arXiv:2611.29747v1 [econ.EM] 22 Nov 2026

# 机器学习后的夏普比率推断*

Sander Barendse[1] 和 Ming Cheng[1,2]

[1]阿姆斯特丹经济学院，阿姆斯特丹大学[2]廷伯根研究所

本版本：2026年11月20日

**摘要**

收益信号的经济价值通常通过基于这些信号的投资组合的夏普比率进行评估。在现代资产定价中，收益信号通常取决于许多特征，包括状态变量和公司特定特征，而机器学习方法正日益被用于估计这些信号。但对于估计的夏普比率，我们能够有多大程度的信心？我们发现，即使在大型资产面板中，信号估计误差对于推断仍可能具有重要影响。此外，当预测变量包含共同特征或表现出显著的横截面依赖时，合并不同资产的观测值并不一定会使这种不确定性变得可以忽略。为此，我们提出一种去偏夏普比率估计量，该估计量能够考虑信号估计误差，并针对基准提供有效的置信区间和检验。我们的目标是：对于给定的投资组合规则，在真实的、给定特征条件下的条件期望收益处进行评估所得的总体夏普比率。该框架涵盖市场择时、多空及其他基于信号的投资组合，允许资产数量随时间序列样本量增长，并大幅超过时间序列样本量。该方法将正交投资组合矩与时间序列交叉拟合相结合；在机器学习预测值满足适当的估计误差界限时，该方法仍然有效。在实证方面，我们使用收益信号的机器学习估计量，重新考察 Gu 等人（2020）的研究设定。相对于市场基准，最强的多空夏普比率增益在去偏后仍然显著，为其特征集的经济价值提供了统计证据。传统夏普比率估计值与去偏夏普比率估计值通常较为相近，但相关的置信区间和检验可能存在显著差异。这些差异因信号估计量和投资组合规则而异，并不会一致地削弱有关投资组合表现的证据。

**关键词：** 夏普比率；机器学习；投资组合表现；估计误差；半参数推断；双重／去偏推断；Neyman 正交性

**JEL 分类：** C12,C14, C58, G11

*我们非常感谢 Peter Boswijk、Yi He，以及在 2026年荷兰计量经济学研究组、2026年东京大学计量经济学理论与应用研讨会、2026年艾克斯-马赛大学 QFFE 活动、2026年塞萨洛尼基非参数统计国际研讨会、2026年维也纳大学维也纳-哥本哈根金融计量经济学会议、2026年日内瓦大学金融计量经济学与机器学习会议，以及阿姆斯特丹大学的研讨会和会议参与者，感谢他们提出的宝贵意见和建议。文中所有剩余错误均由我们自行负责。本文的早期版本曾以《机器学习投资组合夏普比率的去偏检验》为题流传并发表。电子邮箱： s.c.barendse@uva.nl, m.cheng@uva.nl

# 1 引言

当收益可预测性能够转化为更好的投资组合决策时，它对投资者才有意义。现代资产定价研究从大量状态变量和公司特征中寻找这类信息，并日益借助机器学习方法来估计收益信号。基于这些信号构建的投资组合有助于评估其经济价值，例如通过夏普比率进行评估。但估计出的夏普比率很高，并不一定能为信号的经济价值提供有力的统计证据。当收益信号本身也是估计得出的时，我们对所估计的夏普比率能有多大把握？本文提出了置信区间和检验，以考虑收益信号估计中的不确定性。

收益预测模型通过依赖于下一期预测收益的投资组合权重，将信息转化为投资决策。市场择时投资组合根据指数的预测收益调整对该指数的敞口（例如，Campbell and Thompson，2008年；Welch 和 Goyal，2008年；Gu 等人（2020））。具体而言，Campbell and Thompson（2008年）表明，即使预测回归的样本外决定系数低至 1%，相较于静态投资组合也能带来具有经济意义的收益。Gu 等人（2020），下文简称 GKX，使用机器学习方法预测大规模资产面板数据中的股票超额收益。通过合并估计（这种方法正日益普及，参见 Hellum 等人（2025年）及其中的参考文献），他们利用大型特征集获得个股收益预测，并将这些预测汇总起来，以指导对指数投资组合的敞口配置。他们发现，与买入并持有策略相比，某些预测方法能显著提升夏普比率。GKX 还根据个股收益预测构建多空投资组合，买入预测收益较高的股票，并做空预测收益较低的股票。这些应用属于资产定价领域中关于使用大型特征集和灵活模型的更广泛研究的一部分（例如，Rapach 等人，2010年；Kelly 等人，2019年；Feng 等人，2020年；Freyberger 等人，2020年；Kozak 等人，2020年；Bianchi 等人，2021年；Leippold 等人，2022年；Chen 等人，2024年）；

Kelly 和 Xiu（2023年）提供了近期综述。

特征集的经济价值，可以通过使用真实收益信号的特定投资组合规则的总体夏普比率来评估；这里的收益信号是指给定这些特征时的条件期望收益。固定特征集和投资组合规则后，不同的预测方法会对相同信号以及相应的同一夏普比率给出不同估计。当信号估计误差影响显著时，传统置信区间未必能对这一总体夏普比率提供有效推断。我们提出一种能够考虑信号估计误差的方法；当机器学习预测值满足适当的准确性要求时，该方法可以提供有效的置信区间。

我们表明，收益信号估计带来的不确定性，可能使基于标准夏普比率估计量对这一目标值进行的传统推断失效。资产横截面规模大并不一定能解决这一问题。例如，GKX采用合并回归方法，增加可用于训练的观测值数量。然而，合并并非万灵药：共同特征或资产特定特征之间不受限制的横截面依赖，都可能使信号估计误差过大，以致无法对总体夏普比率进行传统推断。关键不在于某种机器学习方法的预测表现是否优于低维基准，而在于其估计误差是否足够小，从而在评估信号的经济价值时可以忽略。

本文提出了一种去偏夏普比率估计量，对于市场择时、多空及其他基于信号的投资组合，它对信号估计误差的敏感度较低。在信号估计满足适当条件时，该估计量可为总体夏普比率以及相对于基准的差异提供有效的置信区间。基准比较可用于检验基于信号的规则的夏普比率是否不同于买入并持有投资组合或其他投资策略。即使信号估计误差过大，以致传统方法无法忽略，该程序仍能进行有效推断。我们的框架适用于大型资产面板数据，其中资产数量随时间序列样本量增加，且可能远远

超过时间序列样本量。

在实证分析中，我们使用收益信号的机器学习估计量重新考察了 GKX 设定。结果表明，通过机器学习估计底层收益信号来应用 GKX特征集时，本文考察的投资组合规则所体现的经济依据得到进一步支持。这些收益主要集中在多空投资组合中；而在两种方法下，均没有任何择时策略显著优于任何基准。纳入预测估计会实质性地改变基准比较结果，尤其是对于一些更灵活的设定。尽管去偏可能会显著扩大置信区间，尤其是在灵活性很高的设定下，但在其他设定中也可能使置信区间变窄。这些发现表明，在现代资产定价应用中评估高维特征集的经济价值，需要仔细考虑信号估计不确定性。

我们的新方法将 Neyman正交投资组合矩与适用于时间序列数据的交叉拟合设计相结合。在适当条件下，我们为投资组合收益的一阶矩和二阶矩构造正交矩条件，从而消除信号估计误差的一阶影响。这些矩产生了一个 Sharpe 比率的去偏估计量，而该比率是两个投资组合矩的函数。我们的样本分割设计适用于标准资产定价设定，可采用固定、移动和扩展的估计窗口。我们证明了该方法在独立观测和时间依赖三角阵列下的渐近有效性与一致推断。此外，我们表明，在独立同分布观测下，该去偏估计量具有半参数有效性。这说明，考虑信号估计误差与有效估计并不矛盾；这一点与我们的发现一致，即去偏置信区间并非总是比传统置信区间更宽。我们还验证了合并神经网络估计量满足所需的信号估计界。模拟研究表明，对于市场择时和多空投资组合，去偏能够减少错误拒绝，并使置信区间覆盖率更接近其预期水平

用于市场择时和多空投资组合。

我们的分析与金融学文献中关于估计风险的研究相关。Kan and Zhou (2007) 刻画了使用估计均值和协方差进行最优投资组合选择所带来的影响。Kan 等人 (2024) 研究了估计风险如何造成多因子模型的样本内与样本外夏普比率之间的差异。Barillas 等人 (2020) 提出了基于最大夏普比率平方的模型比较检验，并将模仿投资组合权重中的估计误差纳入考量。我们研究权重取决于高维估计收益信号的基于信号的规则。这使得估计风险问题延伸到由灵活的条件均值模型构建的投资组合。

我们的去偏估计量也推广了经典的夏普比率推断（Jobson and Korkie, 1981; Memmel, 2003; Lo, 2002; Ledoit and Wolf, 2008）。这些现有方法基于观测到的投资组合收益，对总体夏普比率进行推断。我们还考虑了估计收益信号所带来的不确定性，目标是估计在真实信号下评估的投资组合规则的夏普比率。尽管重抽样方法有可能对我们的夏普比率目标进行有效推断，但据我们所知，此类方法尚未得到开发，而且计算成本高昂，因为每次重抽样迭代都需要重新拟合机器学习方法。Liao 等人 (2025) 针对神经网络预测提出了一种自助法，通过从原始拟合结果出发并在每个重抽样样本上进行更新来降低计算成本，但未将其与夏普比率推断联系起来。我们的程序无需自助法重复抽样，因此也无需为此反复拟合或调整模型。一旦获得满足样本分割要求的预测值以及所需的创新协方差输入，实施过程便包括评估投资组合规则及其导数，并计算去偏样本矩和标准误。最后，我们以双重/去偏机器学习文献（Chernozhukov 等人, 2018）为基础。近期在金融领域的应用包括对条件潜在因子的推断，

风险溢价和随机贴现因子（Gagliardini and Ma，2026年），以及对

企业排放数据中的选择偏差进行校正（Chen 等人，2026年）。在一篇相关且同期发表的论文中，Quaini（2026年）针对 Diebold-Mariano 框架下单变量预测的比较，提出了正交推断方法，适用于评估损失不同于训练目标的情况，但未涉及夏普比率。

本文其余部分的组织如下。第 2 节介绍收益模型和基于信号的投资组合，发展针对夏普比率及其基准差异的去偏推断，并给出独立同分布渐近理论、神经网络速率条件和效率结果。第 3 节和第 4 节分别报告模拟结果和实证证据，第 5 节作出结论。补充材料提供技术细节、证明和补充结果。

我们使用以下符号说明。粗体符号表示向量或矩阵，撇号表示转置。我们用 0、1（也用 $\boldsymbol{\iota}_N$ 表示全为 1 的 $N$ 维向量）以及 $\mathrm{I}_d$ 分别表示零向量、全 1 向量和 $d\times d$ 单位矩阵，其维度由上下文确定。集合 $\mathbb{R}$、$\mathbb{N}_0$ 和 $\mathbb{N}_+$ 分别表示实数、非负整数和正整数；$\lfloor x \rfloor$ 和 $\lceil x \rceil$ 分别表示下取整和上取整。对于向量 $\mathbf{a} = (a_1, \ldots, a_d)'$，$\|\mathbf{a}\|_q = (\sum_{j=1}^{d} |a_j|^q)^{1/q}$ for $1 \le q < \infty$ and $\|\mathbf{a}\|_\infty = \max_j |a_j|$；未指定的向量范数均指欧几里得范数。对于标量或向量值的 $f$ 和 $1 \le q < \infty$，$\|f(X)\|_{P,q} = (\mathbb{E}_P[\|f(X)\|_2^q])^{1/q}$ and $\|f(X)\|_{P,\infty} = \operatorname{ess\,sup}_P \|f(X)\|_2$，其中对于标量 $f$，内层范数替换为绝对值。这里，$\mathbb{E}_P$ 和 $\mathbb{V}_P$ 表示在概率分布 $P$下的期望和方差。对于矩阵 $\mathbf{A}$，$\|\mathbf{A}\|_{\mathrm{op}} = \sqrt{\lambda_{\max}(\mathbf{A}'\mathbf{A})}$；$\lambda_{\min}$ 和 $\lambda_{\max}$ 表示对称矩阵的最小和最大特征值。我们用 $\operatorname{diag}(\mathbf{a})$ 表示对角线为 $\mathbf{a}$ 的对角矩阵，$\odot$ 表示哈达玛积。对于非负序列，$a_T \lesssim b_T$ 表示对所有充分大的 $T$均有 $a_T \le C b_T$，其中 $C > 0$ 为与 $T$无关的绝对常数；$a_T \asymp b_T$ 表示 $a_T \lesssim b_T$ 和 $b_T \lesssim a_T$ 同时成立。符号 $O$ 和 $o$ 具有通常的确定性含义；对于正的确定性 $a_T$，$O_P(a_T)$ 和 $o_P(a_T)$ 分别表示除以 $a_T$ 后在 $P$-概率意义下有界以及收敛于零。小写下标 $p$ 省略概率分布，对 $P \in \mathcal{P}$ 的一致性将明确说明。

符号 $\rightsquigarrow$ 表示弱收敛。

# 2 理论

## 2.1 收益模型

对于样本中的每个 $T$ 期，我们观测 $N = N_T$ 项资产的超额收益，以及状态变量和资产特定特征。我们允许 $N$ 随 $T$增长，并且其规模可能远大于后者（$N \gg T$），以反映常见的资产定价设定：每月可用的资产数量与总样本量成正比，或大于总样本量。为强调 $N$ 对 $T$的依赖关系，我们在所有受其影响的变量和参数上加下标 $T$。

对于每个时间段 $t = 1, \ldots\ldots, T$，我们将超额收益汇集为向量

$$\mathbf{R}_{T,t} = (R_{1,t}, \ldots, R_{N,t})',$$

其中，资产 $i$的超额收益记为 $R_{i,t}$，对于 $i = 1, \ldots\ldots, N$。对于每项资产，我们将共同特征和资产特定特征汇集为 $p$维向量 $\mathbf{Z}_t^{(i)}$，其中$p = p_T$ 可随 $T$增长，并将每个资产特定向量堆叠为 $\mathbf{Z}_{T,t} := \left(\mathbf{Z}_t^{(1)\prime} \cdots \mathbf{Z}_t^{(N)\prime}\right)'$。在预测场景中，我们根据过去信息预测收益，因此令 $\mathbf{Z}_{T,t}$ 包含严格滞后的观测值（但为便于说明而保留时间下标 $t$）。最后，我们将收益和向量汇集为 $\boldsymbol{W}_{T,t} = (\mathbf{R}'_{T,t}, \mathbf{Z}'_{T,t})'$。

设 $\mathcal{P}$ 为支配资产收益与特征联合分布的一个分布类。对于每个分布 $P \in \mathcal{P}$，以下条件均值和协方差函数可能取决于 $P$，但为减少符号说明上的冗余，我们省略这一依赖关系。我们引入分布类 $\mathcal{P}$，因为我们的目标是一致推断。在双重/去偏机器学习文献中，数据生成过程是未知的，通常旨在获得对一类合理的数据生成过程均一致有效的推断，而不仅仅是

在某个固定分布 $P$下。

以下假设规定了收益模型，并施加了两项限制。首先，假设所有资产具有共同的信号函数 $\mu_0(\cdot)$，这为合并回归提供了依据。其次，假设条件协方差矩阵不依赖于特征，且其特征值在上方一致有界并且与零保持正距离。如下方备注所述，这两项限制都可以放宽。令 $\mathrm{M}_T$ 表示一个可测信号函数类，该函数类在 $\mathrm{Z}_T$的所有实现处都有定义，且返回 $N_T$维信号向量。令$\mathcal{S}_T$ 表示一个由 $N_T \times N_T$ 半正定对称矩阵组成的类。

**假设1**（条件收益模型）。 对于每个 $P \in \mathcal{P}$、样本量 $T \geq 1$和时间指数 $t$，收益满足

$$\begin{aligned}
\mathbf{R}_{T,t} &= \boldsymbol{\mu}_{T,0}(\mathbf{Z}_{T,t}) + \varepsilon_{T,t}, \\
\boldsymbol{\mu}_{T,0}(\mathbf{Z}_{T,t}) &= \left(\mu_0(\mathbf{Z}_t^{(1)}), \ldots, \mu_0(\mathbf{Z}_t^{(N_T)})\right)', \\
\mathbb{E}_P[\varepsilon_{T,t} \mid \mathbf{Z}_{T,t}] &= \mathbf{0}, \\
\mathbb{E}_P[\varepsilon_{T,t}\varepsilon_{T,t}' \mid \mathbf{Z}_{T,t}] &= \boldsymbol{\Sigma}_{T,0},
\end{aligned}$$

其中 $\boldsymbol{\mu}_{T,\,0} \in \mathrm{M}_T$ 以及 $\boldsymbol{\Sigma}_{T,\,0} \in \mathcal{S}_T$。信号函数 $\mu_0(\cdot)$ 和协方差矩阵 $\boldsymbol{\Sigma}_{T,\,0}$ 可能依赖于 $P$。此外，存在常数 $0 < c \leq \bar{c} < \infty$ 使得 $\boldsymbol{\Sigma}_{T,\,0}$ 满足

$$\underline{c} \leq \lambda_{\min}(\boldsymbol{\Sigma}) \leq \lambda_{\max}(\boldsymbol{\Sigma}) \leq \bar{c},$$

对于所有 $P \in \mathcal{P}$ 和 $T$均一致。

**注 1 (**共同信号函数的逼近与特征依赖的协方差矩阵**)**。 通过允许资产特定的条件均值函数 $\mu_0^{(i)}(\cdot)$ 由共同信号函数 $\mu_0(\cdot)$ 很好地逼近，可以放宽共同信号限制。下文用于控制信号估计误差的$L_2$ 距离，也自然地度量了共同信号限制所导致的这种近似误差。综合而言，

近似误差和估计误差必须满足下文对估计误差所施加的条件。常数条件协方差限制也可以放宽为允许特征依赖的条件协方差矩阵几乎处处正定，但代价是必须对其估计误差的算子范数施加 $L_2$ 界限，这会大幅增加符号说明的繁杂程度。

## 2.2 基于信号的投资组合及其夏普比率

我们考虑这样一类投资组合策略：其权重是信号 $\boldsymbol{\mu}_{T,0}(\mathbf{Z}_{T,t})$ 以及 $\mathbf{Z}_{T,t}$ 中其他信息（例如资产特定的指数权重）的已知函数。市场择时和多空投资组合，以及它们的受限版本，都属于这类投资组合策略。本节第二部分将给出这些投资组合的设定。

我们将投资组合权重向量定义为

$$\boldsymbol{\omega}_T(\mathbf{Z}_{T,t}, \boldsymbol{\mu}_{T,0}(\mathbf{Z}_{T,t})) = [\omega_1(\mathbf{Z}_{T,t}, \boldsymbol{\mu}_{T,0}(\mathbf{Z}_{T,t})), \ldots, \omega_N(\mathbf{Z}_{T,t}, \boldsymbol{\mu}_{T,0}(\mathbf{Z}_{T,t}))]',$$

因此，该策略的夏普比率为

$$\mathrm{SR}_{T,0} = \frac{\mathbb{E}_P[\boldsymbol{\omega}_T(\mathbf{Z}_{T,t}, \boldsymbol{\mu}_{T,0}(\mathbf{Z}_{T,t}))'\mathbf{R}_{T,t}]}{\sqrt{\mathbb{V}_P[\boldsymbol{\omega}_T(\mathbf{Z}_{T,t}, \boldsymbol{\mu}_{T,0}(\mathbf{Z}_{T,t}))'\mathbf{R}_{T,t}]}}.$$

指数 $T$ 反映出，夏普比率可能会通过可用资产的数量 $N$ （该数量随 $T$增长）而依赖于样本量。

以下假设将投资组合策略限定为投资组合权重函数连续可微的策略。由此得到的雅可比矩阵是本文所引入的去偏夏普比率估计量的关键组成部分。

**假设 2** （平滑的投资组合权重）。 对于每个 $T \geq 1$，投资组合权重函数$\boldsymbol{\omega}_T(\cdot, \cdot)$在其参数上联合可测。它关于信号变量连续可微

在一个开集上，并且其雅可比矩阵，

$$\mathbf{J}_T(\mathbf{z},\mathbf{m}) := \left.\frac{\partial \boldsymbol{\omega}_T(\mathbf{z},\widetilde{\mathbf{m}})}{\partial \widetilde{\mathbf{m}}'}\right|_{\widetilde{\mathbf{m}}=\mathbf{m}} \in \mathbb{R}^{N\times N},$$

作为(**z**,**m**)的函数是联合可测的。

### 2.2.1 择时投资组合

择时策略依据指数投资组合的信号来确定其未来头寸（Campbell and Thompson, 2008年）。令 $\mathbf{a}_{T,t} = (a_{T,1,t}, \ldots, a_{T,N,t})'$表示在时期 $t$开始时观测到的事前指数权重向量，其中 $0 \le a_{i,t} \le 1$ 和 $\sum_{i=1}^{N} a_{i,t} = 1$。此外，我们施加分散化条件 $\max_i a_{T,i,t} \le C/N$。指数权重包含在 $\mathbf{Z}_{T,t}$中，或者说是确定性的，因此无需估计。例如，对于等权重市场投资组合，$a_{i,t} = 1/N$；对于市值加权市场投资组合，$a_{i,t} = V_{i,t}/\sum_{i=1}^{N} V_{i,t}$，其中$V_{i,t}$ 是资产 $i$ 在时期 $t$ 开始时的市值。对于其他指数（例如标普500指数），权重取决于指数编制规则。

该市场择时策略为

$$\omega_i(\mathbf{Z}_{T,t}, \boldsymbol{\mu}_{T,0}(\mathbf{Z}_{T,t})) = a_{i,t}\left(\sum_{i=1}^{N} a_{i,t}\,\mu_{0,i}(\mathbf{Z}_{T,t})\right),$$

其雅可比矩阵为

$$\mathbf{J}_T(\mathbf{Z}_{T,t}, \boldsymbol{\mu}_{T,0}(\mathbf{Z}_{T,t})) = \mathbf{a}_{T,t}\mathbf{a}'_{T,t}.$$

由于我们关注的是市场择时投资组合的夏普比率，因此无需像 Campbell and Thompson（2008年）那样，将头寸除以指数收益率的条件方差。该条件方差在各时间段内保持不变，因此会在夏普比率中相互抵消。

### 2.2.2 多空投资组合

多空策略根据信号表现，在表现最好的标的上做多，在表现最差的标的上做空。通常使用顶部和底部十分位组来对表现最好和最差的标的进行分类，并对投资组合采用等权重配置。多空投资组合权重正式表示为

$$\omega_i(\mathbf{Z}_t, \boldsymbol{\mu}_0(\mathbf{Z}_t)) = h_i(\boldsymbol{\mu}_0(\mathbf{Z}_t)),$$

其中，对于任意 $\boldsymbol{\mu} \in \mathbb{R}^N$,

$$h_i(\boldsymbol{\mu}) = \frac{10}{N}\left(\mathbb{1}\left[\sum_{j=1}^{N} \mathbb{1}[\mu_i - \mu_j \leq 0] \leq 0.1N\right] - \left(1 - \mathbb{1}\left[\sum_{j=1}^{N} \mathbb{1}[\mu_i - \mu_j \leq 0] \leq 0.9N\right]\right)\right),$$

其中，$\mathbb{1}[A]$ 表示指示函数：若事件 $A$ 成立，则取值为 1，否则为 0。投资于范围更窄的顶部/底部表现者的多空策略，其设定可直接得出。

由于投资组合权重函数的设定依赖于指示函数，该策略的权重是不光滑的。我们基于校准逻辑函数引入多空投资组合的光滑近似；校准逻辑函数是指示函数的一种常见光滑逼近。考虑逻辑函数 $s(z;\tau) = (1 + e^{-z/\tau})^{-1}$，其带宽为 $\tau > 0$。对于每个 $\boldsymbol{\mu} \in \mathbb{R}^N$ 和带宽 $\tau > 0$，我们都可以找到唯一的阈值函数 $c_+(\boldsymbol{\mu})$ 和 $c_-(\boldsymbol{\mu})$，使得

$$p_i^+(\boldsymbol{\mu};\tau) = s(\boldsymbol{\mu}_i - c_+(\boldsymbol{\mu});\tau), \qquad \sum_i p_i^+(\boldsymbol{\mu};\tau) = \lfloor 0.1N \rfloor$$

$$p_i^-(\boldsymbol{\mu};\tau) = s(c_-(\boldsymbol{\mu}) - \boldsymbol{\mu}_i;\tau), \qquad \sum_i p_i^-(\boldsymbol{\mu};\tau) = N - \lfloor 0.9N \rfloor.$$

平滑多空投资组合的投资组合权重函数随后表示为

$$\tilde{h}_i(\boldsymbol{\mu};\tau) = \frac{10}{N}\left(p_i^+(\boldsymbol{\mu};\tau) - p_i^-(\boldsymbol{\mu};\tau)\right).$$

雅可比矩阵的第（$i$, $j$）项为

$$\frac{\partial \omega_i(\mathbf{Z}_t, \boldsymbol{\mu})}{\partial \mu_j} = \frac{10}{N\tau} \sum_{\sigma\in\{+,-\}} p_i^\sigma(1-p_i^\sigma)\left(\mathbb{1}[i=j] - \frac{p_j^\sigma(1-p_j^\sigma)}{\sum_{\ell=1}^N p_\ell^\sigma(1-p_\ell^\sigma)}\right), \qquad i,j = 1,\ldots,N.$$

对于每次实现的 $\boldsymbol{\mu}_{T,\,0}(\mathbf{Z}_{T,\,t})$，都会通过数值方法重新校准唯一阈值 $c_-(\boldsymbol{\mu}_{T,\,0}(\mathbf{Z}_{T,\,t}))$ 和 $c_+(\boldsymbol{\mu}_{T,\,0}(\mathbf{Z}_{T,\,t}))$。 $\tau$ 越小，逼近越精确。图1直观比较了非平滑多空投资组合和平滑多空投资组合的权重函数。实际操作中，我们会选取一组候选 $\tau$ 值，并选择一个较小的值，使得估计的投资组合方差仍为正。

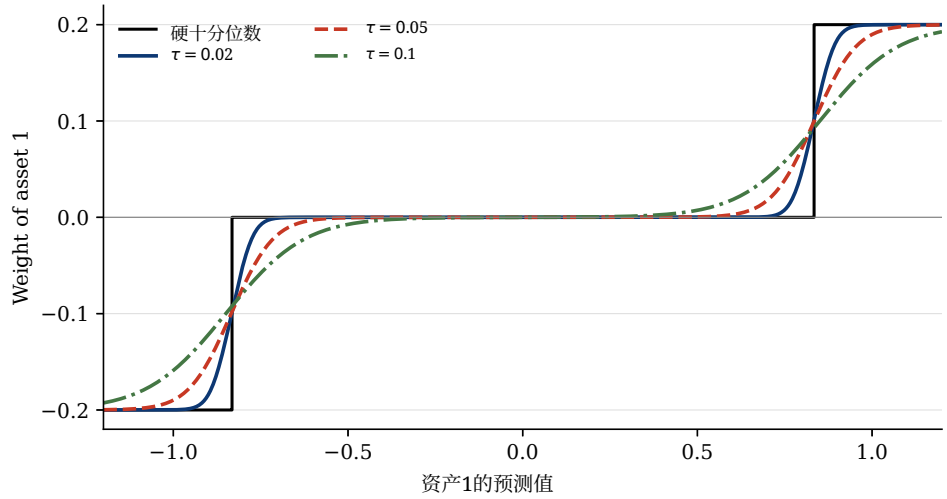



图*1.* 单个资产的非平滑与平滑多空权重。共有 $N = 50$ 个资产。其余 49 个预测值固定且在 $[-1, 1]$ 上等距分布，而资产1的预测值则有所变化。

### 2.2.3 其他投资组合

其他投资组合若是信号 $\boldsymbol{\mu}_{T,\,0}(\mathbf{Z}_{T,\,t})$ 和 $\mathbf{Z}_{T,\,t}$ 的已知函数，也可以在下文采用类似技术处理。例如，我们可以对市场择时投资组合施加总敞口约束，或对个别头寸设置（平滑）上限。

## 2.3 动机示例：估计误差会导致标准夏普比率推断产生偏差

$\widehat{\ }_{\text{std}}$后文第 2.4 节定义的标准夏普比率估计量 $\text{SR}_T$ ，在对 $\text{SR}_{T,0}$进行推断时可能存在偏差，导致置信区间无效。为说明这一点，我们考虑一个资产等权市场择时投资组合，其中的资产满足稀疏收益模型，该模型包含依赖于资产特定特征和共同特征的线性和非线性成分，反映了现代资产定价情境：模型使用股息价格比等资产特定特征，以及市场波动率等共同特征。在此示例中，标准$\widehat{\ }_{\text{DB}}$估计量给出的推断存在偏差，而后文第 2.5 节定义的新去偏版本 $\text{SR}_T$ 则能提供有效推断和置信区间。

在此示例中，我们使用合并普通最小二乘法，基于 $p$ 特征估计模型，其中 $p < T$ ，且设计矩阵具有满列秩，因此普通最小二乘法定义良好。随着样本量增大，非线性成分趋于消失，因此普通最小二乘估计量的近似误差也趋于消失。这与神经网络回归相似：即使对于成分不会消失的非线性模型，近似误差仍可能趋于消失。这种设定的优点是，可以解析地推导夏普比率估计量的一阶偏差。在后续章节中，我们将通过模拟展示，神经网络示例中也会出现偏差。

本节采用固定窗口估计策略来估计夏普比率，以减少符号说明上的繁杂，但类似的结果也适用于移动窗口和扩展窗口方法，例如 S1.4节中讨论的方法。具体而言，我们首先在训练集 $t = 1, \ldots, T$ 上估计未观测到的信号函数 $\boldsymbol{\mu}_{T,0}(\mathbf{Z}_{T,t})$。随后，我们计算每个

投资组合权重 $\boldsymbol{\omega}_{T,0}(\mathbf{Z}_{T,t})$ 对应的估计值，并将其记为 $\hat{\boldsymbol{\omega}}_T(\mathbf{Z}_{T,t})$。最后，我们按照通常的方式，使用测试集上的代入样本矩来估计夏普比率：

$$\widehat{\mathrm{SR}}_T^{\mathrm{std}} = \frac{\frac{1}{T}\sum_{t=T+1}^{2T} \hat{\boldsymbol{\omega}}_T(\mathbf{Z}_{T,t})'\mathbf{R}_{T,t}}{\sqrt{\frac{1}{T-1}\sum_{t=T+1}^{2T}\left(\hat{\boldsymbol{\omega}}_T(\mathbf{Z}_{T,t})'\mathbf{R}_{T,t} - \frac{1}{T}\sum_{t=T+1}^{2T}\hat{\boldsymbol{\omega}}_T(\mathbf{Z}_{T,t})'\mathbf{R}_{T,t}\right)^2}}.$$

**例1。（采用合并普通最小二乘法的非线性模型）**。令 $p < T$，并且对每个日期 $t \geq 1$，令 $C_t$ 为一个跨日期相互独立的标准正态因子。令 $Z_{i,t}$ 和 $W_{i,t,k}$（其中 $k = 1, \ldots, p-2$）为相互独立的标准正态资产特定回归变量，且独立于 $C_t$，并在资产和月份之间相互独立。最后，定义三阶多项式 $H(c) = \frac{c^3-3c}{\sqrt{6}}$，使得对于标准正态分布的 $Z$，$H(Z)$ 的均值为零、方差为单位方差。收益模型如下：

$$R_{i,t} = \mu_{i,t} + \varepsilon_{i,t}, \qquad \varepsilon_{i,t} \sim \mathrm{NID}\left(0, \sigma^2_{\varepsilon,T}\right),$$

其中 $\mu_{i,t} = \sqrt{\lambda}\, C_t + \sqrt{1-\lambda}\, Z_{i,t} + a_T H(C_t)$ 和 $\sigma^2_{\varepsilon,T} = 19(1 + a_T^2)$。收益误差独立于所有回归变量。此外，$0 < \lambda < 1$ 和 $a_T = c_a T^{-\alpha}$ 表示非线性项上趋于零的系数，其中 $1/4 < \alpha < 1/2$ 且 $0 < c_a$。误差方差水平确保总体 $R^2$ 为 5%。合并设计为 $\mathbf{Z}_t^{(i)} = (C_t, Z_{i,t}, W_{i,t,1}, \ldots, W_{i,t,p-2})'$，系数向量为 $\boldsymbol{\beta}_0 = (\sqrt{\lambda}, \sqrt{1-\lambda}, 0, \ldots, 0)'$。因此，条件均值仅取决于前两个回归变量；$a_T H(C_t)$ 是遗漏的非线性成分，而回归变量 $W_{i,t,k}$ 是多余的。

