arXiv:2303.01651v1 [q-fin.ST] 3 Mar 2023

风险管理的最优概率预测[a]

Yuru Sun、Worapree Maneesoonthorn[b]、Rub´en Loaiza-Maya 和 Gael M. Martin

2023年3月6日

摘要

本文探讨了根据与金融风险管理相关的评分规则进行优化后生成预测分布所带来的影响。我们评估了可能存在设定错误的模型所生成的最优预测的预测性能，包括：i）风险价值和预期损失预测；以及 ii）用于涉及 VIX 期货的对冲策略的 VIX 波动率指数预测。实证结果表明，使用能够奖励对极端收益率进行准确预测的评分来校准预测分布，可提高风险价值和预期损失预测的表现。研究还表明，侧重尾部的预测分布能够在使用 VIX 期货的对冲策略中带来更好的结果。

关键词：预测分布；评分规则；风险价值；预期损失；VIX期货；动态风险管理

*MSC2010*主题分类：60G25, 62M20, 91G70*JEL*分类：
C18,C53, C58.

[a]四位作者均隶属于澳大利亚维多利亚州3800莫纳什大学计量经济学与商业统计系。Sun曾获莫纳什商学院研究生研究奖学金（2023年2月起）和墨尔本大学商学与经济学院研究生研究奖学金（2022年7月至2023年1月）的资助；Maneesoonthorn和Martin曾获澳大利亚研究理事会（ARC）Discovery Grant DP200101414资助；Loaiza-Maya曾获澳大利亚研究理事会职业早期研究人员奖DE230100029资助。[b]通讯作者；电子邮箱：Ole.Maneesoonthorn@monash.edu

1 引言

概率预测——通过完整的预测概率分布生成——提供了关于随机变量未来取值的完整信息。传统的概率预测方法假定：预测模型（或似然函数）能够正确指定真实的数据生成过程（DGP），或者某组指定的预测模型能够涵盖真实情况；然而，在实践中，这两种情况都不太可能发生。[1]

最优 概率预测则基于预测分布，这些预测分布在用户指定的某种适当评分规则下达到最优（Gneiting 等，2005年；Gneiting and Raftery，2007年；Patton，2019年；Martin 等，2022年；Zischke et al，2022年）。也就是说，这类预测不再默认采用传统的（设定错误的）似然函数，而是利用由评分规则构建的准则函数，对符合当前问题所关注的预测准确性进行奖励。优化可以针对给定模型的参数、预测模型组合的权重，或两者同时进行。对于加权组合，这种方法是更广泛努力的一部分：依据贝叶斯和频率学派原则，构建明确旨在根据特定预测指标取得良好表现的预测组合（近期贡献及相关文献综述，参见 Loaiza-Maya 等，2021年；更全面的综述，参见 Aastveit 等，2018年和 Wang 等，2022年）。

简言之，最优概率预测旨在生成以重要方式准确的预测，而不受预测模型（或模型集合）设定的影响。例如，在金融风险管理中，金融收益预测分布的尾部尤其值得关注：下尾反映多头（买入）头寸的损失风险，上尾则反映卖空头寸的损失风险。迄今为止，风险预测领域的文献主要聚焦于开发能够更好捕捉收益率尾部特征的替代模型，方法包括：采用不同的波动率设定（例如 Ding 等，1993年；Hansen 等，2012年）、采用不同的收益条件分布（例如 Chen 等，2008年；BenSa¨ıda 等，2018年）、增加更多潜在因子（Maheu and McCurdy，2004年；Maneesoonthorn 等，2017年），或采用仅对特定分位数（Engle and Manganelli，2004年）或特定尾部区域（Brooks 等，2005年）进行建模的设定。随后可以使用与尾部相关的评分规则来评估 预测，但这些评分规则并不参与预测本身的构建（参见 Giacomini and Komunjer，2005年；Jensen and Maheu，2013年；Chiang 等，2021年，以及许多其他文献，包括上文已引用的研究）。

本文采取不同的立场。我们采用一种已知无法捕捉

[1]我们一开始就明确说明，为了语言表达的多样性，我们将“预测”和“预测”（以及它们的各种语法派生形式）视为同义词并交替使用。需要指出的是，“预测”是一个更广泛的术语，例如也用于非时间性场景。

金融收益率某些关键特征的预测模型——即具有高斯创新项的广义自回归条件异方差（GARCH）模型（Bollerslev，1987年）——但该模型简单且估计迅速。随后，我们使用与所关注尾部区域相关的适当评分规则对该模型进行校准，并据此构建最优预测分布。我们表明，利用聚焦于尾部的评分规则校准高斯 GARCH(1,1) 模型，能够对极端收益作出准确预测，并说明结果如何随模型设定错误程度而变化。我们还评估了最优预测分布能否对风险价值（VaR）和预期损失（ES）作出准确预测，并将其准确性与基于似然（或者等价地，基于对数评分（LS））的概率预测进行比较。我们采用两种特定的聚焦于尾部的规则来构建最优预测：Diks 等人（2011年）的删失似然评分（CLS），以及分位数评分（QS）（Giacomini and Komunjer，2005年；Gneiting and Raftery，2007年）。CLS 是一种适当评分规则，使预测者能够关注任何特定关注区域中的准确性，包括预测分布的尾部。作为适用于分位数的适当评分规则，QS 对于生成 VaR 预测也尤其相关；它此前还被用作估计分位数回归的准则函数（Koenker and Bassett，1978年；Engle and Manganelli，2004年）。VaR 和 ES 预测分别采用 Christoffersen（1998）的条件覆盖检验，以及 Ziegel 等人（2020年）最近提出的 ES 预测优势检验进行评估。

随后，我们将最优预测的应用扩展至预测期权隐含波动率指数（VIX）这一问题，并基于 VIX 衍生品制定有效的对冲策略（Moran 和 Liu，2020年；Konstantinidi 等，2008年；Taylor，2019年）。我们再次不再寻求该指数的理想预测模型（Konstantinidi 等，2008年；Fernandes等，2014年；Taylor，2019年），而是使用 Corsi（2009年）的简单异质自回归（HAR）模型和高斯 GARCH(1,1) 误差来构建最优预测。此处，我们使用 CLS 评分校准模型并生成最优预测分布，同时将“关注”区域设定为与相关交易策略的某些特征相匹配。

本文结构安排如下。第2节阐述如何利用评分规则构建最优分布预测，并概述与金融风险管理相关的特定尾部聚焦型规则。第3节在模拟环境中考察最优预测带来的益处，尤其关注通过评估金融收益的风险价值和预期损失预测，最优预测在预测尾部风险方面的表现。模拟设计明确纳入模型设定错误，以期重现真实的实证情景。随后，第4节对12个标准普尔市场及行业指数展开广泛的实证分析，结果证实了第3节的模拟发现：也就是说，与基于似然的方法相比，最优预测分布能够显著提升尾部风险预测效果。第5节利用近期的VIX指数和VIX期货数据，探讨如何将VIX指数的最优预测用于构建有效的对冲策略。

定制的基于评分的预测能够带来更好的投资结果，这凸显了最优预测在风险管理中的实际重要性。我们在第6节作出总结。

2 风险预测中的评分规则

2.1 评分规则与最优预测分布

评分规则（Gneiting and Raftery，2007年）是衡量概率预测表现的指标。因此，人们常依据已实现事件或数值，通过赋予数值评分来评估和排序相互竞争的概率预测。也就是说，如果预测者生成一个预测分布 P，且事件 y 得以实现，那么评分规则为该预测给出的奖励记为 $S(P, y)$。对于正向评分，在评分规则适当的前提下，两个相互竞争的预测中表现较好的一个会获得较高的评分。评分规则的适当性在此至关重要，因为非适当评分规则可能会给错误的预测赋予较高的平均评分，即与真实数据生成过程不一致的预测。

更形式化地说，假设 P 是预测分布，而 Q_0 是真实数据生成过程。$\mathbb{S}(P, Q_0)$ 表示在 Q_0 下 $S(P, \cdot)$ 的期望值。也就是说，

$$\mathbb{S}(P, Q_0) = \int_{y \in \Omega} S(P, y) dQ_0(y) \tag{1}$$

其中 Ω 为样本空间。如果对于所有 P 和 Q_0，$\mathbb{S}(Q_0, Q_0) \geq \mathbb{S}(P, Q_0)$，则称评分规则为适当评分规则；仅当 $P = Q_0$ 时，若 $\mathbb{S}(Q_0, Q_0) = \mathbb{S}(P, Q_0)$，则称其为严格适当评分规则（Gneiting and Raftery，2007年）。这确保了适当评分规则会在预测揭示真相时达到最大值，前提是预测分布类 $\mathcal{P}$ 包含 Q_0。

当然，在实际中无法获得期望评分 $\mathbb{S}(\cdot, Q_0)$，但我们可以将观测评分的样本均值作为合理的估计量。渐近而言，如果真实预测分布包含在进行最大化所依据的预测分布类中，那么通过优化基于任意适当评分规则的样本准则，就可以恢复真实预测分布。即使预测者提出的预测分布类 $\mathcal{P}$ 不包含 Q_0，这也不会改变这样一个事实：适当评分规则会奖励特定形式的预测准确度，因此，优化将根据该评分规则，在预测分布类中选出最佳预测。

因此，可以使用评分规则来定义准则函数，并通过优化该函数获得参数估计，进而生成在该评分规则下最优的概率预测。假设评分规则为正向取值，且假定预测模型的未知参数记为 $\boldsymbol{\theta} \in \Theta$，令 $P_{\boldsymbol{\theta}}^{t-1} := P(\cdot|\mathcal{F}_{t-1}, \boldsymbol{\theta})$ 表示基于假定模型的一步预测分布函数，其中 $\mathcal{F}_{t-1}$ 表示时间 $t-1$ 可用的所有信息，而 $p(\cdot|\mathcal{F}_{t-1}, \boldsymbol{\theta})$ 表示相应的预测密度

函数。对于满足 $T \geq \tau \geq 1$ 的 τ，令 $\{y_t\}_{t=1}^{T-\tau}$ 表示用于优化的一个大小为 $T-\tau$ 的序列，其中 τ 表示留出样本的大小。定义：

$$\bar{S}(\boldsymbol{\theta}) = \frac{1}{T-(\tau+1)} \sum_{t=2}^{T-\tau} S(P_{\boldsymbol{\theta}}^{t-1}, y_t), \tag{2}$$

那么，通过最大化 $S(\boldsymbol{\theta})$ 得到的估计量 $\boldsymbol{\theta}$：

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\Theta} \bar{S}(\boldsymbol{\theta}) \tag{3}$$

