在使用聚合因子提高预测性能时。

3 样本外预测评估

3.1 数据

我们的实证分析依托 FRED-MD 月度数据库。该数据库由 McCracken 和 Ng (2016) 建立，是用于宏观经济预测和商业周期研究的标准数据集。样本涵盖 1960年1月至2024年12月，包含 126 条时间序列（剔除部分序列后为 122 条），涉及经济中的关键行业，例如产出与收入、住房活动和金融市场。所有序列均按照 McCracken 和 Ng (2016) 的方法进行转换。目标变量（y_t）是美国国家经济研究局的衰退指标 USRECM，在经济衰退月份取值为 1，其他月份取值为 0。

3.2 模型设定与调优参数

模型设定。 我们的基准模型在逻辑回归和 XGBoost 设定中，均将 Z_t 及其滞后项作为预测变量。我们将纳入的滞后项集合 $\mathcal{L}$ 定义为 $\mathcal{L} = \{3, 6, 12\}$，其中包括短期、中期和长期滞后项。我们使用一组固定的滞后项，以避免对因预测期限而异的最优滞后项作出任何假设；这样模型便可从中选出最佳特征。对于采用 X_t 的模型，以及使用 Z_t 和 X_t聚合版本的模型，我们也采用相同的模型设定。

调优参数。 风险状态转换包含两个调优参数，即 τ_g 和 h_g。分位数水平 τ_g 决定了将连续数据转换为二元指标时所使用的阈值。该分位数根据移动平均 $\bar{x}_{it}^{h_g}$ 的经验分布计算，其中 h_g 为

窗口大小。为避免前视偏差，我们在预测性能评估中按如下方式设定 τ_g：基于初始训练样本（1960年1月至1989年12月），取衰退时期内各个体分位数中位数的中位数。该过程利用了关于预警信号何时会被激活的历史信息。随后，在评估样本的其余部分固定该数值，以确保不使用任何未来信息。我们在附录A.3.1中给出了选择 τ_g 的正式算法。窗口大小的默认选择为 $h_g = 1$（即不进行平滑）。较大的取值允许模型将滞后信息纳入预测变量本身，原则上可能改善预测值。然而，实际上我们发现，在保持 $h_g = 1$不变的情况下纳入 Z_t 的滞后项，能够带来较强的预测性能。

这些规则刻意设得简单而实用，为模型提供了自然的基准。在附录报告的稳健性检验中，我们表明，通过微调这些参数，预测增益还有进一步提升的空间。例如，τ_g 可以根据不同的预测变量类别或预测变量本身而变化（附录C.1）。此外，h_g 可以大于一，以纳入更多滞后信息（附录C.2）。尽管这些改进可能提高预测精度，我们仍保持基准设定简洁，以凸显即使采用这些简约的选择，模型也能表现出色。

计算。当逻辑回归包含许多变量时，我们采用 ℓ_2 正则化，并通过初始训练样本上的时间序列交叉验证来选择惩罚强度；为避免前视偏差，在评估期的剩余时间内将其固定不变（完整流程见附录A.3.2）。对于主成分分析，我们按照 McCracken and Ng (2016) 的做法，将 X_t 的因子数量固定为八个，并对 Z_t采用相同数量。所有计算均使用 Python 实现：逻辑回归模型通过带有 ℓ_2 惩罚项的最大似然法估计（使用 `scikit-learn`），非线性基准则采用带有叶权重 ℓ_2正则化的梯度提升法估计（使用 `xgboost`）。除非另有说明，XGBoost 均使用默认超参数，在本研究的应用中表现稳定且具有竞争力。

3.3 评估方法

为确保结果反映真实的预测能力，我们的整个评估都严格处于样本外情境中。我们采用递归预测设计，初始训练窗口为1960年至1989年。我们据此生成1990年1月的第一个预测值。此后，每生成一个预测值，训练数据便扩展一个月；并且在每一步都重新估计所有模型参数和聚合权重，以防止任何前视偏差。

在全文中，我们使用两个主要指标来评估二分类问题的预测质量：

精确率-召回率曲线下面积（PR AUC）。我们的模型会生成预测的衰退概率$\hat{p}_t$，并可通过决策阈值$\delta \in [0,1]$将其转换为点预测：

$$\hat{y}_t(\delta) = \mathbf{1}\{\hat{p}_t \geq \delta\}.$$

对于每个δ，我们在评估样本上计算召回率（真正率）和精确率（阳性预测值）：

$$R(\delta) = \frac{TP(\delta)}{TP(\delta) + FN(\delta)}, \qquad P(\delta) = \frac{TP(\delta)}{TP(\delta) + FP(\delta)},$$

其中，TP、FN和FP分别表示真正例、假负例和假正例。改变δ即可描绘出精确率-召回率曲线。精确率-召回率曲线下面积随后定义为

$$\text{PR AUC} = \int_0^1 P(\delta)\,\frac{dR(\delta)}{d\delta}\,d\delta = \int P(R)\,dR,$$

即随着决策阈值的变化，将精确率对召回率进行积分。

精确率-召回率曲线下面积尤其适用于不平衡数据集，例如衰退预测，因为它强调模型在罕见正类上的表现。与以假正率衡量表现的 ROC AUC 不同，精确率-召回率曲线下面积直接赋予精确率权重。这意味着

它对误报的惩罚要严厉得多；在不平衡情形下，这是一项重要优势，例如衰退预测中，非衰退时期的数量多于衰退月份。有关这一问题的详细讨论，以及 ROC AUC 与精确率-召回率曲线下面积在衰退预测及相关研究中的比较，请参阅 Lahiri and Yang（2023a）及其中的参考文献。ROC AUC 的说明和初步比较所得分数见附录 B.2。

基准精确率-召回率曲线下面积等于经济衰退的无条件概率，

$$\frac{\#\{\text{recession months}\}}{\#\{\text{total months}\}} = \frac{36}{420} \approx 0.086,$$

这意味着随机猜测的精确率-召回率曲线下面积为 0.086。该值提供了一个自然下限：任何模型都应超过这一基准，且精确率-召回率曲线下面积越高越好，其中 1 表示完美模型。

布里尔分数。 布里尔分数是一种严格适当评分规则，用于衡量概率预测的均方误差：

$$\text{BS} = \frac{1}{n}\sum_{t=1}^{n}(\hat{p}_t - y_t)^2,$$

其中，$\hat{}p_t$ 是预测概率，y_t 是实际结果（0 表示经济扩张，1 表示经济衰退）。与点预测的均方误差类似，布里尔分数由反映可靠性（校准，类似于偏差）和分辨率（区分能力，类似于方差）的项构成（Murphy，1973年；Diebold 和 Rudebusch，1989年）。随机猜测 $\hat{}p_t = 0.5$ 会得到 0.25 的布里尔分数，这可作为一个有用的上限。数值越小，概率预测的准确率和校准效果越好，其中 0 表示完美模型。

4 主要实证发现

本节提供了有关风险状态转换的核心实证证据。我们首先将 Z_t 的预测能力与连续输入和因子模型等替代方法进行基准比较。随后，我们评估其他聚合方案或非线性模型能否进一步提升表现。

表1：样本外风险状态转换（Z_t）与基准模型的表现 ks

模型配置	精确率-召回率曲线下面积			布里尔分数		
	$h=3$	$h=6$	$h=12$	$h=3$	$h=6$	$h=12$
所提出的方法（Z_t，Logit-ℓ_2）	**0.718**	0.370	**0.398**	**0.049**	**0.082**	**0.087**
替代设定						
X_t，Logit-ℓ_2	0.501	0.286	0.170	0.069	0.096	0.121
X_t的主成分分析，Logit-ℓ_2	0.552	**0.408**	0.314	0.064	0.083	0.108
X_t，XGBoost	0.584	0.338	0.351	0.062	0.085	0.098

注：所有模型均使用完整的预测变量集，并纳入同期值以及 3、6 和 12 个月的滞后项。“所提出的方法”模型是一个采用 ℓ_2 惩罚项、在二值化的“风险状态”矩阵（Z_t）上训练的逻辑回归模型。基准方法则在标准连续数据矩阵（X_t）上训练。粗体数值表示各项指标和预测期下表现最佳的模型。

最后，我们使用预测包容性检验，评估竞争方法是否能在所提出的方法框架之外提供额外信息。

4.1 “风险暴露”变换的预测价值

我们首先在传统线性框架下评估风险状态转换的表现：将基于未聚合 Z_t 矩阵训练、采用 ℓ_2 （岭回归）正则化惩罚项的逻辑回归模型（我们的基准模型），与基于完整 X_t 预测变量集训练的模型进行比较，并以采用主成分分析的标准因子模型作为基准。表1展示了3、6和12个月预测期的完整样本外结果。对于每种模型与输入的组合，我们报告精确率-召回率曲线下面积和布里尔分数。其他指标见附录。

样本外结果表明，我们所提出的“风险状态”转换（Z_t）在所有预测期均具有明显优势。概率准确率的提升尤为显著，所提模型的布里尔分数始终较低便是明证。所提模型的区分能力也显著增强，尤其是在 $h=3$上。

与以往研究一致，在将 X_t 纳入预测模型之前先提取共同成分，似乎能够增强整体预测能力。此外，我们还

发现，允许 X_t 以非线性方式进入预测模型，相较于其线性对应形式（例如， X_t 与Logit-ℓ_2），能够提升表现（例如， X_t 与XGBoost）。然而，我们所提模型优于这些替代设定与传统设定。这表明，我们的“风险暴露”变换似乎捕捉到了一种对衰退预测高度相关的非线性形式。

中期（$h = 6$）预测期的表现揭示了更为复杂的动态特征。所提模型的精确率-召回率曲线下面积为0.370，略低于其在12个月预测期的表现（0.398）。不同预测期之间的这种非单调模式已在其他研究中得到观察（Vrontos 等人，2021年）。值得注意的是，我们发现，在替代设定下，例如对 Z_t使用主成分分析时，这种凹性会消失，这表明性能曲线的形状对风险信号的聚合方式较为敏感。尽管如此，即使在 $h = 6$预测期，我们所提出的方法仍能提供校准程度更高的概率，并且总体准确率高于传统替代方法。