合并普通最小二乘估计量为

$$\hat{\boldsymbol{\beta}}_T = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(R_{i,t} - \mathbf{Z}_t^{(i)\prime}\boldsymbol{\beta}\right)^2.$$

$\widehat{\text{SR}}_T^{\text{DB}}$对于去偏估计量 SR$_T$，条件协方差矩阵必须已知，或仅使用训练样本进行估计。在此，我们要求其满足 $N^{-2}\mathbf{1}'\widehat{\boldsymbol{\Sigma}}_T\mathbf{1} = O_p(1)$，这足以使协方差估计在根-$T$ 尺度下的影响可以忽略；即使$N$ 大于 $T$，训练残差的样本协方差矩阵也可以满足这一要求（见第2.9节）。

下面的命题区分了标准夏普比率估计量的一阶偏差（以 $a_T$ 表示）与去偏估计量的二阶偏差。该结果表明，当 $a_T = T^{-\alpha}$ 时，对于任意 $\alpha > 1/4$，基于去偏估计量的推断在渐近意义下仍以零为中心；而对于任意 $\alpha < 1/2$，标准估计量的根-$T$ 偏差会发散。后者会造成问题，因为机器学习估计量的收敛速度可能慢于 $\sqrt{T}$ 速率（$\alpha < 1/2$），即使能够保证 $\alpha > 1/4$ 的速率也是如此。以神经网络估计量为例：根据 Farrell 等人（2021）的神经网络收敛速率，在特征的横截面相关性不受限制的情形下，我们可以保证相应的合并回归误差至多为 $T^{-\alpha}$ 阶，其中某个 $1/4 < \alpha < 1/2$ 成立（参见第2.9节）。

**命题1**。 假设例*1*中的数据生成过程成立，并令 $p/T \to \pi \in (0, 1)$，$N/T \to \nu \in (0, \infty)$。那么，对于等权市场择时投资组合的标准夏普比率估计量和去偏夏普比率估计量，分别有：

$$\sqrt{T}\left(\widehat{\text{SR}}_T^{\text{std}} - \text{SR}_{T,0}\right) = Z_T + \frac{\sqrt{3}}{2\sqrt{\lambda}}\sqrt{T}\, a_T + o_p(1),$$
$$\sqrt{T}\left(\widehat{\text{SR}}_T^{\text{DB}} - \text{SR}_{T,0}\right) = Z_T + \frac{29}{4\sqrt{2}\lambda}\sqrt{T}\, a_T^2 + o_p(1),$$

其中 $Z_T \rightsquigarrow N(0, 3/4)$，去偏夏普比率估计量定义于后续内容中 *l.*

### 2.3.1 横截面依赖下的夏普比率展开

在附录S1.2节中，我们证明，对于横截面相关性不受限制的特征，也会出现与命题1相同的定性结果。事实上，我们可以将共同的

因子 $C_t$ 替换为一个因子暴露型资产特定回归变量 $C_{i,\ t}$,

$$C_{i,t} = \sqrt{\rho}\, F_t + \sqrt{1-\rho}\, \varsigma_{i,t}, \qquad \rho \in [0,1],$$

其中， $F_t$ 和 $\varsigma_{i,\ t}$ 是相互独立的标准正态随机变量，且在不同日期之间相互独立；$\varsigma_{i,\ t}$ 也在不同资产之间相互独立。对于任意固定的 $0 < \rho \leq 1$ ，关于 $a_t$ 的一阶偏差和二阶偏差仍然存在。

## 2.4 标准夏普比率估计量

标准夏普比率估计量是关于一阶和二阶代入样本矩的光滑映射。令 $\boldsymbol{\theta}_{T,\ 0}$ 表示包含一阶和二阶神谕投资组合矩的向量:

$$\boldsymbol{\theta}_{T,0} = \begin{bmatrix} \theta_{T,0,1} \\ \theta_{T,0,2} \end{bmatrix} = \mathbb{E}\begin{bmatrix} \boldsymbol{\omega}_T(\mathbf{Z}_{T,t}, \boldsymbol{\mu}_{T,0}(\mathbf{Z}_{T,t}))'\mathbf{R}_{T,t} \\ (\boldsymbol{\omega}_T(\mathbf{Z}_{T,t}, \boldsymbol{\mu}_{T,0}(\mathbf{Z}_{T,t}))'\mathbf{R}_{T,t})^2 \end{bmatrix}.$$

由此可得，神谕夏普比率 $\mathrm{SR}_{T,\ 0}$ 为

$$\mathrm{SR}_{T,0} = g(\boldsymbol{\theta}_{T,0}), \quad \text{where } g(\boldsymbol{\theta}) = \frac{\theta_1}{\sqrt{\theta_2 - \theta_1^2}},$$

对于满足 $\theta_2 - \theta_1^2 > 0$的任意 $\boldsymbol{\theta} \in \mathbb{R}^2$ 。

$\boldsymbol{\theta}_{T,\ 0}$的样本矩估计量记为 $\hat{\boldsymbol{\theta}}_T^{\mathrm{std}}$ ，它是基于得分函数的代入式广义矩估计量

$$\boldsymbol{\psi}_T^{\mathrm{std}}(\mathbf{w}, \boldsymbol{\theta}, \boldsymbol{\mu}) = \begin{bmatrix} \boldsymbol{\omega}_T(\mathbf{z}, \boldsymbol{\mu}(\mathbf{z}))'\mathbf{r} \\ \{\boldsymbol{\omega}_T(\mathbf{z}, \boldsymbol{\mu}(\mathbf{z}))'\mathbf{r}\}^2 \end{bmatrix} - \boldsymbol{\theta},$$

对于 $\mathrm{W}_{T,t}$ 和 $\boldsymbol{\mu} \in \mathrm{M}_T$ 的定义域中的一般观测值 $\mathrm{w} = (\mathrm{z}', \mathrm{r}')'$ 。通常使用样本分割来估计 $\boldsymbol{\mu}_{T,0}$，随后在

评估样本上计算代入样本矩。由此得到标准夏普比率估计量：

$$\widehat{\mathrm{SR}}_T^{\mathrm{std}} = g\left(\hat{\boldsymbol{\theta}}_T^{\mathrm{std}}\right).$$

在上面的命题1中，我们证明，基于该估计量的推断可能存在渐近偏差。在 S1.3节中，我们证明，该得分函数通常不满足 Neyman正交性，而这正是实现去偏推断的关键要素。

## 2.5 去偏夏普比率估计量

为便于进行有效推断，我们引入一个基于新型得分函数的 $\boldsymbol{\theta}_{T,\,0}$ 的去偏GMM估计量

$$\boldsymbol{\psi}_T^{\mathrm{DB}}(\mathbf{w}, \boldsymbol{\theta}, \boldsymbol{\eta}) = \boldsymbol{\psi}_T^{\mathrm{std}}(\mathbf{w}, \boldsymbol{\theta}, \boldsymbol{\mu}) - \begin{bmatrix} \boldsymbol{\mu}(\mathbf{z})' \mathbf{J}_{T,0}(\mathbf{z}, \boldsymbol{\mu}(\mathbf{z}))\{\boldsymbol{\mu}(\mathbf{z}) - \mathbf{r}\} \\ 2\boldsymbol{\omega}_{T,0}(\mathbf{z}, \boldsymbol{\mu}(\mathbf{z}))' \{\boldsymbol{\mu}(\mathbf{z})\boldsymbol{\mu}(\mathbf{z})' + \boldsymbol{\Sigma}\} \mathbf{J}_{T,0}(\mathbf{z}, \boldsymbol{\mu}(\mathbf{z}))\{\boldsymbol{\mu}(\mathbf{z}) - \mathbf{r}\} \end{bmatrix},$$

针对 $\boldsymbol{\mu} \in \mathrm{M}_T$ 和 $\boldsymbol{\Sigma} \in \mathcal{S}_T$，其中 $\boldsymbol{\eta} = (\boldsymbol{\mu},\ \boldsymbol{\Sigma})$。去偏项依赖于收益的信号函数和条件协方差矩阵。两者都需要估计；我们将在第2.9节证明，采用神经网络回归估计信号，并使用高维收益残差样本协方差矩阵，效果良好，且满足有效推断所需的理论限制。

在假设1下，矩条件

$$\mathbb{E}\left[\boldsymbol{\psi}_T^{\mathrm{DB}}(\mathbf{W}_{T,t}, \boldsymbol{\theta}, \boldsymbol{\eta}_{T,0})\right] = \mathbf{0}$$

在 $\boldsymbol{\theta}$ 中，由于期望得分函数是仿射的，该条件唯一地在 $\boldsymbol{\theta} = \boldsymbol{\theta}_{T,\,0}$ 处得到满足。此外，得分函数相对于完整干扰参数 $\boldsymbol{\eta}$ 满足 Neyman正交性质，在该点处计算为

$$\boldsymbol{\eta}_{T,0} = (\boldsymbol{\mu}_{T,0}, \boldsymbol{\Sigma}_{T,0}).$$

记为 $\hat{\boldsymbol{\theta}}_T^{\mathrm{DB}}$ 的去偏GMM估计量是基于去偏矩的代入式广义矩估计量，而去偏夏普比率估计量则定义为

$$\widehat{\mathrm{SR}}_T^{\mathrm{DB}} = g\left(\hat{\boldsymbol{\theta}}_T^{\mathrm{DB}}\right).$$

**注2**。 我们的所有结果都可以直接推广到任何能够表示为适当函数 $g(\hat{\boldsymbol{\theta}}_T^{\mathrm{DB}})$ 的其他统计量。示例包括均值—方差效用下的确定性等价收益、参数正态风险价值、预期损失和/或风险调整后资本收益。下方表*1*概述了它们的函数。

表1作为投资组合矩函数的绩效和风险度量。

| 度量 | 函数 $g(\boldsymbol{\theta})$ |
|---|---|
| 均值—方差确定性等价收益 | $\theta_1 - \frac{\gamma}{2}(\theta_2 - \theta_1^2)$ |
| 高斯风险价值 | $-\theta_1 + z_q \sqrt{\theta_2 - \theta_1^2}$ |
| 高斯预期损失 | $-\theta_1 + \frac{\phi(z_q)}{1-q} \sqrt{\theta_2 - \theta_1^2}$ |
| 风险调整资本收益率 | $\frac{\theta_1}{z_q \sqrt{\theta_2 - \theta_1^2}}$ |

注： $\gamma > 0$ 是均值–方差风险厌恶系数。对于置信水平 $q \in (1/2,\ 1)$，$z_q$ 是标准正态分布的 $q$ 分位数，$\phi$ 是其密度。风险价值和预期损失均使用损失收益计算。风险调整资本是超出预期损失的高斯风险价值。

## 2.6 样本分割与最终去偏估计量

我们的最终估计量 $\hat{\boldsymbol{\theta}}_T^{\mathrm{DB}}$ ，用于估计 $\boldsymbol{\theta}_{T,0}$ ，是对样本分割设计中得到的去偏插件广义矩估计量取平均；该设计允许采用标准的时间序列训练–测试划分，例如固定窗口、移动窗口和扩展窗口估计方案。我们在 S1.4节提供完整细节。为严格保持训练–评估分离，我们仅使用评估样本之前的观测值来估计信号和协方差矩阵。

## 2.7 夏普比率差异的估计

我们通常关心策略之间夏普比率差异的估计。我们用 $\boldsymbol{\nu}_{T,0}(\mathbf{Z}_{T,t})$ 表示已知且无需估计的投资组合权重向量，该投资组合是我们要与市场择时策略进行比较的对象，例如买入并持有策略和等权重策略。与上一小节一样，我们可以使用 GMM估计量来估计一阶和二阶投资组合矩，其总体向量记为 $\boldsymbol{\tau}_{T,0} = (\tau_{T,0,1}, \tau_{T,0,2})'$。鉴于此类投资组合的 $\boldsymbol{\nu}_{T,0}(\mathbf{Z}_{T,t})$ 已知，我们使用基准策略的标准得分函数：

$$\boldsymbol{\psi}_T^{\text{bench}}(\mathbf{w}, \boldsymbol{\tau}) = \begin{bmatrix} \boldsymbol{\nu}_{T,0}(\mathbf{z})'\mathbf{r} \\ \{\boldsymbol{\nu}_{T,0}(\mathbf{z})'\mathbf{r}\}^2 \end{bmatrix} - \boldsymbol{\tau}.$$

基准矩的估计量为

$$\hat{\boldsymbol{\tau}}_T = \frac{1}{T_{\text{eval}}} \sum_{t \in \mathcal{T}_{\text{eval}}} \begin{bmatrix} \boldsymbol{\nu}_{T,0}(\mathbf{Z}_{T,t})'\mathbf{R}_{T,t} \\ \{\boldsymbol{\nu}_{T,0}(\mathbf{Z}_{T,t})'\mathbf{R}_{T,t}\}^2 \end{bmatrix},$$

其中 $\mathcal{T}_{\text{eval}}$ 表示　评估期索引。夏普比率差异为

$$\Delta\text{SR}_{T,0} = g(\boldsymbol{\theta}_{T,0}) - g(\boldsymbol{\tau}_{T,0}),$$

其去偏估计量为

$$\widehat{\Delta\text{SR}}_T^{\text{DB}} = g\left(\hat{\boldsymbol{\theta}}_{T,0}^{\text{DB}}\right) - g(\hat{\boldsymbol{\tau}}_T)\,.$$

### 2.7.1 有效置信区间

在接下来的章节中，我们将正式证明，在满足充分条件时，以下置信区间是有效的；例如，使用合并神经网络收益信号估计量即可满足这些条件。这些置信区间基于新去偏估计量的渐近正态性。我们分别针对独立同分布数据设定，以及

具有 $\beta$-混合三角阵列行的时间序列设定推导结果。本文中使用的所有 oracle 方差、估计方差及协方差矩阵，包括其异方差和自相关一致版本与非异方差和自相关一致版本，均在 S1.5 节中定义。

令 $z_{1-\alpha/2}$ 表示标准正态分布的 $(1-\alpha/2)$ 分位数，例如，对于 $\alpha = 5\%$，其值为$z_{1-\alpha/2} \approx 1.96$。对于固定的非零向量 $\mathbf{c} \in \mathbb{R}^2$，$\mathbf{c}'\boldsymbol{\theta}_{T,0}$、$\mathrm{SR}_{T,0}$ 和 $\Delta\mathrm{SR}_{T,0}$ 的渐近双侧 $(1-\alpha)$ 置信区间分别为

$$\begin{aligned}
\mathbf{c}'\widehat{\boldsymbol{\theta}}_T &\pm z_{1-\alpha/2}\sqrt{\frac{\mathbf{c}'\widehat{\mathbf{V}}_T\mathbf{c}}{T_{\mathrm{eval}}}},\\
\widehat{\mathrm{SR}}_T^{\mathrm{DB}} &\pm z_{1-\alpha/2}\sqrt{\frac{\widehat{v}_{T,\mathrm{SR}}}{T_{\mathrm{eval}}}},\\
\widehat{\Delta\mathrm{SR}}_T^{\mathrm{DB}} &\pm z_{1-\alpha/2}\sqrt{\frac{\widehat{v}_{T,\Delta}}{T_{\mathrm{eval}}}},
\end{aligned}$$

其中，$T_{\mathrm{eval}}$ 表示评估折的总样本量（见 S1.4节），方差和协方差估计量的定义见 S1.5 节。在定理 1、引理 1 和推论 S1.1 的条件下，这些区间分别在 $P \in \mathcal{P}$ 上一致地具有 $1-\alpha+o(1)$ 的覆盖率。

对于 S1.8 节所述的时间序列设定，使用相应的异方差和自相关一致协方差估计量。置信区间变为

$$\begin{aligned}
\mathbf{c}'\widehat{\boldsymbol{\theta}}_T^{\mathrm{DB}} &\pm z_{1-\alpha/2}\sqrt{\frac{\mathbf{c}'\widehat{\mathbf{V}}_{T,0}^{\mathrm{HAC}}\mathbf{c}}{T_{\mathrm{eval}}}},\\
\widehat{\mathrm{SR}}_T^{\mathrm{DB}} &\pm z_{1-\alpha/2}\sqrt{\frac{\widehat{v}_{T,\mathrm{SR}}^{\mathrm{HAC}}}{T_{\mathrm{eval}}}},\\
\widehat{\Delta\mathrm{SR}}_T^{\mathrm{DB}} &\pm z_{1-\alpha/2}\sqrt{\frac{\widehat{v}_{T,\Delta}^{\mathrm{HAC}}}{T_{\mathrm{eval}}}}.
\end{aligned}$$

在 S1.8 节所述附加条件下，这些区间同样在 $P \in \mathcal{P}$ 上具有 $1-\alpha+o(1)$ 的一致覆盖率。

## 2.8 去偏估计量的渐近正态性

现在，我们建立投资组合矩和夏普比率的去偏估计量的一致渐近正态性，从而为相关的渐近置信区间提供依据。在本节中，我们针对独立同分布数据（假设 4）证明该结果，而 S1.8 节将其扩展至时间序列数据。具体而言，该节允许 $\beta$-混合三角阵列行，并发展基于 HAC的推断。

下文给出一组充分条件，其中假设 3 尤为重要，因为它为信号向量的估计量以及收益的条件协方差矩阵的估计量提供了充分的误差界。在第2.9节中，我们将证明，对于任意固定的 $\xi > 0$，这些速率可由信号的合并神经网络估计量和训练残差的样本协方差矩阵满足，且 $N$ 关于 $T$、 $N \asymp T^{\xi}$, 呈多项式增长。该假设还对数据施加了一致界。这类有界性条件在机器学习估计量的理论分析中很常见（参见 Farrell 等人（2021））。**假设 3** （信号与干扰项估计界）。 对于所有 $P \in \mathcal{P}$ 和 $T \geq 1$，存在常数 $C < \infty$，其与 $P$ 和 $T$无关，使得：

$$P(\|\mathbf{W}_T\|_\infty \leq C) = P(\|\boldsymbol{\mu}_{T,0}(\mathbf{W}_T)\|_\infty \leq C) = 1.$$

此外，存在确定性序列 $r_{\mu,\,T} \geq 0$、 $r_{\Sigma,\,T} \geq 0$ 和 $\Delta_T \geq 0$，它们与 $P$ 无关且满足 $\Delta_T \to 0$，使得对于所有 $P \in \mathcal{P}$ 和各折 $k = 1$、.....、$K$，以至少 $1 - \Delta_T$ 的 $\mathbb{P}_P$-概率，有

$$P(\|\widehat{\boldsymbol{\mu}}_T^{(k)}(\mathbf{W}_T)\|_\infty \leq C) = 1,$$

$$\left\|\widehat{\boldsymbol{\mu}}_{T,k}(\mathbf{W}_T) - \boldsymbol{\mu}_{T,0}(\mathbf{W}_T)\right\|_{P,2} \leq r_{\mu,T},$$

$$\left\|\widehat{\boldsymbol{\Sigma}}_{T,k} - \boldsymbol{\Sigma}_{T,0}\right\|_{\mathrm{op}} \leq r_{\Sigma,T}.$$

其中， $\mathbb{P}_P$ 表示相对于训练样本的概率，而 $\|\cdot\|_{P,\,2}$ 是总体

$L_2$（$P$）范数，对固定的拟合估计量而言。

最后，这些速率满足

$$\frac{r_{\Sigma,T}}{N_T} \longrightarrow 0, \qquad \sqrt{T}\left(\frac{r_{\mu,T}}{\sqrt{N_T}}\right)^2 \longrightarrow 0, \qquad \sqrt{T}\left(\frac{r_{\Sigma,T}}{N_T}\right)\left(\frac{r_{\mu,T}}{\sqrt{N_T}}\right) \longrightarrow 0.$$

**假设 4 (**独立同分布行和协方差矩阵正则性**)**。 对于每个 $T \geq 1$，序列 $\{\mathbf{W}_{T,t} : t \geq 1\}$ 在每个 $P \in \mathcal{P}$ 下均为独立同分布。

**假设5**。 渐近协方差矩阵 $\mathbf{V}_{T,0}$ （定义见 *S1.5* 节）的特征值一致地有正下界和有限上界。存在一个常数 $0 < c \leq \bar{c} < \infty$，与 $P$ 和 $T$无关，使得 $c \leq \lambda_{\min}(\mathbf{V}_{T,0}) \leq \lambda_{\max}(\mathbf{V}_{T,0}) \leq \bar{c}$，对所有 $P \in \mathcal{P}$以及充分大的 $T$均成立。

以下结果表明，市场择时投资组合和平滑多空投资组合的去偏投资组合矩，在经^一致协方差矩阵估计量 $\mathbf{V}_T$ 进行学生化后渐近服从标准正态分布（定义见 S1.5 节）。

**定理 1 (**交叉拟合平滑得分估计量的一致弱收敛**)**。 假设假设 *1-5* 成立。则对于满足 $\max_{i} a_{T,i,t} \leq C/N_T$的市场择时投资组合，以及固定 $\tau \in (0, \infty)$的平滑多空投资组合，对所有 $P \in \mathcal{P}$均一致成立，

$$\widehat{\mathbf{V}}_T^{-1/2}\sqrt{T_{\text{eval}}}\{\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_{T,0}\} \rightsquigarrow N(\mathbf{0}, \mathbf{I}_2), \tag{1}$$

在有界-利普希茨度量下。

最后，正如引理 1 所证明的，去偏夏普比率估计量渐近服从正态分布。此外，这还要求投资组合方差远离零，且投资组合收益的一阶和二阶矩有界。正如推论 S1.1 所证明的，夏普比率差异的估计量也具有类似结果。

**假设 6 (**夏普比率的一致正则性**)**。 存在常数 $v_{\mathrm{SR}} > 0$ 和$C < \infty$，其与 $P$ 和 $T$ 无关，使得

$$\theta_{T,0,2} - \theta_{T,0,1}^2 \geq \underline{v}_{\mathrm{SR}} \qquad and \qquad \|\boldsymbol{\theta}_{T,0}\|_2 \leq C,$$

对所有 $P \in \mathcal{P}$ 且 $T$ 充分大时。

**引理 1** （夏普比率的一致性 Delta 方法）。 假设假设 *1-6* 成立。则，

$$\sqrt{T_{\mathrm{eval}}} \frac{\widehat{\mathrm{SR}}_T^{\mathrm{DB}} - \mathrm{SR}_{T,0}}{\sqrt{\widehat{v}_{T,\mathrm{SR}}}} \rightsquigarrow N(0,1),$$

在有界-利普希茨度量下均匀成立。

## 2.9 示例：通过合并神经网络回归估计的信号

我们将展示，结合收益残差的样本协方差矩阵，信号的合并神经网络估计量满足假设 3 中去偏推断的条件。为便于说明，考虑包含 $K = 2$ 折的固定窗口方案，此时 $T = 2T_{\mathrm{block}}$。第一折 $t = 1, \ldots\ldots, T_{\mathrm{block}}$ 用于训练，第二折 $t = T_{\mathrm{block}} + 1, \ldots\ldots, T$ 用于评估。由于只有一个干扰项估计，我们省略折的上标。

我们采用广泛使用的 ReLU 多层感知机神经网络类（参见 Farrell 等人（2021）），其宽度和深度可以随 $T$增加。我们将该文中关于信号函数 $\mu_0$ 及其神经网络估计量的条件调整至合并回归设定。我们通过合并最小二乘回归估计假设1中的共同信号函数 $\mu_0$ ，

平方神经网络估计量

$$\widehat{\mu}_T \in \arg\min_{\mu \in \mathcal{N}_T} \frac{1}{N T_{\text{block}}} \sum_{t=1}^{T_{\text{block}}} \sum_{i=1}^{N} \left\{ R_{i,t} - \mu\left(\mathbf{Z}_t^{(i)}\right) \right\}^2 . \tag{2}$$

信号向量的相应估计量为

$$\widehat{\boldsymbol{\mu}}_T(\mathbf{Z}_{T,t}) := \left( \widehat{\mu}_T(\mathbf{Z}_t^{(1)}), \ldots, \widehat{\mu}_T(\mathbf{Z}_t^{(N)}) \right)' . \tag{3}$$

以下假设给出了 Farrell 等人（2021）中假设 1 和 2 的合并回归对应形式。请注意，该假设对特征的横截面依赖不作限制。在现代资产定价情境下，这一点十分重要，因为特征集通常包含共同状态变量以及资产特定特征，而许多资产对的这些特征之间存在中度至高度相关性。

**假设 7.** 对于所有 $t = 1$、……、$T$，以及当$P \in \mathcal{P}$ 和 $T$ 充分大时，一致地满足以下条件。

*1.* 真实信号函数 $\mu_0$ 满足 ***(i)*** $\max_{i=1,\ldots,N} \|\mu_0(\mathbf{Z}_T^{(i)})\|_{P,\infty} \le M$，其中$M$有限；且 ***(ii)*** 对某个 $\beta \in \mathbb{N}_+$,

$$\mu_0 \in \mathcal{W}_C^{\beta,\infty}([-1,1]^p) := \left\{ \mu : \max_{|\boldsymbol{\alpha}| \le \beta} \operatorname{ess\,sup}_{\mathbf{z} \in [-1,1]^p} \left| D^{\boldsymbol{\alpha}} \mu(\mathbf{z}) \right|_\infty \le 1 \right\},$$

其中 $D^{\boldsymbol{\alpha}}\mu$ 表示与多重指标 $\boldsymbol{\alpha} \in \mathbb{N}_0^p$ 对应的弱偏导数，且 $|\boldsymbol{\alpha}| = \sum_{j=1}^{p} \alpha_j$;

*2.* 估计量 $\widehat{\ }\mu_T$ 属于 $\mathcal{N}_T$，其中 $\mathcal{N}_T$ 是一个标量输出 ***ReLU*** 多层感知机类，其架构的宽度为 $H \asymp T^{p/[2(\beta+p)]} \log^2 T$、深度为 $L \asymp \log T$，且仅包含满足$\sup_{\mathbf{z}} \|\mu(\mathbf{z})\|_\infty \le 2M$的函数;

*3.*收益误差 $\varepsilon_{T,i}$ 相互独立。特征向量 $\mathbf{Z}_t^{(i)}$ 服从连续分布，且特征维度 $p$ 固定。数据满足以下有界条件

$P(\|\mathbf{Z}_T\|_\infty \le 1) = 1$ *and* $P(\|\mathbf{R}_T\|_\infty \le M) = 1$.

去偏估计量需要一个$\mathbf{\Sigma}_{T,0}$的估计量，我们使用收益残差的样本协方差矩阵。定义拟合收益残差向量及其训练样本均值如下

$$\widehat{\varepsilon}_{T,t} := \mathbf{R}_{T,t} - \widehat{\boldsymbol{\mu}}_T(\mathbf{Z}_{T,t}) \quad \text{and} \quad \overline{\varepsilon}_T := \frac{1}{T_{\text{block}}} \sum_{t=1}^{T_{\text{block}}} \widehat{\varepsilon}_{T,t}.$$

收益残差的样本协方差矩阵定义为

$$\widehat{\mathbf{\Sigma}}_T := \frac{1}{T_{\text{block}}} \sum_{t=1}^{T_{\text{block}}} (\widehat{\varepsilon}_{T,t} - \overline{\varepsilon}_T)(\widehat{\varepsilon}_{T,t} - \overline{\varepsilon}_T)'. \tag{4}$$

下述引理给出了这两个估计量的估计率保证。证明采用了 Farrell 等人（2021）中神经网络收敛率结果的一个新版本，即合并回归版本，该结果见 S4.3 节；此外，还采用了 Vershynin（2018）定理 4.7.1 和习题 4.7.3 中关于次高斯随机向量的样本协方差矩阵界。该引理表明，对于任意固定的 $N \asymp T^{\xi}$，只要 $\xi > 0$，去偏推断的速率限制就满足假设 3。这一限制要求预言信号函数 $\mu_0$ 相对于特征数量足够平滑。利用 Brown（2025）的结果，可以将这些结果推广到具有时间序列依赖的特征。

**引理 2**（合并神经网络和残差协方差收敛率）。 假设假设 *1*、*4* 和 *7* 成立。那么，以至少 $1 - \Delta_T$ 的概率，

$$P(\|\widehat{\boldsymbol{\mu}}_T(\mathbf{Z}_T)\|_\infty \le 2M) = 1, \quad \|\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_{T,0}\|_{P,2} \le r_{\mu,T}, \quad \left\|\widehat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_{T,0}\right\|_{\text{op}} \le r_{\Sigma,T},$$

其中

$$
\begin{aligned}
r_{\mu,T} &:= \sqrt{N q_T}, \\
r_{\Sigma,T} &:= b_{\Sigma,T} + 2\sqrt{(C + b_{\Sigma,T}) N q_T} + N q_T, \\
\Delta_T &:= \exp\left\{-T_{\text{block}}^{p/(\beta+p)} \log^8 T_{\text{block}})\right\} + 4\exp(-N), \\
q_T &:= C\left\{T_{\text{block}}^{-\beta/(\beta+p)} \log^7 T_{\text{block}} \ \log(N T_{\text{block}}) + \frac{\log\log T_{\text{block}}}{T_{\text{block}}}\right\}, \\
b_{\Sigma,T} &:= C\left\{\sqrt{\frac{N}{T_{\text{block}}}} + \frac{N}{T_{\text{block}}}\right\}.
\end{aligned}
$$

## 2.10 去偏估计量的半参数效率

在 S1.7节中，我们证明了 $\widehat{\boldsymbol{\theta}}_T^{\mathrm{DB}}$ 对独立同分布数据是半参数有效的，并且这种$\widehat{\ }_{\mathrm{DB}}$有效性也适用于夏普比率估计量 $\mathrm{SR}_T$ 。这一结果意味着，在相关模型中，没有其他正则估计量能够比$\widehat{\ }_{\mathrm{DB}}\mathrm{SR}_T$ 具有更小的渐近方差。它有力地支持了使用去偏估计量：它不仅能给出有效的渐近置信区间，还能给出渐近最短的 Wald$\widehat{\ }_{\mathrm{std}}$置信区间。重要的是，即使标准估计量 $\mathrm{SR}_T$ 在相关模型中是正则的，它给出的渐近置信区间也不会比去偏估计量更小。

# 3 蒙特卡洛模拟

为评估检验程序的有限样本性能，我们在汇总神经网络回归设定下进行蒙特卡洛模拟。我们对例1中的数据生成过程（DGP）进行调整，使其满足上一节所要求的理论条件。具体而言，我们确保收益、回归变量和条件均值函数均有界。此外，我们将条件均值函数改为关于两个活跃回归变量的非线性函数。

对于 $(\mathrm{R}_t)_{t=1}^T$，我们进行模拟

$$R_{it} = \mu_0(\mathbf{x}_{it}) + \varepsilon_{it}, \qquad \varepsilon_{it} \overset{\text{iid}}{\sim} \sigma\text{clip}(N(0,1), -10, 10),$$

其中，扰动项（$\varepsilon_t$）$_{t=1}^T$ 在每个指数上均相互独立。我们固定 $p = 25$ 和$T = 780$，并对预测变量向量施加以下结构

$$\mathbf{x}_{it} = (U_{t1}, U_{it2}, \ldots, U_{it,25})'.$$

$\mathbf{x}_{it}$ 的每个元素均服从边际均匀分布于 $[0,1]$。第一个预测变量 $U_{t1}$ 对所有资产 $i$ 在每个 $t$ 上都是共同的。其余 $p-1$ 个预测变量在资产之间及预测变量之间相互独立，且所有日期相互独立。

我们使用

$$L(u) = \sqrt{3}(2u-1), \qquad H(u) = \frac{\sqrt{7}}{2}\{5(2u-1)^3 - 3(2u-1)\},$$

表示一阶和三阶标准化勒让德多项式。条件均值函数的函数形式如下：

$$\begin{aligned}\mu_0(\mathbf{x}_{it}) = &\left\{\sqrt{\lambda}L(U_{t1}) + aH(U_{t1})\right\} \\ &+ \frac{\sqrt{1-\lambda}}{\sqrt{1+a^2}}\left\{\sqrt{0.99}L(U_{it2}) + 0.1L(U_{it3}) - aH(U_{it2})\right\},\end{aligned}$$

其中，第一个括号内的项表示共同因子成分，而第二个括号内的项旨在捕捉特质性变异。注意，数据生成过程是稀疏的；只有 $p$ 个特征中的前三个实际进入数据生成过程。

共同成分的方差为 $\lambda + a^2$，而特质信号成分的方差为 $1-\lambda$。参数 $a$ 经过校准，使得预测 $R^2$ 为 5%，即

$$\text{Var}(\mu_{0,it}) = 1 + a^2, \qquad \sigma^2 = 19(1+a^2) \implies R^2_{\text{pred}} = \frac{\text{Var}(\mu_{0,it})}{\text{Var}(R_{it})} = 0.05.$$

