高频尾部风险测度：基于 L_1 正则化极值回归方法和单位根预测变量

Julien Hambuckers†、Li Sun†⋆ 和 Luca Trapin‡

† 比利时列日大学列日高等商学院
‡ 博洛尼亚大学，意大利

摘要。 我们研究高频金融市场中的尾部风险动态及其与交易活动和市场不确定性的联系。我们提出一种动态极值回归模型，该模型兼容平稳和局部单位根预测变量，从而恰当地捕捉高频极端损失分布随时间变化的特征。为刻画交易活动和市场不确定性，我们考察若干波动率和流动性预测变量，并提出一种两步自适应L_1正则化最大似然估计量，以筛选最合适的变量。我们证明，所提出的估计量在筛选平稳和局部单位根预测变量方面具有预言机性质，并通过广泛的模拟研究表明其具有良好的有限样本性质。我们利用 42 个流动性和波动率预测变量研究九只流动性较高的大型美国股票的高频极端损失，发现当流动性波动率和波动率较高时，较低的价格冲击能够很好地预测极端损失的严重程度。

关键词： 高频金融数据 · 超阈值峰值法（POT）· 时变广义帕累托分布· L_1-正则化最大似然估计 · 非平稳变量选择

arXiv:2301.01362v1 [econ.EM] 3 Jan 2023

1 引言

对高频尾部风险进行测量，对于市场参与者和监管机构而言已变得至关重要（Weller，2017年）。尽管人们已投入大量精力研究低频尾部风险的测量（Nieto 和 Ruiz，2016），但针对高频风险测量的研究却寥寥无几，相关研究可见 Giot（2005年）、Dionne 等（2009年）以及 Chavez-Demoulin 和 Davison（2012）。此外，尽管这些模型可能非常准确，但它们以“约化形式”解释尾部风险的演变，即利用时间序列的持久性，采用自回归项。因此，这些模型未能深入揭示驱动尾部风险的因素及其结构。正如在低频下理解尾部风险的宏观经济决定因素是一个重要问题（Massacci，2017），理解市场不确定性和交易活动如何影响高频尾部风险也十分重要。

从市场微观结构的角度来看，尽管高频交易的日益普及改善了交易成本和流动性（Hendershott et al., 2011），但人们也怀疑它是短时间内极端价格变动更加频繁的原因（Brogaard et al., 2018）。这类极端波动往往源于市场已经承压时发起的激进方向性做市活动。Brogaard et al. (2018) 发现，市场范围内的极端冲击可能触发高频流动性提供者的风险控制措施，促使他们退出市场以降低风险敞口。同样，Kirilenko et al. (2017) 发现，在2010年闪电崩盘引发的市场动荡期间，许多高频流动性提供者退出市场，进而加剧了价格下跌。因此，研究市场不确定性和交易活动如何影响极端损失，有助于更深入地理解高频尾部风险的演变；本文提出了相应的计量经济学方法来开展这项研究。

我们考虑一个动态极值回归框架（Chavez-Demoulin et al., 2016; Massacci, 2017; Schwaab et al., 2021），其中假设极端损失的分布可由广义帕累托分布（GPD）很好地近似，该分布的时变参数由外生预测变量和自回归

⋆ 通讯作者：ir.li.sun@gmail.com

项驱动。为评估市场不确定性和交易活动对极端损失的影响，我们考虑若干波动率预测变量，用以表征市场不确定性；同时考虑流动性预测变量，用以刻画交易活动。尽管极值回归技术已广泛应用于金融领域（Chavez-Demoulin et al., 2016; Hambuckers et al., 2018; Bee et al., 2019），我们的研究仍面临新的挑战：（i）金融文献提出了多种波动率和流动性度量，因此我们需要解决变量选择问题，以识别能够捕捉影响极端损失的交易活动关键方面的预测变量，并提高尾部风险的预测准确性；（ii）高频观测到的波动率和流动性度量具有很强的持久性和季节性，从而违反了使用最大似然估计量（MLE）进行推断所需的经典平稳性假设。为解决这些问题，我们提出一种两步自适应 L_1-正则化最大似然估计量（ALMLE），可使用平稳和局部单位根预测变量进行变量选择（Lee et al., 2022），并证明其具有预言机性质。

我们研究了 42 个流动性和波动率指标对 2006 至 2014 年间观测到的九只大型高流动性美国股票的高频极端损失分布的影响。我们发现，在波动率的波动率和流动性波动率较高的时期，价格冲击较低（Goyenko et al., 2009）能够很好地预测尾部风险的严重程度，该程度由广义帕累托分布的形状参数衡量。这一发现与 Brogaard et al. (2018) 的证据一致：发生大规模不确定性冲击后，做市商的流动性供给不足以满足流动性需求，而做市商急于退出市场以降低风险敞口，进一步放大了极端价格变动。只有采用两步 ALMLE，才能揭示这一模式，因为标准最大似然估计会发现几乎所有预测变量都显著。为验证我们的估计策略，我们进行了样本外 VaR 预测分析，发现估计模型在样本外表现良好。

本文其余部分安排如下：第2节介绍时变广义帕累托分布模型，该模型纳入平稳和局部单位根预测变量以及自回归成分；第3节介绍最大似然估计，并说明模型纳入局部单位根预测变量时，其具有渐近非正态性；第 4 节介绍两步 ALMLE，并证明该估计量在选择平稳和局部单位根预测变量时均具有预言机性质；第 5 节提供一项全面的模拟研究，将两步 ALMLE 的表现与最大似然估计进行比较，结果显示前者在有限样本中的表现更优。第6节讨论实证研究结果，第7节则作总结。其他结果和数学证明见附录。

2 极值回归

我们分别用$\{l_t\}_{t=1}^T$ 和 $\{r_t\}_{t=1}^T$表示某金融资产的对数损失和收益率时间序列，其中 $l_t = -r_t$；用 $\boldsymbol{z}_t$ 表示在时间 t观测到的外生预测变量向量。

假设 M.1. $\{l_t\}_{t=1}^T$和 $\{\boldsymbol{z}_t\}_{t=1}^T$定义在完备概率空间 $(\Omega, \mathcal{F}, P)$上。在每个时刻 $t \in \{1, 2, \ldots, T\}$，都有一个可用的信息集 $\mathcal{F}_{t-1}$，它是由 $\{\boldsymbol{z}_{t-1}, l_{t-1}, \boldsymbol{z}_{t-2}, l_{t-2}, \ldots\}$生成的 σ-代数。

假设 $\{l_t\}_{t=1}^T$ 是独立同分布的，其累积分布函数（c.d.f.）为 $F(\cdot)$。极值理论中的概率结果表明，若存在实数序列 $a_T > 0$和 β_T，使得 $\lim_{T\to\infty} F^T(a_T\, x + \beta_T)$ 收敛于一个非退化分布 $G(\cdot)$，则 $F(\cdot)$ 属于 $G(\cdot)$ 的最大吸引域，即 $F \in \mathcal{D}(G)$，并且 $G(\cdot)$ 必须是广义极值（GEV）分布（参见 Coles（2001）的定理 3.1.1）。

令 $\{y_t\}_{t=1}^T$ 为超过高阈值 u的超额损失删失序列，使得超额损失$y_t = l_t - u$，若 $l_t > u$，否则为 $y_t = 0$。定义超额损失的条件分布，

$$F_{|u}(y) := P\{l_t - u \le y | l_t > u\} = P\{y_t \le y | y_t > 0\}, \quad 0 < y_t \le L^F - u,$$