综合来看，表1中的证据为我们的核心假设提供了有力支持。“风险暴露”变换将大量连续数据转换为二元状态指标矩阵，从而为衰退预测构建了强大且稳健的特征集。我们的所提框架即使采用线性设定和简单的正则化，也被证明优于或至少具有很强竞争力，能够与包含标准因子模型和更复杂非线性分类器的基准方法相媲美。在确立该 Z_t 矩阵的基本价值后，我们接下来将转向通过聚合来改进这一信号的方法。

4.2 高级建模策略评估

在确立了 Z_t 矩阵的内在价值之后，我们现在探究其信号是否能够被提炼和增强。在本节中，我们对方法论中概述的不同聚合与建模策略进行比较，以确定风险状态特征集的最优建模方法。

表 2 展示了三种主要聚合策略的样本外表现：分解式基准、简单平均和无监督主成分分析方法。为考察

Table 2：替代聚合策略的样本外表现 Z_t

聚合方法	模型	精确率-召回率曲线下面积			布里尔分数		
		$h=3$	$h=6$	$h=12$	$h=3$	$h=6$	$h=12$
未聚合（基准）	Logit-ℓ_2	**0.718**	0.370	0.398	**0.049**	0.082	0.087
	XGBoost	0.583	0.300	0.269	0.065	0.089	0.099
简单平均	Logit-ℓ_2	0.541	0.365	0.190	0.151	0.188	0.229
	XGBoost	0.466	0.257	0.106	0.115	0.133	0.141
Z_t上的主成分分析	Logit-ℓ_2	0.688	0.528	**0.404**	0.062	0.081	0.106
	XGBoost	0.686	**0.567**	0.315	0.052	**0.063**	**0.086**

注： 此表比较了不同方法对 Z_t 矩阵进行建模的表现。“未聚合”行对应第4.1节中的基准模型。粗体数值表示各项指标和预测期下的最佳表现。

聚合与模型复杂度之间的相互作用，我们在全部三个预测期限内，分别使用带有 ℓ_2 正则化的逻辑回归和 XGBoost 模型评估每种策略。与上一节一样，我们展示了逻辑回归和 XGBoost 下各聚合方法的精确率-召回率曲线下面积和布里尔分数。

综合来看，表 2 的结果揭示了不同聚合策略如何与模型复杂度相互作用，从而影响预测性能。由此可以看出几种重要模式。

首先，Logit-ℓ_2 模型表现出显著更强的区分能力，且通常不会明显损失布里尔分数，在较短预测期内尤其如此。这表明，简单的线性分类器与我们的风险状态转换相结合后，仍然十分有效。

其次，正如前文对分解式基准的分析所指出的，XGBoost 并未持续改善表现。事实上，加入额外的非线性往往会降低预测准确率。这与表1中的结果形成对比：在原始 X_t 数据集上应用 XGBoost，大幅提升了其表现。这种差异表明，“风险暴露”变换有效地将预测问题相对于衰退指标线性化。通过将关键的非线性嵌入特征本身，该变换使简约且稳健的线性模型能够胜过更复杂的替代方案。

第三，简单 le平均聚合，类似于扩散指数或“计数规则 ,”

表现不佳。通过平等对待所有指标，它实际上抹去了各预测变量之间有价值的异质性。这一结果凸显了允许权重在不同信号之间有所差异的重要性，而不是仅仅计算当前处于风险中的指标数量。

最后，基于 Z_t 的 PCA 在中长期预测期（$h = 6$，12）中表现尤其出色。就精确率-召回率曲线下面积而言，它通常可达到或超过分解式基准模型的水平，同时还能生成具有竞争力布里尔分数的校准概率。对于短期限（$h = 3$），分解式基准模型的表现仍然最佳，但 PCA 也十分接近。值得注意的是，在与 XGBoost 搭配时，基于 Z_t 的 PCA 在 6 个月预测期取得了最低的布里尔分数，这表明它在某些设定下能够提高预测准确率。从实用的实时预测角度来看，基于 Z_t 的 PCA 具有显著优势。宏观经济数据以异步方式发布，导致样本末尾出现“锯齿边缘”。PCA 框架为处理这些缺失数据提供了一种自然且成熟的方法。因此，基于 Z_t 的 PCA 成为一种颇具吸引力的实时预测工具，可补充分解式设定在短期限预测中的出色表现。

4.3 预测值包容性检验

从各项指标本身来看，我们可以得出结论：我们的方法（Z_t 和采用 PCA 的 Z_t ）优于其连续输入对应模型。然而，重要的是正式检验这些方法是否提供了标准基准之外具有统计显著性的信息。为此，我们实施了预测包容检验。这类检验用于评估一个预测值是否包含另一个预测值中的全部相关信息，因此效率更高（Granger和Newbold，1973年）。早期应用包括 Fair和Shiller（1990），后来的研究则将这一框架扩展到概率预测（Clements和Harvey，2010）。我们的实施遵循相同思路，但具体做法有所不同：我们不将目标变量对预测概率进行回归，而是以竞争性概率预测的对数优势比为自变量，对二元衰退指标进行 Probit 回归。

更具体地说，我们估计一个概率单位模型，其中因变量 y_t 是时间 t 的 NBER 衰退指标。回归变量是我们

表 3：预测包容性检验结果

所提模型 (A)	基准模型 (B)	β_A 系数	β_A p值	β_B 系数	β_B p-值
$h = 3$					
Z_t (Logit-ℓ_2)	X_t (Logit-ℓ_2)	0.630^{***}	0.000	0.003	0.958
Z_t (Logit-ℓ_2)	X_t (XGBoost)	0.592^{***}	0.000	0.026	0.689
主成分分析 Z_t (Logit-ℓ_2)	对 X_t (Logit-ℓ_2)进行主成分分析	0.548^{***}	0.000	0.040	0.405
$h = 6$					
Z_t (Logit-ℓ_2)	X_t (Logit-ℓ_2)	0.331^{***}	0.000	0.049	0.444
Z_t (Logit-ℓ_2)	X_t (XGBoost)	0.298^{***}	0.000	0.049	0.288
对 Z_t (Logit-ℓ_2)进行主成分分析	对 X_t 进行主成分分析(Logit-ℓ_2)	0.587^{***}	0.000	0.020	0.730
$h = 12$					
Z_t (Logit-ℓ_2)	X_t (Logit-ℓ_2)	0.595^{***}	0.000	-0.097	0.405
Z_t (Logit-ℓ_2)	X_t (XGBoost)	0.350^{***}	0.005	0.077^{*}	0.069
对 Z_t 进行主成分分析(Logit-ℓ_2)	对 X_t 进行主成分分析（Logit-ℓ_2）	0.746^{***}	0.000	0.308	0.243

注： 该表报告了公式2所设定的Probit包容回归的估计系数（β）及对应的 p值。 Z_t 指分解后的“风险”矩阵，X_t 指完整的FRED-MD数据集。显著性水平分别为10%、5%和1%，对应以*、**和***表示。

所提模型和竞争基准的样本外预测概率。为确保回归变量无界，我们对每个概率序列应用对数优势比（Logit）变换， $L(\hat{p}_t) = \log\left(\frac{\hat{p}_t}{1-\hat{p}_t}\right)$。估计模型的设定如下：

$$P(y_t = 1 \mid \hat{p}_{t,A}, \hat{p}_{t,B}) = \Phi(\beta_0 + \beta_A \cdot L(\hat{p}_{t,A}) + \beta_B \cdot L(\hat{p}_{t,B})), \tag{2}$$

其中，$\hat{p}_{t,A}$ 是所提模型（A）的预测值，$\hat{p}_{t,B}$ 是基准模型（B）的预测值，而 $\Phi(\cdot)$ 表示标准正态累积分布函数。由于我们的目标是检验模型 A （所提出的方法）是否包容模型 B（基准），因此我们关注系数 β_B的显著性。如果 $\beta_B = 0$，那么在以所提模型的预测为条件后，基准模型的预测不会带来额外的预测能力。

表3所列的预测包容性检验结果，为我们所提出的方法具有优越性提供了有力且一致的统计证据。在大多数情况下，基准模型的系数（β_B）在统计上与零无显著差异，这表明基准

模型的预测值没有带来额外信息。虽然在少数情况下，β_B 与零存在显著差异，但即便如此，β_A 的幅度也远大于 β_B，这意味着基于 X_t的所提模型预测值优于基准模型的预测值。综合来看，这些结果进一步印证了前文的发现：与传统设定相比，风险状态转换能够提供更优的预测信息。

5 理解预测表现的驱动因素

为了更好地理解基准“风险状态”转换模型与连续预测变量基准模型各自的相对优势，本节将考察这两种模型生成的预测值及其背后的经济驱动因素。我们首先分析样本外概率，以评估各模型在发出经济衰退信号时的及时性和明确程度。接着，我们考察预测重要性在各变量和行业之间的构成，以评估表现差异是否源于不同的经济基础。