基于这一评分规则，该估计量被称为“最优”，而 P^{t-1} $\hat{\boldsymbol{\theta}}$ 称为最优预测分布。在某些条件下（包括评分规则是适当的，且预测模型设定正确），当 $\hat{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0$ 时，随着$T \to \infty$ 趋于无穷大（对于固定的 τ），$\boldsymbol{\theta}_0$ 代表真实的（向量）参数。若预测模型设定错误，则 $\boldsymbol{\theta}$ 收敛于 $\boldsymbol{\theta}_*$; 后者表示极限准则函数的最大值，而当 T 趋于无穷大时，$S(\boldsymbol{\theta})$ 收敛于该函数。（Gneiting and Raftery，2007年；Martin *et al.*，2022年；Zischke et al，2022年。）

2.2 金融风险预测的评分规则

在风险管理应用中，关注区域通常位于分布的尾部，其中下尾与多头头寸的亏损风险相关，上尾则与空头头寸的亏损风险相关。Diks *et al.*（2011年）提出的删失对数评分（CLS）允许用户对特定关注区域（或多个区域）内的准确预测给予奖励，例如上尾或下尾。其定义如下：

$$S_{CLS}(P_{\boldsymbol{\theta}}^{t-1}, y_t) = I(y_t \in A)\log\left(p(y_t|\mathcal{F}_{t-1}, \boldsymbol{\theta})\right) + I(y_t \in A^c)\log\left(\int_{y\in A^c} p(y|\mathcal{F}_{t-1}, \boldsymbol{\theta})dy\right), \tag{4}$$

其中，A 是关注区域，A^c 是其补集。CLS 已用于评估面向风险管理的预测分布（Jensen and Maheu，2013；Chiang 等，2021），也用于针对特定数据区域的推断程序（Gatarek等，2014；Borowska 等，2020）。Opschoor 等（2017年）还将 CLS 用于预测组合的估计，并发现基于优化 CLS 得到的密度预测优于那些并非旨在奖励尾部准确性的其他预测（另见 Loaiza-Maya 等，2021）。

在金融风险管理中，预测分布有一个至关重要的特征，即 VaR：在一步预测的情形下，它是一步预测分布 $P_{\boldsymbol{\theta}}^{t-1}$ 的 $p\times 100\%$ 分位数，记作 $VaR_p(P_{\boldsymbol{\theta}}^{t-1})$。分位数评分（QS）是针对分位数的适当评分，因此特别适合用于奖励对 VaR 的准确预测（例如，参见 Giacomini and Komunjer，2005年，Bao 等，2006年，以及 Laporta 等，2018年）。QS 定义如下：

$$S_{QS}(P_{\boldsymbol{\theta}}^{t-1}, y_t) = \left[I(y_t \leq VaR_p(P_{\boldsymbol{\theta}}^{t-1})) - p\right]\left(y_t - VaR_p(P_{\boldsymbol{\theta}}^{t-1})\right). \tag{5}$$

本文探讨，针对尾部的适当评分规则（例如经过恰当设定的 CLS 和 QS）校准模型，能否为金融风险管理提供准确的概率预测。作为比较基准，我们采用 LS，其定义如下：

$$S_{LS}(P_{\boldsymbol{\theta}}^{t-1}, y_t) = \log\left(p(y_t|\mathcal{F}_{t-1}, \boldsymbol{\theta})\right), \tag{6}$$

作为一种“局部”评分规则，如果 y_t 的实现值落在 p （$y_t|\mathcal{F}_{t-1}$， $\boldsymbol{\theta}$）的高密度区域内，该规则将给出较高的值；否则则给出较低的值。使用对数评分定义的样本准则等价于对数似然函数。因此，对于 LS 进行优化可得 $\boldsymbol{\theta}$ 的最大似然估计量，即

$$\hat{\boldsymbol{\theta}}_{MLE} = \arg\max_{\boldsymbol{\theta}\in\Theta} \left\{ \frac{1}{T-(\tau+1)} \sum_{t=2}^{T-\tau} \log\left(p(y_t|\mathcal{F}_{t-1}, \boldsymbol{\theta})\right) \right\}, \tag{7}$$

在正确设定（且满足正则性）的假设下，该估计量是 $\boldsymbol{\theta}_0$ 的渐近有效估计量；因此，有必要记录这种可视为构造预测密度的传统方法的表现。

最后，我们指出，预期损失（ES）是一种点预测，通常与 VaR 一同报告，用于评估金融投资中的尾部风险。鉴于其具有一致性等性质，ES 被认为是一种稳健的风险度量（参见 Acerbi 和 Tasche，2002年；Tasche，2002年；以及 Yamai 和 Yoshiba，2005年）。根据定义，由一步预测分布 $P_{\boldsymbol{\theta}}^{t-1}$ 构造的 p^{th} ES 定义为

$$ES_p(P_{\boldsymbol{\theta}}^{t-1}) = \frac{1}{p}\int_0^p VaR_\alpha(P_{\boldsymbol{\theta}}^{t-1})d\alpha, \tag{8}$$

其中， $VaR_\alpha(P_{\boldsymbol{\theta}}^{t-1})$ 表示预测分布的 $\alpha \times 100\%$ 分位数。预期损失无法作为点预测被诱导，这意味着不存在仅根据预期损失就能对模型表现进行一致排序的评分函数（Gneiting，2011年）。然而，Fissler 等人 （2015年）提出了一类评分函数，可联合 诱导风险价值和预期损失的预测准确度，因此可用于衡量预期损失预测准确度（结合风险价值的预测准确度）。 $VaR_p(P_{\boldsymbol{\theta}}^{t-1})$ 和 $ES_p(P_{\boldsymbol{\theta}}^{t-1})$ 的联合评分函数（正向取值）定义为

$$\begin{aligned} S_\eta(P_{\hat{\boldsymbol{\theta}}}^{t-1}, y_t) = -I(\eta \le ES_p(P_{\hat{\boldsymbol{\theta}}}^{t-1}))\Bigg(&\frac{1}{p} I(y_t \le VaR_p(P_{\hat{\boldsymbol{\theta}}}^{t-1}))(VaR_p(P_{\hat{\boldsymbol{\theta}}}^{t-1}) - y_t) \\ &- (VaR_p(P_{\hat{\boldsymbol{\theta}}}^{t-1}) - \eta)\Bigg) - I(\eta \le y_t)(y_t - \eta). \end{aligned} \tag{9}$$

这里，评分函数由实数 η 索引，该实数涵盖了所关注随机变量的取值范围；一个预测模型相对于另一个模型的优越性（就预期损失准确性而言），通过在 η 的相关范围内比较二者各自的评分值来衡量（参见 Ehm 等，2016年，以及 Ziegel 等，2020年）。虽然我们不使用这一评分函数来确定用于生成 $\boldsymbol{\theta}$ 估计值的准则，并最终确定相应的最优预测分布，但我们确实采用

这一评分函数来比较相关尾部聚焦预测分布所生成的预期损失（以及风险价值）预测的准确性，与对数评分（或基于最大似然估计的）预测分布所生成的预测准确性。

3 基于评分的风险预测的表现：模拟证据

明确设计为奖励某些类型准确性的预测——无论是通过本文倡导的优化方法，还是通过适当的（“聚焦型”）贝叶斯后验更新——确实能够优于传统方法生成的预测；参见 Opschoor 等（2017年）、Loaiza-Maya 等（2021年）和 Martin 等（2022年）。在模拟环境中已有研究表明，模型设定错误的程度会显著影响最优预测的表现。简言之，模型设定错误的程度越高，最优（或聚焦型）预测所带来的增益就越大，至少在所假定的预测模型总体上与真实数据生成过程“相容”的条件下是如此。[2] 这一结果预示着，此类有针对性的预测在实证环境中具有良好的实用性；在实证环境中，模型设定错误不可避免，这也为本文开展的分析提供了有力动机——即在实证风险管理环境中寻求最优预测。本节将在模拟环境中，依据预测尾部风险的能力，评估最优尾部聚焦预测相对于传统基于极大似然估计的预测分布的表现。尽管基于极大似然估计的预测分布相对于对数评分而言在形式上是“最优”的，但下文中我们将“最优”一词专门用于指基于尾部侧重评分的预测分布。

3.1 模拟设计

我们从三个不同的真实数据生成过程中生成 T 个随机变量 y_t 的观测值，每个过程都刻画了金融收益及其波动率的主要典型特征。在所有实验中，我们使用带有高斯误差的(1,1)阶 GARCH 模型来定义预测类别 $\mathcal{P}$ 。情景(i)代表正确设定的情形，即在该情形下，真实数据生成过程与预测模型相匹配。情景（ii）以带有 Student-t 误差的 GARCH(1,1)模型作为真实数据生成过程，从而代表四阶矩模型设定错误的情形。最后，在情景（iii）中，我们采用 Smith 和 Maneesoonthorn（2018）提出、并由 Loaiza-Maya 等（2021）使用的偏态随机波动率模型作为真实数据生成过程。该模型假设偏态正态边际分布，使我们能够衡量数据三阶矩模型设定错误时最优预测的表现，同时还可以考察条件确定性波动率设定错误的影响。在情景（ii）和（iii）下，我们分别通过改变 Student-t分布的自由度和偏态正态分布中的形状参数来改变模型设定错误的程度，形状

[2]关于这一点的讨论，请参阅 Martin 等人（2022年）。

参数。各情景的详细信息见表1。

表1：数值分析中使用的三种模拟情景。所有情景均假设预测模型为 GARCH(1,1)，真实数据生成过程则因情景而异。情景(i)表示模型设定正确。情景（ii）表示第一种模型设定错误形式，其中 t_ν 表示 Student-t分布，自由度参数 ν取值为 3 和 12。情景（iii）表示第二种模型设定错误形式，偏态正态边际收益分布的形状参数取值为 0（表示对称边际分布）、-3 和 -5；这些取值依次使边际分布的负偏度逐渐增大。

	情景(i)	**情景（ii）**	**情景（iii）**
真实数据生成过程	$y_t = \sigma_t \epsilon_t$ $\sigma_t^2 = 1 + 0.2 y_{t-1}^2 + 0.7 \sigma_{t-1}^2$ $\epsilon_t \sim i.i.d. N(0,\ 1)$	$y_t = \sigma_t \epsilon_t$ $\sigma_t^2 = 1 + 0.2 y_{t-1}^2 + 0.7 \sigma_{t-1}^2$ $\epsilon_t \sim i.i.d. (\frac{\nu-2}{\nu})^{0.5} * t_\nu$ $\nu \in \{3, 12\}$	$y_t = D^{-1}(F_z(z_t))$ $z_t = \exp(h_t/2)\epsilon_t,\ \epsilon_t \sim i.i.d. N(0,1)$ $h_t = -0.4581 + 0.9(h_{t-1} + 0.4581) + 0.4172\eta_t,\ \eta_t \sim i.i.d. N(0,1)$ $D(.)$ 是偏态正态分布的累积分布函数 形状参数为 $\in \{0, -3, -5\}$
假定模型	$y_t = \mu + \sigma_t \epsilon_t$ $\sigma_t^2 = \alpha_0 + \alpha_1 (y_{t-1} - \mu)^2 + \beta_1 \sigma_{t-1}^2$ $\epsilon_t \sim i.i.d. N(0,\ 1)$	$y_t = \mu + \sigma_t \epsilon_t$ $\sigma_t^2 = \alpha_0 + \alpha_1 (y_{t-1} - \mu)^2 + \beta_1 \sigma_{t-1}^2$ $\epsilon_t \sim i.i.d. N(0,\ 1)$	$y_t = \mu + \sigma_t \epsilon_t$ $\sigma_t^2 = \alpha_0 + \alpha_1 (y_{t-1} - \mu)^2 + \beta_1 \sigma_{t-1}^2$ $\epsilon_t \sim i.i.d. N(0,\ 1)$