其中 $L^F := \sup\{x : F(x) < 1\}$ 为 $F(\cdot)$的右端点。Pickands（1975年）以及 Balkema 和 De Haan（1974）证明，若 $F(\cdot) \in \mathcal{D}(G)$，则 $F_{|u}(y)$的极限分布为广义帕累托分布，即

$$\lim_{u \to +\infty} \sup_{0<y<+\infty} |F_{|u}(y) - \mathrm{GPD}(y; k, \sigma_u)| = 0, \tag{1}$$

w此处，GPD$(\cdot; k, \sigma)$ 表示形状参数为 $k \in \mathbb{R}$、尺度参数为 $\sigma > 0$ 的广义帕累托分布 ，

$$\mathrm{GPD}(y; k, \sigma) = 1 - \left(1 + k\frac{y}{\sigma}\right)^{-1/k}. \tag{2}$$

式（1）表明，当 u 足够大时， $F_{|u}(y)$ 可由 GPD$(\cdot;\ k,\ \sigma_u)$ 近似表示，其中尺度参数 σ_u 取决于 u。超阈值峰值法假设该关系在固定阈值 u 以上严格成立，并使用超过该阈值的观测值来估计广义帕累托分布参数 σ 和 k（参见 Coles（2001）第 4.3 节）。

2.1 时变超阈值峰值法（POT）

经典超阈值峰值法假设 $\{l_t\}$ 独立同分布。然而，金融数据通常具有时变异方差和极值聚集等依赖特征，因而不满足这一假设。为捕捉这些特征，我们采用动态超阈值峰值法。设 $\{y_t\}_{t=1}^T$ 为超过阈值时间序列 $\{u_t\}_{t=1}^T$ 的删失超额损失序列，我们使用广义帕累托分布对给定信息集 $\mathcal{F}_{t-1}$,$F_{t|u_t}(y) := P\{y_t \le y | y_t > 0, \mathcal{F}_{t-1}\}$ 条件下的超额损失分布进行建模，并令参数 k_t 和 σ_t 随时间变化。参见例如 Chavez-Demoulin 等（2014）；Massacci（2017）；Bee 等（2019）。

考虑由 $p \in \mathbb{N}$ 解释变量 $\{\boldsymbol{z}_t := [z_{1,t}, \ldots, z_{p,t}]'\}_{t=1}^T$构成的向量值时间序列。给定信息集 $\mathcal{F}_{t-1}$，我们考虑 $\{(k_t,\ \sigma_t)\}$ 的以下设定，

$$\log\left(\frac{k_t}{0.5 - k_t}\right) = \beta_{1,0} + \sum_{j=1}^{p} \beta_{1,j}\, z_{j,t-1}, \tag{3}$$

$$\log(\sigma_t) = \beta_{2,0} + \sum_{j=1}^{p} \beta_{2,j}\, z_{j,t-1} + \beta_{2,p+1}\, \log(\sigma_{t-1}). \tag{4}$$

我们施加条件 $0 < k_t < 0.5$ 和 $\sigma_t > 0$ （参见 Hosking 和 Wallis（1987年）），以确保 y_t 的条件方差有限，并保证估计的数值稳定性。由于尺度参数 σ_t 可与底层分布 $F_t(\cdot)$ 的方差相关联，我们借鉴 GARCH模型（Engle，2001）的思路，在 log(σ_t) 中纳入一个自回归项。(3) 和 (4) 中既允许平稳解释变量，也允许单位根解释变量，因此能够纳入持久性预测变量。

3 最大似然估计

令 $\boldsymbol{\beta} := [\beta_{1,0}, \beta_{1,1}, \ldots, \beta_{1,p}, \beta_{2,0}, \beta_{2,1}, \ldots, \beta_{2,p+1}]'$ 表示 (3)–(4) 中的模型系数向量，并将 $\boldsymbol{\beta}$ 的系数空间 Θ 定义为 $\mathbb{R}^{2p+2} \times (-1,\ 1)$ 的一个子空间，其中包含所有允许的系数向量 $\boldsymbol{\beta}$。我们给出 (3)–(4) 中模型系数的最大似然估计，并证明该估计是一致的；但当模型中包含局部单位根解释变量时，它是渐近非正态的。

3.1 最大似然估计量

假设 M.2。 我们假设，对于给定的 $\{u_t\}$， l_t 在给定 $\mathcal{F}_{t-1}$ 时的条件累积分布函数 $F_t(\cdot)$ 对于$t \in \{1, 2, \ldots, T\}$ 和 $y_t := l_t - u_t > 0$ 存在，且服从时变广义帕累托分布，即

$$F_{t|u_t}(y_t) = GPD(y_t; k_t, \sigma_t) = 1 - \left(1 + k_t\frac{y_t}{\sigma_t}\right)^{-\frac{1}{k_t}}, \tag{5}$$

其中，$\{k_t\}$ 和 $\{\sigma_t\}$ 由(3)-(4) 给出，真实系数向量为 $\boldsymbol{\beta}^o \in \boldsymbol{\Theta} \subset \mathbb{R}^{2p+2} \times (-1, 1)$。此外，$\{u_t\}$ 对应一个恒定的无条件超越率，即对所有 $t \in \{1, \ldots, T\}$， $P\{y_t > 0\} = \tau$均成立，且τ 是一个接近零的常数。

假设 M.3. 在模型 (3)-(4)的解释变量中，我们假设 $\{z_{i,t}, i = 1, \ldots, p_0\} \in I(0)$ 和 $\{z_{j,t}, j = p_0 + 1, \ldots, p\} \in I(1)$ 满足 $\epsilon_{j,t} := z_{j,t} - z_{j,t-1}$ 且 $\{\epsilon_{j,t}, j = p_0 + 1, \ldots, p\} \in I(0)$。我们分别用 $I(0)$ 和 $I(1)$ 表示平稳预测变量集和单位根预测变量集。

在假设 $M.2$ 下， $y_t|\{y_t > 0, \mathcal{F}_{t-1}\}$ 的条件概率密度函数（p.d.f.）为

$$f_t(y_t) = \frac{1}{\sigma_t} \left(1 + k_t \frac{y_t}{\sigma_t}\right)^{-\frac{1}{k_t} - 1}, \tag{6}$$

且 t he 对数似然函数 $L(\cdot)$ of $\{y_t|y_t > 0, \mathcal{F}_{t-1}\}$ 可定义为（Schwaab 等人，2021） ，

$$\begin{aligned} \mathcal{L}(\boldsymbol{\beta}; \{y_t\}, \{\boldsymbol{z}_{t-1}\}) &= \sum_{t=1}^{T} \mathbb{1}\{y_t > 0\} \log(f_t(y_t)) \\ &= \sum_{t=1}^{T} \mathbb{1}\{y_t > 0\} \left(-\log(\sigma_t) - \left(\frac{1}{k_t} + 1\right) \log(1 + k_t \frac{y_t}{\sigma_t})\right), \end{aligned} \tag{7}$$

其中

$$\begin{cases} k_t(\boldsymbol{\beta}) = 0.5 \left(1 + \exp\left(-(\beta_{1,0} + \sum_{j=1}^{p} \beta_{1,j}\, z_{j,t-1})\right)\right)^{-1}, \\ \sigma_t(\boldsymbol{\beta}) = \exp\left(\beta_{2,0} + \sum_{j=1}^{p} \beta_{2,j}\, z_{j,t-1} + \beta_{2,p+1}\, \log(\sigma_{t-1})\right), \end{cases} \tag{8}$$