因此，在各种设定下，我们保持了相同的信噪比。对于 $i \neq j$，我们进一步有：

$$\mathrm{Corr}(\mu_{0,it}, \mu_{0,jt}) = \frac{\lambda + a^2}{1 + a^2}, \qquad \mathrm{Corr}(R_{it}, R_{jt}) = \frac{\lambda + a^2}{20(1 + a^2)}.$$

我们改变资产数量 $N \in \{50, 150, 250, 500\}$以及共同成分解释的方差比例 $\lambda \in \{0.25, 0.40, 0.55\}$。这些载荷对应的信号相关性约为 0.252、0.402 和 0.551，收益相关性则分别为 0.0126、0.0201 和 0.0276。

## 3.1 神经网络训练

我们使用 360 个日期进行拟合，60 个日期进行验证，并使用最后 360 个日期进行样本外评估。我们拟合一个包含 1 个隐藏层、由 32 个 ReLU 单元组成的合并前馈神经网络（FNN）。为加快计算，我们将其权重和偏置初始化为使初始预测与合并普通最小二乘法的预测相同。初始化的详细信息以及训练流程的具体内容见附录 S5.1。

我们采用第 2 节中的投资组合规则和推断程序；在估计夏普比率之差时，两种策略均以被动等权重（EW）指数为基准。MT 投资组合规则跟踪 EW 指数。收益协方差这一干扰参数采用 420 个拟合和验证日期上的未中心化残差二阶矩进行估计。我们报告相对于相应总体夏普比率的双侧 5% 拒绝频率，计算方法如附录 S5.2 所述。

## 3.2 拒绝频率

表 2 报告了 $N$ 和 $\lambda$的所有组合，并始终使用相同的数据生成过程和训练规则。为便于正文阐述，我们选择 $\tau = 0.05$。每项结果均基于 1,000 次蒙特卡罗重复试验计算得出，传统检验和去偏检验使用相同的模拟样本进行评估。附录 S5.3 报告了其他若干

$\tau$取值下的拒绝频率。多空结果对带宽较为敏感，这与图1所示的权衡一致。 $\tau$ 的较小取值会使权重在截断点附近发生剧烈变化，从而可能产生较大的雅可比矩阵元素，并使去偏校正在有限样本中变得不稳定。 $\tau$ 的较大取值会将权重平滑到偏离原始十分位规则较大的程度。

采用 FNN 预测时，传统独立 MT检验拒绝率从 $N = 50$ 时的 6.9–10.5% 上升至 $N = 500$ 时的 34.0–54.4%。在完整网格范围内，去偏MT检验拒绝率为 4.5–7.4%。例如，在 $\lambda = 0.40$ 时，随着 $N$ 增加，传统拒绝率分别为 10.5、24.9、35.3 和 42.0%，而去偏程序对应的拒绝率分别为 5.3、4.5、5.1 和 6.2%。固定数量的拟合日期和共同的非线性信号意味着，增加资产数量并不会机械地消除共同预测误差。同时，指数的特质收益噪声方差随 $\sigma^2/N$ 增加而下降。

表 2模拟拒绝频率

| | | MT | | MT–EW | | LS | | LS–EW | |
|---|---|---|---|---|---|---|---|---|---|
| $\lambda$ | $N$ | C | D | C | D | C | D | C | D |
| 0.25 | 50 | 6.9 | 5.3 | 5.5 | 4.5 | 18.6 | 9.9 | 14.3 | 8.2 |
| 0.25 | 150 | 25.3 | 5.8 | 11.7 | 6.6 | 15.8 | 6.4 | 12.5 | 6.7 |
| 0.25 | 250 | 37.9 | 6.1 | 15.9 | 4.9 | 12.0 | 6.9 | 10.6 | 6.1 |
| 0.25 | 500 | 54.4 | 7.4 | 22.3 | 5.5 | 13.2 | 6.6 | 11.4 | 5.5 |
| 0.40 | 50 | 10.5 | 5.3 | 7.2 | 5.3 | 19.5 | 10.6 | 12.8 | 8.0 |
| 0.40 | 150 | 24.9 | 4.5 | 10.1 | 4.7 | 14.6 | 7.4 | 13.4 | 7.1 |
| 0.40 | 250 | 35.3 | 5.1 | 14.9 | 5.9 | 13.4 | 6.0 | 10.9 | 5.7 |
| 0.40 | 500 | 42.0 | 6.2 | 17.2 | 5.5 | 14.4 | 4.8 | 12.6 | 5.0 |
| 0.55 | 50 | 10.2 | 4.5 | 6.3 | 4.6 | 19.8 | 10.9 | 15.2 | 7.9 |
| 0.55 | 150 | 22.4 | 5.6 | 9.2 | 4.6 | 18.4 | 6.2 | 13.2 | 5.5 |
| 0.55 | 250 | 30.2 | 6.2 | 13.6 | 6.8 | 16.5 | 7.5 | 12.2 | 5.9 |
| 0.55 | 500 | 34.0 | 4.9 | 14.4 | 6.7 | 15.2 | 4.9 | 13.5 | 5.3 |

各项数值为基于 1,000 次蒙特卡罗重复试验的双侧 5 % 检验拒绝百分比。C 和 D 分别表示传统检验和去偏检验。

对于 $\tau = 0.05$ 的独立 LS，传统拒绝率为 12.0–19.8%。去偏拒绝率

对于 $N \geq 150$，去偏拒绝率为 4.8–7.5%；对于 $N = 50$，为 9.9–10.9%。在 $\lambda = 0.40$ 和 $N = 500$ 下，传统值和去偏值分别为 14.4% 和 4.8%。因此，该网格清楚地展示了估计对 LS 推断以及 MT 推断的影响均有所减弱，同时最小横截面还存在额外的有限样本变异。

同一张表还报告了与等权指数进行夏普比率差异检验的结果。传统MT–EW检验的拒绝率介于5.5%至22.3%之间，而去偏后为4.5–6.8%。传统LS–EW检验的拒绝率介于10.6%至15.2%之间，而去偏后为5.0– 8.2%。

总体而言，模拟实验表明，我们的检验对于所考察的两种投资组合确实仍然有效，尤其是在以资产数量衡量的有效样本量 $N$ 增加时。在这种情况下，标准检验的规模未必会更接近其名义值；而对于去偏检验，我们通常发现，随着 $N$ 增加，覆盖率会更接近正确的覆盖率。

## 3.3 标准和去偏夏普比率估计量的分布

为进一步说明标准检验及其相应置信区间可能存在的偏差，我们展示了图2中若干突出设定下检验统计量的分布。随着 $N$ 增加，传统MT分布向右偏离标准正态参照分布，而去偏分布则更接近假定的渐近分布。在这些示例中，传统LS统计量则向相反方向移动。校正减小了这两种情况下的偏移。符号上的差异凸显了偏差项的符号通常未知，因为它是两个投资组合矩的非线性变换。结果表明，我们提出的程序消除了这种偏差。附录中的图S1和图S2展示了其他数据生成过程取值下的结果。

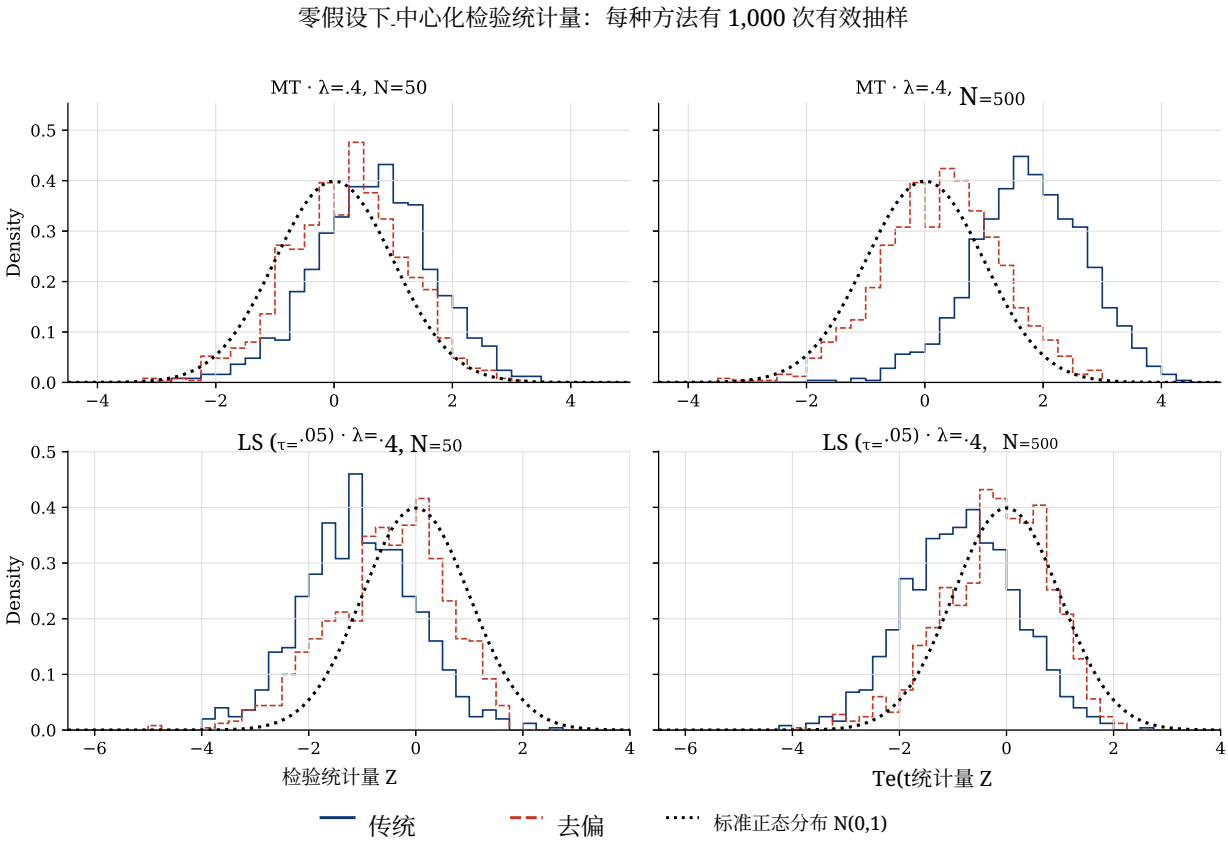



图*2*。 MT 和平滑 LS 在 $\tau = 0.05$ 下的检验统计量直方图，其中 $\lambda = 0.40$ 且$N \in \{50,500\}$。虚线曲线为标准正态密度。

# 4 实证应用

## 4.1 数据与机器学习方法

我们遵循 GKX 的做法，将月度 CRSP 超额收益与 94 个公司特征、这些特征与八个滞后宏观经济预测变量的交互项（Welch 和 Goyal，2008年），以及两位数 SIC 指示变量结合起来[1]。我们采用他们对特征变量进行月度中位数插补和秩变换的方法。样本延伸至2021年12月。样本外时期为 1987年1月至2021年12月，共覆盖 420 个月和约 286 万个股票—月份观测值。对于缺失的 SIC 代码，我们在每家公司内部向前沿用最近的代码，根据第一个可用代码填补开头的缺失值，并将其余缺失值归入未知行业类别。按照 GKX 的年度验证方案，我们最初使用 1957—1974 年的数据进行拟合，在 1975—1986 年进行验证，并在 1987年进行评估。每年扩大拟合样本，并将 12 年验证窗口向前滚动。我们根据合并后的验证均方误差选择超参数，并将选定的拟合结果带入下一评估年份。

为估计条件均值，我们使用 GKX 所述的 OLS-3、弹性网络、主成分回归、偏最小二乘以及 NN1–NN5 模型类别。我们采用平方损失，并沿用 GKX 的弹性网络混合参数和惩罚项范围。受 Shen 和 Xiu（2026）的启发，我们将 GKX 的神经网络$l_1$ 惩罚项替换为$l_2$ 惩罚项。我们保留其网络宽度和深度，对三个拟合后的网络进行集成，并采用 S6.1节中的调参网格。对于每种预测方法和评估年份，我们将所选模型在合并后的训练和验证样本中得到的月度残差向量的未中心化二阶矩，作为同期创新协方差的估计值。我们将缺失的股票—月份残差设为零，并对随后评估年份的全部 12 个月使用相同的协方差估计值。

图3显示，公司特征在相当大比例的情况下呈现出显著的相关性

[1]公司特征和宏观经济数据可从Dacheng Xiu和Amit Goyal的网站获取。

股票对。例如，在至少有 120 个共同月份且两段历史记录均存在非零变异的股票对中，61.9% 的市场权益相关性和 48.4% 的市场贝塔相关性的绝对值不小于 0.3。股息价格比（49.0%）、账面市值比（40.0%）和特质波动率（51.6%）也呈现类似模式。这些结果在各行业中均成立，表明特征普遍存在横截面依赖。因此，在第2.9节的假设及所述推断条件下，我们的神经网络误差界保证去偏置信区间渐近有效，但无法保证传统区间有效。

## 4.2 M市场择时和多空投资组合与基准的比较 s

我们使用第 2 节中的市场择时和平滑多空规则构建投资组合。我们利用匹配成分股的预测值，并按照滞后市场权益加权，对标普500指数进行择时。对于多空投资组合，我们采用平滑常数 $\tau = 0.001$；该常数是根据表 S16所报告的结果选定的。我们的被动基准包括CRSP等权重指数、CRSP市值加权指数，以及基于观测面板构建的匹配成分股市值加权标普500指数投资组合。我们报告年化夏普比率，并使用五阶滞后Newey–West标准误（Newey和West，1987）进行名义5%显著性水平下的双侧检验。我们同时报告 $t$统计量及其变化 $\Delta t$，其中 $\Delta t$计算为去偏统计量减去传统统计量。

### 4.2.1 基准

CRSP等权重、CRSP市值加权和匹配标普500指数基准的年化夏普比率分别为0.517、0.564和0.608（表 S12）。水平检验将这些基准的夏普比率与零进行比较，结果表明它们均显著异于零。

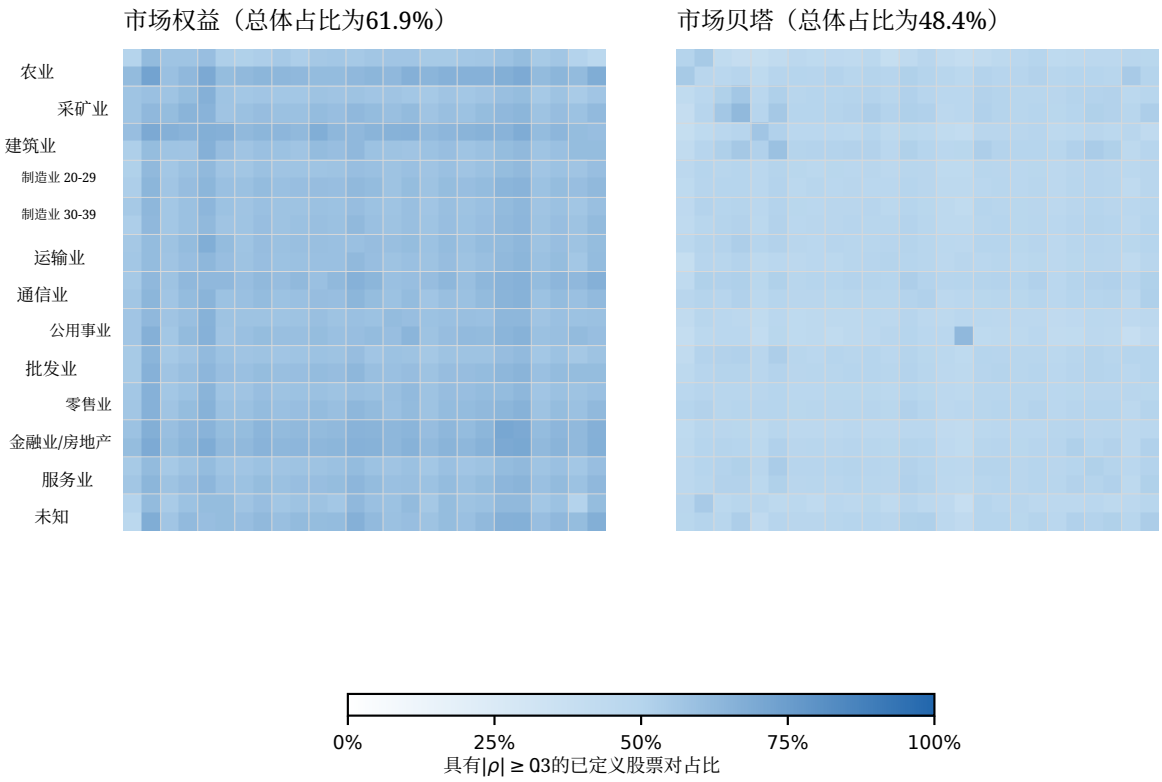



图*3*。 1987年至2021年公司特征的跨股票共同变动。左右面板分别显示市场权益和市场贝塔。每个单元格表示行业—规模组内或组间股票对中，时间序列相关性的绝对值至少为0.3的占比。我们使用至少120个共同月份，对股票间相同的按月秩变换的特征进行相关性计算。每个行业在两个坐标轴上均依次划分为小型股组和大型股组。灰色单元格表示没有已定义的相关性。

### 4.2.2 标普500指数择时

表3显示，传统推断拒绝了全部九种标普500指数择时策略的零夏普比率假设，而去偏推断拒绝了其中六种。尽管年化点估计的变动幅度至多为0.017，但NN2、NN4和NN5的检验结论发生了变化。例如，NN2的估计值从0.492变为0.487，其标准误则从0.196增至0.273。因此， $t$统计量从2.50降至1.78。这一模式符合预测灵活性与估计误差之间的权衡。灵活的设定能够捕捉更丰富的收益信号，而最终投资组合绩效估计的精确度取决于模型对信号中与经济相关部分的学习效果。在这些择时结果中，若将预测估计纳入考量，若干灵活性较高的神经网络所表现出的不确定性增幅最大。

表3标普*500*指数择时的夏普比率与零夏普比率检验

| | $R^2_{\text{oos}}$ | 传统 | | 去偏 | |
|---|---|---|---|---|---|
| 方法 | (%) | $\widehat{SR}$ | $t$ | $\widehat{SR}$ | $t$ |
| OLS-3 | 0.20 | 0.455 | 2.55 | 0.442 | 3.59 |
| ENet | 0.30 | 0.525 | 2.77 | 0.549 | 2.92 |
| PCR | 0.24 | 0.509 | 3.18 | 0.547 | 3.70 |
| PLS | 0.24 | 0.470 | 2.74 | 0.475 | 3.22 |
| NN1 | 0.33 | 0.613 | 2.78 | 0.665 | 3.24 |
| NN2 | 0.12 | 0.492 | 2.50 | 0.487 | 1.78 |
| NN3 | 0.16 | 0.589 | 3.06 | 0.588 | 2.36 |
| NN4 | 0.07 | 0.512 | 2.57 | 0.496 | 1.75 |
| NN5 | 0.08 | 0.521 | 2.64 | 0.505 | 1.80 |

注：样本外期间为1987年1月至2021年12月（$T = 420$）。夏普比率按$\sqrt{12}$进行年化。所有 $t$统计量均使用包含五阶滞后的Newey–West HAC标准误。传统推断使用实现收益及其平方项；去偏推断使用正交一阶和二阶矩得分。$R^2_{\text{oos}}$ 是相对于零收益率预测的合并个股层面预测 $R^2$，以百分比表示；其使用完整的预测样本总体，而不仅仅是匹配的指数成分股。

在各项择时策略中，夏普比率点估计仍然相近。对于 NN2–NN5，去偏标准误大 29–43%；对于 NN1 和线性设定，去偏标准误则更小。去偏后，一些 $t$-统计量有所上升，其中包括 OLS-3 的统计量。

图4展示了基于传统和去偏夏普比率估计值及其标准误的正态参照密度。尽管两种密度的中心值大致相同，但对于架构更复杂的神经网络，去偏后离散程度的增加十分显著。这说明，即使点估计几乎没有变化，推断结果也可能发生改变。

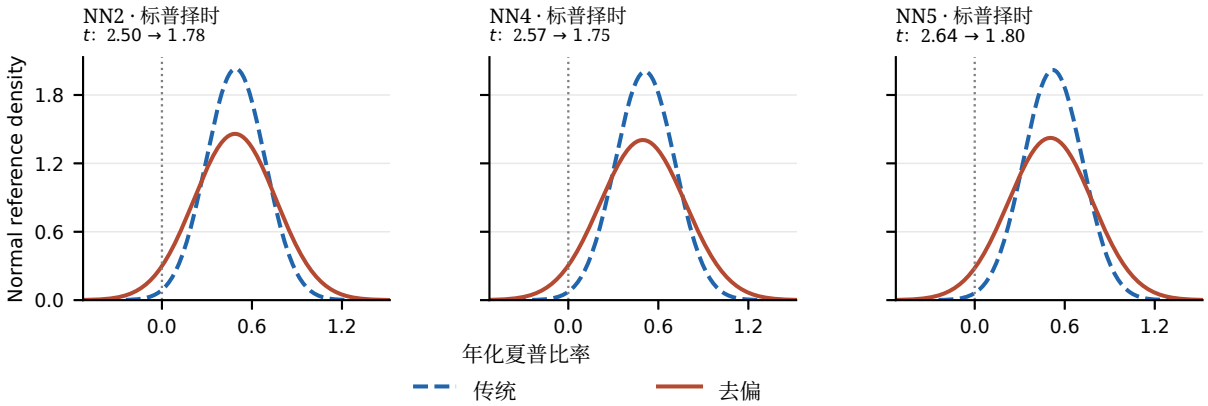



图*4.* 标普500指数择时的夏普比率推断。各面板展示了三种策略：在名义双侧5%显著性水平下，其关于夏普比率为零的传统检验与去偏检验得出的结论不一致。每个正态参考密度的均值等于年化夏普比率估计值，标准差等于其Newey–West标准误。竖线表示零。

表4中的基准比较对择时的经济价值进行了更严格的检验。在任一程序下，这些择时策略均未显著优于三个被动基准中的任何一个。对于 NN4 和 NN5，传统检验表明，相对于匹配标普500指数的表现显著不佳，而去偏检验则无法拒绝相等的假设。对于 NN4，估计差距从 $-0.096$ 扩大至 $-0.111$，其标准误从 0.045 增加至 0.103。因此，$t$-统计量从 $-2.15$ 变为 $-1.08$。补充材料中的表 S13 报告了完整的估计差距集合。

表4*t*-统计量：标普*500*指数择时与被动基准之间的夏普差异

| | CRSP等权 | | | CRSP市值加权 | | | 观测面板标准普尔 | | |
|---|---|---|---|---|---|---|---|---|---|
| 方法 | 常规 | Deb. | $\Delta t$ | 常规 | Deb. | $\Delta t$ | 常规 | Deb. | $\Delta t$ |
| OLS-3 | −0.31 | −0.42 | −0。11 | −0。61 | −0。76 | −0。15 | −0.86 | −1。04 | −0。18 |
| ENet | 0.03 | 0.12 | +0.09 | −0。15 | −0。06 | +0.09 | −0。33 | −0。23 | +0。09 |
| PCR | −0。03 | 0.13 | +0。16 | −0。24 | −0。08 | +0。16 | −0。44 | −0.29 | +0。15 |
| PLS | −0。18 | −0。17 | +0。01 | −0。37 | −0.37 | 0.00 | −0。55 | −0。57 | −0。01 |
| NN1 | 0.58 | 0.99 | +0。41 | 0.37 | 0.88 | +0.51 | 0.04 | 0.53 | +0。49 |
| NN2 | −0。19 | −0。17 | +0。01 | −0。92 | −0。66 | +0。27 | −1。75 | −1。11 | +0.64 |
| NN3 | 0.59 | 0.48 | −0。11 | 0.36 | 0.25 | −0。12 | −0。27 | −0。21 | +0。06 |
| NN4 | −0。05 | −0。13 | −0。08 | −0.94 | −0。62 | +0.31 | −2。15 | −1.08 | **+1.06** |
| NN5 | 0.03 | −0.08 | −0。11 | −0。81 | −0。57 | +0。24 | −2.09 | −1。05 | **+1.05** |

注： 常规和去偏分别报告零策略减基准夏普比率差下的常规和去偏 $t$ 统计量。标准误采用具有五阶滞后的联合策略–基准 Newey–West 协方差。 $\Delta t$ 表示描述性变化，即去偏值减去常规值，在四舍五入前计算。加粗的变化表示在双侧 5% 检验中作出了不同的决策。

### 4.2.3 多空投资组合

多空投资组合对收益信号的经济价值呈现出更为乐观的图景。表5显示，对于相同的七种平滑多空策略，传统程序和去偏程序均拒绝零夏普比率。七种策略的估计夏普比率均高于基准，尽管在两种程序下，只有五种策略显著优于全部三个基准。传统和去偏点估计通常较为相近，但 NN2 的年化夏普比率在去偏后从 1.467 上升至 1.938。更深的神经网络架构 NN4 和 NN5 再次得到较低的估计夏普比率，这可能反映了利用现有数据估计高度参数化模型的困难。

表6中的基准检验显示，在两种方法下，弹性网络、主成分回归、偏最小二乘、NN1 和 NN2 的表现均显著优于三个被动基准中的每一个。对于 NN1 相对于匹配标普500指数的表现，年化夏普比率差在传统方法下为 1.594，在去偏后为 1.564，相应的 $t$ 统计量分别为 5.52 和 5.72。因此，去偏后，NN1 估计出的巨大优势以及其表现优越的有力证据基本保持不变。

表5平滑多空的夏普比率和零夏普比率检验（$\tau = 0.001$）

| | $R^2_{\text{oos}}$ | 传统 | | 去偏 | |
|---|---|---|---|---|---|
| 方法 | (%) | $\widehat{SR}$ | $t$ | $\widehat{SR}$ | $t$ |
| OLS-3 | 0.20 | 0.791 | 4.01 | 0.764 | 4.04 |
| ENet | 0.30 | 1.692 | 7.80 | 1.697 | 8.82 |
| PCR | 0.24 | 1.624 | 7.02 | 1.601 | 7.64 |
| PLS | 0.24 | 1.879 | 7.43 | 1.815 | 8.39 |
| NN1 | 0.33 | 2.202 | 9.90 | 2.172 | 10.95 |
| NN2 | 0.12 | 1.467 | 5.60 | 1.938 | 6.63 |
| NN3 | 0.16 | 0.952 | 4.51 | 0.937 | 5.20 |
| NN4 | 0.07 | 0.111 | 0.55 | 0.343 | 0.37 |
| NN5 | 0.08 | -0.007 | -0.04 | -0.084 | -0.22 |

注：样本外期间为1987年1月至2021年12月（$T = 420$）。夏普比率按 $\sqrt{12}$年化。所有 $t$统计量均使用五阶滞后的Newey–West HAC标准误。传统推断使用实现收益及其平方项；去偏推断使用正交一阶和二阶矩得分。$R^2_{\text{oos}}$ 是合并的个股层面预测 $R^2$，以百分比表示，相对于零收益率预测；它使用完整的预测样本总体，而非仅使用匹配的指数成分股。

对于弹性网络、主成分回归、偏最小二乘和 NN1，在报告的 $\tau$ 带宽网格范围内，其显著的优越表现始终成立。表 S14 报告了主要带宽下的全部点估计，S6.3节报告了敏感性结果。综合来看，这些结果表明，基于 GKX特征集的收益信号能够带来显著且具有统计显著性的风险调整后表现提升，优于被动基准。

传统检验发现，NN4 相对于匹配标普500指数表现显著不佳，而 NN5 相对于全部三个基准均表现显著不佳。去偏检验在这四项比较中均无法拒绝两者相等的假设，这表明这些方法的估计误差过大，无法对这两种策略进行有意义的排序。

图5展示了六项比较中夏普比率差异的正态参照密度；在这些比较中，传统检验与去偏检验在名义双侧5%显著性水平上结论不一致，其中包括四项多空比较和两项择时比较。对于更深的神经网络架构 NN4 和 NN5，去偏参照密度明显宽于相应的传统参照密度。

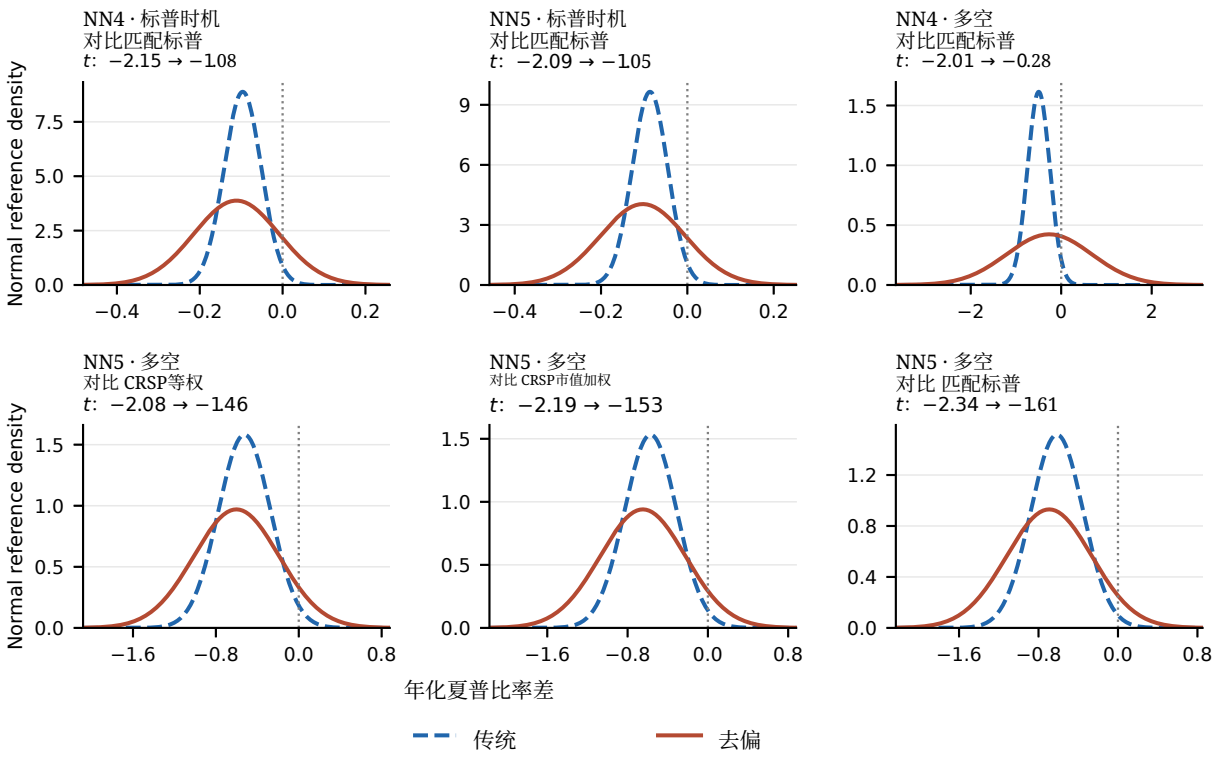



图*5.* 被动基准的夏普差异推断。各面板展示了六组比较，在这些比较中，传统检验与去偏检验在名义双侧5%显著性水平下结论不一致。多空规则使用 $\tau = 0$。001。每个高斯参考密度均以年化策略与基准夏普比率差估计为中心，标准差等于根据联合得分计算的相应 Newey–West 标准误。竖线表示差异为零。

表6*t*基于被动基准的多空夏普差异统计量（$\tau = 0.001$）

| | CRSP等权 | | | CRSP市值加权 | | | 观测面板标准普尔 | | |
|---|---|---|---|---|---|---|---|---|---|
| 方法 | 常规 | Deb. | $\Delta t$ | 常规 | Deb. | $\Delta t$ | 常规 | Deb. | $\Delta t$ |
| OLS-3 | 0.88 | 0.79 | −0.08 | 0.76 | 0.67 | −0.08 | 0.61 | 0.53 | −0。08 |
| ENet | 3.58 | 3.83 | +0.25 | 3.67 | 3.94 | +0。27 | 3.59 | 3.86 | +0。27 |
| PCR | 3.31 | 3.46 | +0。16 | 3.26 | 3.42 | +0.16 | 3.17 | 3.32 | +0.15 |
| PLS | 3.88 | 4.05 | +0。17 | 3.93 | 4.13 | +0。20 | 3.87 | 4.07 | +0。20 |
| NN1 | 5.43 | 5.65 | +0.22 | 5.59 | 5.80 | +0.21 | 5.52 | 5.72 | +0.20 |
| NN2 | 2.56 | 3.67 | +1.11 | 2.56 | 3.73 | +1。18 | 2.49 | 3.68 | +1。19 |
| NN3 | 1.41 | 1.55 | +0。14 | 1.29 | 1.48 | +0。19 | 1.17 | 1.34 | +0。17 |
| NN4 | −1。66 | −0。18 | +1。48 | −1.87 | −0。23 | +1。64 | −2。01 | −0。28 | **+1.73** |
| NN5 | −2。08 | −1.46 | **+0.61** | −2。19 | −1.53 | **+0.66** | −2.34 | −1。61 | **+0.72** |