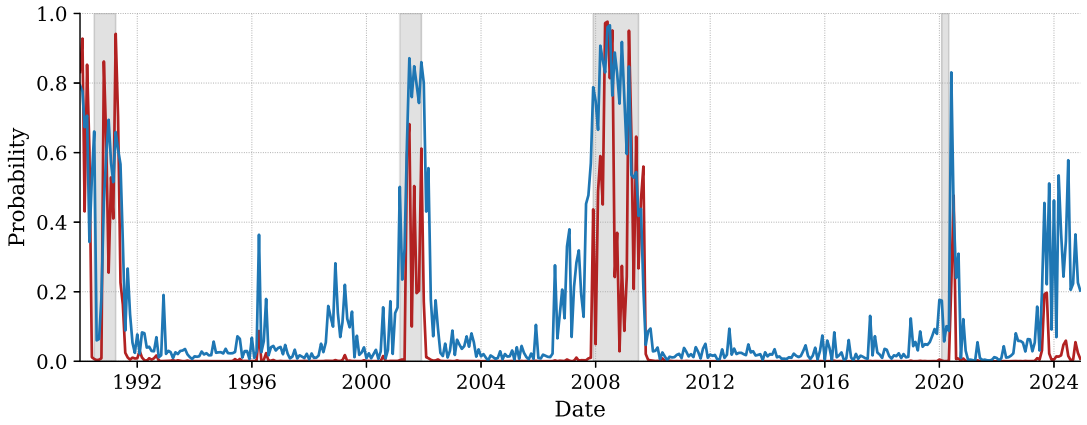
5.1 样本外概率

首先，我们从考察模型生成的概率预测入手。为直观展示这一点，图1呈现了基于标准“风险暴露”变换得到的样本外月度衰退概率——未聚合 Z_t (Logit-L2) 模型和完整 X_t (XGBoost) 模型。图中展示了三个预测期（$h = 3$，6，12 个月）的预测值，从而可以直接比较二元变量与其连续对应变量的表现。

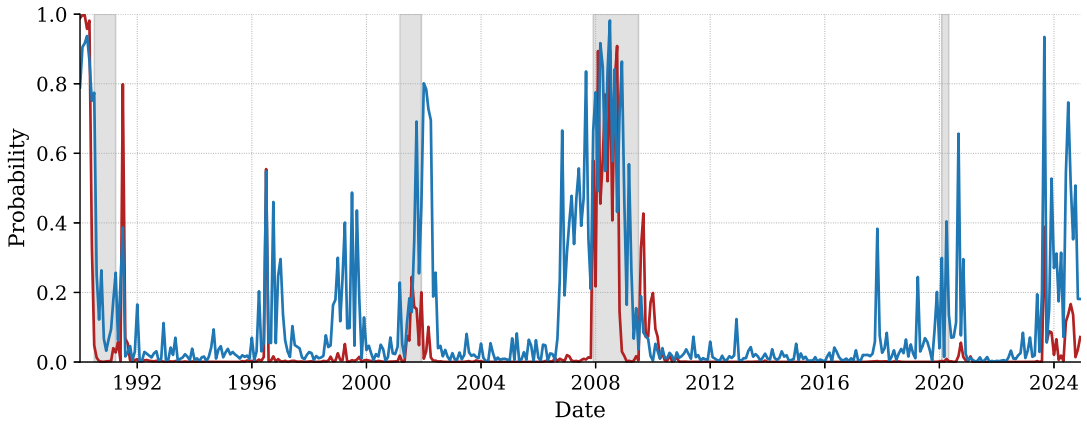
这些图展示了我们所提模型的若干特点，以及它与连续型基准之间的关系。首先，在所有预测期内，Z_t 的概率上升速度都远快于 X_t 的概率。该模型给出的概率通常会在1990年、2001年和2008年等经济衰退期间急剧上升，这表明，与连续变量相比，二值化转换能更有效地捕捉经济衰退的开始及其动态变化。

另一个值得注意的现象是，基于完整 FRED-MD 训练的 XGBoost 模型表现得较为谨慎

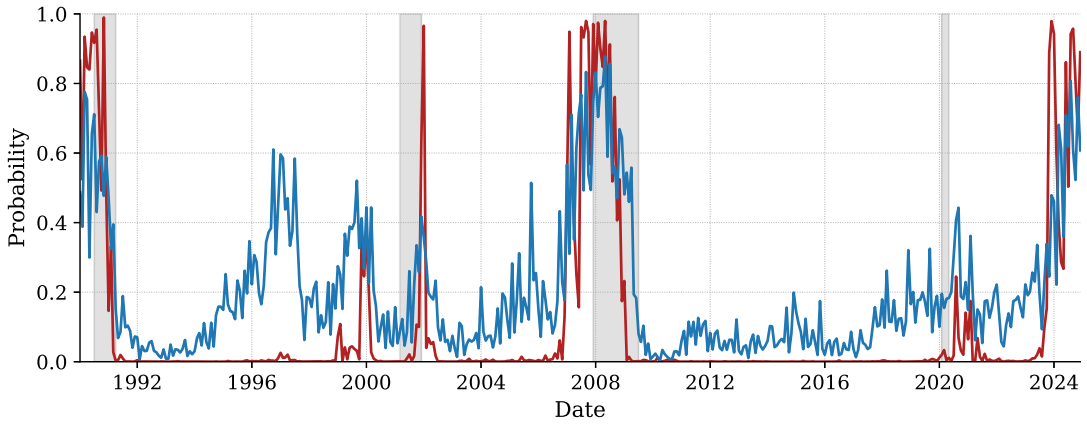
图1：未聚合 Z_t的样本外衰退概率



(a) 3个月期预测（$h = 3$）



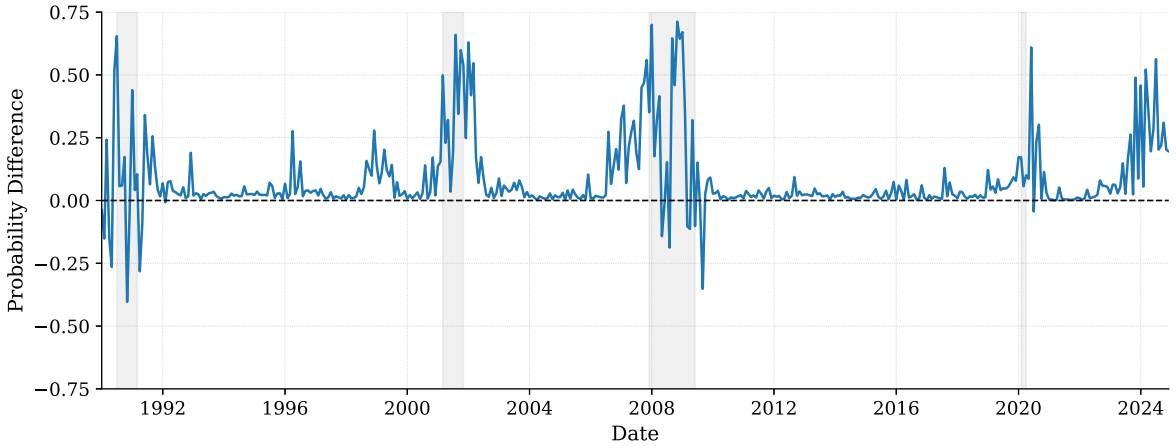
(b) 6个月期预测（$h = 6$）



(c) 12个月期预测（$h = 12$）

注： 该图绘制了我们主要 Z_t (Logit-L2)模型（蓝线）和完整 X_t (XGBoost)模型（红线）的样本外月度衰退概率。阴影竖条表示美国国家经济研究局官方认定的衰退时期。

图 2：所提模型与基准模型之间的预测分歧（$h = 3$）



注：该图绘出了我们主要的 Z_t (Logit-L2) 模型与 X_t (XGBoost) 基准模型在3个月期预测中的样本外衰退概率之差。正值表示所提模型给出的经济衰退概率较高。阴影竖条表示美国国家经济研究局官方认定的衰退时期。

数据集。在 $h = 3$，6，12 中，该模型在经济扩张时期给出的概率始终很低，即使出现了未导致官方认定衰退的经济动荡也是如此。然而，模型的这种犹豫不决正是其表现欠佳的原因。它往往十分谨慎，只有在出现最明确的经济下行信号时才会提高概率。例如，在 $h = 6$，面对2001年衰退（以相对温和著称），它给出的概率仅略高于0.2；而在 $h = 12$，它完全错过了这次衰退，直到衰退之后才提高概率。

另一方面，我们的方法在这种权衡中实现了更优的表现。在经济明显扩张的时期（例如2010年代），它在所有预测期的概率都保持在很低的水平，因此布里尔分数更优。它的敏感度也更高，反应更迅速，并且在经济衰退期间保持较高概率的时间更长，因此区分能力（精确率-召回率曲线下面积）更强。因此，我们的框架在敏感度和校准之间实现了很好的平衡，而连续预测变量模型难以做到这一点。

为进一步说明这一点，我们首先对我们的 Z_t (Logit-L2)模型与基准模型 X_t (XGBoost)在3个月预测期的预测分歧进行可视化分析。图2绘出了两者样本外衰退概率之差的时间序列。正值表示我们的“风险状态”模型给出的经济衰退概率高于XGBoost基准模型；负值则表示相反。

表 4： $h = 3$ 预测值的均方误差（MSE）分解

模型配置	**所提模型基准模型**	
	(Z_t, Logit-l2)	(X_t, XGBoost)
经济衰退期间的均方误差	0.222	0.513
经济扩张期的均方误差	0.031	0.015
全样本的均方误差	**0.049**	**0.062**

注： 该表报告了未来 3 个月的样本外预测的均方误差（MSE/布里尔分数），并按美国国家经济研究局确定的衰退期和扩张期进行分解。数值越低，表示预测校准效果越好。

该图揭示了使用二元预测变量和连续预测变量的特点。在美国国家经济研究局界定的1990年、2001年、2008年和2020年经济衰退开始前的几个时期，分歧序列持续飙升至正值区域。这表明，我们的“风险状态”转换比基准模型更早发出预警信号，且信号更为明确。此外，尽管存在一定波动，分歧序列在衰退时期往往为正值。

在经济扩张时期，分歧序列通常在 0 附近波动，表明这些概率之间并无显著差异。不过，该序列确实位于正值一侧，这印证了我们从图1得出的发现： X_t 和 XGBoost 的组合在经济扩张期校准良好。为进一步说明这一现象，我们按衰退期和经济扩张期分解了各模型的布里尔分数（均方误差），结果见表4。