对于 $t = 1$， 2， ....., T，其中 $\mathbb{1}\{\cdot\}$ 为指示函数：当输入为真时取值为一，否则取值为零。

我们在估计中考虑标准化预测变量 $\{\boldsymbol{z}_t^*\}$ ，以获得随机有界变量，即对于每个 $t \in \{1, \ldots, T\}$，我们按如下方式对 $\boldsymbol{z}_t$ 进行标准化：

$$\boldsymbol{z}_t^* := [\, z_{1,t}, \ldots, z_{p_0,t}, \frac{z_{p_0+1,t}}{\sqrt{T}}, \ldots, \frac{z_{p,t}}{\sqrt{T}}\,]'. \tag{9}$$

将 $\{\boldsymbol{z}_t\}$ 替换为 $\{\boldsymbol{z}_t^*\}$ 并代入式（7）的似然函数，然后求最大值，可得

$$\widehat{\boldsymbol{\beta}}^{\text{mle}} = \underset{\boldsymbol{\beta} \in \boldsymbol{\Theta}}{\arg\max}\, \mathcal{L}(\boldsymbol{\beta}; \{y_t\}, \{\boldsymbol{z}_{t-1}^*\}), \tag{10}$$

其中 $\boldsymbol{\Theta} \subset \mathbb{R}^{2p+2} \times (-1,\ 1)$。我们将相应的真实系数向量记为 $\boldsymbol{\beta}^{o*}$。

注。 假设 $M.2$ 假定超越事件 $\mathbb{1}\{y_t > 0\}$ 的无条件概率对于$t \in \{1, \ldots, T\}$为常数，这比假设 $\mathbb{1}\{y_t > 0|\mathcal{F}_{t-1}\}$ 的条件概率为常数更一般。由于 $\mathbb{1}\{y_t > 0\}$ 有界且不是模型系数的函数，因此这不会给我们求得最大似然估计的极限行为带来额外负担。假设 $M.3$ 允许$\boldsymbol{z}_t$中既有平稳预测变量，也有单位根预测变量。

3.2 最大似然估计的渐近性质

Smith（1985）在独立同分布设定下，确立了常数 k 和 σ 的广义帕累托分布最大似然估计的渐近性质。我们扩展了 Smith（1985）的研究，确立了式（10）中含有平稳和单位根预测变量的动态广义帕累托分布最大似然估计的一致性和极限分布。下文将列出推导最大似然估计渐近行为所需的假设，并确立 $\boldsymbol{\beta}$的一致性和极限分布。

假设 M.4。我们假设：

$$\begin{cases} \beta_{s,i}^{o*} := \beta_{s,i}^{o} = O(1), \quad for \quad i = 0, 1, \ldots, p_0, \ and \ s = 1, 2; \\ \beta_{s,j}^{o*} := \sqrt{T}\beta_{s,j}^{o} = O(1), \quad for \quad j = p_0 + 1, \ldots, p_0 + p \ and \ s = 1, 2; \\ \beta_{2,p+1}^{o*} := \beta_{2,p+1}^{o} \in (-1, 1), \end{cases} \tag{11}$$

且 $\boldsymbol{\beta}^{o*} := [\beta_{1,0}^{o*}, \ldots, \beta_{1,p}^{o*}, \beta_{2,0}^{o*}, \ldots, \beta_{2,p+1}^{o*}]' \in \mathbb{R}^{2p+2} \times (-1, \ 1)$。

假设 M.5。 $\{\boldsymbol{\epsilon}_t := [z_{1,t}, \ldots, z_{p_0,t}, \epsilon_{p_0+1,t}, \ldots, \epsilon_{p,t}]'\}_{t=1}^{T}$ 被假定为独立同分布$(\mathbf{0}, \ \Sigma^{(0)})$，均值为$\mathbf{0}$ 且协方差矩阵 $\Sigma^{(0)}$ 正定。对于 $\left\{\boldsymbol{z}_t^* := [\ z_{1,t}, \ldots, z_{p_0,t}, \ \frac{z_{p_0+1,t}}{\sqrt{T}}, \ldots, \frac{z_{p,t}}{\sqrt{T}}\ ]'\right\}$，~~我们~~假设当 ~~$T$~~ $\to \infty$ 时，有

$$(1) \begin{cases} \dfrac{1}{\sqrt{T}} \displaystyle\sum_{t=1}^{T} z_{i,t}^* = O_p(1), \quad \dfrac{1}{T} \displaystyle\sum_{t=1}^{T} (z_{i,t}^*)^2 = O_p(1), \quad i = 1, 2, \ldots, p_0; \\ \dfrac{1}{T} \displaystyle\sum_{t=1}^{T} z_{j,t}^* = O_p(1), \quad \dfrac{1}{T} \displaystyle\sum_{t=1}^{T} (z_{j,t}^*)^2 = O_p(1), \quad j = (p_0 + 1), \ldots, p; \\ \dfrac{1}{T} \displaystyle\sum_{t=1}^{T} \boldsymbol{z}_t^* \boldsymbol{z}_t^{*'} \ is \ positive \ definite \ in \ probability \ one; \end{cases}$$

并且存在一个正定矩阵 $\Sigma := [\Sigma_{i,j}]_{i, \ j=1, \ \ldots p}$，使得

$$(2) \begin{cases} \lim\limits_{T\to\infty} \dfrac{1}{T} \displaystyle\sum_{t=1}^{T} z_{i,t}^* = 0, \quad \lim\limits_{T\to\infty} \dfrac{1}{T} \displaystyle\sum_{t=1}^{T} (z_{i,t}^*)^2 = 1, \quad \dfrac{1}{\sqrt{T}} \displaystyle\sum_{t=1}^{t_0} z_{i,t}^* \overset{D}{\sim} W_i(\frac{t_0}{T}), \quad i = 1, 2, \ldots, p_0; \\ \dfrac{1}{T} \displaystyle\sum_{t=1}^{T} z_{j,t}^* \overset{D}{\sim} \int_0^1 \Sigma_{j,j}^{1/2} W_j(t) \, dt, \quad \dfrac{1}{T} \displaystyle\sum_{t=1}^{T} (z_{j,t}^*)^2 \overset{D}{\sim} \int_0^1 \Sigma_{j,j} W_j^2(t) \, dt, \qquad j = (p_0 + 1), \ldots, p; \\ \dfrac{1}{T} \displaystyle\sum_{t=1}^{T} \boldsymbol{z}_t^* \boldsymbol{z}_t^{*'} \overset{D}{\sim} \int_0^1 \left(\Sigma^{1/2} \boldsymbol{W}_{\boldsymbol{z}*}(t)\right) \left(\Sigma^{1/2} \boldsymbol{W}_{\boldsymbol{z}*}(t)\right)' \, dt, \\ \dfrac{1}{T} \displaystyle\sum_{t=1}^{T} \boldsymbol{z}_t^* \boldsymbol{z}_t^{*'} \ is \ positive \ definite \ in \ probability \ one; \end{cases}$$

其中 $1 \le t_0 \le T$，且 $W_i(\cdot)$ $W_j(\cdot)$ 是相互独立的布朗运动。记“$\overset{D}{\sim}$”表示依分布收敛，并记 $\boldsymbol{W}_{\boldsymbol{z}*}(t) := [\ ^{dW_1(t)\sqrt{dt}}, \ \ldots, \ ^{dW_{p_0}(t)\sqrt{dt,}} \ W_{p_0+1}(t) \ \ldots, \ W_p(t)\ ]'$。

假设 M.6。 Θ 是 $\mathbb{R}^{2p+2} \times (-1, \mathbf{1})$ 中的一个紧子空间，且包含真实系数向量 $\boldsymbol{\beta}^{o*}$，使得 $H_{\mathcal{L}}(\cdot)$ 在 Θ 中几乎必然为正定。

假设 M.7。给定假设 $M.3$ 和 $M.5$，我们进一步假设当 $T \to \infty$时，满足 t

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*}) \overset{\mathcal{D}}{\sim} S_{\psi}, \tag{12}$$