注： 常规和去偏分别报告零策略减基准夏普比率差的常规和去偏 *t*统计量。标准误采用策略与基准联合的 Newey–West 协方差，并使用五阶滞后。 $\Delta t$ 表示描述性变化，即去偏值减去常规值，计算时未进行四舍五入。粗体变化表示双侧 5% 检验得出了不同的决策。

# 5 结论

我们为基于估计收益信号构建的市场择时和多空投资组合开发了夏普比率推断方法。针对投资组合矩的 Neyman 正交校正可计入估计误差，而按时间顺序交叉拟合则适用于存在时间依赖性的情况。在大型资产面板数据中，只要学习率满足适当条件，由此得到的置信区间便适用于夏普比率及其与观测基准的差值。

模拟结果说明，较大的横截面并不一定能够消除估计效应，也展示了去偏如何改善夏普比率推断的校准。在实证应用中，传统点估计与去偏点估计通常较为相近，但二者的标准误和显著性结论可能存在显著差异。估计得到的正态参考分布使这些差异清晰可见，而无需将比较简化为检验是否跨越显著性阈值。结果支持在评估投资组合表现时，对估计信号加以考虑，同时结合传统程序所度量的收益变动。

# 参考文献

Barillas, F.、R. Kan、C. Robotti 和 J. Shanken（2020）：“基于夏普比率的模型比较”，《金融与定量分析杂志》，55，1840–1874。Bianchi, D.、M. Büchner 和 A. Tamoni（2021）：“利用机器学习研究债券风险溢价”，《金融研究评论》，34，1046–1089。Brown, C.（2025）：“相关数据下深度神经网络的统计性质”，arXiv:2410.11113v3。Campbell, J. Y. 和 S. B. Thompson（2008年）：“样本外预测股票超额收益：有什么能够战胜历史平均值吗？”，《金融研究评论》，21，1509–1531。Chen, C. Y.-H.、A. Lioui 和 O. Scaillet（2026年）：“Green Silence：样本选择偏差下碳排放的双重机器学习”，瑞士金融研究院研究论文第25-66号，2026年4月修订版。Chen, L.、M. Pelger 和 J. Zhu（2024）：“资产定价中的深度学习”，《管理科学》，70，714–750。Chernozhukov, V.、D. Chetverikov、M. Demirer、E. Duflo、C. Hansen 和 W. Newey（2017）：“处理效应的双重/去偏/Neyman机器学习”，《美国经济评论》，107，261–265。Chernozhukov, V.、D. Chetverikov、M. Demirer、E. Duflo、C. Hansen、W. Newey 和 J. Robins（2018）：“处理效应与结构参数的双重/去偏机器学习”，《计量经济学杂志》，21，C1–C68。Farrell, M. H.、T. Liang 和 S. Misra（2021）：“用于估计与推断的深度神经网络”，《计量经济学》，89，181–213。Feng, G.、S. Giglio 和 D. Xiu（2020）：“驯服因子动物园：新因子检验”，《金融学杂志》，75，1327–1370。Freyberger, J.、A. Neuhierl 和 M. Weber（2020）：“非参数地剖析特征变量”，《金融研究评论》，33，2326–2377。Gagliardini, P. 和 H. Ma（2026年）：“条件潜在因子的双重机器学习推断”，瑞士金融研究院研究论文第19-65号，2026年5月修订版。Gu, S.、B. Kelly 和 D. Xiu（2020）：“通过机器学习进行实证资产定价”，《金融研究评论》，33，2223–2273。Hellum, O.、T. I. Jensen、B. Kelly 和 L. H. Pedersen（2025）：“The power of the

共同任务框架”，SSRN工作论文第5242901号。Jobson, J. D. 和 B. M. Korkie（1981）：“使用夏普比率和特雷诺测度进行绩效假设检验”，《金融学杂志》，36，889–908。Kan, R.、X. Wang 和 X. Zheng（2024）：“多因子资产定价模型的样本内和样本外夏普比率”，《金融经济学杂志》，155，103837。Kan, R. 和 G. Zhou（2007）：“参数不确定性下的最优投资组合选择”，《金融与定量分析杂志》，42，621–656。Kelly, B.、S. Malamud 和 K. Zhou（2024）：“收益预测中复杂性的价值”，《金融学杂志》，79，459–503。Kelly, B.、S. Pruitt 和 Y. Su（2019）：“特征变量就是协方差：风险与收益的统一模型”，《金融经济学杂志》，134，501–524。Kelly, B. 和 D. Xiu（2023）：“金融机器学习”，《金融基础与趋势》，13，205–363。Kozak, S.、S. Nagel 和 S. Santosh（2020）：“收缩横截面”，《金融经济学杂志》，135，271–292。Ledoit, O. 和 M. Wolf（2008）：“基于夏普比率的稳健绩效假设检验”，《实证金融杂志》，15，850–859。Leippold, M.、Q. Wang 和 W. Zhou（2022）：“中国股票市场中的机器学习”，《金融经济学杂志》，145，64–82。Liao, Y.、X. Ma、A. Neuhierl 和 L. Schilling（2025）：“资产定价中机器学习预测的不确定性”，arXiv:2503.00549。Lo, A. W.（2002）：“夏普比率的统计学”，《金融分析师杂志》，58，36–52。Memmel, C.（2003）：“基于夏普比率的绩效假设检验”，《金融快报》，1，21–23。Newey, W.（1990）：“半参数效率界”，《应用计量经济学杂志》，5，99–135。Newey, W. 和 K. D. West（1987）：“一种简单、正半定的异方差与自相关一致协方差矩阵”，《计量经济学》，55，703–708。Quaini, A.（2026）：“机器学习预测的正交推断”，SSRN工作论文第7687819号。Rapach, D. E.、J. K. Strauss 和 G. Zhou（2010）：“样本外股权溢价预测：组合预测及其与实体经济的联系”，《金融评论》

研究》，23，821–862。Shen, Z. 和 D. Xiu（2026）：“机器能学习弱信号吗？”，芝加哥大学布斯商学院研究论文第24-03号，修订于2026年11月18日。Van der Vaart, A. W.（2000）：《渐近统计学》，剑桥统计与概率数学丛书，剑桥大学出版社。Vershynin, R.（2018）：《高维概率：数据科学中的引言及其应用》，第47卷，剑桥统计与概率数学丛书，剑桥大学出版社。Welch, I. 和 A. Goyal（2008）：“股票风险溢价预测实证表现的全面考察”，《金融研究评论》，21，1455–1508。White, H.（2001）：《计量经济学家的渐近理论》，圣迭戈：学术出版社，修订版。

# 补充材料

**补充附录：** 补充附录包含技术细节和证明、更多理论结果，以及补充的模拟和实证分析。

**S1 技术细节 S.3**

S1.1 符号说明.S.3 S1.2 横截面因子依赖下的夏普比率展开 . . . . . S.3 S1.3 标准得分函数并非 Neyman正交.S.4 S1.4 样本外估计方案.S.4 S1.5 协方差和方差设定.S.5 S1.5.1 Oracle设定.S.5 S1.5.2 时间序列 Oracle设定 . . . . . S.6 S1.5.3 协方差估计.S.7 S1.6 夏普比率差异估计量的渐近正态性 . . . . . S.8 S1.7 半参数效率.S.8 S1.7.1 验证 MT 和平滑 LS 投资组合的条件 (E1) - (E3).S.13 S1.8 时间序列推断.S.13

**S2 证明** **S.15**

S2.1 命题1.S.15 S2.2 定理 1.S.15 S2.3 引理 1.S.17 S2.4 定理 S1.1.S.19 S2.5 引理 2.S.20

**S3 中间引理** **S.22**

**S4 附加结果** **S.26**

S4.1 其他时间序列结果.S.26 S4.2 依赖于训练的经验过程可忽略性.S.26 S4.3 合并神经网络收敛率.S.34

**S5补充模拟结果 S.37**

S5.1 神经网络实现.S.37 S5.1.1 OLS初始化.S.37 S5.1.2 优化与验证.S.38 S5.2 总体夏普比率.S.38 S5.3 $\tau$ 其他取值的结果.S.39 S5.4 全网格检验统计量直方图. S.49

**S6其他实证结果 S.51**

S6.1 预测调优.S.51 S6.2 被动基准与夏普比率比较推断 . . . . . S.51 S6.3 多空带宽敏感性.S.53

# 《机器学习后的夏普比率推断》补充附录

作者：Sander Barendse & Ming Cheng
本版本：2026年11月20日

# S1 技术细节

## S1.1 符号说明

我们沿用正文中的符号，并采用以下补充约定。对于矩阵 $\mathrm{A} = (a_{ij})$，记 $\|\mathrm{A}\|_{\max} = \max_{i,j} |a_{ij}|$；对于对称矩阵，$\mathrm{A} \succeq \mathrm{B}$ 表示 $\mathrm{A} - \mathrm{B}$ 是半正定的。我们用 $Q \ll P$ 表示绝对连续，$dQ/dP$ 表示拉东–尼科迪姆导数，$P^{\otimes n}$ 表示 $n$ 折乘积概率律，$P_{\mathbf{Z}}$ 表示 $\mathbf{Z}$ 在 $P$ 下的边缘概率律。对于方阵，$\mathrm{tr}(\mathrm{A})$ 表示其迹。对于给定的函数类 $\mathcal{P}$，$X_T = o_P(1)$ 关于 $P \in \mathcal{P}$ 一致成立是指，对每个 $\varepsilon > 0$，$\sup_{P \in \mathcal{P}} P(\|X_T\| > \varepsilon) \to 0$；一致 $O_P(1)$ 表示 $\lim_{M \to \infty} \limsup_{T \to \infty} \sup_{P \in \mathcal{P}} P(\|X_T\| > M) = 0$。正的确定性尺度下的一致阶通过除以该尺度来定义。

## S1.2 横截面因子依赖下的夏普比率展开

考虑例1，将 $C_t$ 替换为

$$C_{i,t} = \sqrt{\rho}\, F_t + \sqrt{1-\rho}\, \varsigma_{i,t}, \qquad \rho \in [0,1],$$

其中，$F_t$ 和 $\varsigma_{i,t}$ 是相互独立的标准正态随机变量；它们在不同日期之间相互独立，并且在特质误差的情况下，在不同资产之间也相互独立。我们保留该例的其余假设，包括 $a_T = c_a T^{-\alpha}$，其中 $1/4 < \alpha < 1/2$，$p/T \to \pi \in (0, 1)$，以及 $N/T \to \nu \in (0, \infty)$，并对样本协方差矩阵施加相同的收敛速率。对于任意固定的 $\rho \in (0, 1]$，，可以证明

$$\sqrt{T}\left(\widehat{\mathrm{SR}}_T^{\mathrm{std}} - \mathrm{SR}_{T,0}\right) = Z_T + \frac{\sqrt{3}\,\rho}{2\sqrt{\lambda}} \sqrt{T}\, a_T + o_p(1),$$
$$\sqrt{T}\left(\widehat{\mathrm{SR}}_T^{\mathrm{DB}} - \mathrm{SR}_{T,0}\right) = Z_T + \frac{29\rho^2}{4\sqrt{2}\lambda} \sqrt{T}\, a_T^2 + o_p(1),$$

其中 $Z_T \rightsquigarrow N(0,\ 3/4)$。因此，特征中的任何固定因子暴露都会使标准估计量保留一阶偏差，而去偏估计量则渐近保持中心化。

## S1.3 标准得分函数并非Neyman正交

标准得分函数 $\psi_T^{\mathrm{std}}$ 不满足Neyman正交性，而这是去偏推断的关键要素（例如参见 Chernozhukov 等人（2017））。事实上，在 $\boldsymbol{\mu}_{T,\,0}$ 处的Gateaux导数取值为

$$\begin{aligned}&\frac{d}{d\kappa}\mathbb{E}\left[\psi_T^{\mathrm{std}}(\mathbf{W}_{T,t},\boldsymbol{\theta},\boldsymbol{\mu}_{T,0}+\kappa\mathbf{h})\right]\Big|_{\kappa=0}\\&\quad=\mathbb{E}\begin{bmatrix}\mathbf{h}(\mathbf{Z}_{T,t})'\mathbf{J}_{T,0}(\mathbf{Z}_{T,t},\boldsymbol{\mu}_{T,0}(\mathbf{Z}_{T,t}))'\,\mathbf{R}_{T,t}\\2\left\{\boldsymbol{\omega}_T(\mathbf{Z}_{T,t},\boldsymbol{\mu}_{T,0}(\mathbf{Z}_{T,t}))'\,\mathbf{R}_{T,t}\right\}\mathbf{h}(\mathbf{Z}_{T,t})'\mathbf{J}_{T,0}(\mathbf{Z}_{T,t},\boldsymbol{\mu}_{T,0}(\mathbf{Z}_{T,t}))'\,\mathbf{R}_{T,t}\end{bmatrix}.\end{aligned}$$

在现实的建模假设下，对于至少某些可容许的h，这个导数通常不为零。

## S1.4 样本外估计方案

我们将样本划分为固定数量 $K$ 的连续折。为便于说明，假设 $T = K\,T_{\mathrm{block}}$，其中折大小为 $T_{\mathrm{block}} \geq 1$，并令

$$I_{T,k} = \{(k-1)T_{\mathrm{block}}+1,\ldots,kT_{\mathrm{block}}\},\qquad k=1,\ldots,K.$$

为减少符号上的繁杂，我们不作此处理；不过，这种划分策略可以放宽，允许各折大小有所不同，只要每折的大小都与 $T$成比例即可。

设固定训练窗口长度满足 $1 \leq K_{\mathrm{train}} \leq K-1$，并定义评估窗口长度和评估折总数如下：

$$T_{\mathrm{eval}} = K_{\mathrm{eval}}T_{\mathrm{block}},\qquad K_{\mathrm{eval}} = K - K_{\mathrm{train}},$$

对于每个评估折，干扰项估计量

$$\widehat{\boldsymbol{\eta}}_T^{(k)} = \left(\widehat{\boldsymbol{\mu}}_T^{(k)},\widehat{\boldsymbol{\Sigma}}_T^{(k)}\right) \in \mathcal{M}_T \times \mathcal{S}_T$$

相对于 h 相对于由其训练集索引的观测值 $I_{T\ ,\ k}^{-}$，其中

三种示例性训练方案如下

$$
\begin{aligned}
I_{T,k}^{-} &= \bigcup_{\ell=1}^{K_{\text{train}}} I_{T,\ell} && \text{(fixed window)},\\
I_{T,k}^{-} &= \bigcup_{\ell=k-K_{\text{train}}}^{k-1} I_{T,\ell} && \text{(moving window)},\\
I_{T,k}^{-} &= \bigcup_{\ell=1}^{k-1} I_{T,\ell} && \text{(expanding window)}.
\end{aligned}
$$

中间估计量针对每一折求解 $k = 1, \dots, K$

$$
\frac{1}{T_{\text{block}}} \sum_{t \in I_{T,k}} \psi_T^{\text{DB}}(\mathbf{W}_{T,t}, \breve{\boldsymbol{\theta}}_T^{\text{DB}}(I_{T,k}), \widehat{\boldsymbol{\eta}}_T(I_{T,k}^{-})) = \mathbf{0}.
$$

由于得分函数关于其目标参数是仿射的,

$$
\breve{\boldsymbol{\theta}}_T^{\text{DB}}(I_{T,k}) = \boldsymbol{\theta}_{T,0}(P) + \frac{1}{T_{\text{block}}} \sum_{t \in I_{T,k}} \psi_T^{\text{DB}}(\mathbf{W}_{T,t}, \boldsymbol{\theta}_{T,0}(P), \widehat{\boldsymbol{\eta}}_T(I_{T,k}^{-})).
$$

$\boldsymbol{\theta}_{T,0}$ 的最终估计量为

$$
\widehat{\boldsymbol{\theta}}_T^{\text{DB}} = \frac{1}{K_{\text{eval}}} \sum_{k \geq K_{\text{train}}+1} \breve{\boldsymbol{\theta}}_T^{\text{DB}}(I_{T,k}).
$$

## S1.5 协方差与方差设定

### S1.5.1 Oracle设定

Let

$$
\begin{aligned}
\mathbf{V}_{T,0} &:= \mathbb{E}_P\left[\psi_T^{\text{DB}}\left(\mathbf{W}_{T,t}, \boldsymbol{\theta}_{T,0}, \boldsymbol{\eta}_{T,0}\right)\psi_T^{\text{DB}}\left(\mathbf{W}_{T,t}, \boldsymbol{\theta}_{T,0}, \boldsymbol{\eta}_{T,0}\right)'\right],\\
\mathbf{B}_{T,0} &:= \mathbb{E}_P\left[\psi_T^{\text{bench}}\left(\mathbf{W}_{T,t}, \boldsymbol{\tau}_{T,0}\right)\psi_T^{\text{bench}}\left(\mathbf{W}_{T,t}, \boldsymbol{\tau}_{T,0}\right)'\right],\\
\mathbf{C}_{T,0} &:= \mathbb{E}_P\left[\psi_T^{\text{DB}}\left(\mathbf{W}_{T,t}, \boldsymbol{\theta}_{T,0}, \boldsymbol{\eta}_{T,0}\right)\psi_T^{\text{bench}}\left(\mathbf{W}_{T,t}, \boldsymbol{\tau}_{T,0}\right)'\right]
\end{aligned}
$$

分别表示策略的协方差矩阵、基准的协方差矩阵和交叉协方差矩阵。以下对这些矩阵以及其他总体方差和协方差的规定可能取决于 $P$。

对于夏普比率的方差，定义

$$\mathbf{d}_{T,\theta} := \nabla g(\boldsymbol{\theta}_{T,0}), \qquad \mathbf{d}_{T,\tau} := \nabla g(\boldsymbol{\tau}_{T,0}), \qquad \mathbf{d}_{T,\Delta} := \begin{bmatrix} \mathbf{d}_{T,\theta} \\ -\mathbf{d}_{T,\tau} \end{bmatrix},$$

其中

$$\nabla g(\mathbf{x}) = \frac{1}{(x_2 - x_1^2)^{3/2}} \begin{bmatrix} x_2 \\ -x_1/2 \end{bmatrix}, \qquad x_2 - x_1^2 > 0.$$

表 S1 提供了与投资组合矩相对应的协方差矩阵，以及与夏普比率和夏普比率差异相对应的方差的 Oracle设定。

表 S1

独立同分布观测下矩和夏普比率推断的 *Oracle*渐近协方差矩阵与方差。

| 目标值 | Oracle设定 |
|---|---|
| 策略矩 $\boldsymbol{\theta}_{T,\,0}$ | $\mathbf{V}_{T,0}$ |
| 扩展矩（$\boldsymbol{\theta}'_T$ $^{0,\ \tau'_T}_{,\ \ ,0}$)′ | $\boldsymbol{\Xi}_{T,0} = \begin{bmatrix} \mathbf{V}_{T,0} & \mathbf{C}_{T,0} \\ \mathbf{C}'_{T,0} & \mathbf{B}_{T,0} \end{bmatrix}$ |
| 夏普比率 $\mathrm{SR}_{T,\,0}$ | $v_{T,\,0,\,\mathrm{SR}} := \mathrm{d}'_{T\ \theta,} \mathbf{V}_{T,0}\mathbf{d}_{T,\theta}$ |
| 夏普比率差 $\Delta\mathrm{SR}_{T,0}$ | $v_{T,\,0,\,\Delta} := 天'_{T\ ,\Delta} \boldsymbol{\Xi}_{T,0}\mathbf{d}_{T,\Delta}$ |

### S1.5.2 时间序列Oracle设定

时间序列对应项将独立同分布协方差矩阵替换为表S1中的长期对应项。

我们首先定义自协方差矩阵

$$\begin{aligned}
\mathbf{V}_{T,0}(h) &:= \mathbb{E}_P\left[\psi_T^{\mathrm{DB}}(\mathbf{W}_{T,t}, \boldsymbol{\theta}_{T,0}, \boldsymbol{\eta}_{T,0})\psi_T^{\mathrm{DB}}(\mathbf{W}_{T,t+h}, \boldsymbol{\theta}_{T,0}, \boldsymbol{\eta}_{T,0})'\right], \\
\mathbf{B}_{T,0}(h) &:= \mathbb{E}_P\left[\psi_T^{\mathrm{bench}}(\mathbf{W}_{T,t}, \boldsymbol{\tau}_{T,0})\psi_T^{\mathrm{bench}}(\mathbf{W}_{T,t+h}, \boldsymbol{\tau}_{T,0})'\right], \\
\mathbf{C}_{T,0}(h) &:= \mathbb{E}_P\left[\psi_T^{\mathrm{DB}}(\mathbf{W}_{T,t}, \boldsymbol{\theta}_{T,0}, \boldsymbol{\eta}_{T,0})\psi_T^{\mathrm{bench}}(\mathbf{W}_{T,t+h}, \boldsymbol{\tau}_{T,0})'\right], \\
\mathbf{D}_{T,0}(h) &:= \mathbb{E}_P\left[\psi_T^{\mathrm{bench}}(\mathbf{W}_{T,t}, \boldsymbol{\tau}_{T,0})\psi_T^{\mathrm{DB}}(\mathbf{W}_{T,t+h}, \boldsymbol{\theta}_{T,0}, \boldsymbol{\eta}_{T,0})'\right].
\end{aligned}$$

在零阶滞后时，这些矩阵满足

$$\mathbf{V}_{T,0}(0) = \mathbf{V}_{T,0} \quad \mathbf{B}_{T,0}(0) = \mathbf{B}_{T,0}, \quad \mathbf{C}_{T,0}(0) = \mathbf{D}'_{T,0}(0) = \mathbf{C}_{T,0}.$$

长期对应量为

$$\begin{aligned}
\mathbf{V}^{\mathrm{LR}}_{T,0} &:= \mathbf{V}_{T,0}(0) + \sum_{h=1}^{\infty} \{\mathbf{V}_{T,0}(h) + \mathbf{V}_{T,0}(h)'\}, \\
\mathbf{B}^{\mathrm{LR}}_{T,0} &:= \mathbf{B}_{T,0}(0) + \sum_{h=1}^{\infty} \{\mathbf{B}_{T,0}(h) + \mathbf{B}_{T,0}(h)'\}, \\
\mathbf{C}^{\mathrm{LR}}_{T,0} &:= \mathbf{C}_{T,0}(0) + \sum_{h=1}^{\infty} \{\mathbf{C}_{T,0}(h) + \mathbf{D}_{T,0}(h)'\}.
\end{aligned}$$

### S1.5.3 协方差估计

我们首先讨论长期协方差矩阵的HAC 估计量。IID估计量是其特殊情形。我们处理 $\mathbf{V}^{\mathrm{LR}}_{T\,0}$，的情形。其他情形可通过类似步骤得到。

对于 S1.4节 中概述的交叉拟合设计，令每一折 $k > K_{\mathrm{train}}$ 的自协方差估计对应量为

$$\begin{aligned}
\widehat{\mathbf{V}}_{T,k}(h) := \frac{1}{T_{\mathrm{block}}} \sum_{t=(k-1)T_{\mathrm{block}}+h+1}^{kT_{\mathrm{block}}} & \psi^{\mathrm{DB}}_T\big(\mathbf{W}_{T,t}, \widehat{\boldsymbol{\theta}}_T, \widehat{\boldsymbol{\eta}}_T(I^-_{T,k})\big) \\
& \times \psi^{\mathrm{DB}}_T\big(\mathbf{W}_{T,t-h}, \widehat{\boldsymbol{\theta}}_T, \widehat{\boldsymbol{\eta}}_T(I^-_{T,k})\big)', \qquad h = 0, \ldots, m_T.
\end{aligned}$$

对于预定义滞后长度 $m_T$ 和权重序列 $\{w_{T_{\mathrm{block}},h}\}$，分块HAC估计量为

$$\widehat{\mathbf{V}}^{\mathrm{HAC}}_T := \frac{1}{K_{\mathrm{eval}}} \sum_{k > K_{\mathrm{train}}} \widehat{\mathbf{V}}^{\mathrm{HAC}}_{T,k},$$

其中

$$\widehat{\mathbf{V}}^{\mathrm{HAC}}_{T,k} = \widehat{\mathbf{V}}_{T,k}(0) + \sum_{h=1}^{m_T} w_{T_{\mathrm{block}},h} \{\widehat{\mathbf{V}}_{T,k}(h) + \widehat{\mathbf{V}}_{T,k}(h)'\}.$$

在独立同分布观测下，我们省略所有正滞后项，得到这一特殊情形

$$\widehat{\mathbf{V}}_T := \frac{1}{K_{\text{eval}}} \sum_{k>K_{\text{train}}} \widehat{\mathbf{V}}_{T,k}(0).$$

相同的估计策略给出$\widehat{\mathrm{B}}_T^{\mathrm{HAC}}$、$\widehat{\mathrm{B}}_T$、$\widehat{\mathrm{C}}_T^{\mathrm{HAC}}$和$\widehat{\mathrm{C}}_T$。对表 S1 作相应替换后，我们还可得到$v_{T,0,\mathrm{SR}}$和$v_{T,0,\Delta}$的异方差和自相关一致方差估计量及 IID 方差估计量，分别记为$\widehat{v}_T^{\mathrm{HAC}}$、$\widehat{v}_{T,\mathrm{SR}}$、$\widehat{v}_{T,\mathrm{SR}}^{\mathrm{HAC}}$、$\widehat{v}_{T,\Delta}$和$\widehat{v}_{T,\Delta}$。

## S1.6 夏普比率差异估计量的渐近正态性

**推论 S1.1**（夏普比率差异的去偏估计量的渐近正态性）。设假设 *1—6* 成立。此外，假设以下条件对 $P \in \mathcal{P}$ 且对充分大的 $T$一致成立。

*1.* 已知基准权重满足

$$P\left(\|\boldsymbol{\nu}_{T,0}(\mathbf{Z}_{T,t})\|_\infty \le \frac{C}{N}\right) = 1;$$

*2.* 基准投资组合具有一致的正方差：

$$\tau_{T,0,2} - \tau_{T,0,1}^2 \ge \underline{v}_{\text{bench}} > 0, \qquad \|\boldsymbol{\tau}_{T,0}\|_2 \le C;$$

*3.* 对某些常数 $0 < \underline{v}_\Delta \le v_{T,\Delta} \le \overline{v}_\Delta < \infty$，

$$\underline{v}_\Delta \le v_{T,\Delta} \le \overline{v}_\Delta.$$

那么，

$$\sqrt{T_{\text{eval}}} \frac{\widehat{\Delta\mathrm{SR}}_T^{\mathrm{DB}} - \Delta\mathrm{SR}_{T,0}}{\sqrt{\widehat{v}_{T,\Delta}}} \rightsquigarrow N(0,1).$$

## S1.7 半参数效率

我们讨论独立同分布情形下去偏矩条件的半参数效率。我们遵循 Newey（1990）中的类似论证，并将其扩展到三角阵列。令 $P_T$ 表示一个样本外观测的真实分布。比较类允许独立同分布评估

观测值，其联合分布为 $Q_T$，由以下方式生成：

$$\mathbf{R}_{T,t} = \boldsymbol{\mu}_{T,Q}(\mathbf{Z}_{T,t}) + \boldsymbol{\varepsilon}_{T,Q,t}, \qquad \mathbb{E}_{Q_T}[\boldsymbol{\varepsilon}_{T,Q,t} \mid \mathbf{Z}_{T,t}] = \mathbf{0},$$

满足下文将详细说明的条件。假设1在真实分布 $P_T$ 下成立。我们比较在局部备择 $Q_T \ll P_T$ 下正则的估计量，其中 $Q_T^{\otimes T_{\text{eval}}}$ 与 $P_T^{\otimes T_{\text{eval}}}$ 相邻。只有评估观测分布受到扰动，独立的初始训练分布则保持不变。所有极限均取 $T \to \infty$，且样本分割方案固定。

我们施加以下正则性条件，并将在下文验证这些条件对于固定 $\tau > 0$ 的 MT 和平滑 LS 均成立。

**假设 S1.1**（独立同分布效率的正则性）。 *(E1)* 独立同分布观测且收益有界。 每个观测到的三角阵列行 $\{\mathbf{W}_{T,t} = (\mathbf{R}_{T,t}', \mathbf{Z}_{T,t}')' : t = 1, \ldots, T\}$ 均服从共同分布 $P_T$，$T_{\text{eval}} \to \infty$，且在共同收益支撑集上，$\|\mathbf{R}_{T,t}\|_\infty \le M$，其中常数 $M$ 与 $T$ 无关。

*(E2)* 投资组合变换的一致界。 已知规则在 $[-M, M]^{N_T}$ 的某个邻域内连续可微。存在一个与 $T$ 无关的常数 $C$，使得对于每个相关的 $\mathbf{z}$ 和 $\mathbf{m}$，$\mathbf{m} + \mathbf{v} \in [-M, M]^{N_T}$，

$$\begin{aligned} \|\boldsymbol{\omega}_T(\mathbf{z}, \mathbf{m})\|_1 &\le C, \\ \|\mathbf{J}_T(\mathbf{z}, \mathbf{m})\mathbf{v}\|_1 &\le C\|\mathbf{v}\|_\infty, \\ \|\boldsymbol{\omega}_T(\mathbf{z}, \mathbf{m} + \mathbf{v}) - \boldsymbol{\omega}_T(\mathbf{z}, \mathbf{m}) - \mathbf{J}_T(\mathbf{z}, \mathbf{m})\mathbf{v}\|_1 &\le C\|\mathbf{v}\|_\infty^2. \end{aligned}$$

所有投资组合缩放都包含在 $\boldsymbol{\omega}_T$ 中：收益为 $\boldsymbol{\omega}_T'\mathbf{R}$，权重之外没有额外的因子。

*(E3)* 渐近线性与得分协方差。记

$$\boldsymbol{\psi}_{T,0}(\mathbf{w}) = \boldsymbol{\psi}_T^{\mathrm{DB}}(\mathbf{w}, \boldsymbol{\theta}_{T,0}, \boldsymbol{\eta}_{T,0}), \qquad \mathbf{V}_{T,0} = \mathbb{E}_{P_T}[\boldsymbol{\psi}_{T,0}\boldsymbol{\psi}_{T,0}'].$$

该估计量满足

$$\sqrt{T_{\text{eval}}}\{\widehat{\boldsymbol{\theta}}_T^{\mathrm{DB}} - \boldsymbol{\theta}_{T,0}\} = \frac{1}{\sqrt{T_{\text{eval}}}} \sum_{t \in \mathcal{T}_{\text{eval}}} \boldsymbol{\psi}_{T,0}(\mathbf{W}_{T,t}) + o_P(1), \tag{S1.1}$$

并且 $\lambda_{\min}(\mathbf{V}_{T,0}) \ge c > 0$ 对所有充分大的 $T$ 均成立，其中 $c$ 与 $T$ 无关。协方差上界由*(E1)–(E2)*推出。

在(E1)–(E2)条件下，预言机投资组合收益以及$\psi_{T,0}$的两个成分均一致有界。(E3)中的特征值界也使$\mathrm{V}_T^{-1/2}$，$_0$ $\psi_{T,0}$一致有界。

**命题 S1.1** (变换投资组合的效率)。 **</style}设 为任意已知的投资组合权重函数。在 处假设假设 1 和 2，并满足假设 S1.1。则** $\mathbf{b}\boldsymbol{\theta}_T^{\mathrm{DB}}$ 对于所述比较模型中的前两个神谕投资组合矩是正则且半参数有效的。特别地，对于每个可容许局部替代，

$$\mathbf{V}_{T,0}^{-1/2}\sqrt{T_{\mathrm{eval}}}\{\widehat{\boldsymbol{\theta}}_T^{\mathrm{DB}} - \boldsymbol{\theta}_{T,0}(Q_T)\} \rightsquigarrow N(\mathbf{0}, \mathbf{I}_2).$$

**证明**。 令 $\mathrm{U}_T = \mathrm{V}_T^{-1/2},_0$ $\psi_{T,0}$。因此 $\mathbb{E}_{P_T}\mathrm{U}_T = \mathbf{0}$且 $\mathbb{E}_{P_T}[\mathrm{U}_T\mathrm{U}_T'] = \mathrm{I}_2$。对于概率分布 $Q$，令 $\boldsymbol{\mu}_{T,Q}$和 $\boldsymbol{\Omega}_{T,Q}$ 分别表示其条件均值和条件二阶矩矩阵。其神谕矩在相同的已知投资组合规则中使用 $\boldsymbol{\mu}_{T,Q}$，并在 $Q$下取期望。