如图2所示，使用 X_t 训练的 XGBoost 模型在经济扩张期的校准效果略好，但差异很小。相比之下，在严格的衰退期内， X_t 和 XGBoost 模型的布里尔分数显著高于我们的方法，这表明我们的方法在敏感度与校准之间实现了更好的权衡。

5.2 预测性能的经济驱动因素

前文各节已证实，采用“风险暴露”变换的模型具有更优的样本外表现。接下来，我们将分析推动这一表现提升的经济驱动因素。

表5：各框架中最重要的前10个预测变量（$h = 3$）

“风险状态”（Z_t） +Logit-L2		**连续型（X_t） + XGBoost**	
变量	重要性	变量	重要性
3个月联邦基金利差	0.704	5年期联邦基金利差	93.733
1年联邦基金利差	0.694	10年期联邦基金利差	71.358
6个月联邦基金利差	0.653	1年联邦基金利差	70.255
实际M2	0.649	6个月联邦基金利差	68.517
5年期联邦基金利差	0.623	3个月联邦基金利差	26.535
商业票据与联邦基金利差	0.622	总就业	26.478
10年期联邦基金利差	0.594	制造业平均工时	17.736
标普股息收益率	0.575	Aaa级与联邦基金利差	17.389
Aaa级与联邦基金利差	0.520	非农住房开工	17.373
标普500	0.491	就业：制造业	14.949

注： 重要性得分是对每个基础变量的所有相关滞后项进行汇总后得到的。Logit-ℓ_2模型的重要性是系数绝对值的平均值。XGBoost 模型的重要性是“增益”指标的平均值。这两项指标的绝对尺度不可直接比较；分析重点在于预测变量集的相对排名和构成。

我们首先考察哪些具体特征对所考察的 Z_t和 X_t 模型所做的预测值有所贡献，以了解“风险暴露”变换提升预测表现的关键驱动因素。表5列出了平均系数/重要性最高的前10个预测变量（对所有滞后项求和）。我们选取 $h = 3$ 作为本次分析的示例。

结果表明，各模型所学策略既有根本性的相似之处，也存在差异。两个模型都将利率和期限利差置于预测经济衰退的核心地位，这与Estrella and Mishkin（1996年）的经典发现一致。然而，重要性的分布却截然不同：在 X_t + XGBoost 模型中，特征重要性从利差变量的约70迅速降至领先劳动力市场指标的约27，表明该模型将权重集中于狭窄的一组预测变量。相比之下， Z_t + Logit-l2模型的重要性在各预测变量之间逐渐递减，货币总量（例如广义货币 M2）和股市指标仍保持不可忽视的影响。尽管两个模型都纳入了 ℓ_2惩罚，XGBoost却形成了陡峭得多的层级，对一小部分变量给予不成比例的强调，而我们的基准模型则在预测变量集上更均匀地分配权重。

表6：衰退前时期各行业对预测的贡献

衰退前一年对预测的贡献（%）				
‘风险状态’ (Z_t) +Logit-ℓ_2 模型				
经济部门	1990	2001	2008	2020
产出与收入	10.8	8.8	9.7	10.1
劳动力市场	26.9	28.0	27.2	26.5
住房	7.1	8.0	8.1	7.8
消费与订单	4.5	4.3	4.5	5.1
货币与信贷	10.5	10.6	10.1	10.8
利率与利差	27.1	25.9	25.6	24.3
价格	8.1	9.3	9.4	9.2
股票市场	5.1	5.2	5.4	6.2
连续型 (X_t) + XGBoost 模型				
经济部门	1990	2001	2008	2020
产出与收入	3.4	2.4	3.8	6.3
劳动力市场	19.5	20.2	30.4	20.2
住房	2.5	11.9	13.2	13.7
消费与订单	0.7	1.7	2.8	1.0
货币与信贷	0.8	5.3	1.6	2.1
利率与利差	66.1	52.6	43.8	53.8
价格	0.8	0.6	1.2	1.0
股票市场	6.3	5.4	3.2	1.9

注：本表报告了在每次经济衰退由美国国家经济研究局确定的峰值之前紧邻的12个月窗口内，各经济部门所占总特征重要性的百分比。数值根据“Z_t，Logit-l2”和“X_t，XGBoost”模型的样本外特征重要性计算得出。

我们在表6中进一步展示了这一模式。该表报告了在每次美国国家经济研究局经济衰退峰值之前的12个月内，各经济部门——遵循McCracken and Ng (2016)的分类——对预测的平均贡献。该表强化了我们此前的发现： X_t 模型过度依赖利率和利差，往往为其分配超过总重要性一半的权重，而对其他行业仅赋予微弱权重。相比之下，“风险暴露”变换形成了更为均衡的结构。利率和劳动力市场指标成为两大主要贡献来源，各自所占重要性份额相近，而价格、货币与信贷以及产出与收入等其他类别也发挥着更有意义的作用。这种多元化凸显了“风险暴露”

与连续设定相比，风险状态转换提升了预测性能。

表6还揭示了另一项重要发现：在我们的基准模型中，四次经济衰退期间各行业贡献十分稳定。相比之下，X_t + XGBoost 模型在重要性分配上出现了显著波动。例如，利率与利差以及股票市场变量的贡献在前三次经济衰退期间逐渐下降，而劳动力市场和住房变量的影响则逐渐上升。鉴于我们的基准模型表现更佳，X_t 模型中的这些变化更可能源于过拟合，而非底层预测关系随时间发生了真实变化。二值化似乎通过防止预测框架对 X_t 中的短期波动反应过度，转而聚焦于携带最强经济衰退信号的尾部行为，从而增强了框架的稳健性。

6 运用风险状态转换进行简约建模

我们对完整高维模型的分析揭示了两项关键发现：第一，我们的“风险暴露”变换在样本外表现更优；第二，从高维 Z_t 中提取因子比 X_t 是更好的建模选择。然而，这引出了一个至关重要的最终问题：“风险暴露”变换的优势是否仅仅是“大数据”带来的现象，还是说它代表了一项根本性改进，也适用于更简单、更传统的预测任务?

为回答这一问题，我们在低维环境中开展了与主要结果相同的样本外分析。我们考察两组预测变量集。第一组仅包含一个变量——10年期美国国债与联邦基金利率利差。第二组包含经济衰退预测文献所强调的十个典型预测变量，我们将其称为简约模型。[2]对于这两组精简的预测变量集，我们分别评估应用于风险状态转换（ℓ_2）的 Logit-ℓ_2 模型与应用于相应连续变量（ℓ_2）的

[2]为构建公平且稳健的比较，我们使用的指标并非直接根据特征重要性得出。相反，我们遵循学术文献，选择能够模拟传统预测分析的变量：两种利差（10年期美国国债与联邦基金利率利差和Baa级债券与联邦基金利率利差）、三项劳动力市场指标（总就业人数、失业保险申请人数和失业率），以及其他每个主要类别中各一项代表性变量：工业生产（实际经济活动）、非农住宅开工量（住房）、零售销售（消费与订单）、标准普尔500指数（股票市场）和实际M2（货币与信贷）。

表7：样本外表现：单变量模型与简约模型

	精确率-召回率曲线下面积			布里尔分数		
模型设定	$h=3$	$h=6$	$h=12$	$h=3$	$h=6$	$h=12$
完整特征空间						
连续型 (X_t) +XGBoost	0.584	0.338	0.351	0.062	0.085	0.098
“风险状态” (Z_t) + Logit-ℓ_2	**0.718**	**0.370**	**0.398**	**0.049**	**0.082**	**0.087**
单变量模型（仅期限利差）						
连续型 (X_t) +XGBoost	**0.367**	0.242	0.255	**0.092**	**0.111**	**0.109**
“风险状态” (Z_t) + Logit	0.264	**0.334**	**0.384**	0.130	0.124	0.121
简约模型（*10* 项核心指标）						
连续型 (X_t) +XGBoost	0.529	0.314	0.343	0.072	0.093	**0.090**
“风险状态”（Z_t） + Logit-ℓ_2	**0.721**	**0.485**	**0.451**	**0.057**	**0.088**	0.117

注：“完整特征空间”对应于前文各节使用的完整 FRED-MD 面板数据。“单变量”模型设定仅使用10年期美国国债与联邦基金利率利差。“简约”模型使用文献中确立的十个核心指标：两个利差（10年期美国国债与联邦基金利率利差、Baa债券与联邦基金利率利差）、三个劳动力市场指标（总就业人数、失业保险申请人数、失业率）、工业生产（实际经济活动）、非农住宅开工量（住房）、零售销售（消费与订单）、标准普尔500指数（股票市场）以及实际M2（货币与信贷）。

XGBoost 模型的表现。

结果见表7。作为参考，顶部面板重现了上一节的完整特征空间结果；在上一节中，我们表明，即使将连续 X_t 表示与非线性XGBoost模型结合，风险状态转换的表现仍优于连续表示。第二个面板考察了两种方法在使用10年期美国国债与联邦基金利率期限利差时的单变量表现，结果较为微妙。在所有预测期内，连续表示在整体校准方面均更优，但二值化期限利差在 $h=6$，12时展现出显著更强的区分能力（精确率-召回率曲线下面积）。此外，在较长预测期内，结合长提前期指标时，“风险状态”也很有效。然而，我们发现，当依赖单变量特征空间时，二值化的影响会因指标和预测期而异。

然而，最后一个面板揭示了我们方法的真正优势。在一个贴近传统衰退预测实践的现实多变量情境中，风险状态转换（Z_t）在几乎所有预测期和指标上都持续优于连续设定（X_t），