我们使用第 2.2 节中介绍的两种适当评分规则——CLS 和 QS，在所有三种情景下估计最优预测。对于 CLS，我们关注 10% 下尾，这对应于持有多头金融头寸所面临的风险，以下将其记为 CLS10。为模拟实践中的情形——真实预测分布未知——用于评估 CLS 的区域 A 根据边际经验分位数而非条件分位数确定。QS 用作评分规则，以衡量 $p \times 100\%$ 条件分位数的预测准确性。我们考虑与实践中计算风险价值和预期损失相关的分位数，具体设定为 $p \times 100\% = \{2.5\%, 5\%, 10\%\}$。我们分别将这三个版本的 QS 记为 QS2.5 、 QS5 和 QS10。由于基于 LS 的预测分布能够帮助我们了解传统极大似然估计法的表现，因此将其用作基准，以评估最优预测带来的改进。

令 $P^{t-1}(\cdot|\boldsymbol{\theta}_i)$ 表示基于高斯 GARCH(1,1) 模型、按照评分规则 S_i 优化的一步预测分布函数。对于表1中指定的每个真实数据生成过程，我们从 Q_0 中抽取 $T = 6000$ 个 y_t 观测值。给定 $\tau = 5000$，我们随后采用扩展估计窗口，从样本量为999的样本开始，得到 $t = T - \tau + 1, ..., T$ 的最优一步预测 $P^{t-1}(\cdot|\boldsymbol{\theta}_i)$。对于每个 t，$\hat{\boldsymbol{\theta}}_i = \{\hat{\mu}, \hat{\alpha}_0, \hat{\alpha}_1\}_i$ 均按式（3）和式（2）针对 $S_i \in \{LS, CSL10, QS2.5, QS5, QS10\}$计算得到。随后，我们通过两种方式记录基于最大似然估计的预测与最优预测的相对表现。首先，我们通过实施 Christoffersen（1998）的条件覆盖检验，评估所有方法产生准确风险价值预测的表现。若预测产生的风险价值超越与名义p 水平之间的差异不显著，且在时间上不相关，则认为这些预测是理想的。其次，我们评估

采用式（9）中定义的评分规则评估预期损失预测的准确性，同时使用 Ehm 等（2016）和 Ziegel 等（2020）的 Murphy 图；如前所述，该图旨在评估一种预测相对于另一种预测的预期损失预测优势。

3.2 风险价值评估

本节评估基于预测密度 $P^{t-1}\cdot\boldsymbol{\theta}_i$ 构建的 2.5%、5% 和 10% 风险价值预测（针对 $S_i \in \{LS, CLS10, QS2.5, QS5, QS10\}$）的准确性。表2列出了经验尾部覆盖率和条件覆盖检验的结果。加粗数值表示在 5% 显著性水平下，条件覆盖检验的零假设未被拒绝的情形。给定列中的加粗斜体 数值表示经验尾部覆盖率最接近名义覆盖率的（不显著）结果。

表2的面板 A 报告了设定正确情形下的结果。正如预期，由于所有评分规则都是适当的，且（扩展的）估计样本规模较大，因此对每一种评分的选择进行式（2）优化，都会得到一个接近 $\boldsymbol{\theta}_0$ 真实值的 $\hat{}\boldsymbol{\theta}$；因此，所有预测分布都非常相似，其样本外表现也具有可比性（无论采用何种衡量方式）——这正是我们在面板 A 的总体结果中所看到的。

相比之下，若考察可能是模型设定错误最严重的情形——形状参数为 -5 的情景（iii），其结果见面板 F，我们可以看到 Martin 等人（2022年）所称的“严格相干性”的证据。也就是说，依据某一特定度量设计、旨在产生最优预测的评分规则，在样本外就该度量而言表现最佳。具体而言，基于 QS($p \times 100$) 的预测分布，对该名义水平下的风险价值预测最准确，且对 $p.$ 的每个取值均如此。此外，基于 CLS10 的预测分布——其本身侧重于下尾准确性——表现几乎与每一种基于 QS($p \times 100$) 的预测分布不相上下；对于所考察的三个风险价值，其经验覆盖率与名义值均无显著差异。在情景（ii）（面板 C）中，模型设定错误体现为极厚尾的 Student-t 创新项，而这与（条件）高斯预测模型不匹配；对于 5% 和 10% 的风险价值，也同样可以看到严格相干性的证据。

对于其余设定错误的情形，确有一些证据表明，尾部侧重评分在样本外的风险价值预测方面有所裨益。最值得注意的是，基于 QS 的最优预测始终 能准确预测相关风险价值（即表中相应数值始终以粗体标出），且不受模型设定错误程度的影响；也就是说，即使基于 QS 的预测分布并非在每一个 个例中都最准确，也从不会表现不佳，这为在实证场景中普遍采用基于 QS 的预测分布提供了有力依据，因为此类场景的目标是准确覆盖风险价值。

表2：表1中情景 (i)–(iii) 下风险价值的预测性能。报告的数字为超越次数的比例。粗体表示超越次数与名义水平无显著差异，且时间上不相关，因此在 5% 的显著性水平下，条件覆盖检验未能拒绝零假设。给定列中以粗体标出的斜体数值，是未能拒绝条件覆盖检验零假设且最接近名义 p 水平的超越次数。

.

	样本外超越次数					
优化器	2.5%的风险价值	5%的风险价值	10%的风险价值	2.5%的风险价值	5% 风险价值	10% 风险价值
	面板 A：情景(i)			**面板 D：情景（iii），形状参数=0**		
MLE	***2.94%***	**5.24%**	**10.18%**	**2.30%**	***5.12%***	**10.82%**
CLS10	3.26%	**5.52%**	**9.98%**	**2.42%**	***5.12%***	10.46%
QS2.5	**2.98%**	***4.86%***	**9.36%**	**2.56%**	**5.38%**	10.76%
QS5	3.38%	**5.70%**	***10.00%***	***2.46%***	**5.14%**	10.72%
QS10	3.50%	5.90%	**10.60%**	**2.26%**	**4.66%**	***10.46%***
	面板 B：情景（ii），带有 $\nu = 12$			**面板 E：情景（iii），形状参数=-3**		
MLE	2.98%	**5.38%**	**9.72%**	4.10%	7.08%	11.76%
CLS10	***2.78%***	**5.28%**	***10.00%***	***2.50%***	***5.10%***	**9.54%**
QS2.5	**2.82%**	***4.88%***	9.06%	**2.54%**	**4.62%**	8.60%
QS5	3.10%	**5.32%**	**9.40%**	**2.88%**	**5.24%**	9.16%
QS10	3.80%	6.06%	**10.40%**	3.52%	6.06%	***10.38%***
	面板 C：情景（ii）下 $\nu = 3$			**面板 F：情景（iii）下形状参数=-5**		
MLE	***2.70%***	4.00%	7.06%	4.54%	7.66%	12.06%
CLS10	**2.14%**	4.24%	10.96%	**2.66%**	**5.24%**	**9.48%**
QS2.5	**2.92%**	4.32%	7.24%	***2.58%***	**4.70%**	8.22%
QS5	4.10%	***5.28%***	7.96%	**2.80%**	***5.20%***	8.88%
QS10	6.20%	7.78%	***10.14%***	3.72%	6.26%	***10.38%***

3.3 预期损失评估

基于（9），Ziegel 等人（2020年）提出了“预测优势”这一概念，用于比较评估 VaR 和预期损失的联合配对。当对于任意 η 值， $E[S_\eta(P^{t-1}\text{-}\boldsymbol{\theta_i}, y_t)] \geq E[S_\eta(P^{t-1}\text{-}\boldsymbol{\theta_j}, y_t)]$ 时，预测分布 $P^{t-1}\text{-}\boldsymbol{\theta_i}$ 被认为弱支配预测分布 $P^{t-1}\text{-}\boldsymbol{\theta_j}$。我们按照 Ziegel 等人（2020年）的方法，使用 Murphy 图评估相互竞争的预测分布之间的支配关系：横轴为 η，纵轴为平均评分差，计算方式如下：

$$\Delta_\eta \left(P^{t-1}_{\hat{\boldsymbol{\theta}}_i}, P^{t-1}_{\hat{\boldsymbol{\theta}}_{MLE}} \right) = \widehat{E}\left[S_\eta(P^{t-1}_{\hat{\boldsymbol{\theta}}_i}, y_t) \right] - \widehat{E}\left[S_\eta(P^{t-1}_{\hat{\boldsymbol{\theta}}_{MLE}}, y_t) \right], \tag{10}$$

其中，期望值按照通常方法，以相关样本外评分的样本均值进行估计。为使分析易于处理，我们仅针对 $p = 0.1$ 计算（10），并分别使用基于 CLS10 和 QS10 计算得到的最优预测分布 $P^{t-1}\text{-}\boldsymbol{\theta_i}$。在每种情况下，比较对象（从（10）所用的符号可清楚看出）都是基于极大似然估计的预测分布。同样，表1列出了所有情景的结果。

情景（i）和情景（ii）的墨菲图见图 1，情景（iii）的图见图 2。每种情形下，左列展示最优 QS10 预测分布与 MLE预测值之间的差异；右列展示 CLS10 预测分布与 MLE预测值之间的差异。在每张图中，平均预期损失评分差异以黑色实线表示，自助法得到的 95% 置信区间以灰色阴影表示，零基准则以红色水平虚线表示。

在图 1 中，对于情景（ii）中设定错误最严重的情形，在合理范围的 η 内，QS10 最优预测分布优于 MLE预测值，且 $\nu = 3$ （面板 E）。另一方面，CLS10 预测分布产生的平均评分绝对值大于 MLE预测值，但差异在统计上并不显著（面板 F）。在设定正确和设定错误程度较轻的情形中（面板 A - D），我们未观察到任何一种侧重尾部的预测分布与 MLE预测值之间存在显著差异。

在情景（iii）下，与 MLE 预测值相比，使用 QS10 和 CLS10 最优预测分布总体上带来的益处更加明显。正如预期，当真实预测分布具有对称性时（面板 A 和 B），最优预测分布与 MLE 预测值的平均评分并无显著差异。然而，随着真实预测分布变得更加不对称（形状参数非零），我们观察到 QS10 和 CLS10 最优预测分布明显优于 MLE 预测值，如图 2 的面板 C–F 所示。此外，在这些情况下，QS10 预测分布的平均评分差异幅度大于 CLS10 预测分布相应的平均评分差异。这表明，与使用 CLS10 评分相比，使用 QS 评分校准预测分布，在准确预测（联合）预期损失和风险价值方面带来的益处更大。另一项观察也进一步支持了这一结论：相较于 CLS10 预测分布，QS10 预测分布（相对于 MLE 预测值）的优势更大；这与前者在风险价值方面的整体表现优于后者这一事实相吻合