步骤 *1*：沿正则参数子模型求导。 对于此计算，固定一行$T$，并取一个经过 $P_T$ 的正则参数子模型 $Q_{T,\rho}$，其密度为 $p_{T,\rho}$，得分函数为$s_T = \partial \log p_{T,\rho}/\partial\rho\big|_0$。这里只需从有界得分函数开始。求导

$$\mathbb{E}_{Q_{T,\rho}}\left[\mathbf{b}(\mathbf{Z})'\{\mathbf{R} - \boldsymbol{\mu}_{T,Q_{T,\rho}}(\mathbf{Z})\}\right] = 0$$

对于任意有界检验函数 $\mathbf{b}$，莱布尼茨积分法则给出

$$\mathbb{E}_{P_T}[\mathbf{b}(\mathbf{Z})'\{\varepsilon_T s_T - \dot{\boldsymbol{\mu}}_{T,s}\}] = 0.$$

因此

$$\dot{\boldsymbol{\mu}}_{T,s}(\mathbf{Z}) = \mathbb{E}_{P_T}[\varepsilon_T s_T \mid \mathbf{Z}].$$

令 $q_{T,\rho}(\mathrm{w}) = \boldsymbol{\omega}_T(\mathrm{z}, \boldsymbol{\mu}_{T,Q_{T,\rho}}(\mathrm{z}))'\mathrm{r}$。对于 $k = 1,2$，再次应用莱布尼茨法则可得

$$\begin{aligned}\frac{d}{d\rho}\theta_{T,0,k}(Q_{T,\rho})\Big|_0 &= \frac{d}{d\rho}\int q_{T,\rho}(\mathbf{w})^k p_{T,\rho}(\mathbf{w})\, d\lambda(\mathbf{w})\Big|_0 \\ &= \mathbb{E}_{P_T}[q_{T,0}^k s_T] + \mathbb{E}_{P_T}[k q_{T,0}^{k-1}\dot{q}_{T,0}],\end{aligned}$$

其中链式法则给出 $\dot{q}_{T,0} = \mathrm{R}'\mathrm{J}_{T,0}\dot{\boldsymbol{\mu}}_{T,s}$。条件 (E1)–(E2) 证明了这些变换的合理性

对于本文所使用的有界子模型。现在对 Z 进行条件化可得

$$\begin{aligned}\dot{\theta}_{T,1} &= \mathbb{E}_{P_T}[q_{T,0}s_T] + \mathbb{E}_{P_T}[\boldsymbol{\mu}'_{T,0}\mathbf{J}_{T,0}\dot{\boldsymbol{\mu}}_{T,s}] \\ &= \mathbb{E}_{P_T}[\psi_{T,0,1}s_T], \\ \dot{\theta}_{T,2} &= \mathbb{E}_{P_T}[q^2_{T,0}s_T] + 2\mathbb{E}_{P_T}[\boldsymbol{\omega}'_{T,0}\boldsymbol{\Omega}_{T,0}\mathbf{J}_{T,0}\dot{\boldsymbol{\mu}}_{T,s}] \\ &= \mathbb{E}_{P_T}[\psi_{T,0,2}s_T].\end{aligned}$$

在真值处，$\boldsymbol{\Omega}_{T,\,0} = \boldsymbol{\mu}_{T,\,0}\boldsymbol{\mu}'_{T,\,0} + \boldsymbol{\Sigma}_{T,\,0}$，因此这些恰好就是我们引入的去偏得分。它们的期望为零。

每个有界零均值函数都是子模型 $dQ_{T,\,\rho}/dP_T = 1 + \rho s_T$ 在足够小的 $|\rho|$ 下的得分。它们的均方闭包为 $L^0_2(P_T)$。因此，这些得分属于模型的切空间，且它们在该空间上的投影等于得分本身。这验证了 Newey（1990）定理 2.2 和 3.1 中的影响函数和投影计算。以下步骤将建立所需的极限结论，其中 $N_T$ 可以随 $T$ 增长。

步骤*2*：关于资产数量的一致界。固定 $T$，取 $Q \ll P_T$，并在本步骤中省略行指数。利用 (E2) 展开投资组合规则，并减去步骤 1 中确定的线性项，得到

$$\|\boldsymbol{\theta}_0(Q) - \boldsymbol{\theta}_0(P) - \mathbb{E}_Q\boldsymbol{\psi}_0\| \le C\mathbb{E}_{Q_\mathbf{z}}\left[\|\boldsymbol{\mu}_Q - \boldsymbol{\mu}_0\|^2_\infty + \|\boldsymbol{\mu}_Q - \boldsymbol{\mu}_0\|_\infty\|\boldsymbol{\Omega}_Q - \boldsymbol{\Omega}_0\|_{\max}\right].$$

下文中，$C$ 与 $T$ 和 $N_T$相互独立。记 $H(P,Q)^2 = \mathrm{fl}\,(\sqrt{dP} - \sqrt{dQ})^2,$，并令$h_\mathbf{z} = H(P(\mathrm{R} \mid \mathbf{z}), Q(\mathrm{R} \mid \mathbf{z}))$。有界收益和柯西–施瓦茨不等式意味着

$$\|\boldsymbol{\mu}_Q(\mathbf{z}) - \boldsymbol{\mu}_0(\mathbf{z})\|_\infty + \|\boldsymbol{\Omega}_Q(\mathbf{z}) - \boldsymbol{\Omega}_0(\mathbf{z})\|_{\max} \le Ch_\mathbf{z}.$$

平方根密度的三角不等式给出

$$\mathbb{E}_{Q_\mathbf{z}}h^2_\mathbf{Z} \le \{H(P,Q) + H(P_\mathbf{Z}, Q_\mathbf{Z})\}^2 \le 4H(P,Q)^2.$$

因此，

$$\|\boldsymbol{\theta}_{T,0}(Q) - \boldsymbol{\theta}_{T,0} - \mathbb{E}_Q\boldsymbol{\psi}_{T,0}\| \le CH(P_T,Q)^2, \tag{S1.2}$$

其中，$C$ 与 $T$ 和 $N_T$相互独立。

步骤 *3*：局部备择下的极限。渐近相邻性意味着 $T_{\mathrm{eval}}H(P_T,\ Q_T)^2 = O(1)$，并且还确保（S1.1）中的可忽略余项在 $Q_T$ 下仍可忽略（参见第 14.5 节及 Van der Vaart（2000）的引理6.4——

节）。结合步骤2和（E3）可得

$$\sqrt{T_{\text{eval}}}\,\mathbf{V}_{T,0}^{-1/2}\{\widehat{\boldsymbol{\theta}}_T^{\text{DB}} - \boldsymbol{\theta}_{T,0}(Q_T)\} = \frac{1}{\sqrt{T_{\text{eval}}}}\sum_t \{\mathbf{U}_T(\mathbf{W}_{T,t}) - \mathbb{E}_{Q_T}\mathbf{U}_T\} + o_{Q_T}(1).$$

由于 $H(P_T,\ Q_T) \to 0$且 $\mathrm{U}_T$一致有界，$\mathrm{Var}_{Q_T}(\mathrm{U}_T) = \mathrm{I}_2 + o(1)$。因此，有界三角阵中心极限定理给出了所述正态极限，并确立了正则性。

步骤 *4*：信息界及其达到。 我们构造一个局部子模型，并验证其满足 Van der Vaart (2000) 的定义7.14所述的局部渐近正态性。对每个固定的 $\mathrm{h} \in \mathbb{R}^2$，考虑子模型

$$\frac{dQ_{T,\mathbf{h}}}{dP_T} = 1 + \frac{\mathbf{h}'\mathbf{U}_T}{\sqrt{T_{\text{eval}}}}.$$

由于 $\mathrm{U}_T$ 有界且中心化，当 $T$充分大时，这些都是正密度。对有界的 h，对数似然展开一致成立：

$$\log \frac{dQ_{T,\mathbf{h}}^{\otimes T_{\text{eval}}}}{dP_T^{\otimes T_{\text{eval}}}} = \mathbf{h}'\frac{1}{\sqrt{T_{\text{eval}}}}\sum_t \mathbf{U}_T(\mathbf{W}_{T,t}) - \tfrac{1}{2}\mathbf{h}'\mathbf{h} + o_P(1).$$

该和收敛到 $N(0,\ \mathrm{I}_2)$，因此这些是极限信息量为 $\mathrm{I}_2$ 的相邻局部备择。此外，$\mathbb{E}_{Q_{T,\,\mathrm{h}}}\mathrm{U}_T = \mathrm{h}/\sqrt{T_{\text{eval}}}$ 精确成立，而步骤2给出

$$\sqrt{T_{\text{eval}}}\,\mathbf{V}_{T,0}^{-1/2}\{\boldsymbol{\theta}_{T,0}(Q_{T,\mathbf{h}}) - \boldsymbol{\theta}_{T,0}\} = \mathbf{h} + o(1).$$

根据 Van der Vaart (2000) 的定理9.3–9.4和命题8.4，任意正则竞争估计量的标准化极限误差与 $\mathrm{G} + \mathrm{B}$ 同分布，其中 $\mathrm{G} \sim N(0,\ \mathrm{I}_2)$，且 B 是与 G 独立的随机向量。因此，只要极限协方差有限，

$$\mathrm{Cov}(\mathbf{G} + \mathbf{B}) = \mathbf{I}_2 + \mathrm{Cov}(\mathbf{B}) \succeq \mathbf{I}_2.$$

步骤 3 中的正态极限表明，去偏估计量达到了这一界限。 ∎

**推论 S1.2**（夏普比率与基准差异）。 设 $\omega_T$ 为任意满足命题 *S1.1* 条件的已知投资组合规则。假设其矩向量 $\theta_{T,\,0}$满足假设 *6*。则相应的去偏夏普比率估计量在同一比较模型中是正则且半参数有效的。在推论 *S1.1* 的附加条件下，对该估计量与已知基准之差也同样成立。方差界序列为相应的影响函数方差。

**证明**。 将引理 1 的一致泰勒论证应用于上述矩展开，并与差分的基准矩联合进行。这给出相应的渐近线性表示，并保留步骤2的一致 $O(H(P_T, Q)^2)$总体余项。标准化的标量影响函数一致有界，因此步骤3–4确立了正则性，并达到了相应的方差界。

■

**S1.7.1 机器翻译和光滑 LS 投资组合的条件 (E1)–(E3) 验证**。

假设 3–5 以及定理 1 证明中的渐近线性展开给出了 (E1) 和 (E3)。引理 S3.1 证明中的投资组合界蕴含 $\mathrm{k}\omega_T\mathrm{k}_1 \le C$ 和 $\mathrm{kJ}_T\mathrm{vk}_1 \le C\mathrm{kvk}_\infty$。MT 泰勒余项为零。对于平滑 LS，引理 S3.1 证明中的对称性、$l_1$ 对偶性和收缩导数界给出

$$\|D\mathbf{J}_T[\mathbf{v}]\mathbf{v}\|_1 \le \frac{C_\tau}{N_T}\|\mathbf{v}\|_2^2 \le C_\tau\|\mathbf{v}\|_\infty^2.$$

泰勒公式给出（E2），对于固定的 $\tau > 0$，该结果对 $T$ 和 $N_T$ 一致成立。因此，命题 S1.1 适用于这两种规则。在假设 6 下，推论 S1.2 适用；对于差分，还需满足推论 S1.1 的附加条件。

## S1.8 时间序列推断

本节将说明，定理 1 中的推断结果可推广至具有绝对正则行的三角阵结构数据。若假设满足定理 S1.1 的条件，并将 IID 方差估计量替换为其 HAC 版本，则引理 1 中关于夏普比率的 Delta 方法结果仍然成立。

**假设 S1.2**（平稳且一致指数 β 混合的行）。 对于每个样本量，行过程

$$\{\mathbf{W}_{T,t} : t \ge 1\}$$

对所有 $P \in \mathcal{P}$ 都严格平稳。定义绝对正则系数

$$\beta_{P,T}(h) = \sup_{s \geq 1} \beta_P\big(\sigma(\mathbf{W}_{T,t} : 1 \leq t \leq s), \sigma(\mathbf{W}_{T,t} : t \geq s + h)\big), \qquad h \geq 1.$$

存在常数 $C_\beta < \infty$ 和 $0 < \rho_\beta < 1$，使得

$$\sup_{P \in \mathcal{P}} \sup_{T \geq 1} \beta_{P,T}(h) \leq C_\beta \rho_\beta^h, \qquad h \geq 1.$$

**Rem假设 S1.1 (**因子模型论证)。 设可测映射 $g_T$，

$$\mathbf{W}_{T,t} = g_T(\mathbf{F}_t, \mathbf{U}_{T,t}),$$

其中，固定维度的因子过程 $\{\mathbf{F}_t\}$ 严格平稳，且以一致于分布类的常数呈指数 $\beta$ 混合。假设创新向量$\mathbf{U}_{T,t}$ 在时间上独立同分布，并且与整个因子过程相互独立，但其同期横截面依赖不作限制。因此，每个行过程 $\{\mathbf{W}_{T,t}\}_t$ 均严格平稳。由于加入独立同分布的创新不会改变因子过程在正滞后处的 $\beta$ 混合系数，且可测变换不会增大这些系数，因此无论$\mathbf{W}_{T,t}$的维度如何，$\mathbf{W}_{T,t}$ 都具有与$\mathbf{F}_t$相同的指数界。

**假设 S1.3 (**时间序列预言协方差正则性)。 对于所有充分大的 $T$，存在常数 $0 < \underline{l} \leq \overline{l} < \infty$，使得

$$\inf_{P \in \mathcal{P}} \lambda_{\min}(\mathbf{V}_T^{\mathrm{LR}}) \geq \underline{\ell}, \qquad \sup_{P \in \mathcal{P}} \lambda_{\max}(\mathbf{V}_T^{\mathrm{LR}}) \leq \overline{\ell}.$$

**假设 S1.4 (**HAC 带宽和滞后权重)。 令 $\delta_T = \frac{r_\mu}{\sqrt{N_T}} + r_\Sigma N_T^{-T}$。确定性带宽 $m_T \in \{0, \ldots, T_{\mathrm{block}} - 1\}$ 以及确定性权重满足

$$m_T \to \infty, \qquad m_T = o(T^{1/3}), \qquad m_T \delta_T \to 0,$$
$$w_{T_{\mathrm{block}},0} = 1, \qquad \sup_T \max_{0 \leq h \leq m_T} |w_{T_{\mathrm{block}},h}| \leq C, \qquad w_{T_{\mathrm{block}},h} \to 1 \quad \textit{for every fixed } h \geq 0.$$

**定理 S1.1** （HAC 学生化的一致时间序列弱收敛）。 保持定理 *1* 的假设不变，将假设 *4* 替换为假设 *S1.2*。此外，假设假设 *S1.3* 和假设 *S1.4* 成立。则

$$(\widehat{\mathbf{V}}_T^{\mathrm{HAC}})^{-1/2} \sqrt{n_T} \{\widehat{\boldsymbol{\theta}}_{T,\tau} - \boldsymbol{\theta}_0(P)\} \rightsquigarrow \mathcal{N}(\mathbf{0}, \mathbf{I}_2) \tag{S1.3}$$

在有界-利普希茨度量下关于 $P \in \mathcal{P}$ 一致收敛。

# S2 证明

本节给出正文和 S1 节中结果的证明。

## S2.1 命题1

合并普通最小二乘得分恒等式给出 $\hat{\beta}_T - \beta_{T,0} = (\sum_{i,t} \mathbf{Z}_t^{(i)} \mathbf{Z}_t^{(i)\prime})^{-1} \sum_{i,t} \mathbf{Z}_t^{(i)} \{a_T H(C_t) + \varepsilon_{i,t}\}$。结合回归变量的总体正交性以及样本格拉姆矩阵的集中性，我们得到 $\hat{\beta}_{C,T} - \sqrt{\lambda} = O_p(a_T/\sqrt{T} + 1/\sqrt{NT})$、$\hat{\beta}_{Z,T} - \sqrt{1-\lambda} = O_p(1/\sqrt{NT})$ 以及 $\|\hat{\beta}_{W,T}\|_2 = O_p(\sqrt{p/(NT)})$，其中下标标识相应的回归变量。

在 $\beta_T$,0 处，代入样本矩具有精确的线性得分表示，因此独立同分布三角阵中心极限定理和一阶德尔塔法可应用于极限矩 $(\lambda, 3\lambda^2)'$（见定理 1 和引理 1 的证明）。因此，$Z_T = -(4\sqrt{2T})^{-1} \sum_{t=T+1}^{2T} (C_t^4 - 6C_t^2 + 3)$，其极限方差为 3/4。恒等式 $\mathbb{E}[\overline{C_t} H(C_t)] = 0$、$\mathbb{E}[C_t^3 H(C_t)] = \sqrt{6}$ 以及 $\mathbb{E}[C_t^2 H(C_t)^2] = 7$ 给出矩偏差 $(O_p(a_T^2),\ -2\sqrt{6}\lambda^{3/2} a_T + O_p(a_T^2))'$ 以及去偏矩偏差 $(-a_T^2, -35\lambda a_T^2)' + o_p(a_T^2)$。应用 $\nabla g(\lambda, 3\lambda^2) = (3/(2\sqrt{2\lambda}), -1/(4\sqrt{2\lambda^2}))'$ 并使用 $\sqrt{T}\, a_T^2 \to 0$ 即可得到结果。

## S2.2 定理 1

我们引入以下简化符号说明。

$$
\begin{aligned}
\boldsymbol{\xi}_{T,0,t} &:= \psi_T^{\mathrm{DB}}(\mathbf{W}_{T,t}, \boldsymbol{\theta}_{T,0}, \boldsymbol{\eta}_{T,0}), \\
\mathbf{D}_{T,k}(\mathbf{w}) &:= \psi_T^{\mathrm{DB}}(\mathbf{w}, \boldsymbol{\theta}_{T,0}, \widehat{\boldsymbol{\eta}}_T^{(k)}) - \psi_T^{\mathrm{DB}}(\mathbf{w}, \boldsymbol{\theta}_{T,0}, \boldsymbol{\eta}_{T,0}), \\
\mathbf{D}_{T,k,t} &:= \mathbf{D}_{T,k}(\mathbf{W}_{T,t}), \\
\overline{\mathbf{D}}_{T,k} &:= \mathbb{E}_P[\mathbf{D}_{T,k}(\mathbf{W}_T)],
\end{aligned}
$$

其中，最后一行中的期望是相对于新观测 $\mathbf{W}_T$ 的边际分布而言的。

步骤 *1*（预言机中心极限定理）。条件收益模型意味着两个预言机校正项的条件期望均为零，从而有 $\mathbb{E}_P[\boldsymbol{\xi}_{T,0,t}] = \mathbf{0}$。此外，在假设 4- 5 下，得分函数独立同分布，且我们有 $\left\| \mathbf{V}_T^{-1/2,0} \right\|_{\mathrm{op}} \le c^{-1/2}$，在 $\mathcal{P}$ 上一致成立。此外，引理 S3.2 表明李雅普诺夫条件成立，因此林德伯格条件也得到满足。

因此，三角阵中心极限定理（Van der Vaart（2000）中的定理 2.27，结合 Cramér-Wold 设备）给出

$$\mathbf{V}_{T,0}^{-1/2}\frac{1}{\sqrt{T_{\text{eval}}}}\sum_{k>K_{\text{train}}}\sum_{t\in I_{T,k}}\boldsymbol{\xi}_{T,0,t}\rightsquigarrow N(\mathbf{0},\mathbf{I}_2). \tag{S2.1}$$

对所有 $P\in\mathcal{P}$ 均成立。一致假设意味着沿着每个可能漂移的序列 $\{P_T\}$ 都满足中心极限定理，从而序列刻画得到一致收敛。

步骤*2*（分解与中心化余项）。该估计量具有如下分解

$$\begin{aligned}&\sqrt{T_{\text{eval}}}\{\widehat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_{T,0}\}\\&\quad=\frac{1}{\sqrt{T_{\text{eval}}}}\sum_{k>K_{\text{train}}}\sum_{t\in I_{T,k}}\boldsymbol{\xi}_{T,0,t}+\frac{1}{\sqrt{T_{\text{eval}}}}\sum_{k>K_{\text{train}}}\sum_{t\in I_{T,k}}(\mathbf{D}_{T,k,t}-\overline{\mathbf{D}}_{T,k})+\frac{T_{\text{block}}}{\sqrt{T_{\text{eval}}}}\sum_{k>K_{\text{train}}}\overline{\mathbf{D}}_{T,k}.\end{aligned}$$

为控制中心化余项，定义 $\mathrm{S}_{T,\,k}:=\frac{1}{\sqrt{T_{\text{eval}}}}\sum_{t\in I_{T,\,k}}(\mathrm{D}_{T,\,k,\,t}-\mathrm{D}_{T,\,k})$，以及事件$A_{T,\,k}:=\{\left\|\mathrm{D}_{T,\,k}(\mathrm{W}_T)\right\|_{P,\,2}\le\delta_T\}$。每个事件都可测于其所在折对应的训练σ-域。以训练数据为条件时，评估观测是独立同分布的，并且独立于所估计的干扰参数；在 $A_{T,\,k}$ 上，我们有

$$\mathbb{E}_P[\left\|\mathbf{S}_{T,k}\right\|_2^2]\le\frac{T_{\text{block}}}{T_{\text{eval}}}\left\|\mathbf{D}_{T,k}(\mathbf{W}_T)\right\|_{P,2}^2.$$

根据引理 S3.2 给出的界，条件切比雪夫不等式（这里指将切比雪夫不等式应用于条件概率，以及应用 Chernozhukov 等人（2017）中的引理 2.1）、三角不等式和并集界可知，对于任意 $\varepsilon>0$，有

$$\sup_{P\in\mathcal{P}}P\left(\left\|\sum_{k>K_{\text{train}}}\mathbf{S}_{T,k}\right\|_2>\varepsilon\right)=o(1).$$

步骤 *3*（偏差、逼近和结论）。注意，

$$\left\|\frac{T_{\text{block}}}{\sqrt{T_{\text{eval}}}}\sum_{k>K_{\text{train}}}\overline{\mathbf{D}}_{T,k}\right\|_2\le\sqrt{T_{\text{eval}}}\max_{k>K_{\text{train}}}\left\|\overline{\mathbf{D}}_{T,k}\right\|_2\le\sqrt{T}\max_{k>K_{\text{train}}}\left\|\overline{\mathbf{D}}_{T,k}\right\|_2.$$

鉴于引理 S3.1 中的界，利用条件切比雪夫不等式和并集界可得，对于

任意 $\varepsilon > 0$，有

$$\sup_{P\in\mathcal{P}} P\left(\left\|\frac{T_{\text{block}}}{\sqrt{T_{\text{eval}}}}\sum_{k>K_{\text{train}}}\overline{\mathbf{D}}_{T,k}\right\|_2 > \varepsilon\right) = o(1).$$

步骤 *4*（一致协方差估计与学生化）。这可通过采用与时间序列证明（定理 S1.1）中步骤 4 和步骤 5 类似的步骤得到，只需从 HAC 估计量中去除所有自协方差（设为 $m_T = 0$）。

## S2.3 引理 1

以下所有随机界和极限均在 $P \in \mathcal{P}$ 上一致成立。定理 1 的证明给出

$$\begin{aligned}\sqrt{T_{\text{eval}}}\left\|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_{T,0}\right\|_2 &= O_P(1),\\ \left\|\widehat{\mathbf{V}}_T - \mathbf{V}_{T,0}\right\|_{\text{op}} &= o_P(1),\\ \mathbf{V}_{T,0}^{-1/2}\sqrt{T_{\text{eval}}}\{\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_{T,0}\} &\rightsquigarrow N(\mathbf{0}, \mathbf{I}_2).\end{aligned}$$

步骤 *1*（一致泰勒展开）。假设 6 将所有向量 $\boldsymbol{\theta}_{T,\,0}$ 限制在 $\mathcal{D} = \{\boldsymbol{\theta} :\ \theta_2 - \theta_1^2 > 0\}$ 的一个紧子集中。因此，存在一个 $\epsilon_0 > 0$，使得它们的闭 $\epsilon_0$-邻域包含于 $\mathcal{D}$ 中，并且 $g$ 的梯度和海森矩阵在该邻域内有界。随着概率趋于 1，估计量及其与目标值之间的线段一致地落在该邻域内。因此，根据泰勒定理，余项有如下界：

$$\begin{aligned}&\sqrt{T_{\text{block}}}\left|g(\widehat{\boldsymbol{\theta}}_T) - g(\boldsymbol{\theta}_{T,0}) - \nabla g(\boldsymbol{\theta}_{T,0})'(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_{T,0})\right|\\ &\quad\le C\sqrt{T_{\text{block}}}\left\|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_{T,0}\right\|_2^2 = O_P(T_{\text{block}}^{-1/2}) = o_P(1).\end{aligned}$$

此外，海森矩阵有界保证

$$\left\|\nabla g(\widehat{\boldsymbol{\theta}}_T) - \nabla g(\boldsymbol{\theta}_{T,0})\right\|_2 = o_P(1).$$

步骤*2*（标量方差一致性和非退化性）。预言机梯度 $g(\boldsymbol{\theta}_{T,\,0})$一致有界，且有界于远离零的范围之外。对于后一个主张，其第一分量

满足

$$\frac{\theta_{T,0,2}}{\{\theta_{T,0,2}-\theta_{T,0,1}^2\}^{3/2}} \geq \frac{1}{\sqrt{\theta_{T,0,2}-\theta_{T,0,1}^2}},$$

该量由有界的神谕矩保证有一个远离零的下界。$\mathbf{V}_{T,\,0}$ 的特征值上下界因此意味着存在常数 $0 < v_* \leq v_* < \infty$ 使得

$$\underline{v}_* \leq v_{T,0,\mathrm{SR}} \leq \overline{v}_*$$

对所有充分大的 $T$均一致成立。在步骤 1 得到的共同邻域事件上，我们可以展开

$$|\widehat{v}_{T,\mathrm{SR}} - v_{T,0,\mathrm{SR}}| \leq C\left\|\widehat{\mathbf{V}}_T - \mathbf{V}_{T,0}\right\|_{\mathrm{op}} + C\left\|\nabla g(\widehat{\boldsymbol{\theta}}_T) - \nabla g(\boldsymbol{\theta}_{T,0})\right\|_2.$$

右侧的两项均一致为 $o_P(1)$。由于补事件的概率一致趋于零，

$$\widehat{v}_{T,\mathrm{SR}} - v_{T,0,\mathrm{SR}} = o_P(1)$$

对 $P \in \mathcal{P}$ 一致成立，并且意味着相对方差一致性。此外，估计标量方差为非正的概率一致趋于零。因此，

$$\sqrt{T_{\mathrm{eval}}}\{\widehat{\mathrm{SR}}_T^{\mathrm{DB}} - \mathrm{SR}_{T,0}\} = \nabla g(\boldsymbol{\theta}_{T,0})'\sqrt{T_{\mathrm{eval}}}(\widehat{\boldsymbol{\theta}}_T^{\mathrm{DB}} - \boldsymbol{\theta}_{T,0}) + o_P(1) \tag{S2.2}$$

对 $P \in \mathcal{P}$一致成立。

*Step 3*（投影与学生化）。考虑确定性的单位向量 or

$$\mathbf{c}_T(P) := \frac{\mathbf{V}_{T,0}^{1/2}\nabla g(\boldsymbol{\theta}_{T,0})}{\sqrt{v_{T,0,\mathrm{SR}}}}.$$

（S2.2）中的线性项除以 $\sqrt{v_{T,\,0,\,\mathrm{SR}}}$，等于

$$\mathbf{c}_T(P)'\mathbf{V}_{T,0}^{-1/2}\sqrt{T_{\mathrm{eval}}}\{\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_{T,0}\}.$$

对于任意标量有界利普希茨检验函数，与该单位向量投影复合后得到的是一个具有相同界的向量有界利普希茨检验函数。根据 Portmanteau 引理（例如 Van der Vaart（2000）中的引理 2.2），一致向量中心极限定理因此

意味着一致标量标准正态极限，尽管 $c_T(P)$ 可能随 $T$ 和 $P$ 变化。标准化后，泰勒余项可以忽略不计，因为标量方差有正的下界。相对方差一致性与一致斯卢茨基定理现在给出最终结果。

## S2.4 定理 S1.1

我们按照定理 1 的证明进行。回顾该证明中使用的简化符号$\xi_{T,\tau,t}$、$D_{T,k}$和 $D_{T,k}$，其定义完全相同。所有常数和随机界对 $P \in \mathcal{P}$ 和 $T$ 均一致。对 $k$ 的求和遍历 $K_{\text{train}}+1, \ldots, K$。

步骤 *1*（预言混合中心极限定理）。定理 1 中的步骤 1 仍然成立，只需将 IID 三角阵列中心极限定理替换为适用于具有绝对正则行的三角阵列的对应结果。我们使用伍尔德里奇–怀特中心极限定理（参见怀特（2001）中的定理 5.20）。由于 $\beta$-混合蕴含具有相同速率界的$\alpha$-混合，假设 S1.2 与其他施加的假设一起确保满足中心极限定理的条件。一致性可由与前面相同的论证得出。

步骤 *2*（估计量分解与中心化干扰项余项）。回顾定理 1 证明中步骤 2 的估计量分解。应用引理 S4.1，我们可在假设 S1.2 下得到相同结论，而非假设 4。事实上，将该引理逐分量应用于

$$
\begin{aligned}
f_{T,k}(\mathbf{w}) = \psi_T^{\mathrm{DB}}\big(\mathbf{w}, \boldsymbol{\theta}_{T,0}, \widehat{\boldsymbol{\eta}}_T^{(k)}\big) - \psi_T^{\mathrm{DB}}\big(\mathbf{w}, \boldsymbol{\theta}_{T,0}, \boldsymbol{\eta}_{T,0}\big) \\
- \mathbb{E}_P\left[\psi_T^{\mathrm{DB}}\big(\mathbf{W}_T, \boldsymbol{\theta}_{T,0}, \widehat{\boldsymbol{\eta}}_T^{(k)}\big) - \psi_T^{\mathrm{DB}}\big(\mathbf{W}_T, \boldsymbol{\theta}_{T,0}, \boldsymbol{\eta}_{T,0}\big)\right],
\end{aligned}
$$

其中总体期望将拟合干扰项固定不变，得到

$$
\sup_{P \in \mathcal{P}} P\left(\left\|\mathbf{S}_{T,k}\right\|_2 > \varepsilon\right) = o(1)
$$

对每个 $\varepsilon > 0$ 以及每个 $k = K_{\text{train}}+1$、……、$K$均成立。由于 $K$ 固定，并集界和三角不等式给出了对评估折求和的相同结论。

步骤 *3*（偏差）。定理 1 证明中的步骤 3 可原样沿用。

步骤 *4*（一致 *HAC* 一致性）。我们应用引理 S4.2 来

$$
f_{T,\mathbf{a}}(\boldsymbol{\eta}, \mathbf{w}) := \mathbf{a}'\{\psi_T^{\mathrm{DB}}\big(\mathbf{w}, \boldsymbol{\theta}_{T,0}, \boldsymbol{\eta}\big) + \boldsymbol{\theta}_{T,0}\},
$$

由于 $\psi_T^{\mathrm{DB}}$ 的特殊仿射设定，它与 $\boldsymbol{\theta}_{T,\,0}$ 相互独立。这里，$\mathrm{a}$ 遍历坐标向量及其和。相应的全样本估计量为$\widehat{\ }\theta_{T,\,\mathrm{a}} := \mathrm{a}'\widehat{\ }\boldsymbol{\theta}_T^{\mathrm{DB}}$。在既定条件下，$f_{T,\,\mathrm{a}}(\boldsymbol{\eta},\ \mathrm{W}_{T,\,t})$ 的均值在 $\boldsymbol{\eta} = \boldsymbol{\eta}_{T,\,0}$ 处为零，其边际 $4+\nu$ 范数有界，且评估区块上的拟合二阶矩有界。最后，以至少 $1 - K_T = 1 - \Delta_T$ 的概率，边际干扰误差的平方至多为 $\kappa_T = \mathrm{kak}_2^2\,\delta_T$。因此，鉴于对 $m_T$ 的速率限制以及 $\sqrt{T}$ 速率$\widehat{\ }\boldsymbol{\theta}_T^{\mathrm{DB}} - \boldsymbol{\theta}_{T,\,0}$，引理 S4.2 对每个 $\varepsilon > 0$ 及所考虑的每个 $\underline{\mathrm{a}}$ 给出：

$$\sup_{P\in\boldsymbol{\mathcal{P}}} P\left(\left|\mathbf{a}'(\widehat{\mathbf{V}}_{T,k}^{\mathrm{HAC}} - \mathbf{V}_T^{\mathrm{LR}})\mathbf{a}\right| > \varepsilon\right) \longrightarrow 0,$$