在中长期预测期的区分能力（精确率-召回率曲线下面积）方面，优势尤为显著。这证实了我们的主要发现：“风险暴露”变换是提升衰退可预测性的一种稳健有效的方法。

尽管本文测试的简约模型并非经过精细调优的模型设定，而是基于一组因经济直觉而选取的通用指标构建，但研究者事先并不知道预测模型应纳入哪些指标。因此，不应将表中呈现的表现解读为证据，认为这组指标在实践中能够胜过其他聚合方法。本项分析旨在说明，在这一设定下，Z_t 框架持续显著优于其他方法，这有力地证明了其价值并不局限于高维应用（更多组合见附录 C.4）。

尽管这种变换对任何单个变量的影响可能较为复杂，但其主要优势在于能够提供多样而统一的信号，让简单模型得以从多方证据（离散的 1 和 0）中学习。对于依赖少数关键指标的衰退预测问题，“风险暴露”变换能够更有效地表征这些关键指标所包含的预测信息。

7 结论

本文的证据表明，对预测变量进行简单二值化——即“风险暴露”变换——是衰退预测的有力工具。我们以 Keilis-Borok 等人（2000）的奠基性见解为基础；他们最早在一个更简单的设定中，将这一思路应用于少量指标。本文进一步证明，类似方法在现代高维预测环境中同样有效。递归样本外分析显示，采用这些二元特征的模型不仅极具竞争力，而且往往优于使用标准连续数据的基准方法，其中包括 XGBoost 等机器学习方法。我们还发现，从连续变量的二值化表示中提取主成分分析因子，能够显著提升表现。展望未来

，多个方向可以拓展我们的研究发现范围 . 首先，与任何

预测模型一样，样本外时期虽然跨越了三十多年和多个商业周期，但终究是有限的。检验该框架稳健性的最终标准，将是它在实时环境中的持续表现。其次，使用可比数据集在其他国家检验这一想法（例如，Goulet Coulombe 等人，2021b，针对英国），并在分位数回归等其他情境中进行检验（例如，Adrian 等人，2019），将很有意思。最后，若能将二值化“风险中”预测变量及其聚合优于连续输入模型的数据生成设定形式化，将大有裨益。一种自然情形是共同的极端冲击混合：此时均值变化微不足道，但在衰退前状态中，同步尾部实现值的频率激增，使标准化 Z_t 的计数或主成分几乎足够，而基于 X_t 的线性模型则设定错误。这也与二元数据的非线性主成分分析（例如逻辑主成分分析）相关，为我们的实用主成分分析基准提供了一个有理论依据的对应方案。

参考文献

Aastveit, K. A., Anundsen, A. K., 和 Herstad, E. I.（2019年）。住宅投资与衰退可预测性。《国际预测学杂志》, 35(4):1790–1799。

Aastveit, K. A., Jore, A. S., 和 Ravazzolo, F.（2016年）。挪威商业周期的识别与实时预测。《国际预测学杂志》, 32(2):283–292。

Adrian, T., Boyarchenko, N., 和 Giannone, D.（2019年）。脆弱增长。《美国经济评论》, 109(4):1263–1289。

Antunes, A., Bonfim, D., Monteiro, N., 和 Rodrigues, P. M. M.（2018年）。利用动态面板 probit 模型预测银行业危机。《国际预测学杂志》, 34(2):249–275。

Bluedorn, J. C., Decressin, J., and Terrones, M. E.（2016年）。资产价格下跌是否预示着经济衰退？*International Journal ofForecasting*,32(2):518–526。

Burns, A. F. 与 Mitchell, W. C.（1946）。《商业周期测量》。美国国家经济研究局，纽约。

Carriero, A. 与 Marcellino, M.（2007）。英国综合同步指数与先行指数构建方法比较。《国际预测学杂志》，23(2):219–236。

Carstensen, K., Heinrich, M., Reif, M., and Wolters, M. H. (2020). 使用三状态马尔可夫转换动态因子模型预测一般经济衰退和严重衰退：对德国商业周期的应用。《国际预测学杂志》, 36(3):829–850.

Chauvet, M. and Potter, S. (2010). 结构变化下的商业周期监测。《国际预测学杂志》, 26(4):777–793.

Chen, A., Roll, R., and Rossiter, R. (2011). 预测美国经济衰退的概率：一种概率单位与动态因子建模方法。《加拿大经济学杂志》, 44(2):651–672.

Chung, S. (2024). 黑箱内部：基于神经网络的美国经济衰退实时预测。

Clements, M. P. and Harvey, D. I. (2010). 预测包容性检验与概率预测。《应用计量经济学杂志》, 25(7):1028–1062.

Clements, M. P. 和 Harvey, D. I.（2011年）。概率预测组合。《国际预测学杂志》，27(2)：208–223。

Crump, R. K.、Giannone, D. 和 Lucca, D. O.（2020年）。解读美国商业周期的蛛丝马迹——第一部分。自由街经济学博客，纽约联邦储备银行。发表于2020年2月10日。

Daniele, M.、Kronenberg, P. 和 Reinicke, T.（2025年）。用于宏观经济预测的定向变换。KOF工作论文第517号，KOF瑞士经济研究所，苏黎世联邦理工学院。

Davig, T. 和 Smalter Hall, A.（2019年）。使用贝叶斯分类进行衰退预测。《国际预测学杂志》，35(3)：848–867。

De Pace, P. 和 Weber, K. D.（2016年）。美国高收益率利差的时变先行特性。《国际预测学杂志》，32(1):203–230。

Diebold, F. X. 和 Rudebusch, G. D.（1989年）。先行指标的评分。《商业杂志》，62(3):369–391。

D¨opke, J., Fritsche, U. 和 Pierdzioch, C.（2017年）。利用提升回归树预测经济衰退。《国际预测学杂志》，33(3):745–759。

Duarte, A., Venetis, I. A. 和 Paya, I.（2005年）。利用收益率利差预测欧元区实际增长和经济衰退的概率。《国际预测学杂志》，21(2):261–277。

Estrella, A. 和 Mishkin, F. S.（1996年）。收益率曲线作为美国经济衰退的预测变量。《经济与金融当前问题》，2(7)。

Estrella, A. 和 Mishkin, F. S.（1998年）。预测美国经济衰退：金融变量作为先行指标。《经济学与统计学评论》，80(1):45–61。

Fair, R. C. 和 Shiller, R. J.（1990）。比较计量经济模型预测值中的信息。《美国经济评论》，第375–389页。

Fintzen, D. 和 Stekler, H. O.（1999）。为什么预测者未能预测到1990年经济衰退?《国际预测学杂志》，15(3)：309–323。

Giusto, A. 和 Piger, J.（2017年）。使用向量量化实时识别商业周期转折点。《国际预测学杂志》，33(1)：174–184。

Goulet Coulombe, P.、Leroux, M.、Stevanovic, D. 和 Surprenant, S.（2021年a）。宏观经济数据转换至关重要。《国际预测学杂志》，37(4)：1338–1354。

Goulet Coulombe, P.、Marcellino, M. 和 Stevanovic, D.（2021b）。机器学习能否捕捉到新冠疫情经济衰退?《国家经济研究评论》，256：71–109。

Granger, C. W. J. 和 Newbold, P.（1973年）。关于经济预测评估的一些评论。《应用经济学》，5(1):35–47。

Hamilton, J. D.（2011年）。实时判定经济衰退。《国际预测学杂志》，27(4):1006–1026。

Hansen, A. L.（2024年）。使用 VIX—收益率曲线周期预测经济衰退。《国际预测学杂志》，40(2)：409–422。

Harding, D. 和 Pagan, A.（2006年）。周期同步。《计量经济学期刊》，132(1)：59–79。

Huang, Y.-F. 和 Startz, R.（2020年）。利用股市波动率改进经济衰退日期判定。《国际预测学杂志》，36(2)：507–514。

Keilis-Borok, V.、Stock, J. H.、Soloviev, A. 和 Mikhalev, P.（2000年）。美国六项经济指标的衰退前模式。《预测学杂志》，19(1)：65–80。

Lahiri, K. 和 Yang, C.（2023a）。评估经济衰退预测的受试者工作特征和精确率—召回率曲线方法。《商业周期研究杂志》，19(2)：119–148。

Lahiri, K. 和 Yang, C.（2023年b）。两个经济衰退衍生指标的故事。《实证经济学》，65(2):925–947。

Lahiri, K. 和 Yang, L.（2013年）。预测二元结果。载于 Elliott, G. 和 Timmermann, A. 编，《经济预测手册》第2卷B部分，第1025–1106页。爱思唯尔。

Lahiri, K. 和 Yang, L.（2015年）。对世界大型企业联合会新领先经济指数的进一步分析。《国际预测学杂志》，31(2):446–453。

Laurent, R. D.（1988年）。一种基于利率的货币政策指标。《经济展望》，12(1):3–14。

Layton, A. P. 和 Katsuura, M.（2001年）。使用体制转换、概率单位和 Logit 模型对美国商业周期进行定时和预测的比较。《国际预测学杂志》，17(3)：403–417。

Levanon, G.、Manini, J.-C.、Ozyildirim, A.、Schaitkin, B. 和 Tanchua, J.（2015年）。利用金融指标预测商业周期的转折点：美国领先经济指数案例。《国际预测学杂志》，31(2)：426–445。

Li, H.、Sheng, X. S.与Yang, J.（2021年）。监测经济衰退：贝叶斯序贯最快检测方法。《国际预测学杂志》，37(2):500–510。

Liu, L.与Wang, Y.（2025年）。含极端协变量的二元结果模型：估计与预测。https://arxiv.org/abs/2502.16041。预印本，*arXiv*:2502.16041。