其中 S_{ψ} 服从非退化分布：

$$S_{\psi} = \Sigma_{\psi}^{1/2} \int_0^1 \begin{bmatrix} dW_{g_k}(t) \\ \boldsymbol{W}_{\boldsymbol{z}*}(t) dW_{g_k}(t) \\ dW_{g_\sigma}(t) \\ \boldsymbol{W}_{\boldsymbol{z}*}(t) dW_{g_\sigma}(t) \\ dW_{ar}(t) \, dW_{g_\sigma}(t) \end{bmatrix} \tag{13}$$

其中 Σ_{ψ}是正定$(2p + 3) \times (2p + 3)$矩阵，且 $\{W_{g_k}(t)\}, \{W_{g_\sigma}(t)\}, \{W_{ar}(t)\}$均为布朗运动。

假设 M.8. $H_{L}(\boldsymbol{\beta}^{o*};\{l_t\},\{\boldsymbol{z}_t^*\})T$在 $\boldsymbol{\beta}^{o*}$处假定弱收敛于随机积分 Ω_H，即，

$$\frac{H_{\mathcal{L}}(\boldsymbol{\beta}^{o*};\{l_t\},\{\boldsymbol{z}_t^*\}}{T} \overset{D}{\sim} \Omega_H, \qquad when\ T\to\infty, \tag{14}$$

其中， Ω_H^{-1} 的存在取决于假设 $M.5$中 $\boldsymbol{z}_t^*$ 的极限行为。

注：假设 $M.4$ 规定了 $\boldsymbol{\beta}^o$ 的数量级，以确保单位根解释变量 $z_{p_0+1,\,t}$、……、 $z_{p,\,t}$的系数具有趋于零的局部速率1 2，参见 Phillips 和 Lee（2013）以及 Lee 等（2022）。假设 $M.5$ 确保 $\{\boldsymbol{z}_t^*\}$ 和 $\{\boldsymbol{z}_t^*\ \boldsymbol{z}_t^{*\prime}\}$ 的部分和以特定速率收敛。Saikkonen（1993，1995）和 Lee 等（2022）也采用了假设 $M.5$ ，并证明该假设对于时间依赖程度适中且异方差性为 $\{\boldsymbol{\epsilon}_t\}$的时间序列成立。例如，参见 Billingsley（2013）的定理 18.2、Phillips 和 Durlauf（1986）以及 Phillips（1991）。假设 $M.6$限制了最大似然估计的可行参数空间 Θ ，尤其是维持 $H_{\mathcal{L}}(\boldsymbol{\beta})$的正定性；这与 Smith（1985）的假设 9 类似，旨在确保估计量的唯一性。假设 $M.7$ 假定似然梯度函数在 $\boldsymbol{\beta}^{o*}$处的极限分布存在，参见 Lee（2016）的引理 A.1.1。假设 $M.8$假定似然海森矩阵在 $\boldsymbol{\beta}^{o*}$处的极限分布存在，参见 Lee（2016）的引理 A.1.2。

定理 1（最大似然估计的一致性）。
在假设 $M.1$, $M.2$, $M.3$, $M.5$***(1)***、 $M.4$ 和 $M.6$下，对任意 $\epsilon>0$,

$$\lim_{T\to\infty} P\left\{\left\|\widehat{\boldsymbol{\beta}}^{mle}-\boldsymbol{\beta}^{o*}\right\|>\epsilon\right\}=0, \tag{15}$$

证明。 参见附录 A.1。 □

定理 2（最大似然估计的渐近性质）。
在假设 $M.1$ 至 $M.8$下，我们有

$$\sqrt{T}\left(\widehat{\boldsymbol{\beta}}^{mle}-\boldsymbol{\beta}^{o*}\right)\overset{D}{\sim}\Omega_H^{-1}\,S_\psi, \qquad as\ T\to\infty. \tag{16}$$

证明。 参见附录 A.1。 □

4 自适应 L_1正则化最大似然估计

变量选择有助于解释回归模型，并能解决偏差与效率之间的权衡问题，从而实现预测准确性，参见 James 等（2013）。尽管基于最大似然估计渐近正态性进行推断检验以实现变量选择，似乎是可行的解决方案，但由于以下三个问题，这种方法并不适用于我们的情形：(i) 无法控制多预测变量选择时的第一类错误；(ii) 由于极限分布非正态，选择单位根预测变量时会出现严重的尺寸扭曲；(iii) 由于系数的标准误较大，选择形状参数的预测变量时检验功效较低，参见第 5 节中的模拟。

为解决这些问题，我们采用 L_1正则化最大似然估计（MLE）进行自动变量选择（Tibshirani，1996）。由于 L_1正则化具有约束性，该估计量会将部分系数精确地设为零，从而实现变量选择。Zou（2006）探讨了对模型系数采用加权 L_1正则化的优点，并提出了自适应 LASSO。采用适当的自适应权重后，自适应 LASSO 呈现出预言机性质；也就是说，即使一开始就已知真实模型，它仍能得到具有变量选择一致性的渐近有效估计量。Medeiros 和 Mendes（2016）证明，在具有非高斯误差和异方差误差以及高度相关回归变量的高维时间序列中，自适应 LASSO 具有预言机性质。Kock（2016）表明，自适应 LASSO 在平稳和非平稳自回归模型中均具有神谕有效性。Lee 等（2022）证明了自适应 LASSO 在

平稳预测变量和局部单位根预测变量下的预言机性质，并提出一种新颖的后选择自适应 LASSO，用于选择混合根预测变量，即平稳预测变量、局部单位根预测变量和协整预测变量。

借鉴这一文献，我们将自适应 LASSO 扩展到式 (10) 中的最大似然估计，以估计并筛选式 (3)–(4) 中的平稳预测变量和局部单位根预测变量。自适应 L_1正则化最大似然估计量（ALMLE）的一般形式可直接参照 Zou（2006）得出，具体如下:

$$\widehat{\boldsymbol{\beta}}^{\mathrm{al}} = \underset{\boldsymbol{\beta}\in\Theta^*}{\arg\min} \; -\mathcal{L}(\boldsymbol{\beta}; \{y_t\}, \{\boldsymbol{z}_t^*\}) + \lambda_{k,T}\sum_{i=1}^{p} w_{k,i}|\beta_{1,i}| + \lambda_{\sigma,T}\sum_{j=1}^{p+1} w_{\sigma,j}|\beta_{2,j}|, \tag{17}$$

其中， $\mathcal{L}(\boldsymbol{\beta}; \{y_t\}, \{\boldsymbol{z}_t^*\})$ 是式 (7) 中定义的对数似然函数；$\lambda_{k,T}, \lambda_{\sigma,T} > 0$ 是调优参数；$w_{k,i}, w_{\sigma,j}$ 是自适应权重，用于对不同系数施加不同惩罚。我们采用两个调优参数而非一个，以放宽调优参数的选择限制，从而稳定式 (3)–(4) 中形状模型和尺度模型的变量选择。为设定调优参数，我们首先取足够大的 $\lambda_{k,T}$ 和 $\lambda_{\sigma,T}$，使得 $\widehat{\boldsymbol{\beta}}^{\mathrm{al}}$ 不会选出任何预测变量，并将这两个值分别记为 $\lambda_{k,T,\max}$ 和 $\lambda_{\sigma,T,\max}$。随后，我们在定义于区间 $[\lambda_{k,T,\max}, 10^{-6}]$ 和 $[\lambda_{\sigma,T,\max}, 10^{-6}]$ 上、分别含有 n_{λ_k} 和 n_{λ_σ} 个节点的等间隔网格[1] 上，利用信息准则（IC）搜索最优调优参数。具体而言，形状参数和尺度参数的网格分别定义为 $S_{\lambda_{k,T}} := \{\exp(\log(\lambda_{k,T,\max}) - j\log(\lambda_{k,T,\max})/\log(10^{-6})^{n_{\lambda_k}-1}) \; j = 0, 1, \ldots, (n_{\lambda_k}-1)\}$ 和 $S_{\lambda_{\sigma,T}} := \{\exp(\log(\lambda_{\sigma,T,\max}) - j\log(\lambda_{\sigma,T,\max})/\log(10^{-6})^{n_{\lambda_\sigma}-1}) \; j = 0, 1, \ldots, (n_{\lambda_\sigma}-1)\}$。我们考虑不同的信息准则，即贝叶斯信息准则（BIC）、Hannan–Quinn 信息准则（HQ）和赤池信息准则（AIC），并据此按照以下规则选取最优的调优参数 $(\lambda_{k,T}, \lambda_{\sigma,T})$：