第3.2节中的评估（尤其参见表2中的面板 F）。

4 金融收益率的实证分析

在第3节中，我们使用模拟金融收益特征的人工数据，说明了使用适合预测应用场景的评分来校准预测分布所带来的益处。具体而言，我们发现，在这一情境下，模型设定偏误程度越高，最优预测带来的益处就越大。

本节使用高斯 GARCH(1,1) 模型作为预测模型，为一系列实证金融指数的收益率生成最优预测分布：标准普尔S&P500、标普小盘股、标普中型股及其九个行业指数。具体而言，我们将展示，通过按照能够使预测性能聚焦于金融场景中重要准确性的评分规则校准 GARCH(1,1) 模型，可以提升预测性能，尽管预测模型本身很可能存在模型设定错误。我们在所有实证分析中采用以下评分， $S_i \in \{LS, CSL10, CLS20, CLS80, CSL90, QS2.5, QS5, QS10\}$。在第4.2节中，我们评估预测收益率分布与用于校准的评分规则的一致性——即让用于计算样本外平均性能的评分与用于校准的评分相匹配。在第4.2节和第4.4节中，我们分别展示，采用尾部聚焦评分能够提高风险价值和预期损失预测的准确性。

4.1 数据描述

我们的实证分析考察了1996年8月14日至2020年7月30日期间的数据。标准普尔500指数的数据取自雅虎财经，所有指数的数据均来自全球金融数据（GFD）数据库。我们为每个指数构建了6000个连续复合日收益观测值，其中前1000个观测值作为初始训练样本，后5000个观测值用于样本外预测评估。与模拟实验一样，我们采用扩展估计窗口。2000年8月1日至2020年7月30日的样本外期间包含两个波动剧烈的时期：全球金融危机（GFC）和近期的新冠肺炎大流行。

表3列出了我们考察的12个指数的详细信息，以及所有收益率序列的描述性统计。Jarque-Bera检验结果均被拒绝，表明这12个收益率序列均不服从高斯分布。正如日度金融收益率通常呈现的特征，厚尾现象十分显著。除标准普尔45指数外，我们还观察到所有指数均呈负偏度。对平方收益率进行的Ljung-Box检验也表明，所有序列均存在时变波动率。

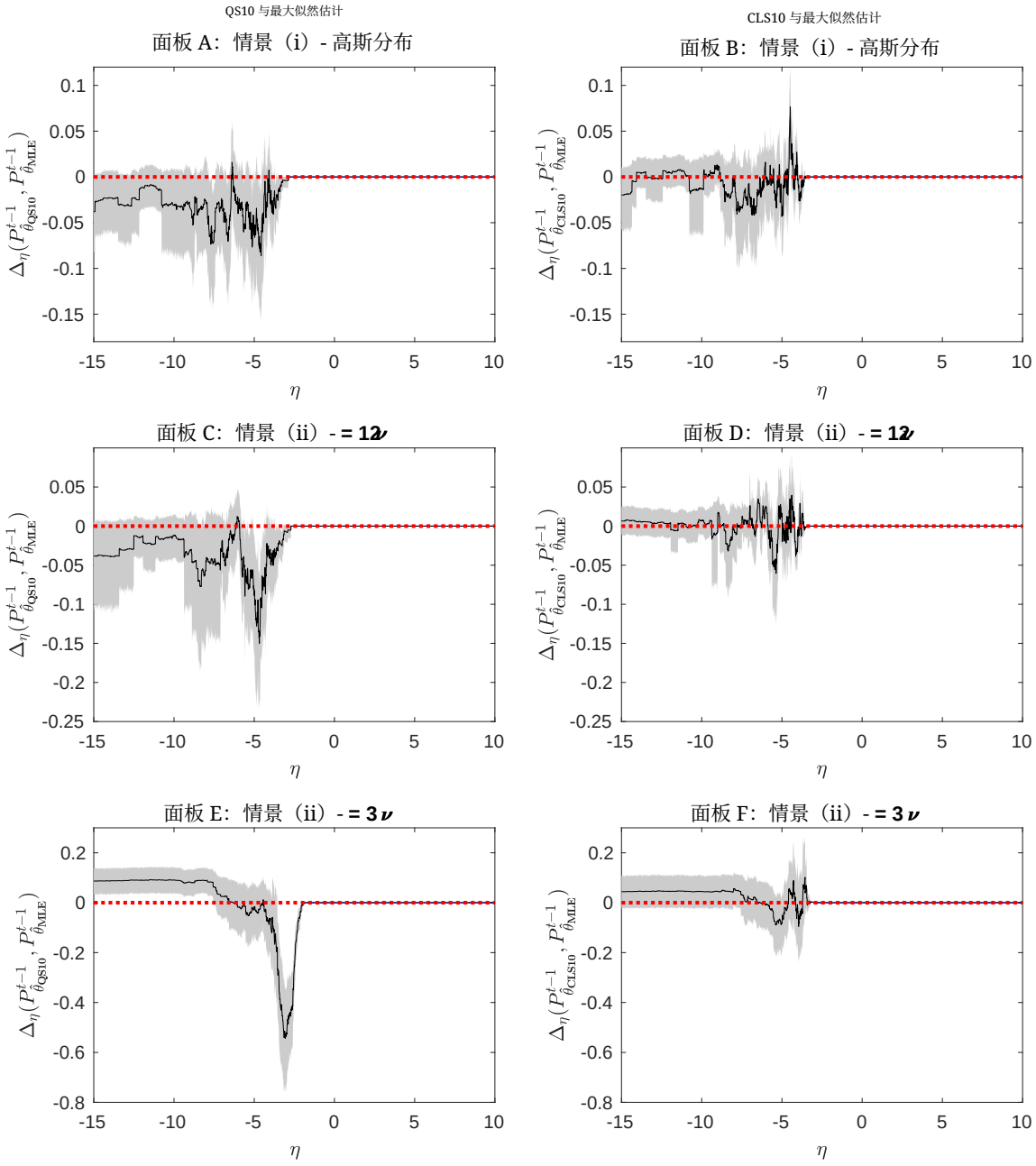


图 1：表1中情景(i)和（ii）下数据生成过程的10%预期损失预测性能。每个面板展示 Murphy 图，绘制 $\Delta_\eta\left(P_{\boldsymbol{\theta_i}}^{t-1}\hat{},P^{t-1}\hat{}_{\boldsymbol{\theta_{MLE}}}\right)$ 随 η 的变化，阴影区域表示95%自助法置信区间。左侧列评估 QS10 预测分布相对于 MLE 预测值的表现，右侧面板评估 CLS10 预测分布相对于 MLE 的表现

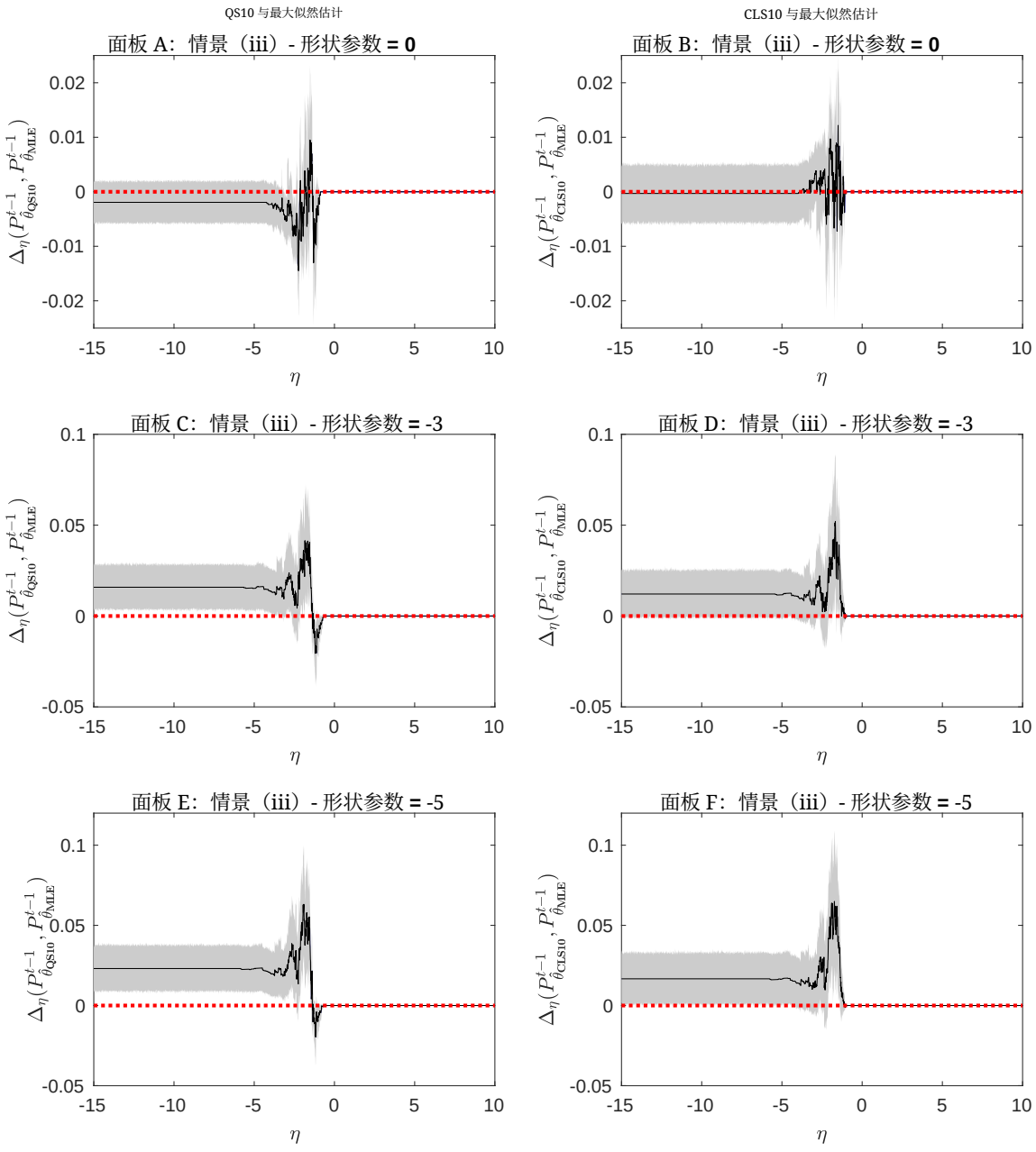


图 2：表1中情景（iii）的数据生成过程的 10% 预期损失预测性能。每个面板均展示 Murphy 图，绘制 $\Delta_\eta\left(P_{\boldsymbol{\theta_i}}^{t-1}\hat{}, P^{t-1}\hat{}_{\boldsymbol{\theta_{MLE}}}\right)$ 与 η 的关系，阴影区域表示 95%自助法置信区间。左侧列评估 QS10 预测分布相对于 MLE预测值的表现，右侧面板评估 CLS10 预测分布相对于 MLE