对每个折 $k = K_{\mathrm{train}}+1, \ldots, K$。由于 $K$ 是固定的，我们由并集界和三角不等式得到

$$\sup_{P\in\boldsymbol{\mathcal{P}}} P\left(\left|\mathbf{a}'(\widehat{\mathbf{V}}_T^{\mathrm{HAC}} - \mathbf{V}_T^{\mathrm{LR}})\mathbf{a}\right| > \varepsilon\right) \longrightarrow 0,$$

标量一致性和极化给出了 $\left\|\widehat{\ }\mathrm{V}_T^{\mathrm{HAC}} - \mathrm{V}_T^{\mathrm{LR}}\right\|_{\mathrm{op}}$ 的相应一致收敛。

步骤 *5*（学生化）。$\widehat{\ }\mathrm{V}_T^{\mathrm{HAC}}$ 的一致一致性与假设 S1.3 蕴含

$$\sup_{P\in\boldsymbol{\mathcal{P}}} P\left(\lambda_{\min}(\widehat{\mathbf{V}}_T^{\mathrm{HAC}} - \mathbf{V}_T^{\mathrm{LR}}) < \underline{\ell}/2\right) \to 1.$$

逆平方根映射在一致正定矩阵上的连续性，结合一致斯卢茨基定理，给出了最终结果。

## S2.5 引理 2

对 $\widehat{\ }\boldsymbol{\mu}_T$ 的界直接由网络输出限制得到。应用定理 S4.1 可得，以至少 $1 - \exp\left\{-T^{p/(\beta+p)}\right.$ 块 $\left.\log^7 T_{\mathrm{block}}\log(NT_{\mathrm{block}})\right\}$ 的概率，

$$\frac{1}{N}\left\|\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_{T,0}\right\|_{P,2}^2 \le q_T, \tag{S2.3}$$

$$\frac{1}{NT_{\mathrm{block}}}\sum_{t=1}^{T_{\mathrm{block}}}\left\|\widehat{\boldsymbol{\mu}}_T(\mathbf{Z}_{T,t}) - \boldsymbol{\mu}_{T,0}(\mathbf{Z}_{T,t})\right\|_2^2 \le q_T. \tag{S2.4}$$

在既定条件下，误差 $\boldsymbol{\varepsilon}_{T,\,t}$ 是中心化的，且一致有界。鉴于其坐标独立同分布，我们可得到该向量具有定理 4.7.1 所要求的次高斯性

以及 Vershynin（2018）中的习题 4.7.3。令

$$\mathbf{U}_T := \frac{1}{T_{\mathrm{block}}} \sum_{t=1}^{T_{\mathrm{block}}} \varepsilon_{T,t}\varepsilon_{T,t}', \qquad \bar{\varepsilon}_T^0 := \frac{1}{T_{\mathrm{block}}} \sum_{t=1}^{T_{\mathrm{block}}} \varepsilon_{T,t},$$

$$\widetilde{\mathbf{\Sigma}}_T := \frac{1}{T_{\mathrm{block}}} \sum_{t=1}^{T_{\mathrm{block}}} (\varepsilon_{T,t} - \bar{\varepsilon}_T^0)(\varepsilon_{T,t} - \bar{\varepsilon}_T^0)' = \mathbf{U}_T - \bar{\varepsilon}_T^0 \bar{\varepsilon}_T^{0\prime}.$$

条件收益模型给出 $\mathbb{E}_P[\varepsilon_{T,t}\varepsilon_{T,t}'] = \mathbf{\Sigma}_{T,\,0}$。Vershynin（2018）中习题 4.7.3 针对次高斯向量的高概率协方差界，在尾部参数$N$下给出

$$\|\mathbf{U}_T - \mathbf{\Sigma}_{T,0}\|_{\mathrm{op}} \le C \left\{ \sqrt{\frac{N}{T_{\mathrm{block}}}} + \frac{N}{T_{\mathrm{block}}} \right\}$$

以至少 $1 - 2\exp(-N)$ 的概率成立，其中，$\varepsilon_{T,\,t}$所需的次高斯性由其一致有界坐标及其在假设 7 下的独立性，以及假设1中的特征值条件推出。相应的样本均值界给出

$$\|\bar{\varepsilon}_T^0\|_2^2 \le C \frac{N}{T_{\mathrm{block}}}$$

以至少 $1 - 2\exp(-N)$ 的概率成立。因此，在概率至少为$1 - 4\exp(-N)$的事件上，

$$\|\widetilde{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_{T,0}\|_{\mathrm{op}} \le \|\mathbf{U}_T - \mathbf{\Sigma}_{T,0}\|_{\mathrm{op}} + \|\bar{\varepsilon}_T^0\|_2^2 \le b_{\Sigma,T},$$

在增大 $C$ 后。

为控制拟合残差，定义

$$\mathbf{e}_{T,t} := \boldsymbol{\mu}_{T,0}(\mathbf{Z}_{T,t}) - \widehat{\boldsymbol{\mu}}_T(\mathbf{Z}_{T,t}), \qquad \bar{\mathbf{e}}_T := \frac{1}{T_{\mathrm{block}}} \sum_{t=1}^{T_{\mathrm{block}}} \mathbf{e}_{T,t}.$$

由于$\widehat{\ }\varepsilon_{T,\,t} = \varepsilon_{T,\,t} + \mathrm{e}_{T,\,t}$，精确的中心化协方差分解为

$$\widehat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_{T,0} = \widetilde{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_{T,0} + \mathbf{H}_T + \mathbf{C}_T + \mathbf{C}_T',$$

$$\mathbf{H}_T := \frac{1}{T_{\mathrm{block}}} \sum_{t=1}^{T_{\mathrm{block}}} (\mathbf{e}_{T,t} - \bar{\mathbf{e}}_T)(\mathbf{e}_{T,t} - \bar{\mathbf{e}}_T)',$$

$$\mathbf{C}_T := \frac{1}{T_{\mathrm{block}}} \sum_{t=1}^{T_{\mathrm{block}}} (\varepsilon_{T,t} - \bar{\varepsilon}_T^0)(\mathbf{e}_{T,t} - \bar{\mathbf{e}}_T)'.$$

半正定性和（S2.4）给出

$$\|\mathbf{H}_T\|_{\mathrm{op}} \le \mathrm{tr}(\mathbf{H}_T) \le \frac{1}{T_{\mathrm{block}}} \sum_{t=1}^{T_{\mathrm{block}}} \|\mathbf{e}_{T,t}\|_2^2 \le N q_T,$$

以至少 $1 - \exp\left\{ -T^{p/(\beta+p)} \text{块} \log^7 T_{\mathrm{block}} \log(N T_{\mathrm{block}}) \right\}$ 的概率。

算子 -范数乘积不等式应用于两个中心化数据矩阵后得到 s

$$\|\mathbf{C}_T\|_{\mathrm{op}} \le \sqrt{\|\widetilde{\mathbf{\Sigma}}_T\|_{\mathrm{op}} \|\mathbf{H}_T\|_{\mathrm{op}}}.$$

因此，在联合事件上，

$$\|\widehat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_{T,0}\|_{\mathrm{op}} \le b_{\Sigma,T} + 2\sqrt{(C + b_{\Sigma,T}) N q_T} + N q_T = r_{\Sigma,T},$$

其中，协方差上界已吸收到 $C$ 中。由并集界可得所述概率 $1 - \Delta_T$。

# S3 中间引理

本节给出去偏得分的中间引理。我们证明，市场择时投资组合和多空投资组合满足所需的估计率和矩界。

**引理 S3.1** (期望去偏得分差异)。 设假设*1–3*成立。那么，对于满足 $\max_i a_{T,i,t} \le C/N_T$ 的市场择时投资组合，以及固定 $\tau \in (0, \infty)$ 的平滑多空投资组合，以至少 $1 - \Delta_T$ 的 $\mathbb{P}_P$-概率，有

$$\left\| \mathbb{E}_P \left[ \psi_T^{\mathrm{DB}} \left( \mathbf{W}_T, \boldsymbol{\theta}, \widehat{\boldsymbol{\eta}}_T^{(k)} \right) - \psi_T^{\mathrm{DB}} (\mathbf{W}_T, \boldsymbol{\theta}, \boldsymbol{\eta}_{T,0}) \right] \right\|_2 \le C \left( \left( \frac{r_{\mu,T}}{\sqrt{N_T}} \right)^2 + \left( \frac{r_{\mu,T}}{\sqrt{N_T}} \right) \left( \frac{\delta_{\Sigma,T}}{N_T} \right) \right),$$

其中，左侧的期望是相对于新观测 $\mathbf{W}_T$ 的边际分布而言的。

**证明**。 对于每个满足 $P(\mathrm{k}\boldsymbol{\mu}(\mathrm{W}_T)\mathrm{k}_\infty \le C) = 1$ 和 $\mathbf{\Sigma} \in \mathcal{S}_T$ 的确定性 $\boldsymbol{\mu} \in \mathrm{M}_T$ s，定义

$$\delta_{\mu,T} := \frac{\|\boldsymbol{\mu}(\mathbf{Z}_T) - \boldsymbol{\mu}_{T,0}(\mathbf{Z}_T)\|_{P,2}}{\sqrt{N_T}}, \qquad \delta_{\Sigma,T} := \|\mathbf{\Sigma} - \mathbf{\Sigma}_{T,0}\|_{\mathrm{op}}$$

并记

$$\mathbf{e}_T(\mathbf{Z}_T) = \boldsymbol{\mu}(\mathbf{Z}_T) - \boldsymbol{\mu}_{T,0}(\mathbf{Z}_T), \qquad \mathbf{m}_{T,\kappa}(\mathbf{Z}_T) = \boldsymbol{\mu}_{T,0}(\mathbf{Z}_T) + \kappa \mathbf{e}_T(\mathbf{Z}_T), \quad 0 \le \kappa \le 1.$$

条件均值和协方差约束给出

$$\begin{aligned}
&\mathbb{E}_P\left[\psi_{T,1}^{\mathrm{DB}}(\mathbf{W}_T, \boldsymbol{\theta}, (\boldsymbol{\mu}, \boldsymbol{\Sigma})) - \psi_{T,1}^{\mathrm{DB}}(\mathbf{W}_T, \boldsymbol{\theta}, (\boldsymbol{\mu}_{T,0}, \boldsymbol{\Sigma}_{T,0}))\right] \\
&\quad = \mathbb{E}_P\left[\{\boldsymbol{\omega}_T(\mathbf{Z}_T, \boldsymbol{\mu}(\mathbf{Z}_T)) - \boldsymbol{\omega}_T(\mathbf{Z}_T, \boldsymbol{\mu}_{T,0}(\mathbf{Z}_T))\}' \boldsymbol{\mu}_{T,0}(\mathbf{Z}_T)\right. \\
&\qquad\qquad \left. - \boldsymbol{\mu}(\mathbf{Z}_T)' \mathbf{J}_T(\mathbf{Z}_T, \boldsymbol{\mu}(\mathbf{Z}_T)) \mathbf{e}_T(\mathbf{Z}_T)\right] \\
&\quad = \mathbb{E}_P\left[\int_0^1 \mathbf{e}_T(\mathbf{Z}_T)' \left\{\mathbf{J}_T(\mathbf{Z}_T, \mathbf{m}_{T,\kappa}(\mathbf{Z}_T))' - \mathbf{J}_T(\mathbf{Z}_T, \boldsymbol{\mu}(\mathbf{Z}_T))'\right\} \boldsymbol{\mu}_{T,0}(\mathbf{Z}_T)\, d\kappa\right. \\
&\qquad\qquad \left. - \mathbf{e}_T(\mathbf{Z}_T)' \mathbf{J}_T(\mathbf{Z}_T, \boldsymbol{\mu}(\mathbf{Z}_T))' \mathbf{e}_T(\mathbf{Z}_T)\right].
\end{aligned}$$

对于第二个分量，相应的分解为

$$\begin{aligned}
&\mathbb{E}_P\left[\psi_{T,2}^{\mathrm{DB}}(\mathbf{W}_T, \boldsymbol{\theta}, (\boldsymbol{\mu}, \boldsymbol{\Sigma})) - \psi_{T,2}^{\mathrm{DB}}(\mathbf{W}_T, \boldsymbol{\theta}, (\boldsymbol{\mu}_{T,0}, \boldsymbol{\Sigma}_{T,0}))\right] \\
&\quad = 2\,\mathbb{E}_P\left[\int_0^1 \mathbf{e}_T(\mathbf{Z}_T)' \begin{Bmatrix} \mathbf{J}_T(\mathbf{Z}_T, \mathbf{m}_{T,\kappa}(\mathbf{Z}_T))' \left[\boldsymbol{\mu}_{T,0}(\mathbf{Z}_T)\boldsymbol{\mu}_{T,0}(\mathbf{Z}_T)' + \boldsymbol{\Sigma}_{T,0}\right] \boldsymbol{\omega}_T(\mathbf{Z}_T, \mathbf{m}_{T,\kappa}(\mathbf{Z}_T)) \\ - \mathbf{J}_T(\mathbf{Z}_T, \boldsymbol{\mu}(\mathbf{Z}_T))' \left[\boldsymbol{\mu}_{T,0}(\mathbf{Z}_T)\boldsymbol{\mu}_{T,0}(\mathbf{Z}_T)' + \boldsymbol{\Sigma}_{T,0}\right] \boldsymbol{\omega}_T(\mathbf{Z}_T, \boldsymbol{\mu}(\mathbf{Z}_T)) \end{Bmatrix} d\kappa\right. \\
&\qquad - \left\{\boldsymbol{\omega}_T(\mathbf{Z}_T, \boldsymbol{\mu}(\mathbf{Z}_T))' \mathbf{e}_T(\mathbf{Z}_T)\right\} \left\{\boldsymbol{\mu}(\mathbf{Z}_T)' \mathbf{J}_T(\mathbf{Z}_T, \boldsymbol{\mu}(\mathbf{Z}_T))\, \mathbf{e}_T(\mathbf{Z}_T)\right\} \\
&\qquad - \left\{\boldsymbol{\omega}_T(\mathbf{Z}_T, \boldsymbol{\mu}(\mathbf{Z}_T))' \boldsymbol{\mu}_{T,0}(\mathbf{Z}_T)\right\} \left\{\mathbf{e}_T(\mathbf{Z}_T)' \mathbf{J}_T(\mathbf{Z}_T, \boldsymbol{\mu}(\mathbf{Z}_T))\, \mathbf{e}_T(\mathbf{Z}_T)\right\} \\
&\qquad \left. - \boldsymbol{\omega}_T(\mathbf{Z}_T, \boldsymbol{\mu}(\mathbf{Z}_T))' (\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{T,0})\, \mathbf{J}_T(\mathbf{Z}_T, \boldsymbol{\mu}(\mathbf{Z}_T))\, \mathbf{e}_T(\mathbf{Z}_T)\right].
\end{aligned}$$

这些表达式将路径余项、信号误差的二次项，以及信号与协方差误差之间的交互项区分开来。

对于市场择时，

$$\mathbf{J}_T(\mathbf{z}, \mathbf{m}) = \mathbf{a}_{T,t}(\mathbf{z}) \mathbf{a}_{T,t}(\mathbf{z})', \qquad \|\mathbf{J}_T(\mathbf{z}, \mathbf{m})\|_{\mathrm{op}} \le \frac{C}{N_T},$$

且雅可比矩阵不依赖于 m。此外，根据分散化和逐坐标信号界，$\|\overline{\boldsymbol{\omega}_T}(\mathbf{z}, \mathbf{m})\|_2 \le C/\sqrt{N_T}$。第一个分量可直接简化为

$$-\mathbb{E}_P\left[\{\mathbf{a}_{T,t}(\mathbf{Z}_T)' \mathbf{e}_T(\mathbf{Z}_T)\}^2\right],$$

其绝对值有界于 $C\|\mathbf{e}_T(\mathbf{Z}_T)\|_{P,2}^2/N_T$。应用相同的算子范数

界于第二分量分解，对两个成分均可得到上界

$$C\left[\frac{\|\mathbf{e}_T(\mathbf{Z}_T)\|_{P,2}^2}{N_T}+\frac{\delta_{\Sigma,T}\|\mathbf{e}_T(\mathbf{Z}_T)\|_{P,2}}{N_T^{3/2}}\right].$$

对于固定的-$\tau$ 平滑多空投资组合，

$$\|\boldsymbol{\omega}_T(\mathbf{z},\mathbf{m})\|_1\le C,\qquad \|\boldsymbol{\omega}_T(\mathbf{z},\mathbf{m})\|_2\le\frac{10}{\sqrt{N_T}},\qquad \|\mathbf{J}_T(\mathbf{z},\mathbf{m})\|_{\mathrm{op}}\le\frac{5}{N_T\tau}.$$

校准逻辑雅可比矩阵也满足收缩导数界

$$|\mathbf{u}'D\mathbf{J}_T(\mathbf{z},\mathbf{m})[\mathbf{u}]\mathbf{v}|\le\frac{C}{N_T\tau^2}\|\mathbf{v}\|_\infty\|\mathbf{u}\|_2^2,$$

其中，u 是扰动方向，也是左侧进行收缩的向量；v 是右侧位置上的向量。为证明此界，令 D $=\mathrm{diag}(d_1,\ldots,d_{N_T})$，d $=(d_1,\ldots,d_{N_T})'$，，且 $\bar{u}_d=(\mathrm{d}'\mathrm{u})/(\mathbf{1}'\mathrm{d})$。每个校准逻辑雅可比矩阵均具有如下形式

$$\mathbf{H}=\mathbf{D}-\frac{\mathbf{d}\mathbf{d}'}{\mathbf{1}'\mathbf{d}},\qquad d_i=\frac{p_i(1-p_i)}{\tau},$$

其中，秩一项用于表示校准截断值 $c_+(\boldsymbol{\mu})$ 和 $c_-(\boldsymbol{\mu})$ 随$\boldsymbol{\mu}$变化。此外，u′Hu $=\sum_i d_i(u_i-\bar{u}_d)^2\le\frac{1}{4\tau}$kuk$_2^2$。对 H 求导并应用此界，即可得到所述收缩关系；这里要回顾 $p_i^+$ 和 $p_i^+$ 的定义，以及它们通过逻辑函数对 $\tau$ 的依赖。对于多空投资组合，分别对两个分支应用相同的论证，将 $p_i$、d、D 和 H 分别替换为 $p_i^\pm$、d$^\pm$、D$^\pm$ 和 H$^\pm$，~其中 $d_i^\pm=p_i^\pm(1-p_i^\pm)/\tau$，然后利用 J $=(10/N_T)(\mathrm{H}^+-\mathrm{H}^-)$ 对所得界求和。逐坐标信号界意味着

$$\left\|\left[\boldsymbol{\mu}_{T,0}(\mathbf{Z}_T)\boldsymbol{\mu}_{T,0}(\mathbf{Z}_T)'+\boldsymbol{\Sigma}_{T,0}\right]\boldsymbol{\omega}_T(\mathbf{Z}_T,\mathbf{m})\right\|_\infty\le C.$$

因此，分解中的路径积分项可由 $C_\tau$ke$_T$(Z$_T$)k$_{P,\,2}^2/N_T$ 界定。

两个信号乘积项具有相同的界，利用 k$\boldsymbol{\mu}$(Z$_T$)k$_2$ +k$\boldsymbol{\mu}_{T,\,0}$(Z$_T$)k$_2=O(\sqrt{N_T})$。协方差误差项的界为

$$\frac{C_\tau\delta_{\Sigma,T}}{N_T^{3/2}}\|\mathbf{e}_T(\mathbf{Z}_T)\|_{P,2}.$$

因此，对于平滑多空投资组合，期望得分差的范数同样有界为

$$C_\tau\left[\frac{\|\mathbf{e}_T(\mathbf{Z}_T)\|_{P,2}^2}{N_T}+\frac{\delta_{\Sigma,T}\|\mathbf{e}_T(\mathbf{Z}_T)\|_{P,2}}{N_T^{3/2}}\right].$$

因此，存在一个与 $P$ 和 $T$ 无关的常数 $C<\infty$，使得

$$\left\|\mathbb{E}_P\left[\psi_T^{\mathrm{DB}}(\mathbf{W}_T,\boldsymbol{\theta},(\boldsymbol{\mu},\boldsymbol{\Sigma}))-\psi_T^{\mathrm{DB}}(\mathbf{W}_T,\boldsymbol{\theta},(\boldsymbol{\mu}_{T,0},\boldsymbol{\Sigma}_{T,0}))\right]\right\|_2\le C\left(\delta_{\mu,T}^2+\frac{\delta_{\Sigma,T}\delta_{\mu,T}}{N_T}\right).$$

在假设 3 的速率条件下，最终结果随即成立；回顾前一个显示式中的期望是关于新观测 $\mathbf{W}_T$ 的边际分布而言的。

■

**引理 S3.2**（得分差与预言得分矩界）。 假设假设*1–3*成立。对于满足 $\max_i a_{T,\,i,\,t}\le C/N_T$ 的市场择时投资组合，或固定-$\tau$平滑多空投资组合，对于所有 $P\in\mathcal{P}$ 以及所有折 $k=1,\ldots,K$，均有

$$\left\|\psi_T^{\mathrm{DB}}\left(\mathbf{W}_T,\boldsymbol{\theta},(\widehat{\boldsymbol{\mu}}_{T,k},\widehat{\boldsymbol{\Sigma}}_{T,k})\right)-\psi_T^{\mathrm{DB}}(\mathbf{W}_T,\boldsymbol{\theta},(\boldsymbol{\mu}_{T,0},\boldsymbol{\Sigma}_{T,0}))\right\|_{P,2}\le C\left(\frac{r_{\mu,T}}{\sqrt{N_T}}+\frac{r_{\Sigma,T}}{N_T}\right),$$

$$\left\|\psi_T^{\mathrm{DB}}\left(\mathbf{W}_T,\boldsymbol{\theta},(\widehat{\boldsymbol{\mu}}_{T,k},\widehat{\boldsymbol{\Sigma}}_{T,k})\right)-\psi_T^{\mathrm{DB}}(\mathbf{W}_T,\boldsymbol{\theta},(\boldsymbol{\mu}_{T,0},\boldsymbol{\Sigma}_{T,0}))\right\|_{P,3}\le C,$$

$$\left\|\psi_T^{\mathrm{DB}}(\mathbf{W}_T,\boldsymbol{\theta},(\boldsymbol{\mu}_{T,0},\boldsymbol{\Sigma}_{T,0}))\right\|_{P,5}\le C,$$

其中，第一个和第二个不等式以至少 $1-\Delta_T$ 的 $\mathbb{P}_P$-概率成立。

**证明**。 逐分量减去实现得分公式，并将它们的乘积展开为望远镜和。投资组合权重界和雅可比矩阵界，连同逐坐标信号界和数据界，依据引理 S3.1 的证明或假设 3 中给出的结果，可得

$$\begin{aligned}&\left\|\psi_T^{\mathrm{DB}}\left(\mathbf{W}_T,\boldsymbol{\theta},(\widehat{\boldsymbol{\mu}}_{T,k},\widehat{\boldsymbol{\Sigma}}_{T,k})\right)-\psi_T^{\mathrm{DB}}(\mathbf{W}_T,\boldsymbol{\theta},(\boldsymbol{\mu}_{T,0},\boldsymbol{\Sigma}_{T,0}))\right\|_2\\&\quad\le C\left[\frac{\|\widehat{\boldsymbol{\mu}}_{T,k}(\mathbf{W}_T)-\boldsymbol{\mu}_{T,0}(\mathbf{W}_T)\|_2}{\sqrt{N_T}}+\frac{\|\widehat{\boldsymbol{\Sigma}}_{T,k}-\boldsymbol{\Sigma}_{T,0}\|_{\mathrm{op}}}{N_T}\right].\end{aligned}$$

取 $L_2(P)$ 范数即可得到第一个不等式。

相同的一致界给出 $\|\boldsymbol{\omega}_T(\mathbf{W}_T,\ \boldsymbol{\mu})\|_1\le C$、$\|\boldsymbol{\omega}_T(\mathbf{W}_T,\ \boldsymbol{\mu})\|_2\le C/\sqrt{N_T}$ 和 $\overline{\|\mathbf{J}_T(\mathbf{W}_T,\ \boldsymbol{\mu})\|}_{\mathrm{op}}\le$

$C/N_T$。因此，得分分量一致有界。例如，

$$\left|\boldsymbol{\mu}(\mathbf{W}_T)'\mathbf{J}_T(\mathbf{W}_T, \boldsymbol{\mu}(\mathbf{W}_T))\left(\boldsymbol{\mu}(\mathbf{W}_T) - \mathbf{R}_T\right)\right| \leq C,$$
$$\left|\boldsymbol{\omega}_T(\mathbf{W}_T, \boldsymbol{\mu}(\mathbf{W}_T))' \left\{\boldsymbol{\mu}(\mathbf{W}_T)\boldsymbol{\mu}(\mathbf{W}_T)' + \boldsymbol{\Sigma}\right\} \mathbf{J}_T(\mathbf{W}_T, \boldsymbol{\mu}(\mathbf{W}_T))\left(\boldsymbol{\mu}(\mathbf{W}_T) - \mathbf{R}_T\right)\right| \leq C.$$

我们得到，得分差在 $L_3(P)$ 中有界，而预言得分在$L_5(P)$ 中有界。对于所有折，干扰项界以至少 $1 - \Delta_T$ 的 $\mathbb{P}_P$-概率成立，因此前两个不等式具有所述概率。

■

# S4 补充结果

## S4.1 其他时间序列结果

本节给出关于依概率收敛以及三角阵列中具有 $\beta$-混合行的可测函数的 HAC 一致性的其他时间序列结果。这些结果可能具有独立的研究价值。

## S4.2 依赖训练的经验过程可忽略性

**引理 S4.1** (依赖训练的经验过程的一致可忽略性)。 假设假设 *S1.2* 成立。设 $T_1 + T_2 = T$，其中 $T_1/T \to \zeta \in (0,1)$。令 $\widehat{\boldsymbol{\eta}}_T$ 为一个相对于 $\sigma(\mathbf{W}_{T,t} : 1 \leq t \leq T_1)$可测的干扰项估计量，并令 $f_T(\boldsymbol{\eta}, \mathbf{w})$ 联合可测且为实值函数。假设，对于每个候选干扰参数以及每个 $P \in \mathcal{P}$，

$$\int f_T(\boldsymbol{\eta}, \mathbf{w})\, dP(\mathbf{W}_T)(\mathbf{w}) = 0.$$

假设存在确定性序列 $\kappa_T \in (0,1)$ 和 $K_T \geq 0$，且满足 $\kappa_T \to 0$ 和$K_T \to 0$，使得

$$\sup_{P \in \mathcal{P}} P\left(\left\|f_T(\widehat{\boldsymbol{\eta}}_T, \mathbf{W}_T)\right\|_{P,2}^2 > \kappa_T\right) \leq K_T.$$

进一步假设存在一个有限常数 $C$，其对 $P \in \mathcal{P}$ 和 $T$ 一致，并存在一个确定性序列 $\Xi_T \geq 0$，满足 $\Xi_T \to 0$，使得

$$\sup_{P\in\mathcal{P}} P\left(\left\|f_T(\widehat{\boldsymbol{\eta}}_T, \mathbf{W}_T)\right\|_{P,3} > C\right) \leq \Xi_T,$$
$$\sup_{P\in\mathcal{P}} \sup_{1\leq t\leq T} \mathbb{E}_P\left[\left|f_T(\widehat{\boldsymbol{\eta}}_T, \mathbf{W}_{T,t})\right|\right] \leq C.$$

边际范数仅对$\mathbf{W}_T$积分，同时固定$\widehat{\boldsymbol{\eta}}_T$，且为训练可测。于是，对每个 $\varepsilon > 0$,

$$\sup_{P\in\mathcal{P}} P\left(\left|\frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^{T} f_T(\widehat{\boldsymbol{\eta}}_T, \mathbf{W}_{T,t})\right| > \varepsilon\right) \longrightarrow 0.$$

**证明**。 固定 $P \in \mathcal{P}$。以下所有常数在 $P$ 和 $T$上均一致。记

$$f_{T,t} := f_T(\widehat{\boldsymbol{\eta}}_T, \mathbf{W}_{T,t}), \qquad \mathcal{F}_{T,\text{train}} := \sigma(\mathbf{W}_{T,t} : 1 \leq t \leq T_1),$$
$$A_T := \{\left\|f_T(\widehat{\boldsymbol{\eta}}_T, \mathbf{W}_T)\right\|_{P,2}^2 \leq \kappa_T\},$$
$$B_T := \{\left\|f_T(\widehat{\boldsymbol{\eta}}_T, \mathbf{W}_T)\right\|_{P,3} \leq C\}.$$

于是 $A_T$， $B_T \in \mathbf{F}_{T,\text{train}}$和 $P((A_T \cap B_T)^c) \leq K_T + \Xi_T$。

步骤 *1*（分离与耦合）。 对所有充分大的 $T$，取 $l_T = \mathsf{dlog}\, T_2\mathsf{e} < T_2$ 。分解

$$\frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^{T} f_{T,t} = D_T + S_T,$$
$$D_T := \frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^{T_1+l_T} f_{T,t}, \qquad S_T := \frac{1}{\sqrt{T_2}} \sum_{t=T_1+l_T+1}^{T} f_{T,t}.$$

注意，可测变换不会增大绝对正则系数。根据 Berbee耦合引理，在概率空间的扩张上，存在未来序列 $\{\mathbf{W}^*_{T,\,t} : t \geq T_1 + l_T + 1\}$ 的一个副本，其联合分布与原序列相同，且独立于 $\mathbf{F}_{T,\,\text{train}}$，使得

$$P\left((\mathbf{W}_{T,t})_{t\geq T_1+l_T+1} \neq (\mathbf{W}^*_{T,t})_{t\geq T_1+l_T+1}\right) \leq \beta_{P,T}(l_T + 1) \leq C_\beta \rho_\beta^{l_T+1}.$$

训练及其干扰项估计量保持不变。定义

$$f^*_{T,t} := f_T(\widehat{\boldsymbol{\eta}}_T, \mathbf{W}^*_{T,t}), \qquad S^*_T := \frac{1}{\sqrt{T_2}} \sum_{t=T_1+l_T+1}^{T} f^*_{T,t}.$$

对于每个 $u > 0$、 $P\left(|S_T| > u\right) \le P\left(|S^*_T| > u\right) + C_\beta \rho_\beta^{l_T+1}$。

步骤*2*（条件协方差界）。在给定训练条件下，估计量是固定的，而耦合未来保留其原始联合分布。在 $B_T$ 上，边际中心化给出 $\mathbb{E}_P[f^*_{T,t} \,|\mathbb{F}_{T,\ \text{train}}] = 0$。Davydov不等式以及用绝对正则系数界定强混合系数的结果给出，在 $B_T$ 上，

$$|\operatorname{Cov}_P(f^*_{T,t}, f^*_{T,t+h} \mid \mathcal{F}_{T,\text{train}})| \le C\rho_\beta^{h/3}, \qquad h \ge 1.$$

在 $A_T \cap B_T$ 上，柯西–施瓦茨不等式还将这些协方差界定为 $\kappa_T$，对于$h \ge 0$。取 $p_T = \mathsf{d3}\ \log(1/\kappa_T)/|\log\ \rho_\beta|\mathsf{e}$。于是，在 $A_T \cap B_T$ 上，

$$\begin{aligned}&\sum_{h=0}^{\infty} |\operatorname{Cov}_P(f^*_{T,t}, f^*_{T,t+h} \mid \mathcal{F}_{T,\text{train}})| \\ &\quad \le (p_T+1)\kappa_T + C\sum_{h=l_T+1}^{\infty} \rho_\beta^{h/3} \le C\kappa_T\{1+\log(1/\kappa_T)\}.\end{aligned}$$

因此，条件平稳性和至多为 $T_2$ 的保留区块长度给出

$$\mathbb{V}\mathrm{ar}_P(S^*_T \mid \mathcal{F}_{T,\text{train}}) \le C\kappa_T\{1+\log(1/\kappa_T)\}$$

在该事件上。

*Step 3*（保留和的概率界）。条件切比雪夫不等式给出 s

$$\begin{aligned}P(|S^*_T| > u) &\le K_T + \Xi_T + \mathbb{E}_P[\mathbf{1}_{A_T\cap B_T} P(|S^*_T| > u \mid \mathcal{F}_{T,\text{train}})] \\ &\le K_T + \Xi_T + \frac{C\kappa_T\{1+\log(1/\kappa_T)\}}{u^2}.\end{aligned}$$