$$\text{AIC:} \quad (\widehat{\lambda}_{k,T}, \widehat{\lambda}_{\sigma,T}) = \underset{\lambda_{k,T}\in S_{\lambda_{k,T}}, \lambda_{\sigma,T}\in S_{\lambda_{\sigma,T}}}{\arg\min} -2\log(\mathcal{L}(\widehat{\boldsymbol{\beta}}^{\mathrm{al}}(\lambda_{k,T}, \lambda_{\sigma,T}); \{y_t\}, \{\boldsymbol{z}_t^*\})) + 2\left(\textstyle\sum_{i=1,\ldots,p}\mathbb{1}\{\widehat{\beta}_{1,i}^{\mathrm{al}} \neq 0\} + \sum_{j=1,\ldots,(p+1)}\mathbb{1}\{\widehat{\beta}_{2,j}^{\mathrm{al}} \neq 0\}\right) \tag{18}$$

$$\text{HQ:} \quad (\widehat{\lambda}_{k,T}, \widehat{\lambda}_{\sigma,T}) = \underset{\lambda_{k,T}\in S_{\lambda_{k,T}}, \lambda_{\sigma,T}\in S_{\lambda_{\sigma,T}}}{\arg\min} -2\log(\mathcal{L}(\widehat{\boldsymbol{\beta}}^{\mathrm{al}}(\lambda_{k,T}, \lambda_{\sigma,T}); \{y_t\}, \{\boldsymbol{z}_t^*\})) + 2\log(\log(T))\left(\textstyle\sum_{i=1,\ldots,p}\mathbb{1}\{\widehat{\beta}_{1,i}^{\mathrm{al}} \neq 0\} + \sum_{j=1,\ldots,(p+1)}\mathbb{1}\{\widehat{\beta}_{2,j}^{\mathrm{al}} \neq 0\}\right) \tag{19}$$

$$\text{BIC:} \quad (\widehat{\lambda}_{k,T}, \widehat{\lambda}_{\sigma,T}) = \underset{\lambda_{k,T}\in S_{\lambda_{k,T}}, \lambda_{\sigma,T}\in S_{\lambda_{\sigma,T}}}{\arg\min} -2\log(\mathcal{L}(\widehat{\boldsymbol{\beta}}^{\mathrm{al}}(\lambda_{k,T}, \lambda_{\sigma,T}); \{y_t\}, \{\boldsymbol{z}_t^*\})) + \log(T)\left(\textstyle\sum_{i=1,\ldots,p}\mathbb{1}\{\widehat{\beta}_{1,i}^{\mathrm{al}} \neq 0\} + \sum_{j=1,\ldots,(p+1)}\mathbb{1}\{\widehat{\beta}_{2,j}^{\mathrm{al}} \neq 0\}\right) \tag{20}$$

Tibshirani 等（2012）的序贯强规则通常用于求解 LASSO 型问题。然而，当 k_t 呈现持久动态时，若调优参数不够小，$\widehat{\boldsymbol{\beta}}^{\mathrm{al}}$ 的序贯强规则会因估计偏差而无法筛选出真正的活跃预测变量和非活跃预测变量，并因此倾向于得到边界解 $k_t = 0.5$。为实现变量选择一致性，必须确保优化器远离参数空间的边界。定理 3 说明了为实现 $\widehat{\boldsymbol{\beta}}^{\mathrm{al}}$ 的模型选择一致性所允许的系数空间限制 Θ，即,

$$\lim_{T\to\infty} P\left\{\mathcal{A}_T^{\mathrm{al}} = \mathcal{A}\right\} = 1, \tag{21}$$

其中 $\mathcal{A}_T^{\mathrm{al}} := \mathcal{A}_{k,T}^{\mathrm{al}} \cup \mathcal{A}_{\sigma,T}^{\mathrm{al}}$，且 $\mathcal{A}_{k,T}^{\mathrm{al}} := \left\{(1,i) : \; i \geq 1, \widehat{\beta}_{1,i}^{\mathrm{al}} \neq 0\right\}$，以及 $\mathcal{A}_{\sigma,T}^{\mathrm{al}} := \left\{(2,j) : \; j \geq 1, \widehat{\beta}_{2,j}^{\mathrm{al}} \neq 0\right\}$；而 $\mathcal{A} := \mathcal{A}_k \cup \mathcal{A}_\sigma$，且 $\mathcal{A}_k := \left\{(1,i) : \; i \geq 1, \beta_{1,i}^{o*} \neq 0\right\}$，以及 $\mathcal{A}_\sigma := \left\{(2,j) : \; j \geq 1, \beta_{2,j}^{o*} \neq 0\right\}$。

定理 3。 在定理 *2* 的假设下，若不存在满足下列条件的 $\widehat{\boldsymbol{\beta}}^{al}(\lambda_{k,T}, \lambda_{\sigma,T})$，其中 $\lambda_{k,T}, \lambda_{\sigma,T} \in O(T^{\frac{1}{2}})$，即 -

$$\det\left(\frac{\partial^2\mathcal{L}(\boldsymbol{\beta})}{\partial[\boldsymbol{\beta}'_{\mathcal{A}_k}, \boldsymbol{\beta}'_{\mathcal{A}_\sigma}]'\partial[\boldsymbol{\beta}'_{\mathcal{A}_k}, \boldsymbol{\beta}'_{\mathcal{A}_\sigma}]}\bigg|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}^{al}(\lambda_{k,T},\lambda_{\sigma,T})}\right) \neq 0, \tag{22}$$

[1] 除非另有说明，本文始终采用$n_{\lambda_k} = 50$ 和 $n_{\lambda_\sigma} = 30$。我们还尝试了$n_{\lambda_k} = 100$ 和 $n_{\lambda_\sigma} = 100$，以检验 $n_{\lambda_k} = 50$ 和 $n_{\lambda_\sigma} = 30$是否充分，并发现结果差异很小。

则 $\lim_{T\to\infty} P\{\mathcal{A}_T^{al} = \mathcal{A}\} \neq 1$，其中 $det(\cdot)$ 表示矩阵行列式算子；$w_{k,i}$ 和 $w_{\sigma,j}$ 通过第3节中的最大似然估计设定，使得 $\sqrt{T}(\hat{\beta}_{1,i}^{k} - \beta_{1,i}^{*o}) = O_p(1)$ 且 $\sqrt{T}(\hat{\beta}_{2,j}^{\sigma} - \beta_{2,j}^{*o}) = O_p(1)$，对于 $i = 1,\ldots,p, j = 1,\ldots,p+1$。