表3：所研究的12个标普指数的描述及汇总统计量。汇总统计量包括均值、标准差（Std）、偏度和峰度统计量。JB.Test列报告Jarque-Bera检验正态性的p值，LB.test列报告对平方收益率进行序列相关检验的p值，滞后阶数为20。

指数	描述	Mean	Std	偏度	峰度	JB.Test	LB.Test
标普500	标普500市场指数	0.026	1.243	-0.394	13.467	0.000	0.000
标普小盘股	标普小型股指数	0.031	1.244	-0.553	10.938	0.000	0.000
标普中型股	标普中型股指数	0.034	1.444	-0.635	13.147	0.000	0.000
标普15	标普材料指数	0.018	1.359	-0.323	10.270	0.000	0.000
标普20	标普工业指数	0.023	1.527	-0.419	10.857	0.000	0.000
标普25	标普非必需消费品指数	0.035	1.373	-0.274	11.194	0.000	0.000
标普30	标普必需消费品指数	0.025	1.378	-0.175	13.683	0.000	0.000
标普35	标普医疗保健指数	0.034	0.997	-0.197	10.152	0.000	0.000
标准普尔40	标普金融指数	0.015	1.216	-0.178	18.230	0.000	0.000
标准普尔45	标普信息技术指数	0.042	1.915	0.063	9.236	0.000	0.000
标准普尔50	标普电信服务指数	0.007	1.772	-0.044	9.941	0.000	0.000
标准普尔55指数	标普公用事业指数	0.016	1.232	-0.096	16.298	0.000	0.000

4.2 最优收益率预测的性能

表 4 汇总了 GARCH(1,1)模型对全部 12 个指数的样本外预测性能，评估依据为全部八种评分规则。对于每个序列，在每一列中我们报告两个数值：i）基于极大似然估计的预测分布的样本外平均得分；ii）针对用于衡量性能的评分规则进行校准的预测分布的样本外平均得分，该预测分布称为“最优”预测分布。对于每个序列，第一列中的这两个数值必然相同，因为该列使用的评分规则是对数评分，而基于极大似然估计的预测分布按照对数评分是最优的。在其他所有列中，这两个数值通常不同。对于某个序列和某一列，加粗的数值表示在样本外期间平均得分较高的预测分布（基于极大似然估计的预测分布或最优预测分布）；星号表示根据 Giacomini 和 White（2006）的检验，在 5% 显著性水平下拒绝预测能力相同的零假设。

总体而言，我们发现，将 GARCH(1,1) 模型校准至所关注的评分规则，能够带来预测增益，而且这种增益往往显著。具体来说，对于全部 12 个序列，将预测分布校准至这些相同的评分规则后，预测表现均有显著改善（CLS10 和 CLS20 均显示了这一点）。此外，使用 CLS80 和 CLS90 校准预测分布，也能提升上尾精度（有时提升显著）。正如先前研究所阐明的（例如，见 Loaiza-Maya等，2021），相对表现存在这种差异并不令人意外；这表明，假定预测模型的设定错误对真实（未知）预测分布支撑集不同区域的影响各不相同，而真实预测分布本身也因序列而异。也就是说，对于所有序列，假定的预测模型在下尾（尤其是这一部分）的表现显然较差，因此按照 CLS10 和 CLS20 进行校准

可带来显著改善。话虽如此，通过下尾分位数评分进行校准所带来的改善并非在各序列间一致，不过，正如我们将在下一节看到的那样，当目标是准确预测风险价值分位数时，这些评分规则确实能带来益处。最后需要指出的是，在所有考察的序列中，标准普尔中型股的收益率偏度最大。因此，相对于基于极大似然估计的预测分布，最优预测分布在所有样本外度量上均有一致且显著的改善。这说明，对称 GARCH(1,1) 模型的设定错误在这一序列中最为明显，因此最优预测的连贯性也最为突出。还应注意，表中零星出现了几种情况：最大似然估计预测值的评分高于最优预测分布，但这些情况下的差异在统计上并不显著。

4.3 风险价值预测

现在，我们来记录基于最大似然估计（MLE）预测分布和最优（QS）预测分布的风险价值预测准确性。最优 QS 指通过式（5）中的 QS 评分进行校准的预测分布，其中，预测名义水平分别为 2.5%、5% 和 10% 时的风险价值所用的 $p \times 100 = 2.5$、5 和 10。经验覆盖率记录于表 5；加粗数值表示在 5% 显著性水平下，Christoffersen（1998）条件覆盖检验的零假设未被拒绝的情形。

与第3.2节模拟实验中的观察结果一致，使用适当的 QS 评分校准风险价值预测，整体上可获得更好的覆盖率结果。在 12 个指数中，除一个以外，最优 QS 预测得出的 2.5% 风险价值覆盖率与名义覆盖水平在统计上没有显著差异；而 12 个 MLE预测值中仅有两个通过了条件覆盖检验。对于 5% 风险价值，最优 QS 预测在所有指数上都获得了适当的覆盖率，而 MLE 预测仅在七个指数上获得了适当的覆盖率。对于 10% 风险价值，MLE 和最优 QS 的表现相当，各自在一半的指数上获得了适当的覆盖率。结果表明，尽管在表 4 的 QS 评估列中，最优 QS 预测分布的表现与 MLE 预测分布不相上下，但若专门用于准确预测相应的风险价值分位数，最优 QS 预测分布整体上确实优于基于极大似然估计的预测分布，尤其是在较低的 p.

4.4 预期损失预测

在第3节第3.3小节中，我们确立了：当使用 QS 评分校准的预测分布在风险价值预测方面优于 MLE预测值时，我们也观察到 QS预测值在预期损失预测方面占优。本节通过评估最优 QS预测值和 MLE预测值生成的预期损失预测，对这一观察结果进行实证检验。

表 4：使用高斯 GARCH(1,1) 模型对 12 个不同标普指数的收益率进行预测时的平均样本外评分。对于每个指数，模型分别使用对数评分（简称“MLE”）以及列标题所示的尾部聚焦评分（简称“最优”）进行校准。列标题表示用于评估概率预测样本外表现的评分。对于给定序列和列，加粗数值表示在样本外期间产生较高平均评分的预测分布（基于 MLE 或最优）；星号表示根据 Giacomini 和 White（2006）的检验，在 5% 显著性水平下拒绝预测能力相等的零假设。

	平均样本外评分							
优化器	LS	CLS10	CLS20	CLS80	CLS90	QS2.5	QS5	QS10
标普500								
MLE	-1.365	-0.358	-0.600	-0.532	-0.300	-0.075	-0.125	-0.202
最优	-1.365	**-0.344**⋆	**-0.584**⋆	**-0.528**⋆	**-0.296**⋆	**-0.074**	-0.125	-0.202
标普小盘股								
MLE	-1.619	-0.420	-0.690	-0.653	-0.395	-0.088	-0.150	-0.247
最优	-1.619	**-0.412**⋆	**-0.682**⋆	**-0.651**	**-0.393**	**-0.086**⋆	**-0.147**⋆	**-0.245**⋆
标普中型股								
MLE	-1.493	-0.382	-0.649	-0.577	-0.346	-0.082	-0.138	-0.227
最优	-1.493	**-0.372**⋆	**-0.638**⋆	**-0.572**⋆	**-0.342**⋆	**-0.079**⋆	**-0.136**⋆	**-0.225**⋆
标普15								
MLE	-1.658	-0.394	-0.654	-0.639	-0.356	-0.094	-0.159	-0.261
最优	-1.658	**-0.386**⋆	**-0.645**⋆	**-0.637**	**-0.354**	**-0.092**	**-0.158**	**-0.260**
标普20								
MLE	-1.521	-0.401	-0.643	-0.565	-0.339	-0.087	-0.144	-0.232
最优	-1.521	**-0.386**⋆	**-0.627**⋆	**-0.563**	**-0.337**⋆	**-0.084**⋆	**-0.142**	**-0.231**
标普25								
MLE	-1.505	-0.375	-0.609	-0.571	-0.316	-0.083	-0.140	-0.228
最优	-1.505	**-0.365**⋆	**-0.597**⋆	**-0.567**⋆	**-0.315**	**-0.081**	**-0.137**⋆	-0.228
标普30								
MLE	-1.157	-0.298	-0.506	-0.474	**-0.251**	-0.058	**-0.095**	-0.154
最优	-1.157	**-0.456**⋆	**-0.285**⋆	-0.474	-0.252	-0.058	-0.096	-0.154
标普35								
MLE	-1.367	-0.317	-0.554	-0.491	-0.237	-0.072	-0.121	-0.195
最优	-1.367	**-0.308**⋆	**-0.542**⋆	**-0.489**⋆	**-0.234**⋆	-0.072	-0.121	-0.195
标准普尔40								
MLE	-1.668	-0.403	-0.630	-0.588	-0.351	-0.107	-0.176	-0.283
最优	-1.668	**-0.393**⋆	**-0.619**⋆	**-0.587**	-0.351	**-0.105**	**-0.175**	-0.283
标准普尔45								
MLE	-1.686	-0.291	-0.547	**-0.507**	**-0.250**	-0.097	-0.166	-0.273
最优	-1.686	**-0.282**⋆	**-0.536**⋆	-0.508	-0.252	-0.097	-0.166	-0.273
标准普尔50								
MLE	-1.566	-0.340	-0.597	-0.543	-0.299	-0.089	-0.147	-0.235
最优	-1.566	**-0.329**⋆	**-0.586**⋆	**-0.542**	-0.299	**-0.088**	-0.147	-0.235
标准普尔55指数								
MLE	-1.419	-0.390	-0.651	-0.574	-0.333	-0.078	-0.130	**-0.210**
最优	-1.419	**-0.381**⋆	**-0.641**⋆	**-0.572**⋆	**-0.330**⋆	-0.078	-0.130	-0.212

表 5：基于高斯 GARCH(1,1) 预测模型的 12 个指数收益率序列的预测风险价值覆盖率统计量。“最优 QS”指使用 $p \times 100 = 2.5$、5 和 10 的 QS 评分校准得到的预测分布。 加粗数值表示在 5% 显著性水平下，Christoffersen（1998）条件覆盖检验的零假设未被拒绝的情形。