Liu, W.与Moench, E.（2016年）。什么可以预测美国经济衰退？《国际预测学杂志》，32(4):1138–1150。

Mathy, G.与Zhao, Y.（2025年）。扩散指数能否预测过大萧条？《预测学杂志》，44(2):320–338。

McCracken, M.W.与Ng, S.（2016年）。FRED-MD：面向宏观经济研究的月度数据库。《商业与经济统计期刊》，34(4):574–589。

Moore, G. H.（1961）。扩散指数、变化率与预测，第282–293页。普林斯顿大学出版社。

Murphy, A. H.（1973年）。概率评分的一种新向量分解。《应用气象与气候学杂志》，12(4)：595–600。

Ng, S.（2014年）。提升经济衰退预测。《加拿大经济学杂志》/《加拿大经济学评论》´经济学，47(1)：1–34。

Politis, D. N. 和 Romano, J. P.（1994年）。平稳自助法。《美国统计学会杂志》，89(428)：1303–1313。

Proa˜no, C. R. 和 Theobald, T.（2014年）。利用复合实时动态概率单位模型预测经济衰退。《国际预测学杂志》，30(4)：898–917。

Qi, M.（2001年）。利用神经网络模型和先行指标预测美国经济衰退。《国际预测学杂志》，17(3)：383–401。

Rudebusch, G. D. 和 Williams, J. C.（2009年）。预测经济衰退：收益率曲线持续发挥作用之谜。《商业与经济统计期刊》，27(4)：492–503。

Sahm, C.（2019年）。向个人直接发放刺激性补助金。载于 Boushey, H.、Nunn, R. 和 Shambaugh, J. 编，《style id='1'>《为经济衰退做好准备：稳定美国经济的财政政策》，第3章。汉密尔顿项目和华盛顿公平增长中心，华盛顿哥伦比亚特区。

Stock, J. H. 和 Watson, M. W.（1993年）。利用先行指标预测经济衰退的方法：计量经济学问题与近期经验。载于 Stock, J. H. 和 Watson, M. W. 编，《style id='1'>《商业周期、指标与预测》。芝加哥大学出版社为美国国家经济研究局出版，芝加哥。

Stock, J. H. 和 Watson, M. W.（2014年）。利用大型数据集估计转折点。《计量经济学期刊》，178(2)：368–381。

Vrontos, S. D.、Galakis, J. 和 Vrontos, I. D.（2021年）。使用机器学习技术对美国经济衰退进行建模和预测。《国际预测学杂志》，37(2)：647–671。

Wright, J. H.（2006年）。收益率曲线与经济衰退预测。金融与经济讨论系列 2006-07，美国联邦储备系统理事会。

附录

A 方法细节

A.1 数据

本研究采用 FRED-MD 宏观经济数据集，其中包含八个类别的 126 个时间序列月度数据。与作者的做法类似，我们剔除 ACOGNO（消费品新订单）、TWEXAFEGSMTHx（名义发达外国经济体美元指数）、UMCSENTx（消费者信心指数）和 OILPRICEx（原油，由西德克萨斯中质原油和库欣数据拼接而成），因为这些数据经过转换后高度不规则，或缺失数据较多（McCracken 和 Ng，2016年），最终得到 122 个宏观经济预测变量。

时间序列分析的一个关键部分是确保预测变量的平稳性。为此，我们遵循 FRED-MD 数据集的标准流程，应用 McCracken and Ng (2016) 推荐的特定变换。

A.2 逆周期变量

我们结合判断与数据驱动分析，确定哪些平稳化序列属于逆周期序列。由于平稳化序列 x_t 的表现不同于 FRED-MD 中的原始序列，我们计算它在样本内期间与美国国家经济研究局衰退指标的相关性。如果相关性显著为负（< -0.10），则将该序列归类为逆周期序列。其余序列则归类为顺周期序列。

归类为逆周期的变量如下：失业率（U：全部）、平均失业持续时间（U：平均持续时间）、失业 <5 周的平民（U <5 周）、失业5至14周的平民（U 5-14周）、失业 15+ 周的平民（U 15+周）、失业15至26周的平民（U 15-26周）、失业 27+ 周的平民（U 27+ 周）、首次申领失业救济人数（UI claims）、库存销售比（M&T invent/sales）以及芝加哥期权交易所波动率指数。

A.3 超参数选择

A.3.1 选择 τ_g

对于正文中使用的全局阈值 τ_g，我们根据初始训练样本（$t = 1$，……，T）仅确定一次，以避免前视偏差。具体步骤如下：

1. 对于每个预测变量 i 以及每个衰退月份 $t \in S_{rec}$，我们计算观测值$\bar{x}_{it}^{h_g}$ 相对于完整训练样本分布$F_{i,\ T}$的经验分位数水平 τ_{it}。

$$\tau_{it} = F_{i,T}(\bar{x}_{it}^{h_g})$$

2. 对于每个预测变量 i，我们随后计算这些衰退期分位数水平的中位数。该值 τ_i^*表示该特定预测变量在历史衰退期的典型分位数水平。

$$\tau_i^* = \text{median}(\{\tau_{it} \mid t \in S_{rec}\})$$

3. 最后，将全局阈值 τ_g 设为这些预测变量特定值的中位数。该方法在数据集中的全部 N 个预测变量之间平衡信号。

$$\tau_g = \text{median}(\{\tau_i^* \mid i = 1, ..., N\})$$

A.3.2 选择 λ

本研究采用带有 ℓ_2（岭回归）正则化惩罚项的逻辑回归模型进行估计，以减轻高维预测变量导致的过拟合，同时处理多重共线性。相应的目标函数为惩罚对数似然：

$$\hat{\beta}(\lambda) = \underset{\beta}{\text{argmax}} \left\{ \sum_{t=1}^{T} \left[y_t X_t' \beta - \log(1 + e^{X_t' \beta}) \right] - \lambda \sum_{j=1}^{N} \beta_j^2 \right\},$$

超参数 $\lambda \geq 0$ 控制惩罚项的强度。由于 scikit-learn 的实现方式，我们对参数 C进行调优，该参数等价于 $\frac{1}{\lambda}$。为稳健地选择 C 并避免前视偏差，我们仅在初始训练样本（来自 $t = 1, .., T$ 的数据）上进行时间序列交叉验证。随后，在整个递归样本外预测过程中，选定值 C^*始终保持不变。

选择流程如下：

1. 我们为超参数指定一个包含 30 个候选值的对数网格，记为集合 Λ，其范围为 $[10^{-3}, 10^{1}]$。

2. 我们采用包含 $K = 5$ 个分割的扩展窗口交叉验证方案，与 scikit-learn 的 TimeSeriesSplit 一致。初始训练样本被划分为 $K+1 = 6$ 个连续区块。对于每个分割 $k \in \{1, ..., K\}$，训练集使用前 k 个区块，验证集使用第 $(k + 1)$ 个区块。

每个验证区块的大小 s 约为 $s = \frac{T}{K+1} \approx 60$ 个月。各集合构造如下：

$$S_k^{\text{train}} = \{t \mid 1 \leq t \leq s \cdot k\}$$
$$S_k^{\text{val}} = \{t \mid s \cdot k < t \leq s \cdot (k + 1)\},$$

确保训练数据始终先于验证数据。

3. 对于每个候选值 $C_m \in \Lambda$ 和每个分割 k，我们执行以下步骤：

ˆ (a) 通过最大化训练数据 S_k^{train}上的惩罚对数似然，估计系数向量 $\beta_k(\lambda_m)$。(b) 为验证集 $t \in S_k^{\text{val}}$中的每个观测值生成样本外概率预测ˆ$p_t(\lambda_m)$。

(c) 计算验证集上的布里尔分数，以衡量预测性能：

$$\mathrm{BS}_k(C_m) = \frac{1}{|S_k^{\mathrm{val}}|} \sum_{t=1}^{S_k^{\mathrm{val}}} (\hat{p}_t(C_m) - y_t)^2,$$

其中， $|S_k^{\mathrm{val}}|$ 是验证集中的观测值数量。

4. 最终值 λ^* 选取为在所有 K 折中平均布里尔分数最低的候选值。

$$C^* = \arg\min_{C_m \in \Lambda} \left(\frac{1}{K} \sum_{k=1}^{K} \mathrm{BS}_k(C_m) \right)$$

Thi随后，对使用 Logit-ℓ_2 分类器训练的每个预测变量集重复此过程 r.

B 其他指标

B.1 自助法置信区间

表8与表1相同，但其置信区间是基于自助法估算的。

我们采用平稳自助法，进行 1,000 次重复抽样[3]，为精确率-召回率曲线下面积和布里尔分数的点估计计算置信区间。我们将平均区块长度设为

$$L = \max\bigl(h,\ \lfloor T^{1/3} \rceil\bigr),$$

其中， h 是预测 horizon， $T = 420$ （我们的样本外规模）。术语 $T^{1/3}$ 遵循Politis和Romano（1994）提出的启发式方法，该方法表明，当区块长度以样本量的三分之一次幂的速率增长时，平稳自助法可实现良好的渐近性质。然而，我们将区块长度设定为至少与 horizon h 一样长，以确保每个自助样本都保留进行 h步预测所需的依赖结构。表8还列出了基准优于

[3]在实际操作中，当 L 较小时，平稳自助法偶尔会因边缘效应而舍弃一次重复抽样，因此可用的重复抽样次数为 999 次，而非 1,000 次。这不会对结果产生实质性影响。

表8： 风险状态转换（Z_t）与基准方法的样本外表现 ks

模型配置	精确率-召回率曲线下面积			布里尔分数		
	$h=3$	$h=6$	$h=12$	$h=3$	$h=6$	$h=12$
所提出的方法（Z_t，Logit-L2）	**0.718** [0.406，0.897]	0.370 [0.154，0.641]	**0.398** [0.107，0.647]	**0.049** [0.025，0.076]	**0.082** [0.044，0.124]	**0.087** [0.055，0.121]
基准（X_t，Logit-L2）	0.501 (2.9%) [0.239, 0.775]	0.286 (16.3%) [0.129，0.526]	0.170 (2.6%) [0.069，0.336]	0.069（0.3%） [0.041，0.099]	0.096（12.3%） [0.063，0.131]	0.121（0.1%） [0.080, 0.161]
基准（X_t 的主成分分析、Logit-L2）	0.552 (12.5%) [0.273, 0.829]	**0.408**（68.4%） [0.196, 0.651]	0.314（24.8%） [0.124, 0.519]	0.064（5.1%） [0.038, 0.094]	0.083（48.8%） [0.054, 0.113]	0.108（1.3%） [0.080，0.136]
基准（X_t，XGBoost）	0.584（2.4%） [0.299，0.806]	0.338（35.6%） [0.131，0.647]	0.351 (32.9%) [0.115, 0.561]	0.062 (11.4%) [0.029, 0.100]	0.085 (43.6%) [0.037，0.138]	0.098（25.1%） [0.048，0.154]