由于 $l_T \to \infty$，最终的显示项一致地消失，耦合误差也同样如此。

步骤 *4*（消除间隙）。无条件一阶矩界给出 $\mathbb{E}_P|D_T| \le$

$Cl_T/\sqrt{T_2}$。结合马尔可夫不等式和前面的步骤,

$$P\left(\left|\frac{1}{\sqrt{T_2}}\sum_{t=T_1+1}^{T} f_{T,t}\right| > \varepsilon\right)$$
$$\leq \frac{Cl_T}{\varepsilon\sqrt{T_2}} + K_T + \Xi_T + \frac{C\kappa_T\{1+\log(1/\kappa_T)\}}{\varepsilon^2} + C_\beta \rho_\beta^{l_T+1} \longrightarrow 0$$

在 $P \in \mathcal{P}$ 上一致成立。 ■

**定义 S4.1** （可行 HAC 估计量）。 令 $f_{T,\,t} := f_T$（$\widehat{\boldsymbol{\eta}}_T$、$\mathbf{W}_{T,\,t}$），对于 $t = T_1+1, \ldots, T$，对于 $f_T$ 某个联合可测标量函数，其中 $T_1 + T_2 = T$ 且 $\widehat{\boldsymbol{\eta}}_T$ 仅使用训练观测值 $t \leq T_1$。令 $\theta_T$ 为一个关于全样本可测的中心化估计量。对于 $h = 0, \ldots, T_2 - 1$，定义

$$\widehat{\gamma}_T(h) := \frac{1}{T_2}\sum_{t=T_1+h+1}^{T} (f_{T,t} - \widehat{\theta}_T)(f_{T,t-h} - \widehat{\theta}_T).$$

令 $m_T \in \{0, \ldots, T_2 - 1\}$ 为确定性滞后截断参数，并令 $w_{T_2,\,h}$ 为确定性权重，其中 $w_{T_2,\,0} = 1$。可行异方差和自相关一致（*HAC*）估计量为

$$\widehat{\Gamma}_T := \widehat{\gamma}_T(0) + 2\sum_{h=1}^{m_T} w_{T_2,h}\widehat{\gamma}_T(h). \tag{S4.1}$$

最后，令

$$\Gamma_{T,0} := \mathbb{V}_P\left(\frac{1}{\sqrt{T_2}}\sum_{t=T_1+1}^{T}\{f_T(\boldsymbol{\eta}_{T,0}, \mathbf{W}_{T,t}) - \theta_{T,0}\}\right),$$
$$\theta_{T,0} := \mathbb{E}_P[f_T(\boldsymbol{\eta}_{T,0}, \mathbf{W}_{T,t})].$$

**引理 S4.2** （可行 HAC 估计量的一致一致性）。 假设假设 *S1.2* 成立。令 $T_1 + T_2 = T$，其中 $T_1/T \to \zeta \in (0, 1)$。令 $\widehat{\boldsymbol{\eta}}_T$ 关于 $\sigma(\mathbf{W}_{T,\,t}: \ 1 \leq t \leq T_1)$ 可测，令 $\boldsymbol{\eta}_{T,\,0}$ 为非随机，并令 $f_T(\boldsymbol{\eta}, \mathbf{w})$ 为联合可测实值函数（可能不同于引理 *S4.1* 中的选择）。

假设存在确定性序列 $\kappa_T \in (0, 1)$ 和 $K_T \geq 0$，且 $\kappa_T \to 0$ 与

$K_T \to 0$，使得

$$\sup_{P\in\mathcal{P}} P\left(\left\|f_T(\widehat{\boldsymbol{\eta}}_T, \mathbf{W}_T) - f_T(\boldsymbol{\eta}_{T,0}, \mathbf{W}_T)\right\|_{P,2}^2 > \kappa_T\right) \le K_T.$$

*Supp*进一步假设，对于某个固定的 $\nu > 0$，存在有限的一致常数 $C$，满足 $s$

$$\sup_{P\in\mathcal{P}} \sup_T \left\|f_T(\boldsymbol{\eta}_{T,0}, \mathbf{W}_T)\right\|_{P,4+\nu} \le C,$$
$$\sup_{P\in\mathcal{P}} \sup_T \sup_{1\le t\le T} \mathbb{E}_P[f_T(\widehat{\boldsymbol{\eta}}_T, \mathbf{W}_{T,t})^2] \le C.$$

设确定性带宽和权重满足

$$m_T \to \infty, \qquad m_T = o(T_2^{1/3}), \qquad m_T\sqrt{\kappa_T} \to 0,$$
$$w_{T_2,0} = 1, \qquad \sup_T \max_{0\le h\le m_T} |w_{T_2,h}| \le C, \qquad w_{T_2,h} \to 1 \quad \textit{for every fixed } h \ge 0.$$

^最后，设 $\theta_T$ 为全样本可测，并假设对于每个 $\varepsilon > 0$ ，

$$\sup_{P\in\mathcal{P}} P\left(m_T|\widehat{\theta}_T - \theta_{T,0}(P)| > \varepsilon\right) \to 0.$$

那么，对于每个 $\varepsilon > 0$,

$$\sup_{P\in\mathcal{P}} P(|\widehat{\Gamma}_T - \Gamma_T(P)| > \varepsilon) \to 0.$$

**证明**。固定 $P \in \mathcal{P}$。下文中的所有常数和随机界均关于 $P$ 和 $T$ 一致。定义

$$f_{T,t} := f_T(\widehat{\boldsymbol{\eta}}_T, \mathbf{W}_{T,t}), \qquad f_{T,0,t} := f_T(\boldsymbol{\eta}_{T,0}, \mathbf{W}_{T,t}).$$

并记

$$Y_{T,t} := f_{T,0,t} - \theta_{T,0}, \qquad e_{T,t} := f_{T,t} - f_{T,0,t}, \qquad d_T := \widehat{\theta}_T - \theta_{T,0}.$$

步骤 *1*（不可行估计量的收敛性）。定义

$$\bar{\Gamma}_T := \frac{1}{T_2}\sum_{t=T_1+1}^{T} Y_{T,t}^2 + \frac{2}{T_2}\sum_{h=1}^{m_T} w_{T_2,h} \sum_{t=T_1+h+1}^{T} Y_{T,t}Y_{T,t-h}.$$

令 $\gamma_T(h) := \mathrm{Cov}_P(Y_{T,t}, Y_{T,t+h})$ 且 $a_\nu := (2+\nu)/(4+\nu) > 0$。平稳性、预言矩界和 Davydov不等式给出

$$|\gamma_T(0)| \le C, \qquad |\gamma_T(h)| \le C\rho_\beta^{a_\nu h}, \quad h \ge 1.$$

此外,

$$\Gamma_T = \gamma_T(0) + 2\sum_{h=1}^{T_2-1}\left(1 - \frac{h}{T_2}\right)\gamma_T(h),$$
$$\mathbb{E}_P[\bar{\Gamma}_T] = \gamma_T(0) + 2\sum_{h=1}^{m_T} w_{T_2,h}\left(1 - \frac{h}{T_2}\right)\gamma_T(h).$$

因此

$$|\mathbb{E}_P[\bar{\Gamma}_T] - \Gamma_T| \le 2\sum_{h=1}^{m_T}|w_{T_2,h} - 1||\gamma_T(h)| + 2\sum_{h=m_T+1}^{T_2-1}|\gamma_T(h)|.$$

最后一项至多为 $C\rho^{a_\nu(m_T+1)\,\beta}$。对于所有充分大的 $T$，定义

$$H_T := \max\left\{j \in \{1,\dots,m_T\} : \max_{1\le h\le j}|w_{T_2,h} - 1| \le \frac{1}{j}\right\}.$$

对于每个固定的 $j$，最终 $m_T \ge j$ 且所示的权重条件成立。因此，该集合最终非空，$H_T \to \infty$，并且截至 $H_T$ 的最大偏差至多为$H_T^{-1}$。在 $H_T$ 处分段可得

$$\begin{aligned}\sum_{h=1}^{m_T}|w_{T_2,h} - 1||\gamma_T(h)| &\le \frac{C}{H_T}\sum_{h=1}^{H_T}\rho_\beta^{a_\nu h} + C\sum_{h=H_T+1}^{\infty}\rho_\beta^{a_\nu h}\\ &\le \frac{C}{H_T} + C\rho_\beta^{a_\nu(H_T+1)} \to 0.\end{aligned}$$

因此,

$$\sup_{P\in\mathcal{P}}|\mathbb{E}_P[\bar{\Gamma}_T] - \Gamma_T| \le \frac{C}{H_T} + C\rho_\beta^{a_\nu(H_T+1)} + C\rho_\beta^{a_\nu(m_T+1)} \to 0.$$

对于波动集合 $U_{T,\,h,\,t} := Y_{T,\,t}Y_{T,\,t-h}$。Hölder不等式给出 $\|U_{T,\,h,\,t}\|_{P,\,2+\nu/2} \le C$，而柯西–施瓦茨不等式将乘积协方差界定为 $C$ （滞后距离为

$0 \le k \le h$。对于 $k > h$，Davydov不等式给出，其中 $b_\nu := \nu/(4+\nu) > 0$，

$$|\operatorname{Cov}_P(U_{T,h,t}, U_{T,h,t+k})| \le C\rho_\beta^{b_\nu(k-h)}.$$

绝对协方差之和由 $C(h+1)$ 界定，因此

$$\left\| \frac{1}{T_2} \sum_{t=T_1+h+1}^{T} \{U_{T,h,t} - \mathbb{E}_P[U_{T,h,t}]\} \right\|_{P,2} \le C\sqrt{\frac{h+1}{T_2}}.$$

Minkowski不等式和有界权重意味着

$$\left\| \bar{\Gamma}_T - \mathbb{E}_P[\bar{\Gamma}_T] \right\|_{P,2} \le \frac{C}{\sqrt{T_2}} \sum_{h=0}^{m_T} \sqrt{h+1} \le \frac{C(m_T+1)^{3/2}}{\sqrt{T_2}} \to 0.$$

因此 $\bar{\Gamma}_T - \Gamma_T = o_P(1)$ 一致成立。

步骤*2*（可行到不可行分解）。展开乘积可得

$$\begin{aligned}
\widehat{\Gamma}_T - \bar{\Gamma}_T = {} & \frac{1}{T_2} \sum_{t=T_1+1}^{T} \{2Y_{T,t}e_{T,t} + e_{T,t}^2\} \\
& + \frac{2}{T_2} \sum_{h=1}^{m_T} w_{T_2,h} \sum_{t=T_1+h+1}^{T} \{Y_{T,t}e_{T,t-h} + Y_{T,t-h}e_{T,t} + e_{T,t}e_{T,t-h}\} \\
& - 2d_T \left\{ \frac{1}{T_2} \sum_{t=T_1+1}^{T} (Y_{T,t} + e_{T,t}) \right. \\
& \qquad \left. + \frac{1}{T_2} \sum_{h=1}^{m_T} w_{T_2,h} \sum_{t=T_1+h+1}^{T} (Y_{T,t} + Y_{T,t-h} + e_{T,t} + e_{T,t-h}) \right\} \\
& + d_T^2 \left\{ 1 + 2 \sum_{h=1}^{m_T} w_{T_2,h} \left(1 - \frac{h}{T_2}\right) \right\}.
\end{aligned}$$

步骤 *3*（估计中心化）。令 $F_T := T_2^{-1} \sum_{t=T_1+1}^{T} \{f_{T,\,t} - \theta_{T,\,0}(P)\}^2$。拟合的二阶矩和预言矩界蕴含 $\mathbb{E}_P[F_T] \le C$，因此 $F_T = O_P(1)$ 一致成立。根据柯西–施瓦茨不等式，我们可以将 $Y_{T,\,t} + e_{T,\,t}$ 的每个标准化子集和界定为 $\sqrt{F_T}$。因此，线性和二次中心化项可由以下内容界定——

$$C(m_T+1)|d_T|\sqrt{F_T} = o_P(1), \qquad C(m_T+1)d_T^2 = o_P(1).$$

利用 $m_T|d_T| = o_P(1)$。

步骤 *4*（干扰误差乘积）。定义 $Q_T := T_2^{-1}\sum_{t=T_1+1}^{T} e_{T,t}^2$。选择 $l_T = \lceil \log T_2 \rceil$并进行拆分

$$
\begin{aligned}
Q_T &= Q_T^{\mathrm{gap}} + Q_T^{\mathrm{rem}},\\
Q_T^{\mathrm{gap}} &:= \frac{1}{T_2}\sum_{t=T_1+1}^{T_1+l_T} e_{T,t}^2, \qquad Q_T^{\mathrm{rem}} := \frac{1}{T_2}\sum_{t=T_1+l_T+1}^{T} e_{T,t}^2.
\end{aligned}
$$

无条件界给出 $\mathbb{E}_P[Q_T^{\mathrm{gap}}] \le C l_T/T_2$。令 $\mathcal{F}_{T,\ \mathrm{train}} := \sigma(\mathbf{W}_{T,\ t}:\ 1 \le t \le T_1)$，并且

$$
A_T := \{\left\| f_T(\widehat{\boldsymbol{\eta}}_T, \mathbf{W}_T) - f_T(\boldsymbol{\eta}_{T,0}, \mathbf{W}_T)\right\|_{P(\mathbf{W}_T),2}^2 \le \kappa_T\}.
$$

于是 $P(A_T^c) \le K_T$。Berbee 引理提供了余项块的一个副本，保留其联合分布并与训练独立，失配概率至多为 $C_\beta \rho_\beta^{l_T+1}$。保持训练不变，定义

$$
\begin{aligned}
e_{T,t}^* &:= f_T(\widehat{\boldsymbol{\eta}}_T, \mathbf{W}_{T,t}^*) - f_T(\boldsymbol{\eta}_{T,0}, \mathbf{W}_{T,t}^*),\\
Q_T^* &:= \frac{1}{T_2}\sum_{t=T_1+l_T+1}^{T} (e_{T,t}^*)^2.
\end{aligned}
$$

在 $A_T$, $\mathbb{E}_P[Q_T^* \,|\mathcal{F}_{T,\ \mathrm{train}}] \le \kappa_T$ 上。条件马尔可夫不等式、耦合和间隔界给出，对于$u > 0$,

$$
P(m_T^2 Q_T > u) \le \frac{C m_T^2 l_T}{u T_2} + K_T + C_\beta \rho_\beta^{l_T+1} + \frac{2 m_T^2 \kappa_T}{u} \to 0.
$$

由于 $m_T\sqrt{\kappa_T} \to 0$ 并且，该收敛成立

$$
\frac{m_T^2 l_T}{T_2} = \left(\frac{m_T}{T_2^{1/3}}\right)^2 \frac{l_T}{T_2^{1/3}} \to 0.
$$

因此 $m_T\sqrt{Q_T} = o_P(1)$一致地。对于每个$h \le m_T$，柯西–施瓦茨不等式给出 s

$$
\frac{1}{T_2}\left|\sum_{t=T_1+h+1}^{T} e_{T,t} e_{T,t-h}\right| \le Q_T.
$$

干扰参数乘积项至多为 $C(m_T+1)Q_T = o_P(1)$。

步骤 5（混合项与结论）。令 $R_T := T_2^{-1}\sum_{t=T_1+1}^{T} Y_{T,t}^2 = O_P(1)$ 一致成立。每个标准化的滞后混合和，无论方向如何且包括零阶滞后，其绝对值均至多为 $\sqrt{R_T Q_T}$。它们的总贡献有界于 $C(m_T+1)\sqrt{R_T Q_T} = o_P(1)$。步骤 3–5 一致地给出 $\Gamma_T - \widehat{\Gamma}_T = o_P(1)$。将其与步骤 1 结合即可证明该引理。 ∎

## S4.3 合并神经网络速率

本节将 Farrell 等人（2021）（下文简称 FLM）的结果推广至合并最小二乘回归。本节始终采用 FLM 的符号说明，这套符号与本文其他部分使用的符号相互独立。无约束的横截面依赖带来的关键结果是，独立抽样单位是 $T$ 日期，而非 $NT$ 日期—资产观测值。采用 FLM 的证明策略，并通过对每个日期内的观测值取平均将其调整为适用于独立日期区块，可保证平方误差速率为$T^{-\beta/(\beta+d)}\log^7 T\log(NT)$。相较于 FLM 的单变量界，主导项多出因子 $\log(NT)/\log T$。其中，$\beta$ 表示总体目标的光滑度，$d$ 表示协变量的维度。

无约束的横截面依赖是资产定价研究中的一个现实假设，因为大多数特征集都包含状态变量，例如市场波动率，而这类变量在不同资产之间保持不变。相比之下，如果观测值在不同时期的横截面上相互独立，那么根据 FLM 的单变量结果，相应的合并回归论证可以用 $NT$ 代替 $T$ 作为样本量。

令 $\mathrm{B} = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_N)$ 表示一个一般横截面区块，其中 $\boldsymbol{Z}_i = (Y_i, \boldsymbol{X}_i')'$。我们将总体目标定义为区块平均期望平方误差损失的最小化者：

$$\ell(f, \boldsymbol{z}) = \frac{1}{2}\{y - f(\boldsymbol{x})\}^2, \qquad f_* \in \operatorname*{arg\,min}_{f:\, \|f\|_{\mathrm{pool}} < \infty} \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} \ell(f, \boldsymbol{Z}_i)\right],$$

其中，最小化是在下文定义的合并总体范数下，对平方可积可测函数进行的，而不是对神经网络类进行的。我们省略一般区块和总体目标对 $T$ 的依赖。

**假设 S4.1** (1P：面板抽样与光滑度)。对于每个 $T$，令 $N = N_T \geq 1$ 并观测 $\boldsymbol{Z}_{i,t} = (Y_{i,t}, \boldsymbol{X}_{i,t}')'$，$1 \leq i \leq N$，$1 \leq t \leq T$。假设：

1. 完整区块 $\mathrm{B}_t = (\boldsymbol{Z}_{1,t}, \ldots, \boldsymbol{Z}_{N,t})$ 是 $t$ 上 B 的独立同分布副本。每个区块内部的依赖关系不受限制。
2. 对于每个 $i$ 和 $t$，$\boldsymbol{Z}_{i,t} \in \mathcal{Y} \times [-1, 1]^d$ 几乎必然成立，其中 $\boldsymbol{X}_{i,t}$ 服从连续分布。

边际分布可能因 $i$而异。

*3.* 对于固定常数 $M > 0$、 $\|f_*\|_\infty \le M$ 和 $\mathcal{Y} \subseteq [-M, M]$。

*4. FLM* 的假设 *2* 具有光滑度参数 $\beta \in \mathbb{N}_+$。

维度 $d$、光滑度参数 $\beta$ 和常数 $M$ 均独立于 $N$ 和 $T$。

**定理 S4.1**（FLM 定理 1 的面板数据版本）。 在假设 *S4.1* 下，令

$$\widehat{f}_{\mathrm{MLP,pool}} \in \operatorname*{arg\,min}_{f \in \mathcal{F}_{\mathrm{MLP}},\, \|f\|_\infty \le 2M} \sum_{t=1}^{T} \sum_{i=1}^{N} \{Y_{i,t} - f(\boldsymbol{X}_{i,t})\}^2,$$

使用宽度为 $H \asymp T^{d/[2(\beta+d)]} \log^2 T$、深度为 $L \asymp \log T$ 的 *ReLU* 多层感知机。

对于所有充分大的 $T$，以至少 $1 - \exp\{-T^{d/(\beta+d)} \log^8 T\}$ 的概率，二者均

$$\text{(a)} \quad \left\| \widehat{f}_{\mathrm{MLP,pool}} - f_* \right\|_{\mathrm{pool}}^2 \le C \left\{ T^{-\beta/(\beta+d)} \log^7 T \log(NT) + \frac{\log\log T}{T} \right\},$$

$$\text{(b)} \quad \mathbb{E}_{NT}\left[ (\widehat{f}_{\mathrm{MLP,pool}} - f_*)^2 \right] \le C \left\{ T^{-\beta/(\beta+d)} \log^7 T \log(NT) + \frac{\log\log T}{T} \right\}.$$

其中，合并总体范数为

$$\|h\|_{\mathrm{pool}}^2 = \mathbb{E}\left[ \frac{1}{N} \sum_{i=1}^{N} h(\boldsymbol{X}_i)^2 \right],$$

以及 $\mathbb{E}_{NT}[h^2] = (NT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} h(\boldsymbol{X}_{i,t})^2$。常数 $C$ 不依赖于 $N$ 或 $T$。

**证明**。 证明沿用了 FLM 的单变量回归证明步骤，但有几处值得注意的变化。下面概述这些变化。按照 FLM 的符号说明，记 $\widehat{f} = \widehat{f}_{\mathrm{MLP,\,pool}}$ 并令 $\mathrm{F}_{\mathrm{DNN}}$ 表示其受 $\|f\|_\infty \le 2M$ 限制的确定性网络类。选取一个确定性近似函数 $f_T \in \mathrm{F}_{\mathrm{DNN}}$，使得 $\|f_T - f_*\|_\infty \le \epsilon_T$，并将分块经验算子和分块损失差定义为

$$\mathbb{E}_T^{\mathrm{block}}[a] = \frac{1}{T} \sum_{t=1}^{T} a(\mathcal{B}_t), \qquad g_f^{\mathrm{block}}(\mathcal{B}_t) = \frac{1}{N} \sum_{i=1}^{N} \{\ell(f, \boldsymbol{Z}_{i,t}) - \ell(f_*, \boldsymbol{Z}_{i,t})\}.$$

由于 $f_*$ 是合并总体平方损失最小化器，对于每个固定的 $f$，

$$\mathbb{E}[g_f^{\mathrm{block}}] = \frac{1}{2} \|f - f_*\|_{\mathrm{pool}}^2.$$

因此，经验最优性给出了 FLM 分解式 (A.1) 的分块对应形式：

$$\frac{1}{2}\left\|\widehat{f}-f_*\right\|_{\text{pool}}^2 \le (\mathbb{E}-\mathbb{E}_T^{\text{block}})g_{\widehat{f}}^{\text{block}} + \mathbb{E}_T^{\text{block}} g_{f_T}^{\text{block}},$$

其中总体　　均值的期望对一个一般区块进行积分，同时将函数视为固定的　　.

对于 FLM 的步骤 1，伯恩斯坦不等式应用于相互独立的横截面平均值 $g_{f_T}^{\text{block}}(\mathrm{B}_t)$，而不是像 FLM 中那样应用于个体观测值。詹森不等式给出 $\mathbb{V}\mathrm{ar}(g_{f_T}^{\text{block}}) \le C\epsilon_T^2$，无需施加日期内独立性；总体均值至多为 $\epsilon_T^2/2$，且包络函数由一个固定常数界定。因此，FLM 的 (A.2) 的对应结果是：以至少 $1-e^{-\tilde{\gamma}}$ 的概率,

$$\mathbb{E}_T^{\text{block}} g_{f_T}^{\text{block}} \le C\left(\epsilon_T^2 + \frac{\tilde{\gamma}}{T}\right).$$

对于 FLM 的步骤2，FLM 的局部化损失类变为

$$\mathcal{G}_r^{\text{block}} = \{g_f^{\text{block}} : f \in \mathcal{F}_{\text{DNN}},\ \|f-f_*\|_{\text{pool}} \le r\},$$

其二次过程则由区块平均值替代

$$q_f^{\text{block}}(\mathcal{B}_t) = N^{-1}\sum_i \{f(\boldsymbol{X}_{i,t}) - f_*(\boldsymbol{X}_{i,t})\}^2,$$

从而我们有

$$(\mathbb{E}_T^{\text{block}} - \mathbb{E})q_f^{\text{block}} = \mathbb{E}_{NT}[(f-f_*)^2] - \left\|f-f_*\right\|_{\text{pool}}^2,$$

并且我们使用 FLM 的“$R_n$F”函数的以下分块版本：

$$R_T^{\text{block}}\mathcal{A} = \sup_{a\in\mathcal{A}} \frac{1}{T}\sum_{t=1}^{T}\eta_t a(\mathcal{B}_t).$$

这里，独立的拉德马赫符号 $\eta_t$ 附加在日期上，而不是单个观测值上。詹森不等式保留了两个区块类依赖半径的方差界，并将它们的经验 $L_2$ 度量比较与合并度量联系起来。具体地，记作 $\mathrm{k}a\mathrm{k}_{T,\text{区块}}^2 = \mathbb{E}_T^{\text{block}}[a^2]$ 以及 $\mathrm{k}h\mathrm{k}_{NT}^2 = \mathbb{E}_{NT}[h^2]$。对于 $a=g,\ q$，有界性和

詹森不等式给出

$$\mathbb{E}[(a_f^{\mathrm{block}})^2] \le C\,\|f - f_*\|_{\mathrm{pool}}^2, \qquad \left\|a_f^{\mathrm{block}} - a_{\widetilde{f}}^{\mathrm{block}}\right\|_{T,\mathrm{block}} \le C\left\|f - \widetilde{f}\right\|_{NT},$$

其中 $C$ 与 $N$ 和 $T$ 相互独立。由于单个符号乘在整个分块上，在应用 FLM 的达德利不等式和对称化引理之前，这些度量比较取代了标量收缩。FLM 原始的引理4可应用于所有 $NT$ 协变量评估，而确定性的覆盖步骤不需要日期内独立性，因此

$$\mathcal{N}\big(\delta, \mathcal{F}_{\mathrm{DNN}}|_{\{\boldsymbol{X}_{i,t}\}_{i,t}}, \infty\big) \le \left\{\frac{4eM\,NT}{\delta\,\mathrm{Pdim}(\mathcal{F}_{\mathrm{DNN}})}\right\}^{\mathrm{Pdim}(\mathcal{F}_{\mathrm{DNN}})}, \qquad NT \ge \mathrm{Pdim}(\mathcal{F}_{\mathrm{DNN}}).$$

分块度量比较将这些覆盖转移到局部化损失类和二次函数类，同时仅对覆盖半径进行固定比例的缩放。对分块二次过程进行对称化，并求解由此得到的经验半径不等式，可提供链式方法所需的半径控制。对于每个固定的 $r \ge T^{-1}$，以至少 $1 - 2e^{-\tilde{\gamma}}$ 的概率，有

$$\sup_{\substack{f \in \mathcal{F}_{\mathrm{DNN}} \\ \|f - f_*\|_{\mathrm{pool}} \le r}} \mathbb{E}_{NT}[(f - f_*)^2] \le C\left\{r^2 + \frac{\mathrm{Pdim}(\mathcal{F}_{\mathrm{DNN}})\log(NT) + \tilde{\gamma}}{T}\right\}.$$

随后，FLM 在附录A.2中的范数比较、一步改进和二进局部化给出总体界和合并经验界；需要指出的是，集中性利用了 $T$ 个独立分块，而熵对数项现在是 $\log(NT)$，而非 $\log T$。逼近和多层感知机特例化沿用 FLM 附录A.3 的推导，不作改动，只需在架构和随机分母中将“$n$”替换为 $T$，并保留修改后的熵对数项。由于 $\log(NT) \ge \log T$，FLM 原有的置信度选取可由修改后的界吸收，从而证明该定理。

■

# S5 其他模拟结果

## S5.1 神经网络实现

### S5.1.1 OLS初始化

令 $f_{\mathrm{OLS}}$ 为包含截距和全部 25 个预测变量、在拟合日期上估计得到的合并OLS拟合。为初始化神经网络，我们使用 16 对带符号的仿射指标

$$h_k(\mathbf{x}) = f_{\mathrm{OLS}}(\mathbf{x}) + \mathbf{d}_k'(\mathbf{x} - \bar{\mathbf{x}}_f) + \kappa_k, \qquad k = 1, \ldots, 16,$$

其中 $\bar{\mathbf{x}}f$ 为拟合样本预测变量均值。我们从均匀分布 $[-1,1]$ 中独立抽取斜率和截距偏移量，在各对之间对每个坐标进行中心化，并统一缩放所有偏移量，使其在各对及拟合观测值上的扰动均方根为 $f_{\mathrm{OLS}}$ 拟合样本标准差的 5%（除数为 $360N$）。对于 $h_k$ 这一对，其由 $\mathrm{ReLU}(h_k)$ 和 $\mathrm{ReLU}(-h_k)$ 组成，输出权重分别为 1/16 和 $-1/16$，且神经网络的输出偏置为零。各对求和后可还原 $f_{\mathrm{OLS}}$。对于斜率向量为 w、偏置为 $b$、输出权重为 $v$ 的单元，我们通过将 $(\mathbf{w}, b, v)$ 重新缩放为 $(s\mathbf{w}, sb, v/s)$ 来平衡初始化，其中 $s = (|v|/\|\mathbf{w}\|_2)^{1/2}$。

### S5.1.2 优化与验证

从相同的初始化开始，我们使用全批量 L-BFGS，将拟合均方误差加上 $\alpha$ 乘以权重平方之和作为目标进行最小化，其中 $\alpha \in \{0,\ 10^{-5},\ 10^{-4},\ 10^{-3}\}$。记忆长度为 10，线搜索采用弱 Wolfe 条件，最多进行 20 次迭代；更新次数上限为 250 次，梯度容差和步长容差均为 $10^{-5}$。我们采用耐心值为 25 的提前停止策略，或在达到训练轮次上限、容差要求或发生更新失败时停止。选择验证误差最低的惩罚项。

## S5.2 总体夏普比率

我们设定 $a = 0。5 \times 360^{-3/8}$。对于市场择时，令 $A_0 = N^{-1}\sum_{i=1}^{N}\mu_0(\mathbf{x}_{it})$，并令 $M_j = \mathbb{E}[A_0^j]$，其中 $j = 2,\ 4$。预言机收益为 $A_0(A_0 + \bar{\varepsilon}_t)$，其中 $\bar{\varepsilon}_t = N^{-1}\sum_i \varepsilon_{it}$。对数据生成过程多项式进行积分可得到有限-$N$ 目标值

$$\mathrm{SR}_{0,\mathrm{MT}} = \frac{M_2}{\sqrt{M_4 - M_2^2 + M_2\sigma_\varepsilon^2/N}}, \qquad \sigma_\varepsilon^2 = \mathrm{Var}(\varepsilon_{it}).$$

精确的截尾分布方差与存储的高斯参考方差 $\sigma^2$ 相比，在相对意义上相差约 3。$02 \times 10^{-23}$。因此，截尾处理不会改变报告中的数值参考值。解析市场择时目标与存储值在数值精度范围内一致。

对于多空策略，使用所选带宽对应的预言机权重 w，并设置 $d = \mathbf{w}'\boldsymbol{\mu}_0$。我们将夏普映射应用于每个 $N$ 中 8,192 个独立预测变量面板上 $d$ 和 $d^2 + \sigma_\varepsilon^2\|\mathbf{w}\|_2^2$ 的平均值。对于不同载荷和带宽，我们复用相同的面板。根据存储的条件矩和协方差重新计算夏普映射及其蒙特卡洛标准误，可以复现报告中的参考值。最大的平滑多空参考标准误为 0。001053。被动等权总体夏普比率为零，因此该策略

并且微分零假设在数值上重合。

## S5.3 其他 $\tau$ 值的结果

表 S2

*LS* 拒绝频率（$\tau = 0.0025$）

| | | LS | | LS–EW | |
|---|---|---|---|---|---|
| $\lambda$ | $N$ | C | D | C | D |
| 0.25 | 50 | 9.2 | 13.3 | 8.2 | 11.5 |
| 0.25 | 150 | 6.1 | 13.2 | 6.6 | 12.3 |
| 0.25 | 250 | 4.8 | 9.2 | 4.1 | 10.0 |
| 0.25 | 500 | 5.7 | 8.9 | 6.5 | 8.2 |
| 0.40 | 50 | 9.0 | 15.0 | 7.1 | 13.5 |
| 0.40 | 150 | 5.9 | 9.4 | 5.7 | 9.3 |
| 0.40 | 250 | 4.5 | 8.9 | 4.4 | 7.9 |
| 0.40 | 500 | 4.7 | 8.6 | 5.5 | 8.3 |
| 0.55 | 50 | 9.0 | 16.4 | 7.3 | 13.2 |
| 0.55 | 150 | 6.9 | 12.0 | 7.3 | 11.1 |
| 0.55 | 250 | 6.2 | 9.7 | 5.5 | 8.8 |
| 0.55 | 500 | 5.6 | 9.5 | 6.1 | 9.5 |