		样本外覆盖率		
指数	优化器	2.5% 风险价值	5%风险价值	10%风险价值
标准普尔 500指数	MLE	3.400%	**5.300%**	**10.040%**
	最优 QS	**2.540%**	**4.480%**	8.980%
标普小盘股	MLE	3.660%	6.420%	11.480%
	最优 QS	3.160%	**5.400%**	10.800%
标普中型股	MLE	3.540%	6.240%	11.120%
	最优 QS	**2.800%**	**4.920%**	**9.820%**
标准普尔 15	MLE	3.280%	5.640%	**9.780%**
	最优 QS	**2.840%**	**5.540%**	**10.200%**
标准普尔 20	MLE	3.560%	**5.280%**	**9.900%**
	最优 QS	**2.580%**	**4.960%**	**10.080%**
标准普尔 25	MLE	**3.220%**	**5.420%**	**9.880%**
	最优 QS	**2.600%**	**5.160%**	**9.520%**
标普 30	MLE	**2.640%**	**4.840%**	8.660%
	最优 QS	**2.400%**	**4.640%**	8.620%
标准普尔 35	MLE	3.100%	**5.000%**	8.480%
	最优 QS	**2.380%**	**4.440%**	8.880%
标准普尔 40	MLE	3.060%	**5.260%**	**9.100%**
	最优 QS	**2.540%**	**5.000%**	**9.240%**
标准普尔 45	MLE	3.360%	5.920%	**9.980%**
	最优 分位数评分	**2.520%**	**4.840%**	**9.220%**
标准普尔 50 指数	MLE	3.240%	**5.240%**	9.040%
	最优 QS	**2.520%**	**5.020%**	**9.420%**
标准普尔 55	MLE	3.520%	5.660%	10.020%
	最优 QS	**2.840%**	**5.380%**	10.320%

图 3—5 分别展示了 12 个收益率序列在 2.5%、5% 和 10% 预期损失下的墨菲图。我们再次强调，如果对于 η 的一系列合理取值，自助法置信区间均高于零，则认为 QS预测值在预期损失预测方面优于 MLE预测值。

在图 3 中，对于标普小盘股、标普中型股和标普20指数的 2.5% 预期损失预测，平均得分差异在统计上显著大于零。在 5% 预期损失的情形下（图 4），标普小盘股、标普中型股和标普25指数均观察到最优 QS预测值占优；而对于 10% 预期损失，则观察到标普小盘股和标普中型股指数占优（图 5）。这些结果与表 4 中相应序列集合的 QS预测值（相较于 MLE预测值）几乎一致的优越表现相吻合。

5 优化预测VIX分布以实现动态风险管理

VIX是芝加哥期权交易所（CBOE）发布的波动率指数，根据期权价格编制，用于衡量美国市场波动率。众所周知，VIX指数与股票市场指数呈负相关；当股票市场指数普遍下跌、市场承压时，VIX通常会上升。股票指数与以VIX指数为标的的可交易期货合约之间也存在这种负相关关系，据报道，这种相关性在期货市场中更强（Moran和Liu，2020年）。因此，VIX期货合约常被用作对冲市场下跌风险的工具。本节将说明，优化VIX指数的预测分布如何有助于改进涉及VIX期货头寸的风险管理策略。

5.1 VIX指数的预测模型

VIX指数是一种无需模型的期权隐含波动率度量，根据以标普500股票市场指数为标的的期权价格提取而得（更多详情请参见CBOE，2021年）。理论上，VIX指数衡量的是标的资产积分方差的风险中性前瞻预期（参见Britten-Jones和Neuberger，2000年，以及Jiang和Tian，2005年）。因此，VIX指数的变动不仅紧密反映波动率本身，还包含衍生品交易者对未来的预期信息，因而常被视为市场走势的领先指标。

我们从雅虎财经收集了1996年9月27日至2020年7月30日期间的每日 VIX指数数据。表6展示了 log(VIX_t) 的描述性统计。统计结果表明，log(VIX_t) 的边际密度呈正偏态，且明显不服从正态分布。与波动率度量的典型特征一致，VIX指数具有很强的持续性，

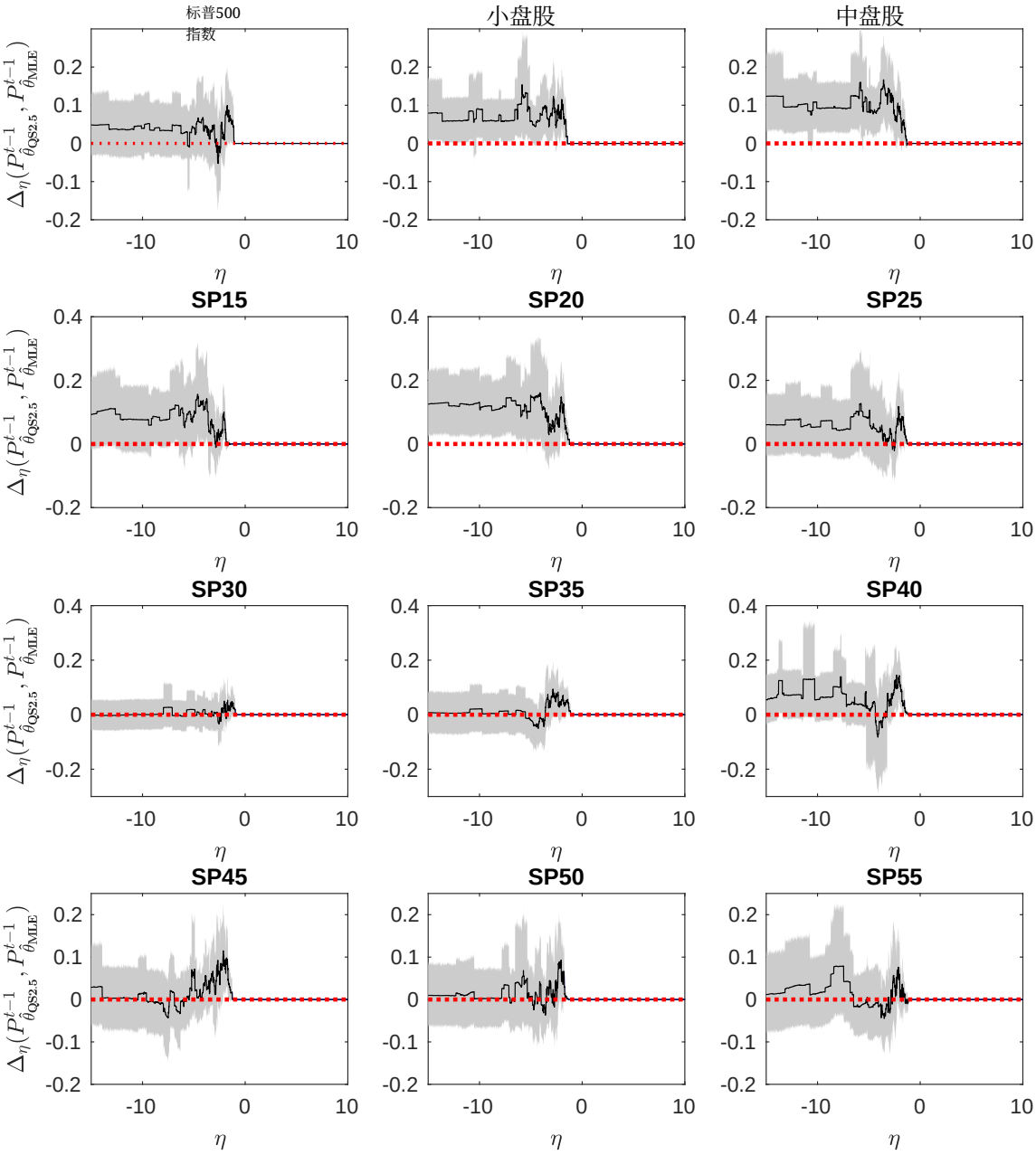


图 3：使用高斯 GARCH(1,1) 作为预测模型时，12 个标普指数的 2.5% 预期损失表现。所有墨菲图均描绘了 $\Delta_\eta\left(P^{t-1}{}_{\hat{\boldsymbol{\theta}}_{QS2.5}}, P^{t-1}{}_{\hat{\boldsymbol{\theta}}_{MLE}}\right)$，阴影区域为 95%自助法置信区间。

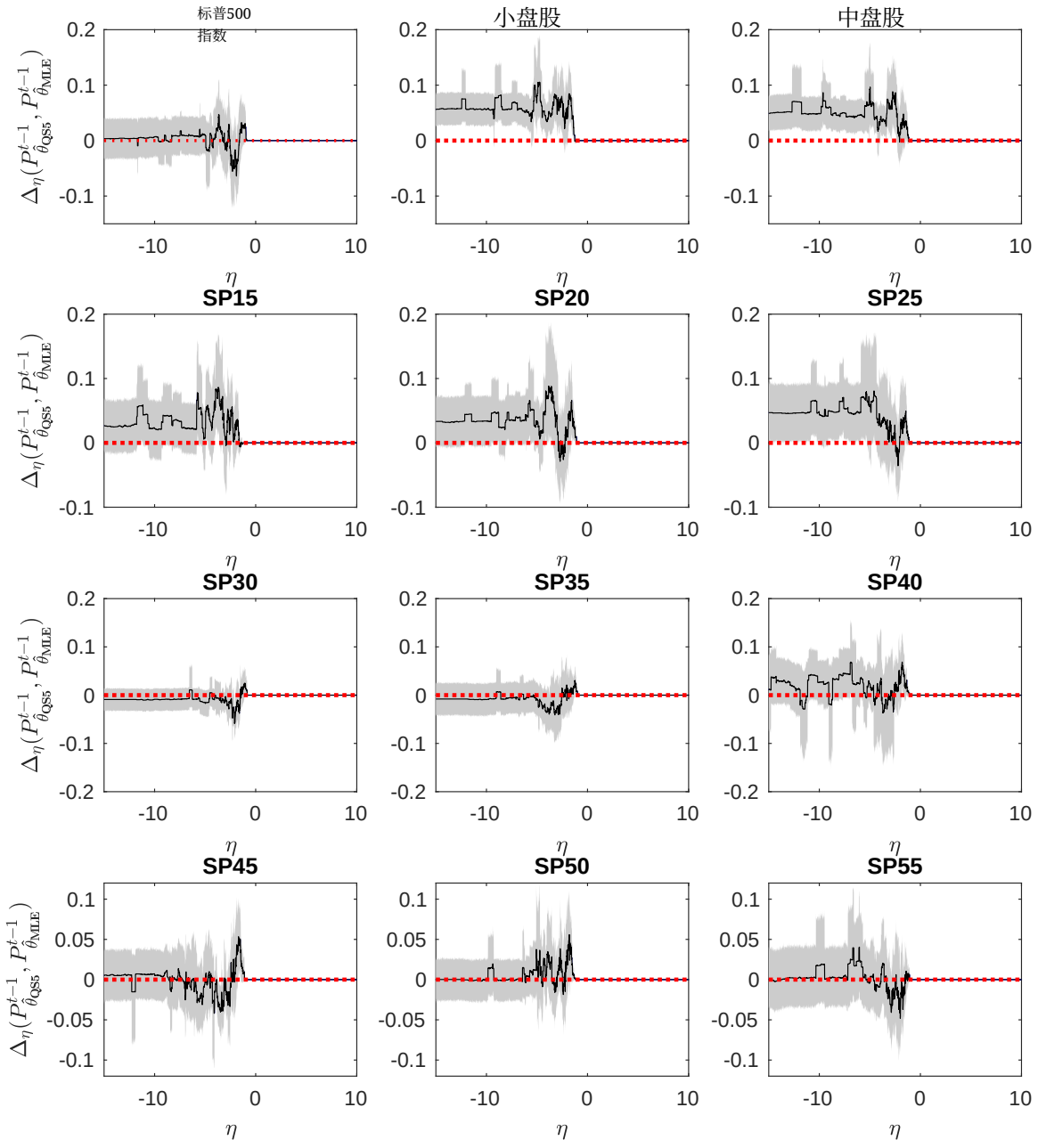


图 4：使用高斯 GARCH(1,1) 作为预测模型时，12 个标普指数的 5% 预期损失表现。所有 Murphy 图均描绘 $\Delta_\eta\left(P^{t-1}_{\hat{\boldsymbol{\theta}}_{QS5}}, P^{t-1}_{\hat{\boldsymbol{\theta}}_{MLE}}\right)$，阴影区域表示 95%自助法置信区间。