括号中的所提模型。

B.2 ROC AUC

现在，我们展示初步比较（表1）中所介绍模型的受试者工作特征（Receiver Operating Characteristic）AUC得分，以便与正文中使用精确率-召回率曲线下面积进行比较。为了计算ROC AUC，我们首先使用决策阈值 $\delta \in [0,1]$将预测的衰退概率$\hat{}p_t$转换为点预测：

$$\hat{y}_t(\delta) = \mathbf{1}\{\hat{p}_t \geq \delta\}.$$

不过，我们没有针对每个 δ计算召回率和精确率，而是直接求出真正率和假正率：

$$T(\delta) = \frac{TP(\delta)}{TP(\delta) + FN(\delta)}, \qquad F(\delta) = \frac{FP(\delta)}{FP(\delta) + TN(\delta)},$$

其中， TP、 TN、 FN和 FP分别表示真正例、真负例、假负例和假正例， T和 F分别表示真正率和假正率。改变 δ即可绘制ROC曲线。曲线下面积定义为

$$\text{ROC AUC} = \int_0^1 T(\delta)\,\frac{dF(\delta)}{d\delta}\,d\delta = \int T(F)\,dF.$$

结果如表9所示。排名基本反映了我们的初步结果，

表9：表1 ROC AUC得分

	ROC AUC		
模型配置	$h = 3$	$h = 6$	$h = 12$
所提出的方法（Z_t，Logit-ℓ_2）	**0.949**	0.861	0.826
替代设定			
X_t，Logit-ℓ_2	0.913	0.801	0.753
主成分分析 X_t、Logit-ℓ_2	0.923	**0.895**	**0.843**
X_t、XGBoost	0.920	0.829	0.822

但在较长预测期内，PCA 在 X_t 上的表现略胜一筹。不过，ROC AUC 未考虑类别不平衡，因此，即使模型在较长的经济扩张时期表现出色，却在短暂且稀少的经济衰退时期表现欠佳，仍可能获得较高的 ROC AUC 分数（这也是我们在正文中考虑精确率-召回率曲线下面积的部分原因）。

对于其他比较，我们发现，与表9一样，结果与原始排名高度一致。

C 稳健性检验

C.1 变化的τ_g

为检验稳健性，我们还使用另外两种阈值策略评估模型表现，分别计算：(1) 每个行业的截断分位数，以及(2) 每个变量的截断点。结果见表10。

结果表明，虽然提高阈值复杂度在一定程度上有所帮助，但也可能增加过拟合的风险。对于未聚合模型和基于主成分分析的模型，采用行业特定阈值可能有益，尤其是在较长的预测期内。然而，采用变量特定截断值会导致过拟合，因为初始训练样本中每个变量的阈值往往会随时间变化。未来研究也可以探索其他专门的阈值设定技术。

表10：对替代阈值设定的稳健性

	精确率-召回率曲线下面积			布里尔分数		
阈值设定	$h=3$	$h=6$	$h=12$	$h=3$	$h=6$	$h=12$
未聚合 Z_t + Logit-ℓ_2 模型						
全局阈值（主要设定）	**0.718**	0.370	0.398	**0.049**	0.082	0.087
行业特定阈值	0.652	**0.397**	**0.508**	0.052	**0.078**	**0.085**
变量特定阈值	0.662	0.370	0.313	0.053	**0.078**	0.098
主成分分析（Z_t） +Logit-ℓ_2 模型						
全局阈值（主要设定）	0.688	0.528	0.404	0.062	0.081	0.106
行业特定阈值	**0.728**	**0.572**	**0.549**	**0.051**	**0.068**	**0.088**
变量特定阈值	0.722	0.453	0.364	0.054	0.079	0.114

C.2 变化 h_g

在第3.2节中，我们简要讨论了 h_g 调优参数，它表示输入信号的移动平均窗口。在主要分析中，我们将 $h_g = 1$ 设定为某值并添加了显式滞后项；现在，我们通过设定 $h_g > 1$ 来探索一种无需显式滞后项、纳入过去信息的替代方法，见表11。

结果表明，尽管使用移动平均（$h_g > 1$）可以纳入过去信息，但与我们的主要设定（$h_g = 1$ 对 Z_t使用显式滞后项）相比，它通常会导致更低的精确率-召回率曲线下面积和更高的布里尔分数。这表明，对于“风险状态”转换而言，由二值化后的“风险状态”的时点与持续性所产生的预测信号（由 Z_t的滞后项捕捉）比在二值化之前对连续变量本身的强度进行平滑更有价值。

C.3 同期预测变量

在我们的主要分析中，研究中的所有模型均使用 $L = \{3, 6, 12\}$ 个滞后项进行评估。作为稳健性检验，我们还评估了未添加任何滞后项的模型。表12展示了仅使用同期信号的基线比较。在大多数情况下，表现有所下降，但某些模型在特定预测期（例如 Z_t 在 $h = 6$ 上）的预测

表11： h_g 移动平均结果

	精确率-召回率曲线下面积			布里尔分数		
移动平均窗口	$h=3$	$h=6$	$h=12$	$h=3$	$h=6$	$h=12$
未聚合 Z_t + Logit-ℓ_2 模型						
$h_g=3$	0.551	**0.370**	0.363	**0.066**	0.102	0.142
$h_g=6$	**0.611**	0.257	**0.479**	0.067	0.110	**0.124**
$h_g=12$	0.563	0.330	0.254	0.074	**0.097**	0.135
主成分分析（Z_t） +Logit-ℓ_2 模型						
$h_g=3$	**0.684**	0.411	0.283	**0.086**	0.121	0.161
$h_g=6$	0.618	**0.481**	**0.288**	0.090	**0.113**	**0.149**
$h_g=12$	0.578	0.254	0.167	0.099	0.134	0.168

表12：无滞后基线性能

	精确率-召回率曲线下面积			布里尔分数		
模型配置	$h=3$	$h=6$	$h=12$	$h=3$	$h=6$	$h=12$
所提出的方法（Z_t，Logit-ℓ_2）	**0.594**	**0.464**	0.274	**0.069**	**0.102**	**0.137**
替代设定						
X_t，Logit-ℓ_2	0.462	0.332	0.311	0.087	0.107	0.162
X_t的主成分分析，Logit-ℓ_2	0.541	0.429	**0.380**	0.105	0.110	0.146
X_t，XGBoost	0.582	0.275	0.212	0.066	0.104	0.120

注： 本表比较了使用带有 ℓ_2 惩罚项的逻辑回归模型在我们的二元风险状态矩阵（Z_t）上训练后，与若干基准方法（使用未滞后的预测变量）的样本外表现。粗体数值表示各指标和预测期下表现最佳的模型。

性能有所提升。不过，我们发现，大多数模型在纳入滞后特征后表现通常更好。下方是仅使用同期特征的表 13（聚合策略）。

对于依赖上述滞后项的 XGBoost，其区分能力显著下降，但采用它的特征集在各个预测期的布里尔分数均最低。如前所述，尽管移除滞后项在某些情形下能提升表现，但总体而言，所有模型在纳入滞后特征后表现更好；即使不使用滞后项， Z_t 在大多数情况下的预测性能仍优于基准方法。

表 13： Z_t 无滞后项的聚合策略

聚合方法	模型	精确率-召回率曲线下面积			布里尔分数		
		$h=3$	$h=6$	$h=12$	$h=3$	$h=6$	$h=12$
未聚合（Z_t）	Logit-ℓ_2	0.594	0.464	0.274	0.069	0.102	0.137
	XGBoost	0.659	0.281	0.239	**0.055**	**0.089**	0.106
简单平均	Logit	0.555	0.362	0.104	0.154	0.205	0.253
	XGBoost	0.322	0.276	0.088	0.191	0.207	0.251
Z_t上的主成分分析	Logit-ℓ_2	**0.668**	**0.507**	0.289	0.083	0.123	0.153
	XGBoost	0.583	0.344	**0.312**	0.068	0.091	**0.091**

注： 此表比较了应用于无滞后项的 Z_t 矩阵的聚合和建模策略。简单平均不使用惩罚项，因为只有一个特征。

C.4 其他简约集合

我们在第 6 节中仅展示一组具有代表性的简约特征集，但也测试了其他备选子集，以了解 Z_t 在这些子集中是否同样优于 X_t 。

对于集合 S_t，我们仅使用利差[4]。对于集合 R_t，我们仅使用实体经济变量[5]。表 14 展示了竞争规格的关键指标。

对于 S_t，尽管所有模型在短期预测期上的表现都有所下降（Estrella and Mishkin（1996年）发现，收益率曲线在2个季度后的表现更佳），但我们发现，较长期预测期的模型表现良好。即使仅使用利差，“风险暴露”变换在 $h=3$、6、12上也展现出更强的区分能力。其概率估计在$h=12$上略不准确，但这通过其更强的两类区分能力得到了弥补。

对于集合 R_t，其表现的变异性略高；在较短预测期内，“风险暴露”集合展现出更强的区分能力，而连续型 R_t 则以区分能力显著下降为代价，表现出更好的校准。尽管连续型 R_t 似乎在较长预测期内表现更好，但其表现过低，因此无法对二者进行有意义的比较——鉴于在这些区间内的表现，甚至不宜将它们用于经济衰退预测模型