证明。见附录 A.3。 □

定理 3 表明，如果并非所有真实活跃预测变量都能以 $\lambda_{k,T}, \lambda_{\sigma,T} \in O(T^{\frac{1}{2}})$ 进入回归模型，那么真实非活跃预测变量便会开始被选中，以弥补缺失的预测变量，因为 $\lambda_{k,T}, \lambda_{\sigma,T} \in O(T^{\frac{1}{2}})$，从而导致变量选择失败。当底层 $\{k_t(\boldsymbol{\beta}^{o*})\}$ 涉及局部单位根预测变量时，定理 3 中的必要条件往往会被破坏。为解决这一问题，我们提出两步 ALMLE，并证明其具有预言机性质。

4.1 两步 ALMLE

根据前面的讨论，我们知道，如果确保估计远离每个 $\lambda_{k,T}$ 对应的 $\{k_t(\cdot) = 0.5\}$，就可以改进 ALMLE。因此，我们提出两步 ALMLE，记为 $\widehat{\boldsymbol{\beta}}^{\text{tal}}$，以避免 $\widehat{\boldsymbol{\beta}}^{\text{al}}$ 的局部极小值问题：第一步选择形状参数的预测变量，第二步使用第一步选出的 $\widehat{\lambda}_{k,T}$ 运行式（17）中的 ALMLE。具体而言，两步 ALMLE $\widehat{\boldsymbol{\beta}}^{\text{tal}}$ 通过以下步骤获得：

步骤1：按如下方式使用信息准则选择最优调节参数 $\lambda_{k,T} \in S_{\lambda_{k,T}}$，

$$\begin{aligned}
\text{AIC:} \quad & \widehat{\lambda}_{k,T} = \underset{\lambda_{k,T}\in S_{\lambda_{k,T}}}{\arg\min} \; -2\log(\mathcal{L}(\widehat{\boldsymbol{\beta}}^{\text{k,al}}(\lambda_{k,T}); \{y_t\}, \{\boldsymbol{z}_t^*\})) + 2\sum_{i=1,\ldots,p} \mathbb{1}\{\widehat{\beta}_{1,i}^{\text{k,al}} \neq 0\} \\
\text{HQ:} \quad & \widehat{\lambda}_{k,T} = \underset{\lambda_{k,T}\in S_{\lambda_{k,T}}}{\arg\min} \; -2\log(\mathcal{L}(\widehat{\boldsymbol{\beta}}^{\text{k,al}}(\lambda_{k,T}); \{y_t\}, \{\boldsymbol{z}_t^*\})) + 2\log(\log(T))\sum_{i=1,\ldots,p} \mathbb{1}\{\widehat{\beta}_{1,i}^{\text{k,al}} \neq 0\} \\
\text{BIC:} \quad & \widehat{\lambda}_{k,T} = \underset{\lambda_{k,T}\in S_{\lambda_{k,T}}}{\arg\min} \; -2\log(\mathcal{L}(\widehat{\boldsymbol{\beta}}^{\text{k,al}}(\lambda_{k,T}); \{y_t\}, \{\boldsymbol{z}_t^*\})) + \log(T)\sum_{i=1,\ldots,p} \mathbb{1}\{\widehat{\beta}_{1,i}^{\text{k,al}} \neq 0\},
\end{aligned}$$

其中 $\widehat{\boldsymbol{\beta}}^{\text{k,al}}(\lambda_{k,T}) := [\widehat{\beta}_{1,0}^{\text{k,al}}, \ldots, \widehat{\beta}_{1,p}^{\text{k,al}}, \widehat{\beta}_{2,0}^{\text{k,al}}, 0, \ldots, 0]'$ 将 $\widehat{\beta}_{2,1}^{\text{k,al}}, \ldots, \widehat{\beta}_{2,p+1}^{\text{k,al}}$ 限制为零，并定义

$$\left[\widehat{\beta}_{1,0}^{\text{k,al}}, \ldots, \widehat{\beta}_{1,p}^{\text{k,al}}, \widehat{\beta}_{2,0}^{\text{k,al}}\right] = \underset{\beta_{1,0},\ldots,\beta_{1,p},\beta_{2,0}}{\arg\min} \; -\mathcal{L}(\boldsymbol{\beta}; \{y_t\}, \{\boldsymbol{z}_t^*\}) + \lambda_{k,T}\sum_{i=1}^{p} \widetilde{w}_{k,i}|\beta_{1,i}|. \tag{23}$$

步骤2：按如下方式使用信息准则以及步骤1中的 $\lambda_{k,T}$ 选择最优调节参数 $\lambda_{\sigma,T} \in S_{\lambda_{\sigma,T}}$，

$$\begin{aligned}
\text{AIC:} \quad & \widehat{\lambda}_{\sigma,T} = \underset{\lambda_{\sigma,T}\in S_{\lambda_{\sigma,T}}}{\arg\min} \; -2\log(\mathcal{L}(\widehat{\boldsymbol{\beta}}^{\text{tal}}(\lambda_{\sigma,T}); \{y_t\}, \{\boldsymbol{z}_t^*\})) + 2\left(\textstyle\sum_{i=1,\ldots,p} \mathbb{1}\{\widehat{\beta}_{1,i}^{\text{tal}} \neq 0\} + \sum_{j=1,\ldots,(p+1)} \mathbb{1}\{\widehat{\beta}_{2,j}^{\text{tal}} \neq 0\}\right) \\
\text{HQ:} \quad & \widehat{\lambda}_{\sigma,T} = \underset{\lambda_{\sigma,T}\in S_{\lambda_{\sigma,T}}}{\arg\min} \; -2\log(\mathcal{L}(\widehat{\boldsymbol{\beta}}^{\text{tal}}(\lambda_{\sigma,T}); \{y_t\}, \{\boldsymbol{z}_t^*\})) + 2\log(\log(T))\left(\textstyle\sum_{i=1,\ldots,p} \mathbb{1}\{\widehat{\beta}_{1,i}^{\text{tal}} \neq 0\} + \sum_{j=1,\ldots,(p+1)} \mathbb{1}\{\widehat{\beta}_{2,j}^{\text{tal}} \neq 0\}\right) \\
\text{BIC:} \quad & \widehat{\lambda}_{\sigma,T} = \underset{\lambda_{\sigma,T}\in S_{\lambda_{\sigma,T}}}{\arg\min} \; -2\log(\mathcal{L}(\widehat{\boldsymbol{\beta}}^{\text{al}}(\lambda_{\sigma,T}); \{y_t\}, \{\boldsymbol{z}_t^*\})) + \log(T)\left(\textstyle\sum_{i=1,\ldots,p} \mathbb{1}\{\widehat{\beta}_{1,i}^{\text{tal}} \neq 0\} + \sum_{j=1,\ldots,(p+1)} \mathbb{1}\{\widehat{\beta}_{2,j}^{\text{tal}} \neq 0\}\right),
\end{aligned}$$

其中 $\widehat{\boldsymbol{\beta}}^{\text{tal}}(\lambda_{\sigma,T}) := [\widehat{\boldsymbol{\beta}}_{1\cdot}^{\text{tal}\prime}, \widehat{\boldsymbol{\beta}}_{2\cdot}^{\text{tal}\prime}]' = [\widehat{\beta}_{1,0}^{\text{tal}}, \ldots, \widehat{\beta}_{1,p}^{\text{tal}}, \widehat{\beta}_{2,0}^{\text{tal}}, \ldots, \widehat{\beta}_{2,(p+1)}^{\text{tal}}]'$，且 $\widehat{\beta}_{1,i}^{\text{tal}} = \widehat{\beta}_{1,i}^{\text{al}} = 0, \forall (1,i) \notin \mathcal{A}_T^{k,al}$ 以及

$$\left[\left[\widehat{\beta}_{1,i}^{\text{tal}}\right]_{(1,i)\in\{(1,0)\}\cup\mathcal{A}_T^{k,al}}, \widehat{\boldsymbol{\beta}}_{2\cdot}^{\text{tal}\prime}\right] = \underset{\left\{\beta_{1,i}|(1,i)\in\{(1,0)\}\cup\mathcal{A}_T^{k,al}\right\},\boldsymbol{\beta}_{2\cdot}}{\arg\min} \; -\mathcal{L}(\boldsymbol{\beta}; \{y_t\}, \{\boldsymbol{z}_t^*\}) + \widehat{\lambda}_{k,T}\sum_{i=1}^{p} \widetilde{w}_{k,i}|\beta_{1,i}| + \lambda_{\sigma,T}\sum_{j=1}^{p+1} \widetilde{w}_{\sigma,j}|\beta_{2,j}|, \tag{24}$$

where $\mathcal{A}_T^{k,al} := \left\{(1,i) : i \geq 1, \widehat{\beta}_{1,i}^{k,al} \neq 0\right\}$.