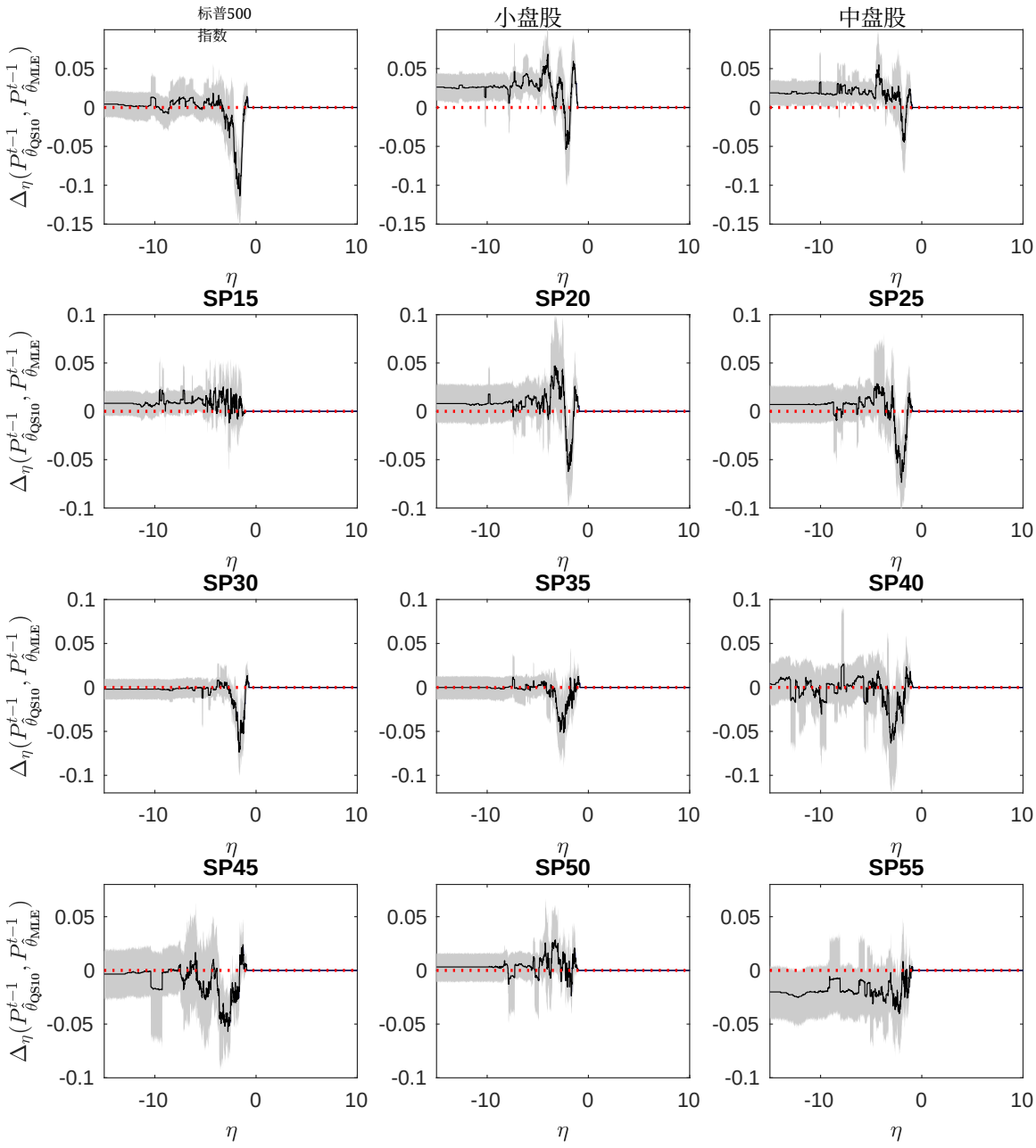


图 5：使用高斯 GARCH(1,1)作为预测模型时，12 个标普指数的 10% 预期损失表现。所有墨菲图均描绘了 $\Delta_\eta\left(P^{t-1}{}_{\hat{}\boldsymbol{\theta}_{QS10}}, P^{t-1}{}_{\hat{}\boldsymbol{\theta}_{MLE}}\right)$，阴影区域表示 95%自助法置信区间。

表6：VIX波动率指数对数的描述性统计。Min和Max分别是样本序列中的最小值和最大值。偏度用于衡量非对称性，峰度用于衡量厚尾程度；正态分布的峰度为3。'JB.Test'是Jarque-Bera正态性检验的统计量，其p值记录在表中。'LB.Test'是基于20个滞后阶数对波动率平方进行的序列相关检验，其p值记录在表中。

序列	Min	中位数	Mean	Max	偏度	峰度	JB.Test	LB.Test
Log(VIX)	2.213	2.916	2.936	4.415	0.604	3.363	0.000	0.000

随着滞后阶数增加，其自相关函数的幅度仍然较大且具有统计显著性（表6中的 Ljung-Box自相关检验采用了20个滞后阶数）。

为捕捉VIX指数的长记忆特征，我们采用异质自回归（HAR）模型来构建VIX指数的预测分布。HAR模型由Corsi（2009年）提出，是一种简单的加性线性模型，它使用相关因变量滞后值的移动平均作为回归变量。它不属于长记忆模型，但通过简单的自回归结构，能够产生与实际金融波动率中观察到的持续性几乎无法区分的持续性。以往研究已将HAR模型及其扩展模型用于研究VIX指数的可预测性；例如，Maneesoonthorn 等人（2012年）和Fernandes 等人（2014年）。我们在HAR模型中加入GARCH成分，以考虑VIX分布二阶矩的时变性。
HAR-GARCH模型定义如下：

$$\log(VIX_{t+1}) = \beta_0 + \beta_1 \log(VIX_t) + \beta_2 \log(VIX_{t-5,t}) + \beta_3 \log(VIX_{t-22,t}) + \sigma_{t+1} z_{t+1} \tag{11}$$

其中

$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1 (\sigma_t z_t)^2 + \alpha_2 \sigma_t^2, \tag{12}$$

$$z_{t+1} \sim i.i.d.N(0,1) \tag{13}$$

and

$$\log(VIX_{t-m,t}) = \frac{1}{m} \sum_{i=1}^{m} \log(VIX_{t+i-1}). \tag{14}$$

5.2 VIX 指数的最优预测的表现

采用与第 3.1 节和第 4.2 节相似的扩展窗口预测评估程序，初始样本量设为 1000，样本外规模为 $\tau = T - 1000 = 5000$，我们基于式 (11) 至 (14) 中定义的 HAR-GARCH 生成并评估预测密度。

表7：VIX指数预测的平均样本外评分，其中预测模型为具有高斯误差的 HAR-GARCH模型。该模型采用对数评分（记为“MLE”）以及列标题中的 CLS（记为“最优”）进行校准。列标题表示用于评估概率预测样本外表现的评分。某一列中的粗体数值表示在样本外期间产生较高平均评分的预测（基于 MLE 或最优），星号表示根据 Giacomini 和 White（2006）的检验，在 5% 显著性水平下拒绝预测能力相等的零假设。

优化器	平均样本外评分				
	LS	CLS10	CLS20	CLS80	CLS90
MLE	1.302	0.344	0.566	0.131	0.059
最优	1.302	**0.350***	**0.569**	**0.134**	**0.061**

我们使用与 VIX 边际分布的 10% 和 20% 下尾以及 80% 和 90% 上尾相对应的 LS 和 CLS 评分来优化预测分布。样本外表现的评估也采用同一组评分。

表7记录了基于MLE的预测分布和最优预测分布的平均样本外评分。再次强调，粗体数值表示给定列中较大的平均样本外评分；星号表示根据Giacomini 和 White（2006）的检验，MLE预测分布与最优预测分布评分之间的差异显著异于零。该表结果显示，在所有 情况下（MLE预测分布与最优预测分布一致的情况除外），最优预测产生了更大的平均样本外评分，其中CLS10预测分布的改进具有统计显著性。

5.3 使用VIX期货进行动态风险管理的应用

Moran和Liu（2020）展示了如何使用VIX期货进行动态风险管理。他们提出了一种回测策略，包括对股票投资组合配置95%的多头头寸，以及对VIX期货配置5%的多头头寸。由于VIX期货的价值来源于VIX现货指数，他们还提出了一种动态配置策略，即VIX期货中的头寸取决于VIX现货指数本身的水平。

为说明在此类情境下基于评分的校准可能带来的益处，我们对 Moran 和 Liu（2020）展示的动态策略进行调整，将基于 HAR-GARCH 模型的 VIX 指数预测分布纳入其中，并采用适当版本的尾部聚焦 CLS 进行优化。我们设计了两种动态交易策略，其中，最短期限 VIX 期货的持仓由基于 VIX 指数最优预测分布的两种决策规则之一触发。[3] 在此期间，VIX 期货合约的开盘价和收盘价

[3]我们注意到 Konstantinidi 等人 （2008年）和 Taylor（2019年）的相关研究，其中涉及与

2008年9月24日至2021年5月28日期间的数据直接取自芝加哥期权交易所网站。

首先，我们对一项预测概率规则进行回测：只有当 Pr($VIX_{t+1} > VIX_t|\mathcal{F}_t$) > 0.5 时，才建立 VIX期货的 5% 多头头寸。也就是说，当 VIX指数很可能在次日上涨时，我们会建立 VIX期货多头头寸，以抵消预期中股票市场可能出现的下跌。为实施该预测概率规则，我们根据与满足 $VIX_t > VIX_{t-1}$ 的区域相关的上尾CLS，对 VIX预测分布进行优化，该指标记为 CLS($VIX_t > VIX_{t-1}$)。

其次，我们利用 80^{th} 预测VIX百分位来制定交易规则。根据此规则，只有当即期VIX的 80^{th} 预测百分位大于 40% 时，才会在 VIX期货市场建立 5% 的多头头寸。也就是说，只有当 VIX预测分布中界定上尾 20% 的数值足够高时，投资者才会建立头寸，以抵消股票市场可能出现的下跌。为实施该规则，我们根据 CLS80 对 VIX预测分布进行优化。