各项数据为 1,000 次蒙特卡罗重复试验所得的拒绝百分比，名义显著性水平为 5%。C 和 D 分别表示传统检验和去偏检验。

表 S3

*LS* 拒绝频率（$\tau = 0.005$）

| | | LS | | LS–EW | |
|---|---|---|---|---|---|
| $\lambda$ | $N$ | C | D | C | D |
| 0.25 | 50 | 9.5 | 12.3 | 8.6 | 10.0 |
| 0.25 | 150 | 6.8 | 10.6 | 6.6 | 9.7 |
| 0.25 | 250 | 5.5 | 7.8 | 4.0 | 6.4 |
| 0.25 | 500 | 6.1 | 7.3 | 6.5 | 6.0 |
| 0.40 | 50 | 9.5 | 13.4 | 7.5 | 10.8 |
| 0.40 | 150 | 6.2 | 7.9 | 5.6 | 7.5 |
| 0.40 | 250 | 4.8 | 8.2 | 4.4 | 6.6 |
| 0.40 | 500 | 4.9 | 7.0 | 6.0 | 6.2 |
| 0.55 | 50 | 9.6 | 13.2 | 7.7 | 11.2 |
| 0.55 | 150 | 7.2 | 9.5 | 7.3 | 8.4 |
| 0.55 | 250 | 6.1 | 7.8 | 5.5 | 7.3 |
| 0.55 | 500 | 5.6 | 7.9 | 6.2 | 7.8 |

数值为基于 1,000 次蒙特卡罗重复试验得到的拒绝百分比，名义显著性水平为 5%。C 和 D 分别表示传统检验和去偏检验。

表 S4*LS* 拒绝频率（$\tau = 0.01$）

| | | LS | | LS–EW | |
|---|---|---|---|---|---|
| $\lambda$ | $N$ | C | D | C | D |
| 0.25 | 50 | 10.0 | 10.8 | 9.3 | 9.4 |
| 0.25 | 150 | 7.3 | 8.4 | 6.9 | 7.2 |
| 0.25 | 250 | 5.9 | 7.1 | 4.5 | 6.9 |
| 0.25 | 500 | 6.2 | 5.8 | 6.5 | 5.5 |
| 0.40 | 50 | 10.2 | 11.1 | 8.1 | 7.8 |
| 0.40 | 150 | 6.4 | 6.9 | 6.2 | 6.3 |
| 0.40 | 250 | 5.8 | 6.6 | 4.9 | 6.0 |
| 0.40 | 500 | 6.5 | 6.3 | 5.6 | 5.9 |
| 0.55 | 50 | 10.8 | 12.0 | 8.8 | 9.3 |
| 0.55 | 150 | 8.2 | 7.3 | 7.6 | 5.8 |
| 0.55 | 250 | 6.8 | 6.8 | 5.8 | 6.5 |
| 0.55 | 500 | 6.4 | 7.2 | 6.8 | 7.5 |

条目表示来自 1,000 次蒙特卡罗重复试验的拒绝百分比，名义显著性水平为 5%。C 和 D 分别表示传统检验和去偏检验。

表 S5*LS* 拒绝频率（$\tau = 0.02$）

| | | LS | | LS–EW | |
|---|---|---|---|---|---|
| $\lambda$ | $N$ | C | D | C | D |
| 0.25 | 50 | 11.6 | 9.6 | 10.6 | 8.6 |
| 0.25 | 150 | 8.8 | 8.2 | 7.4 | 7.3 |
| 0.25 | 250 | 7.0 | 6.8 | 5.7 | 6.7 |
| 0.25 | 500 | 7.6 | 5.7 | 7.0 | 6.2 |
| 0.40 | 50 | 11.4 | 8.4 | 9.3 | 6.9 |
| 0.40 | 150 | 8.1 | 7.0 | 7.5 | 5.8 |
| 0.40 | 250 | 7.0 | 6.0 | 6.1 | 5.3 |
| 0.40 | 500 | 7.6 | 5.4 | 6.7 | 4.9 |
| 0.55 | 50 | 12.8 | 10.1 | 9.7 | 7.3 |
| 0.55 | 150 | 10.4 | 5.3 | 9.8 | 4.8 |
| 0.55 | 250 | 8.8 | 6.7 | 7.5 | 5.8 |
| 0.55 | 500 | 9.0 | 6.1 | 7.6 | 6.1 |

条目为 1,000 次蒙特卡罗重复试验所得的拒绝百分比，名义显著性水平为 5%。C 和 D 分别表示传统检验和去偏检验。

表 S6

*LS* 拒绝频率（$\tau = 0.05$）

| | | LS | | LS–EW | |
|---|---|---|---|---|---|
| $\lambda$ | $N$ | C | D | C | D |
| 0.25 | 50 | 18.6 | 9.9 | 14.3 | 8.2 |
| 0.25 | 150 | 15.8 | 6.4 | 12.5 | 6.7 |
| 0.25 | 250 | 12.0 | 6.9 | 10.6 | 6.1 |
| 0.25 | 500 | 13.2 | 6.6 | 11.4 | 5.5 |
| 0.40 | 50 | 19.5 | 10.6 | 12.8 | 8.0 |
| 0.40 | 150 | 14.6 | 7.4 | 13.4 | 7.1 |
| 0.40 | 250 | 13.4 | 6.0 | 10.9 | 5.7 |
| 0.40 | 500 | 14.4 | 4.8 | 12.6 | 5.0 |
| 0.55 | 50 | 19.8 | 10.9 | 15.2 | 7.9 |
| 0.55 | 150 | 18.4 | 6.2 | 13.2 | 5.5 |
| 0.55 | 250 | 16.5 | 7.5 | 12.2 | 5.9 |
| 0.55 | 500 | 15.2 | 4.9 | 13.5 | 5.3 |

条目为 1,000 次蒙特卡罗重复试验所得的拒绝百分比，名义显著性水平为 5%。C 和 D 分别表示传统检验和去偏差检验。

表 S7*LS* 拒绝频率（$\tau = 0.1$）

| | | LS | | LS–EW | |
|---|---|---|---|---|---|
| $\lambda$ | $N$ | C | D | C | D |
| 0.25 | 50 | 26.3 | 10.4 | 20.1 | 7.2 |
| 0.25 | 150 | 24.2 | 5.1 | 17.6 | 6.1 |
| 0.25 | 250 | 19.2 | 5.1 | 16.4 | 5.2 |
| 0.25 | 500 | 19.2 | 5.8 | 17.4 | 4.6 |
| 0.40 | 50 | 28.7 | 9.8 | 19.0 | 8.5 |
| 0.40 | 150 | 23.7 | 7.1 | 18.6 | 6.1 |
| 0.40 | 250 | 21.2 | 5.0 | 17.5 | 4.6 |
| 0.40 | 500 | 20.4 | 4.8 | 18.4 | 5.3 |
| 0.55 | 50 | 29.4 | 11.1 | 18.8 | 7.5 |
| 0.55 | 150 | 24.8 | 5.6 | 19.1 | 4.7 |
| 0.55 | 250 | 23.5 | 5.6 | 18.0 | 4.5 |
| 0.55 | 500 | 20.2 | 4.2 | 18.6 | 4.6 |

各项数值为基于 1,000 次蒙特卡罗重复试验得出的拒绝百分比，名义显著性水平为 5%。C 和 D 分别表示传统检验和去偏检验。

表 S8

*LS* 拒绝频率（$\tau = 0.2$）

| | | LS | | LS–EW | |
|---|---|---|---|---|---|
| $\lambda$ | $N$ | C | D | C | D |
| 0.25 | 50 | 29.1 | 6.8 | 22.8 | 6.5 |
| 0.25 | 150 | 24.0 | 3.9 | 18.3 | 4.3 |
| 0.25 | 250 | 18.1 | 4.5 | 15.9 | 4.3 |
| 0.25 | 500 | 16.7 | 4.8 | 15.8 | 4.6 |
| 0.40 | 50 | 28.8 | 5.6 | 20.5 | 5.3 |
| 0.40 | 150 | 22.0 | 5.3 | 18.5 | 5.3 |
| 0.40 | 250 | 19.8 | 3.7 | 16.2 | 3.2 |
| 0.40 | 500 | 16.0 | 5.7 | 14.8 | 4.6 |
| 0.55 | 50 | 29.4 | 7.1 | 18.9 | 5.2 |
| 0.55 | 150 | 21.3 | 4.1 | 15.5 | 3.7 |
| 0.55 | 250 | 18.5 | 4.5 | 14.5 | 4.2 |
| 0.55 | 500 | 14.8 | 4.0 | 14.8 | 4.1 |

数值为 1,000 次蒙特卡罗重复试验所得的拒绝百分比，名义显著性水平为 5%。C 和 D 分别表示传统检验和去偏检验。

表 S9

*LS* 拒绝频率（$\tau = 0.3$）

| | | LS | | LS–EW | |
|---|---|---|---|---|---|
| $\lambda$ | $N$ | C | D | C | D |
| 0.25 | 50 | 23.6 | 4.9 | 18.0 | 5.9 |
| 0.25 | 150 | 16.5 | 3.8 | 13.7 | 4.1 |
| 0.25 | 250 | 13.2 | 4.6 | 10.6 | 4.8 |
| 0.25 | 500 | 11.2 | 6.1 | 11.2 | 5.2 |
| 0.40 | 50 | 22.1 | 4.5 | 14.7 | 4.5 |
| 0.40 | 150 | 14.7 | 4.7 | 13.2 | 5.3 |
| 0.40 | 250 | 13.4 | 3.6 | 10.9 | 4.2 |
| 0.40 | 500 | 11.5 | 6.0 | 9.7 | 5.4 |
| 0.55 | 50 | 21.9 | 5.6 | 14.1 | 4.8 |
| 0.55 | 150 | 13.4 | 5.1 | 11.3 | 5.0 |
| 0.55 | 250 | 12.3 | 4.5 | 9.2 | 5.0 |
| 0.55 | 500 | 9.4 | 4.8 | 8.8 | 4.8 |

条目为基于 1,000 次蒙特卡罗重复试验、名义显著性水平为 5% 时的拒绝百分比。C 和 D 分别表示传统检验和去偏检验。

表 S10

*LS* 拒绝频率（$\tau = 0.4$）

| | | LS | | LS–EW | |
|---|---|---|---|---|---|
| $\lambda$ | $N$ | C | D | C | D |
| 0.25 | 50 | 18.4 | 4.7 | 15.4 | 5.3 |
| 0.25 | 150 | 12.5 | 4.0 | 10.2 | 4.6 |
| 0.25 | 250 | 8.3 | 5.3 | 7.5 | 5.7 |
| 0.25 | 500 | 8.3 | 5.9 | 8.7 | 5.4 |
| 0.40 | 50 | 17.4 | 4.8 | 11.5 | 4.7 |
| 0.40 | 150 | 11.0 | 5.3 | 9.6 | 5.3 |
| 0.40 | 250 | 9.6 | 4.4 | 7.7 | 4.7 |
| 0.40 | 500 | 8.5 | 6.1 | 7.7 | 5.4 |
| 0.55 | 50 | 16.6 | 4.6 | 11.7 | 5.1 |
| 0.55 | 150 | 10.7 | 5.4 | 8.6 | 5.9 |
| 0.55 | 250 | 9.2 | 4.7 | 6.8 | 5.1 |
| 0.55 | 500 | 7.0 | 5.7 | 7.1 | 5.9 |

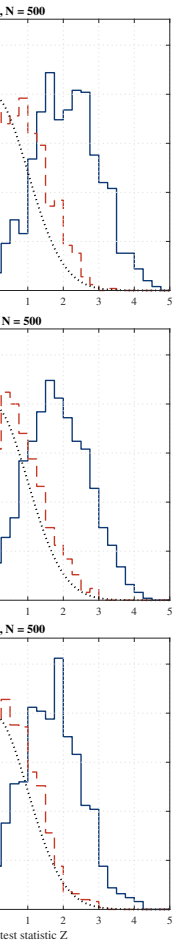
各项数值为 1,000 次蒙特卡罗重复试验所得的拒绝百分比，名义显著性水平为 5%。C 和 D 分别表示传统检验和去偏检验。

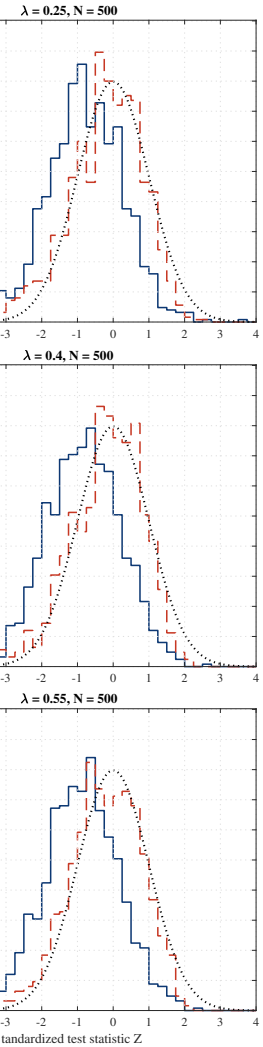
表 S11*LS* 拒绝频率（$\tau = 0.8$）

| | | LS | | LS–EW | |
|---|---|---|---|---|---|
| $\lambda$ | $N$ | C | D | C | D |
| 0.25 | 50 | 10.9 | 5.4 | 9.9 | 4.5 |
| 0.25 | 150 | 6.8 | 5.1 | 7.2 | 5.3 |
| 0.25 | 250 | 4.5 | 5.5 | 6.1 | 5.6 |
| 0.25 | 500 | 6.8 | 5.0 | 7.0 | 5.6 |
| 0.40 | 50 | 10.7 | 5.9 | 8.0 | 4.7 |
| 0.40 | 150 | 7.0 | 4.6 | 6.4 | 5.1 |
| 0.40 | 250 | 5.9 | 4.8 | 5.9 | 4.7 |
| 0.40 | 500 | 7.2 | 6.0 | 6.6 | 5.4 |
| 0.55 | 50 | 12.3 | 4.9 | 8.8 | 4.7 |
| 0.55 | 150 | 6.7 | 4.9 | 6.1 | 6.1 |
| 0.55 | 250 | 6.5 | 5.7 | 4.8 | 4.6 |
| 0.55 | 500 | 5.9 | 5.6 | 6.5 | 5.7 |

表中数据为 1,000 次蒙特卡罗重复试验所得的拒绝百分比，名义显著性水平为 5%。C 和 D 分别表示常规检验和去偏检验。

ndard normal density.






ard normal density.

# S6 其他实证结果

## S6.1 预测调优

对于弹性网络，我们使用混合参数 0.5，以及从 $10^{-4}$ 到 $10^{-1}$的 20 个对数等距惩罚项。主成分回归使用 1、2、3、5、10、20、30、40、50、75 或 100 个成分。偏最小二乘使用 1 至 6、8 或 10 个成分。神经网络使用 Adam，批量大小为 10,000，轮数上限为 100 轮，耐心值为 5 次验证检查。学习率和$l_2$惩罚项的取值范围均为 $10^{-4}$ 0.01,0.1。我们根据三网络集成的验证均方误差选择这一对参数。

## S6.2 被动基准和比较夏普推断

表 S12 报告了三个被动基准的年化夏普比率。表 S13 和表 S14 报告了在 $\tau = 0.001$ 下，标普500指数择时和多空策略的估计的年化策略减基准夏普差距。正文中的表 4 和表 6 报告了相应的$t$统计量及其去偏后的变化。

表 S12被动基准的年化夏普比率

| 基准 | 年度 $\widehat{SR}$ | $t$ 相对于零值 | 月份 |
|---|---|---|---|
| CRSP等权 | 0.517 | 2.57 | 420 |
| CRSP市值加权 | 0.564 | 2.90 | 420 |
| 观测面板标准普尔 | 0.608 | 3.16 | 420 |

注：评估样本为1987年1月至2021年12月（$T = 420$）。夏普比率按$\sqrt{12}$年化。所有 $t$统计量均采用带五阶滞后的Newey–West HAC标准误。每个 $t$统计量均检验相应基准的夏普比率是否为零。观测面板标准普尔是由预测数据中的滞后市场权益加权的匹配标普成分股构成的被动投资组合。CRSP等权和市值加权使用所提供的含股息指数序列。所提供的指数序列使用30天期国库券利率转换为超额收益。

表 S13标普*500*指数择时与被动基准之间的年化夏普差异

| | CRSP等权 | | CRSP市值加权 | | 观测面板标准普尔 | | |
|---|---|---|---|---|---|---|---|
| 方法 | 常规 | Deb. | 常规 | Deb. | 常规 | Deb. | Δ 估计值 |
| OLS-3 | −0。062 | −0.075 | −0。109 | −0。122 | −0。153 | −0。166 | −0。014 |
| ENet | 0.008 | 0.032 | −0。039 | −0。015 | −0。083 | −0。059 | +0.024 |
| PCR | −0.008 | 0.029 | −0.055 | −0.017 | −0。099 | −0。061 | +0。037 |
| PLS | −0.047 | −0.042 | −0。094 | −0。088 | −0.138 | −0。133 | +0。006 |
| NN1 | 0.096 | 0.148 | 0.050 | 0.101 | 0.005 | 0.057 | +0.052 |
| NN2 | −0。026 | −0.030 | −0.072 | −0。077 | −0.116 | −0.121 | −0。005 |
| NN3 | 0.072 | 0.071 | 0.025 | 0.024 | −0。019 | −0.020 | −0。001 |
| NN4 | −0。006 | −0。021 | −0。052 | −0。067 | −0。096 | −0。111 | −0.015 |
| NN5 | 0.004 | −0。013 | −0。043 | −0。059 | −0.087 | −0。103 | −0.017 |

注： 常规和去偏分别报告传统和去偏后的年化策略减基准夏普估计值。正值表示策略更优。变化值为去偏估计减去常规估计，在四舍五入前计算。由于两种方法对每个基准均减去相同的被动夏普比率，因此该变化值对三组比较均相同（表 S12）。

表 S14来自被动基准的年化多空夏普差异（$\tau = 0.001$）

| | CRSP等权 | | CRSP市值加权 | | 观测面板标准普尔 | | |
|---|---|---|---|---|---|---|---|
| 方法 | 常规 | Deb. | 常规 | Deb. | 常规 | Deb. | Δ 估计值 |
| OLS-3 | 0.273 | 0.247 | 0.227 | 0.201 | 0.183 | 0.156 | −0。026 |
| ENet | 1.175 | 1.179 | 1.128 | 1.133 | 1.084 | 1.089 | +0。004 |
| PCR | 1.107 | 1.084 | 1.060 | 1.037 | 1.016 | 0.993 | −0。023 |
| PLS | 1.361 | 1.298 | 1.315 | 1.251 | 1.271 | 1.207 | −0.063 |
| NN1 | 1.685 | 1.655 | 1.638 | 1.608 | 1.594 | 1.564 | −0.030 |
| NN2 | 0.950 | 1.421 | 0.903 | 1.374 | 0.859 | 1.330 | +0.471 |
| NN3 | 0.435 | 0.419 | 0.389 | 0.373 | 0.345 | 0.329 | −0.016 |
| NN4 | −0。407 | −0。174 | −0。453 | −0。221 | −0。497 | −0。265 | +0。232 |
| NN5 | −0。524 | −0。602 | −0.571 | −0。648 | −0.615 | −0。692 | −0。078 |

注： 常规和去偏分别报告传统和去偏后的年化策略与基准夏普比率差估计。正值表示策略更优。变化值为去偏估计减去传统估计，计算时未进行四舍五入。由于两种方法对每个基准均减去相同的被动夏普比率（表 S12），因此三个比较的变化值相同。

## S6.3 多空带宽敏感性

表 S15 和表 S16 报告了所有五种带宽下的夏普估计值，以及针对零值进行检验的 $t$ 统计量。其余表格报告了在另外四种带宽下与三个被动基准进行比较的 $t$ 统计量。表注解释了为什么 NN4 在最小的三种带宽下无法定义去偏统计量。

表 S15不同带宽下的年化多空夏普比率

| 方法 | 0.0001 | 0.00025 | 0.0005 | 0.001 | 0.0025 |
|---|---|---|---|---|---|
| 面板 $A$：传统方法 | | | | | |
| OLS-3 | 0.792 | 0.795 | 0.796 | 0.791 | 0.767 |
| ENet | 1.718 | 1.715 | 1.707 | 1.692 | 1.638 |
| PCR | 1.467 | 1.541 | 1.596 | 1.624 | 1.613 |
| PLS | 1.891 | 1.893 | 1.891 | 1.879 | 1.828 |
| NN1 | 2.204 | 2.201 | 2.205 | 2.202 | 2.112 |
| NN2 | 1.685 | 1.652 | 1.595 | 1.467 | 1.197 |
| NN3 | 1.033 | 1.014 | 1.005 | 0.952 | 0.810 |
| NN4 | 0.200 | 0.152 | 0.132 | 0.111 | 0.102 |
| NN5 | 0.121 | 0.042 | 0.001 | -0.007 | -0.001 |
| 面板 $B$：去偏 | | | | | |
| OLS-3 | 0.858 | 0.814 | 0.795 | 0.764 | 0.610 |
| ENet | 1.761 | 1.723 | 1.709 | 1.697 | 1.599 |
| PCR | 1.530 | 1.564 | 1.598 | 1.601 | 1.478 |
| PLS | 1.794 | 1.798 | 1.819 | 1.815 | 1.676 |
| NN1 | 2.044 | 2.122 | 2.149 | 2.172 | 2.102 |
| NN2 | 1.507 | 1.801 | 1.918 | 1.938 | 1.668 |
| NN3 | 1.364 | 0.837 | 0.824 | 0.937 | 0.944 |
| NN4 | — | — | — | 0.343 | 0.123 |
| NN5 | 0.043 | 0.041 | 0.017 | -0.084 | -0.105 |

注：评估样本为1987年1月至2021年12月（$T = 420$）。夏普比率按年化处理，方法为$\sqrt{12}$。传统推断使用实现收益及其平方项；去偏推断使用正交一阶和二阶矩得分。列标题为平滑带宽 $\tau$。NN4 中最小的三个带宽处的短横线表示负的校正矩方差，即夏普比率估计值未定义。各项均为年化夏普比率估计值。

表 S16按带宽划分的多空投资组合零夏普比率检验

| 方法 | 0.0001 | 0.00025 | 0.0005 | 0.001 | 0.0025 |
|---|---|---|---|---|---|
| 面板 $A$：传统方法 | | | | | |
| OLS-3 | 4.02 | 4.02 | 4.03 | 4.01 | 3.90 |
| ENet | 7.90 | 7.90 | 7.87 | 7.80 | 7.59 |
| PCR | 6.48 | 6.76 | 6.92 | 7.02 | 7.06 |
| PLS | 7.47 | 7.48 | 7.47 | 7.43 | 7.38 |
| NN1 | 9.90 | 9.93 | 9.95 | 9.90 | 9.47 |
| NN2 | 6.14 | 6.12 | 6.01 | 5.60 | 4.69 |
| NN3 | 5.17 | 5.12 | 5.02 | 4.51 | 3.77 |
| NN4 | 1.02 | 0.77 | 0.66 | 0.55 | 0.51 |
| NN5 | 0.77 | 0.23 | 0.00 | -0.04 | -0.01 |
| 面板 $B$：去偏 | | | | | |
| OLS-3 | 4.52 | 4.54 | 4.43 | 4.04 | 2.61 |
| ENet | 8.13 | 8.73 | 8.91 | 8.82 | 7.98 |
| PCR | 6.85 | 7.56 | 7.69 | 7.64 | 6.87 |
| PLS | 8.32 | 8.62 | 8.59 | 8.39 | 7.46 |
| NN1 | 9.88 | 10.29 | 10.61 | 10.95 | 10.00 |
| NN2 | 5.87 | 6.75 | 6.99 | 6.63 | 5.31 |
| NN3 | 0.81 | 3.26 | 4.70 | 5.20 | 4.13 |
| NN4 | — | — | — | 0.37 | 0.31 |
| NN5 | 1.70 | 0.83 | 0.17 | -0.22 | -0.21 |

注：评估样本为1987年1月至2021年12月（$T = 420$）。夏普比率按 $\sqrt{12}$年化。所有 $t$统计量均使用滞后五阶的Newey–West HAC标准误。传统推断使用实现收益及其平方项；去偏推断使用正交一阶和二阶矩得分。列标题为平滑带宽 $\tau$。NN4在最小的三个带宽处的破折号表示校正矩方差为负，即夏普比率估计值未定义。表中数值为检验策略的夏普比率是否为零的 $t$统计量。

表 S17$t$：多空夏普差异相对于被动基准的统计量（$\tau = 0.0001$）

| 方法 | CRSP EW | CRSP VW | 观测面板 S&P |
|---|---|---|---|
| 面板 $A$：传统方法 | | | |
| OLS-3 | 0.87 | 0.75 | 0.61 |
| ENet | 3.65 | 3.75 | 3.67 |
| PCR | 2.91 | 2.78 | 2.66 |
| PLS | 3.91 | 3.97 | 3.91 |
| NN1 | 5.38 | 5.53 | 5.46 |
| NN2 | 3.16 | 3.14 | 3.09 |
| NN3 | 1.75 | 1.63 | 1.50 |
| NN4 | -1.27 | -1.53 | -1.70 |
| NN5 | -1.64 | -1.81 | -1.98 |
| 面板 $B$：去偏 | | | |
| OLS-3 | 1.10 | 1.00 | 0.86 |
| ENet | 3.84 | 3.94 | 3.86 |
| PCR | 3.02 | 2.96 | 2.86 |
| PLS | 3.96 | 4.04 | 3.98 |
| NN1 | 5.04 | 5.16 | 5.07 |
| NN2 | 2.72 | 2.68 | 2.59 |
| NN3 | 0.50 | 0.48 | 0.45 |
| NN4 | — | — | — |
| NN5 | -2.37 | -2.66 | -2.91 |

注：评估样本为1987年1月至2021年12月（$T = 420$）。夏普比率按$\sqrt{12}$年化。所有 $t$统计量均使用五阶滞后的Newey–West HAC标准误。传统推断使用实现收益及其平方项；去偏推断使用正交一阶和二阶矩得分。差值为正表示策略表现更优。每个条目检验 $H_0$： $SR_{\text{strategy}} - SR_{\text{benchmark}} = 0$。这些是双侧检验；通常的渐近5%临界值为 ±1。96。NN4中的短横线表示去偏矩方差非正；其夏普比率及检验未定义。观测面板标普是由预测数据中的滞后市场权益加权的匹配标普成分股被动投资组合。CRSP等权和价值加权使用所提供的含股息指数序列。所提供的指数序列使用30天期国库券利率转换为超额收益。

表 S18

*t*统计量：相对于被动基准的多空夏普差异（$\tau = 0.00025$）

| 方法 | CRSP EW | CRSP VW | 观测面板 S&P |
|---|---|---|---|
| 面板 *A*：传统方法 | | | |
| OLS-3 | 0.88 | 0.76 | 0.62 |
| ENet | 3.65 | 3.74 | 3.67 |
| PCR | 3.12 | 3.02 | 2.91 |
| PLS | 3.92 | 3.98 | 3.92 |
| NN1 | 5.39 | 5.53 | 5.46 |
| NN2 | 3.07 | 3.06 | 3.01 |
| NN3 | 1.69 | 1.55 | 1.43 |
| NN4 | -1.50 | -1.75 | -1.91 |
| NN5 | -1.91 | -2.04 | -2.20 |
| 面板 *B*：去偏 | | | |
| OLS-3 | 0.98 | 0.86 | 0.72 |
| ENet | 3.86 | 3.96 | 3.88 |
| PCR | 3.28 | 3.24 | 3.14 |
| PLS | 4.07 | 4.15 | 4.09 |
| NN1 | 5.31 | 5.43 | 5.34 |
| NN2 | 3.48 | 3.48 | 3.39 |
| NN3 | 1.04 | 0.95 | 0.81 |
| NN4 | — | — | — |
| NN5 | -2.36 | -2.63 | -2.86 |

注：评估样本为1987年1月至2021年12月（$T = 420$）。夏普比率按$\sqrt{12}$年化。所有 *t*统计量均使用五阶滞后的Newey–West HAC标准误。传统推断使用实现收益及其平方项；去偏推断使用正交一阶和二阶矩得分。差值为正表示策略表现更优。每个条目检验 $H_0$： $SR_{\text{strategy}} - SR_{\text{benchmark}} = 0$。这些是双侧检验；通常的渐近5%临界值为 ±1。96。NN4中的短横线表示去偏矩方差非正；其夏普比率及检验未定义。观测面板标普是由预测数据中的滞后市场权益加权的匹配标普成分股被动投资组合。CRSP等权和价值加权使用所提供的含股息指数序列。所提供的指数序列使用30天期国库券利率转换为超额收益。

表 S19

*t*-统计量：多空夏普差异相对于被动基准（$\tau = 0.0005$）

| 方法 | CRSP EW | CRSP VW | 观测面板 S&P |
|---|---|---|---|
| 面板 *A*：传统方法 | | | |
| OLS-3 | 0.89 | 0.77 | 0.63 |
| ENet | 3.63 | 3.72 | 3.64 |
| PCR | 3.25 | 3.18 | 3.09 |
| PLS | 3.92 | 3.97 | 3.91 |
| NN1 | 5.41 | 5.57 | 5.50 |
| NN2 | 2.91 | 2.91 | 2.86 |
| NN3 | 1.63 | 1.50 | 1.38 |
| NN4 | -1.58 | -1.81 | -1.96 |
| NN5 | -2.04 | -2.16 | -2.31 |
| 面板 *B*：去偏 | | | |
| OLS-3 | 0.91 | 0.80 | 0.65 |
| ENet | 3.87 | 3.98 | 3.90 |
| PCR | 3.45 | 3.41 | 3.30 |
| PLS | 4.11 | 4.20 | 4.14 |
| NN1 | 5.45 | 5.58 | 5.49 |
| NN2 | 3.76 | 3.79 | 3.72 |
| NN3 | 1.20 | 1.10 | 0.93 |
| NN4 | — | — | — |
| NN5 | -2.34 | -2.54 | -2.74 |

注：评估样本为1987年1月至2021年12月（$T = 420$）。夏普比率按$\sqrt{12}$年化。所有 *t*统计量均使用五阶滞后的 Newey–West HAC标准误。传统推断使用实现收益及其平方项；去偏推断使用正交一阶和二阶矩得分。差值为正表示策略表现更优。每个条目检验 $H_0$： $SR_{\text{strategy}} - SR_{\text{benchmark}} = 0$。这些是双侧检验；常用的渐近5%临界值为 ±1。96。NN4的破折号表示去偏矩方差非正；夏普比率及其检验未定义。观测面板标准普尔是由预测数据中的滞后市场权益加权的匹配标普成分股被动投资组合。CRSP等权和价值加权采用提供的含股息指数序列。使用30天期国库券利率将提供的指数序列转换为超额收益。

表 S20

*t*统计量：被动基准的多空夏普差异（$\tau = 0.0025$）

| 方法 | CRSP EW | CRSP VW | 观测面板 S&P |
|---|---|---|---|
| 面板 *A*：传统方法 | | | |
| OLS-3 | 0.81 | 0.68 | 0.54 |
| ENet | 3.41 | 3.49 | 3.41 |
| PCR | 3.26 | 3.23 | 3.15 |
| PLS | 3.76 | 3.81 | 3.75 |
| NN1 | 5.14 | 5.31 | 5.24 |
| NN2 | 1.85 | 1.81 | 1.72 |
| NN3 | 0.94 | 0.81 | 0.69 |
| NN4 | -1.70 | -1.90 | -2.03 |
| NN5 | -2.06 | -2.18 | -2.33 |
| 面板 *B*：去偏 | | | |
| OLS-3 | 0.27 | 0.14 | 0.01 |
| ENet | 3.42 | 3.48 | 3.40 |
| PCR | 2.99 | 2.93 | 2.83 |
| PLS | 3.53 | 3.56 | 3.48 |
| NN1 | 5.34 | 5.48 | 5.41 |
| NN2 | 2.81 | 2.85 | 2.79 |
| NN3 | 1.33 | 1.24 | 1.13 |
| NN4 | -0.98 | -1.10 | -1.19 |
| NN5 | -1.22 | -1.27 | -1.35 |

注：评估样本为1987年1月至2021年12月（$T = 420$）。夏普比率按 $\sqrt{12}$年化。所有 *t*统计量均使用五阶滞后的 Newey–West HAC标准误。传统推断使用实现收益及其平方项；去偏推断使用正交一阶和二阶矩得分。差值为正表示策略表现更优。每个条目检验 $H_0$： $SR_{\text{strategy}} - SR_{\text{benchmark}} = 0$。这些是双侧检验；常用的渐近5%临界值为 $\pm 1.96$。观测面板标准普尔是由预测数据中的滞后市场权益加权的匹配标普成分股被动投资组合。CRSP等权和价值加权采用提供的含股息指数序列。使用30天期国库券利率将提供的指数序列转换为超额收益。