最终的两步 ALMLE $\widehat{\boldsymbol{\beta}}^{tal}$ 使用最优调节参数 $\widehat{\lambda}_{k,T}$ 和 $\widehat{\lambda}_{\sigma,T}$ 得出。

我们使用两个最大似然估计来设定 $\widetilde{w}_{k,\ i}$ 和 $\widetilde{w}_{\sigma,\ j}$，因为两步 ALMLE 在每一步都涉及不同的似然函数。具体而言，我们设定

$$\begin{cases} \widetilde{w}_{k,i} = \dfrac{1}{\widehat{\beta}^{\mathrm{k,mle}}_{1,i}} \dfrac{1}{\widehat{\beta}^{\mathrm{mle}}_{1,i}},\ i = 1, \ldots, p, \\ \widetilde{w}_{\sigma,j} = \dfrac{1}{\widehat{\beta}^{\mathrm{mle}}_{2,j}},\ j = 1, \ldots, p+1. \end{cases} \tag{25}$$

其中 $\widehat{\boldsymbol{\beta}}^{\mathrm{mle}} := [\widehat{\beta}^{\mathrm{mle}}_{1,0}, \ldots, \widehat{\beta}^{\mathrm{mle}}_{1,p}, \widehat{\beta}^{\mathrm{mle}}_{2,0}, \ldots, \widehat{\beta}^{\mathrm{mle}}_{2,p+1}]$ 是全模型极大似然估计 (10)，$\widehat{\boldsymbol{\beta}}^{\mathrm{k},\ \mathrm{mle}} := [\widehat{\beta}^{\mathrm{k,mle}}_{1,0}, \ldots, \widehat{\beta}^{\mathrm{k,mle}}_{1,p}, \widehat{\beta}^{\mathrm{k,mle}}_{2,0}, 0, \ldots, 0]$ 是下文定义的部分模型极大似然估计

$$\widehat{\boldsymbol{\beta}}^{\mathrm{k,mle}} = \underset{\{\boldsymbol{\beta} \in \Theta | \beta_{2,j} = 0, j = 1, \ldots, p+1\}}{\arg\min} - \mathcal{L}(\boldsymbol{\beta}; \{y_t\}, \{\boldsymbol{z}^*_t\}). \tag{26}$$

通过这种方式，我们选择 $\widetilde{w}_{k,\ i}$ 和 $\widetilde{w}_{\sigma,\ j}$，使真实活跃预测变量能够被高效选出，并且在步骤1和步骤2中，$S_{\lambda_{k,\ T}}$ 和 $S_{\lambda_{\sigma,\ T}}$ 都排在真实非活跃预测变量之前。因此，我们实现了定理 4 所示的 $\widehat{\boldsymbol{\beta}}^{tal}$ 的预言机性质。

假设 L1。存在 $\lambda_{k,\ T} = O(T^{\frac{1}{2}-\gamma_1})$ 以及 $\lambda_{\sigma,\ T} = O(T^{\frac{1}{2}-\gamma_2})$，满足 $\theta < \gamma_1 < \frac{1}{2}$和 $0 < \gamma_2 < \frac{1}{2}$。

假设 L2。我们假设存在 $\boldsymbol{\beta}^{k,\ o} := [\beta^{k,o}_{1,0}, \beta^{k,o}_{1,1}, \ldots, \beta^{k,o}_{1,p}, \beta^{k,o}_{2,0}, 0, \ldots, 0]' \in \{\boldsymbol{\beta} \in \mathbb{R}^{2p+3} | \beta_{2,j} = 0, j = 1, \ldots, p+1\}$，使得对于任意 $\epsilon > 0$

$$\lim_{T \to \infty} P\left\{ \left| \widehat{\beta}^{k,mle}_{1,i} - \beta^{k,o}_{1,i} \right| > \epsilon \right\} = 0, \quad i = 1, \ldots, p; \tag{27}$$

以及 $\beta^{k,\ o}_{1,\ i} \neq 0$ 对于任意 $(1, i) \in \mathcal{A}_k$。

定理 4（$\widehat{\boldsymbol{\beta}}^{tal}$的预言机性质）。
在假设 $L1$、$L2$ 以及定理 *2* 中的假设下，我们有：*(a)* 模型选择一致性：

$$\lim_{T \to \infty} P\left\{ \mathcal{A}^{tal}_T = \mathcal{A} \right\} = 1, \tag{28}$$

其中 $\mathcal{A}^{tal}_T := \mathcal{A}^{tal}_{k,\ T} \cup \mathcal{A}^{tal}_{\sigma,\ T}$，且 $\mathcal{A}^{tal}_{k,\ T} := \left\{ (1,\ i) :\ \widehat{\beta}^{tal}_{1,\ i} \neq 0、\ i = 1、\ldots\ldots、\ p. \right\}$，以及 $\mathcal{A}^{tal}_{\sigma,\ T} := \{(2,\ j) :\ \widehat{\beta}^{tal}_{2,\ j} \neq 0、\ j = 1、\ldots\ldots、\ p+1.\}$。

(a) $\widehat{\boldsymbol{\beta}}^{tal}$的极限分布：

$$\begin{aligned} \sqrt{T}\left(\widehat{\boldsymbol{\beta}}^{tal}_{\mathcal{A}} - \boldsymbol{\beta}^{o*}_{\mathcal{A}} \right) &\overset{D}{\sim} \Omega^{-1}_{H_{\mathcal{A}}}\ S_{\psi_{\mathcal{A}}}, \\ \sqrt{T}\left(\widehat{\boldsymbol{\beta}}^{tal}_{\mathcal{A}^c} - \boldsymbol{\beta}^{o*}_{\mathcal{A}^c} \right) &\to 0, \end{aligned} \tag{29}$$

当 $T \to \infty$ 时，其中 $S_{\psi_{\mathcal{A}}}$ 和 $\Omega_{H_{\mathcal{A}}}$ 在假设 $M.7$ 和 $M.8$ 中定义；在仅包含真实活跃预测变量的模型设定下，这些变量按照 $\mathcal{A}$ 排序。

证明。见附录 A.3。 □

与 ALMLE (17) 相比，所提出的两步 ALMLE 的优势不仅体现在回归模型包含局部单位根预测变量时具有预言机性质，也体现在计算成本上。为选择最优参数对 $(\lambda_{k,\ T},\ \lambda_{\sigma,\ T}) \in S_{\lambda_{k,\ T}} \times S_{\lambda_{\sigma,\ T}}$，ALMLE (17) 需要在二维调优参数网格上进行计算；而两步 ALMLE 则分别在两个独立的一维调优参数网格上进行计算，先选择最优的 $\lambda_{k,\ T} \in S_{\lambda_{k,\ T}}$，再选择最优的 $\lambda_{\sigma,\ T} \in S_{\lambda_{\sigma,\ T}}$。