在两种动态设定下，投资者在交易日开始时，根据截至前一交易日的数据构建的相关预测分布，决定是否采用VIX期货头寸。如果决定建立VIX期货头寸，则持有最短期限VIX期货合约的多头头寸。VIX期货头寸在日终平仓，投资者实现的收益反映了该交易日VIX期货从开盘价到收盘价的变动。

表8报告了基于上述各项动态策略的投资组合绩效指标。作为比较，表中还报告了静态投资组合的表现；该投资组合由标普500市场投资组合的95%多头头寸和最短期限VIX期货的5%多头头寸构成，与Moran和Liu（2020）中的设定相同。对于每种动态策略，我们还使用基于最大似然估计的预测分布生成结果，以进行比较。我们报告平均超额收益，即投资组合收益超过无风险收益率的部分、投资组合标准差以及夏普比率。

在这两种动态交易规则下，基于上尾 CLS 构建的预测分布所产生的夏普比率明显高于 MLE 预测值，这表明，采用上尾校准的策略相对于所承担的风险，能够获得最高回报。这一观察结果证实，在此类模型设定错误普遍存在的实证情境中，有必要构建针对关注区域的最优预测分布；在本例中，该区域是 VIX 预测分布的上尾。我们还发现，对于这两种预测分布形式，预测百分位规则的表现都优于预测概率规则。事实上，使用预测概率规则并利用基于极大似然估计的预测分布构建的投资组合，表现甚至不如静态投资组合。这凸显了

基于点预测构建 VIX期货。

区域特定校准对于使动态风险管理策略明确优于更简单的静态选项的重要性。

表8：基于动态交易策略的股票与VIX期货投资组合表现。同时报告由95%的标普500市场投资组合和5%的VIX期货构成、在所有交易日均持有的静态投资组合结果，以供比较。

	平均超额收益	标准差	夏普比率
静态投资组合	0.040	0.182	0.219
	预测概率规则		
优化器			
CLS($VIX_t > VIX_{t-1}$)	0.066	0.207	**0.318**
MLE	0.042	0.204	0.206
	80^{th} **预测分位数规则**		
优化器			
CLS80	0.076	0.199	**0.381**
MLE	0.075	0.200	0.375

6 结论

本文研究了尾部导向的最优预测在金融风险管理中的应用。根据模拟和实证结果，我们证实，在假定的预测模型总体上适合对相关数据进行建模的前提下，最优预测在模型设定错误时具有优势。聚焦于最优预测的尾部表现，模拟结果表明，与采用传统的基于极大似然估计的预测相比，模型设定错误越严重，使用最优预测所能获得的收益就越大。具体而言，当假定的预测模型为条件高斯时，与厚尾数据生成过程相比，在偏态数据生成过程中使用最优预测所带来的收益更加显著。我们还证实，当关注尾部风险时，在风险价值覆盖率和预期损失准确性方面，根据分位数评分构建的最优预测优于其他替代方案。

对12个标普指数的实证分析证实了模拟结果在定性方面的结论。我们还探讨了最优预测在VIX预测中的应用，具体而言，利用VIX衍生品对冲股市风险。我们对基于尾部聚焦最优预测构建的交易策略进行了回测，并与基于最大似然估计的预测构建的交易策略进行比较；在这一情境下，尾部聚焦预测带来了显著更优的收益—风险权衡。

参考文献

Aastveit, K. A., Mitchell, J., Ravazzolo, F., and Van Dijk, H. K.（2018年）。经济学中预测密度组合的发展。技术报告，廷伯根研究所讨论论文。Acerbi, C. and Tasche, D.（2002年）。论预期损失的一致性。《银行与金融杂志》，26（7）：1487–1503。Bao, Y., Lee, T.-H., and Saltoglu, B.（2006）。评估新兴市场风险价值模型的预测性能：现实检验。《预测杂志》，25（2）：101–128。BenSa¨ıda, A., Boubaker, S., Nguyen, D. K., and Slim, S.（2018年）。通过高度灵活的模型应对市场变化下的风险价值。《预测杂志》，37（8）：790–804。Bollerslev, T.（1987年）。用于投机价格和收益率的条件异方差时间序列模型。《经济学与统计学评论》，第542–547页。Borowska, A., Hoogerheide, L., Koopman, S. J., and van Dijk, H. K.（2020年）。用于稳健且高效风险评估的部分删失后验。《计量经济学杂志》，217（2）：335–355。Britten-Jones, M. and Neuberger, A.（2000）。期权价格、隐含价格过程与随机波动率。《金融学杂志》，55（2）：839–866。Brooks, C., Clare, A., Dalle Molle, J. W., and Persand, G.（2005年）。确定风险价值的极值理论方法比较。《实证金融杂志》，12（2）：339–352。芝加哥期权交易所（2021年）。芝加哥期权交易所波动率指数。白皮书，第1–20页。Chen, Y., H¨ardle, W., and Jeong, S.-O.（2008年）。基于广义双曲分布的非参数风险管理。《美国统计学会杂志》，103（483）：910–923。Chiang, I.-H. E., Liao, Y., and Zhou, Q.（2021）。在模型池中使用合理模型对股票收益率横截面建模。《实证金融杂志》，60：56–73。Christoffersen, P. F.（1998年）。评估区间预测。《国际经济评论》，第841–862页。Corsi, F.（2009年）。一种简单的已实现波动率近似长记忆模型。《金融计量经济学杂志》，7（2）：174–196。Diks, C., Panchenko, V., and Van Dijk, D.（2011年）。用于比较尾部密度预测的基于似然的评分规则。《计量经济学杂志》，163（2）：215–230。Ding, Z., Granger, C. W., and Engle, R. F.（1993年）。股票市场收益率的长记忆特性及一种新模型。《实证金融杂志》，1（1）：83–106。Ehm, W., Gneiting, T., Jordan, A., and Kr¨uger, F.（2016年）。论分位数与期望分位数：一致评分函数、Choquet表示与预测排名。《英国皇家统计学会杂志：*B*辑（统计方法论）》，78（3）：505–562。Engle, R. F. and Manganelli, S.（2004年）。条件自回归风险价值：基于回归分位数的方法。《商业与经济统计杂志》，22（4）：367–381。

Fernandes，M.、Medeiros，M. C. 和 Scharth，M.（2014）。芝加哥期权交易所市场波动率指数的建模与预测。《银行与金融杂志》，40：1–10。Fissler，T.、Ziegel，J. F. 和 Gneiting，T.（2015）。预期损失与风险价值联合可诱导——对回溯测试的启示。*arXiv*预印本 *arXiv:1507.00244*。Gatarek，L.、Hoogerheide，L.、Hooning，K.、van Dijk，H. K. 等（2014）。用于准确估计风险价值的左尾预测：删失后验与预测似然。技术报告，廷伯根研究所。Giacomini，R. 和 Komunjer，I.（2005）。条件分位数预测的评估与组合。《商业与经济统计杂志》，23（4）：416–431。Giacomini，R. 和 White，H.（2006）。条件预测能力检验。《计量经济学》，74（6）：1545–1578。Gneiting，T.（2011）。点预测的制定与评估。《美国统计学会杂志》，106（494）：746–762。Gneiting，T. 和 Raftery，A. E.（2007）。严格适当的评分规则、预测与估计。《美国统计学会杂志》，102（477）：359–378。Gneiting，T.、Raftery，A. E.、Westveld III，A. H. 和 Goldman，T.（2005）。运用集合模式输出统计和最小连续秩概率评分估计进行校准概率预测。《月度天气评论》，133（5）：1098–1118。Hansen，P. R.、Huang，Z. 和 Shek，H. H.（2012）。已实现GARCH：收益率与已实现波动率度量的联合模型。《应用计量经济学杂志》，27（6）：877–906。Jensen，M. J. 和 Maheu，J. M.（2013）。贝叶斯半参数多元GARCH建模。《计量经济学杂志》，176（1）：3–17。Jiang，G. J. 和 Tian，Y. S.（2005）。无模型隐含波动率及其信息含量。《金融研究评论》，18（4）：1305–1342。Koenker，R. 和 Bassett，G.（1978）。回归分位数。《计量经济学》，第33–50页。Konstantinidi，E.、Skiadopoulos，G. 和 Tzagkaraki，E.（2008）。隐含波动率的演变能否预测？来自欧洲和美国隐含波动率指数的证据。《银行与金融杂志》，32（11）：2401–2411。Laporta，A. G.、Merlo，L. 和 Petrella，L.（2018）。能源大宗商品风险价值模型的选择。《能源经济学》，74：628–643。Loaiza-Maya，R.、Martin，G. M. 和 Frazier，D. T.（2021）。聚焦贝叶斯预测。《应用计量经济学杂志》，36（5）：517–543。Maheu，J. M. 和 McCurdy，T. H.（2004）。个股收益率的新闻到达、跳跃动态与波动率成分。《金融学杂志》，59（2）：755–793。Maneesoonthorn，W.、Forbes，C. S. 和 Martin，G. M.（2017）。利用高频度量推断价格与波动率中的自激跳跃。《应用计量经济学杂志》，32（3）：504–532。

Maneesoonthorn，W.、Martin，G. M.、Forbes，C. S. 和 Grose，S. D.（2012）。波动率及其风险溢价的概率预测。《计量经济学杂志》，171（2）：217–236。Martin，G. M.、Loaiza-Maya，R.、Maneesoonthorn，W.、Frazier，D. T. 和 Ramírez-Hassan，A.（2022）。最优概率预测：何时有效？《国际预测杂志》，38（1）：384–406。Moran，M. T. 和 Liu，B.（2020）。《*VIX*指数、基于波动率的全球指数与交易工具：投资与交易特征指南》。CFA协会研究基金会。Opschoor，A.、Van Dijk，D. 和 van der Wel，M.（2017）。运用聚焦型评分规则组合密度预测。《应用计量经济学杂志》，32（7）：1298–1313页。Patton，A. J.（2019）。比较可能设定错误的预测。《商业与经济统计杂志》，第1–23页。Smith，M. S. 和 Maneesoonthorn，W.（2018）。非线性状态空间模型的反演Copula及其在通胀预测中的应用。《国际预测杂志》，34（3）：389–407。Tasche，D.（2002）。预期损失及其他。《银行与金融杂志》，26（7）：1519–1533页。Taylor，N.（2019）。VIX期货市场的收益率预测。《国际预测杂志》，35（4）：1193–1210。Wang，X.、Hyndman，R. J.、Li，F. 和 Kang，Y.（2022）。预测组合：五十多年综述。《国际预测杂志》。Yamai，Y. 和 Yoshiba，T.（2005）。风险价值与预期损失之比较：实践视角。《银行与金融杂志》，29（4）：997–1015。Ziegel，J. F.、Kr¨uger，F.、Jordan，A. 和 Fasciati，F.（2020）。预期损失的稳健预测评估。《金融计量经济学杂志》，18（1）：95–120。