[4]我们使用五种收益率利差（10年期、5年期、1年期、6个月期和3个月期美国国债—联邦基金利差）、两种信用利差（AAA级和Baa级联邦基金利差），以及商业票据—联邦基金利差。[5]我们使用四项劳动力市场指标（非农就业人数、失业保险申请人数、制造业平均每周工时和失业率）、两项产出指标（工业生产、实际个人收入），以及新屋开工和零售销售。

表 14: 样本外 采用替代指标的简约模型表现

	精确率-召回率曲线下面积			布里尔分数		
模型设定	$h=3$	$h=6$	$h=12$	$h=3$	$h=6$	$h=12$
完整特征空间						
连续型（X_t） + XGBoost	0.584	0.338	0.351	0.062	0.085	0.098
“风险状态”（Z_t） + Logit-ℓ_2	**0.718**	**0.370**	**0.398**	**0.049**	**0.082**	**0.087**
S_t （8 个利差）						
连续型（X_t） + XGBoost	0.490	0.356	0.416	0.079	**0.085**	**0.086**
“风险状态”（Z_t） + Logit-ℓ_2	**0.529**	**0.427**	**0.451**	**0.071**	0.086	0.103
R_t （8 个实体经济指标）						
连续型 (X_t) + XGBoost	0.439	0.208	**0.131**	**0.077**	**0.123**	**0.103**
“风险状态” (Z_t) + Logit-ℓ_2	**0.542**	**0.238**	0.107	0.128	0.183	0.233

注：“完整特征空间”对应前文各节使用的整个 FRED-MD 面板数据。S_t 模型使用五种收益率利差（10年期、5年期、1年期、6个月期和3个月期美国国债—联邦基金利差）、两种信用利差（AAA级和Baa级债券—联邦基金利差）以及商业票据—联邦基金利差。R_t 模型使用四项劳动力市场指标（非农就业人数、失业保险申请人数、制造业平均每周工时和失业率）、两项产出指标（工业生产、实际个人收入）、新屋开工和零售销售。

。在考察一个表现良好且属于更现实的衰退预测练习的代表性集合时，我们发现，“风险暴露”变换有助于提升模型表现。

D 预测变量表示的补充调查

本附录综述了现有基于模型的衰退预测文献中对预测变量的标准处理方法。我们特别关注了该领域的领先期刊《国际预测学杂志》。我们的检索涵盖截至2025年9月所有包含“经济衰退”和“概率单位”这两个术语的文章。在此基础上，还补充了Google Scholar的前100条相关结果（检索式为：source:"international journal of forecasting" "recession"，且不含“概率单位”一词）。最终样本由24篇论文组成，列示如下。这些论文可归为三大主题：（i）使用概率单位/Logit回归进行衰退预测的变量选择，（ii）将新的统计方法或机器学习方法应用于二元经济衰退结果，以及（iii）实时经济衰退判定

（即时预测），使用马尔可夫转换模型。

无论预测模型是线性还是非线性，我们发现这些论文都没有对预测变量应用标准化和平稳化之外的显式非线性变换，这与我们论文所提出的方法形成对比。

1. Liu 和 Moench（2016）；Probit模型将大量金融和宏观指标组成的面板数据映射为美国国家经济研究局的衰退状态；美国月度数据，1959–2011年，预测变量按序列进行平稳化（水平值、对数差分、年度差分、移动平均值）；目标变量为美国国家经济研究局的衰退虚拟变量。

2. Duarte 等人（2005）；动态Probit模型（以及阈值增长模型）使用欧洲经济与货币联盟期限利差预测根据国内生产总值收缩构造的经济衰退；欧元区季度数据，1970年第1季度–2000年第4季度；衰退虚拟变量 = ≥2 5个季度窗口中的负增长季度；国内生产总值采用对数差分（年化），利差采用水平值。

3. Hamilton（2011）；实际经济活动的马尔可夫转换模型利用同步指标实时检测衰退状态；美国季度/月度实际经济活动、失业和期限结构数据；通过对数差分/年化计算增长率；通过体制转换体现非线性。

4. D¨opke 等（2017）；提升回归树用于分类二元衰退状态，并捕捉非线性阈值／交互作用；德国，1973—2014年月度数据；35项指标，其中许多以同比或实际值表示，部分为水平值；已检验平稳性；非线性由回归树刻画。

5. Aastveit 等（2019）；概率单位模型检验住宅投资能否提升相较于标准预测变量的衰退预测能力；12个经济合作与发展组织国家，1960年第1季度至2014年第4季度；美国国家经济研究局／经济周期研究所虚拟变量；预测变量采用季度对数差分，并纳入最多4期滞后项。

6. Qi（2001）；前馈神经网络结合先行指标（尤其是收益率利差），生成未来多个季度的衰退概率；美国，1967年第2季度至1995年第1季度季度数据；27项先行指标，其中许多采用对数差分或增长率，部分为水平值；神经网络可刻画非线性。

7. Aastveit 等（2016年）；马尔可夫转换因子模型识别挪威经济周期

状态与转折点；挪威季度数据，1978–2011年；国内生产总值、宏观经济、调查、金融状况指数；经对数变换/因子提取的数据；状态转换提供非线性。

8. Bluedorn 等人（2016）；Logit 利用资产价格下跌、波动率、利差和石油价格预测衰退开始；七国集团季度数据，1970–2011年；资产价格经通胀调整，石油价格为实际价格，利差采用水平值；仅通过 Logit 引入非线性。

9. Li 等人（2021）；基于隐马尔可夫模型的贝叶斯最快检测（序贯）规则，比动态因子马尔可夫转换模型更早地识别峰值/谷值；美国月度历史数据版本，1967–2013年；四个同步序列；卡尔曼滤波+ 因子提取；没有额外的非线性变换。

10. Davig 和 Smalter Hall（2019）；朴素贝叶斯和马尔可夫转换—朴素贝叶斯利用宏观金融预测变量及其持续性对经济衰退进行分类；美国月度数据，1959–2016年；FRED-MD 标准化变换；10 个滞后项；通过朴素贝叶斯/马尔可夫转换引入非线性。

11. Fintzen 和 Stekler（1999）；决策理论和基于调查的分析解释了预测者为何未能预见1990年经济衰退；美国1989–1990年调查 + 绿皮书；宏观经济数据采用增长率；描述性分析，不进行模型变换。

12. Levanon 等人（2015）；基于主成分分析的领先信贷指数取代领先经济指数中的广义货币 M2，对新的领先经济指数进行 Probit 模型分析可更早发出经济衰退信号；美国，1990–2013年月度数据；对六项标准化金融指标进行插值；基于滞后一期领先经济指数的概率单位模型。

13. Huang 和 Startz（2020）；加入机制转换波动率的马尔可夫转换因子模型提升了实时定年效果；美国，1959–2018年月度数据；四个以对数差分表示的同步序列，以及通过马尔可夫转换波动率处理的股票收益率。

14. Carstensen 等人（2020）；三状态马尔可夫转换动态因子模型区分普通衰退与严重衰退；德国月度数据，1991–2016年；35 个候选值，弹性网络筛选出 6 个；难处理的序列使用增长率，利差使用水平值；均已标准化。

15. Clements和Harvey（2011）；通过对数优势比组合基于 Logit 的概率（利差和工时），可提高预测值评分；美国月度数据，1965–2007年；利差使用水平值，工时使用同比百分比；非线性来自 Logit 链接函数和 KK 组合。

16. Giusto与Piger（2017年）；学习向量量化依据同步增长率并结合持续性规则对月度状态进行分类；美国月度1967–2013年；四个增长同步指标；数据版本模拟实时。

17. Proa˜no与Theobald（2014年）；复合实时动态Probit模型汇集多种滞后结构/指标，以获得稳定概率；德国1991–2011年与美国1969–2011年月度；工业生产/美国国家经济研究局虚拟变量；增长率、水平利差，并经标准化。

18. Hansen（2024年）；静态/动态Probit 模型利用VIX–收益率利差周期，表现优于单独使用利差；美国月度1950–2022年；10年–3个月利差、VIX（延伸至1990之前）；经平滑、标准化和缩尾处理的周期指标。

19. Antunes等人（2018年）；带滞后项 + 过度繁荣虚拟变量的动态面板数据Probit 模型改善了银行业危机预警；22个欧洲国家1970–2012年季度数据；信贷缺口、偿债、房价、股票；比率、同比、HP滤波缺口。

20. Chauvet 和 Potter（2010）；采用反复断点的 Probit 模型（以及带 AR 的潜变量）比固定模型或无断点模型具有更高的分类准确度；美国月度数据，1959–2007；四个同步序列采用100×对数差分；断点特定方差，AR 潜变量增加持久性。

21. Layton 和 Katsuura（2001）；转移概率随时间变化的状态转换模型，在确定时点和预测方面优于概率单位/Logit 模型；美国月度数据，1949–1999；经济周期研究所同步/领先综合指数；月度增长率，LRG 使用移动总和。

22. Carriero 和 Marcellino（2007年）；针对英国转折点，基于精选指标的 Probit 模型优于 VAR／马尔可夫转换模型；英国月度数据，1978–2004年；同步指标（工业生产、销售、就业、收入）+领先指标（CB 集合）；增长率／对数差分，标准化。

23. Lahiri and Yang（2015年）；采用领先指标综合指数扩展领先经济指数的 Probit 模型／马尔可夫转换模型可得出更准确的概率（通过二次概率评分／受试者工作特征诊断）；美国月度数据，1990–2014年；领先经济指数采用对数差分；侧重于分辨率／区分能力，而非预测变量变换。

24. De Pace 和 Weber（2016年）；Probit 模型和时变参数模型显示，2000年后高收益债利差的预测能力减弱，而期限利差重新显现长期预测信号；美国 1982–2011 年季度国内生产总值和 1982–2011 年月度工业生产；高收益债利差/期限利差为平稳序列，增长率按年率计算。