5 模拟研究

我们通过四种数据生成过程（DGP）从偏差、均方误差（MSE）和模型选择的角度评估 $\widehat{\boldsymbol{\beta}}^{mle}$ 和 $\widehat{\boldsymbol{\beta}}^{tal}$ 的有限样本性质。这四种DGP旨在

反映第6节所用高频金融数据的特征。首先，DGP具有异方差性，条件超越率会随时间变化。其次，DGP包含一些预测变量，这些变量是滞后损失率的函数，用以刻画 {中的序列依赖结构（k_t， σ_t）}。第三，我们考虑平稳预测变量、局部单位根预测变量，或两者兼有。

我们根据以下条件分布模拟 $\{l_t\}$ ，

$$l_t = \begin{cases} F^{-1}_{t\left(\frac{1}{k_t}\right)}(\tau_t), & \text{if} \quad \tau_t \leq F_{t\left(\frac{1}{k_t}\right)}(u) \\ u + F^{-1}_{\mathrm{GPD}(k_t,\sigma_t)}\left(\dfrac{\tau_t - F_{t(\frac{1}{k_t})}(u)}{1-\tau_t}\right), & \text{if} \quad \tau_t > F_{t\left(\frac{1}{k_t}\right)}(u), \end{cases} \tag{30}$$

其中， $\{\tau_t\}$ 服从独立同分布的标准均匀分布， $F_{t(\nu)}(\cdot)$ 和 $F^{-1}_{t\ \nu\ }{}_{()}(\cdot)$ 分别表示自由度为 ν 的学生t分布的分布函数和分位数函数。 $\{k_t\}$ 和 $\{\sigma_t\}$ 的过程按照以下设定指定：

$$\begin{cases} \log\left(\dfrac{k_t}{0.5-k_t}\right) = \beta_{10} + \beta_{11}\ \log(|l_{t-1}| + 1 - r_m) + \sum_{j=1}^{14} \beta_{1,1+j}\ z_{j,t-1}\,, \\ \log(\sigma_t) = \beta_{20} + \beta_{21}\ \log(\sigma_{t-1}) + \beta_{22}\ \log(|l_{t-1}| + 1 - r_m) + \sum_{j=1}^{14} \beta_{2,2+j}\ z_{j,t-1}\,, \\ z_{i,t} = \phi_i\, z_{i,t-1} + \epsilon_{i,t-1}, \qquad i = 1, 2, \ldots, 14, \\ \boldsymbol{\phi} := [\phi_1, \ldots, \phi_{14}], \\ \{\boldsymbol{\epsilon}_t := [\epsilon_{1,t}, \ldots, \epsilon_{14,t}]'\} \overset{i.i.d.}{\sim} \mathcal{N}(0, I_{14\times 14})\,, \\ \boldsymbol{\beta^o_{1.}} := [\beta^o_{1,0}, \beta^o_{1,1}, \ldots, \beta^o_{1,15}]'\,, \\ \boldsymbol{\beta^o_{2.}} := [\beta^o_{2,0}, \beta^o_{2,1}, \ldots, \beta^o_{2,16}]'\,, \\ \boldsymbol{\beta^o} := [\boldsymbol{\beta}^{o'}_{1.}, \boldsymbol{\beta}^{o'}_{2.}]'\,. \end{cases} \tag{31}$$

我们设定 $u = F^{-1}_{t\ 3\ }{}_{()}(0.8)$ 和 $r_m = 0.05$，但采用不同的 $\boldsymbol{\phi}$ 和 $\boldsymbol{\beta^o}$ ，以获得不同程度的序列依赖。

数据生成过程 1. 对于 $\{k_t\}$ 和 $\{\sigma_t\}$，共有五个真正活跃的平稳预测变量，即 $\log(|l_{t-1}| + 1 - r_m)$、$z_{1,t-1}, \ldots, z_{4,t-1}$。在真正非活跃的预测变量 $z_{5,t-1}, \ldots, z_{14,t-1}$ 中，有两个是局部单位根变量，即 $z_{13,t-1}$ 和 $z_{14,t-1}$，其余均为平稳变量。

$$\begin{cases} \boldsymbol{\phi} = [0, 0, 0, 0, 0 \ldots, 0, 1, 1], \\ \boldsymbol{\beta^o_{1.}} = [-1, 0.3, -0.4, 0.2, 0.6, 0.6, 0, \ldots, 0]'\,, \\ \boldsymbol{\beta^o_{2.}} = [-1, 0, 0.7, 0.4, 0.3, 0.5, 0.6, 0, \ldots, 0]'\,, \end{cases} \tag{32}$$

数据生成过程 2. 与数据生成过程 1 相同，但有一点不同：$\beta^o_{2,1}$ 被设为非零，因此 $\log(\sigma_{t-1})$ 现在确实处于激活状态。我们设定 $\beta^o_{2,1} = 0.7$ ，其余系数的真实值保持不变。

$$\begin{cases} \boldsymbol{\phi} = [0, 0, 0, 0, 0 \ldots, 0, 1, 1], \\ \boldsymbol{\beta^o_{1.}} = [-1, 0.3, -0.4, 0.2, 0.6, 0.6, 0, \ldots, 0]'\,, \\ \boldsymbol{\beta^o_{2.}} = [-1, 0.7, 0.7, 0.4, 0.3, 0.5, 0.6, 0, \ldots, 0]'\,, \end{cases} \tag{33}$$

DGP 3. 与数据生成过程 1 相同，但区别在于 $\phi_4 = 1$ 和 $(\beta^o_{1,\ 5},\ \beta^o_{2,\ 6}) = (\frac{0.6}{\sqrt{T}}, \frac{0.6}{\sqrt{T}})$。

$$\begin{cases} \boldsymbol{\phi} = [0, 0, 0, 1, 0 \ldots, 0, 1, 1], \\ \boldsymbol{\beta^o_{1.}} = [-1, 0.3, -0.4, 0.2, 0.6, \dfrac{0.6}{\sqrt{T}}, 0, \ldots, 0]'\,, \\ \boldsymbol{\beta^o_{2.}} = [-1, 0, 0.7, 0.4, 0.3, 0.5, \dfrac{0.6}{\sqrt{T}}, 0, \ldots, 0]'\,, \end{cases} \tag{34}$$

DGP 4. 与数据生成过程 3 相同，但区别在于 $\log(\sigma_{t-1})$ 确实处于活跃状态。我们设定 $\beta^o_{2,\ 1} = 0.7$ ，并保持其他系数的真实值不变。

$$\begin{cases} \boldsymbol{\phi} = [0,0,0,1,0\ldots,0,1,1], \\ \boldsymbol{\beta^o_{1\cdot}} = [-1,0.3,-0.4,0.2,0.6,\dfrac{0.6}{\sqrt{T}},0,\ldots,0]', \\ \boldsymbol{\beta^o_{2\cdot}} = [-1,0.7,0.7,0.4,0.3,0.5,\dfrac{0.6}{\sqrt{T}},0,\ldots,0]', \end{cases} \tag{35}$$

在每次模拟中，我们获得一个包含 T 个观测值的样本 $\{l_t\}_{t=1}^T$ ，并使用真实阈值 u提取超额时间序列 $\{y_t = \max(l_t - u,0)\}$ 。我们使用预测变量的经验标准差对其进行标准化。随后，使用标准化后的预测变量，将完整模型设定（31）拟合到 $\{y_t\}$ ，通过 $\widehat{\boldsymbol{\beta}}^{mle}$ 和 $\widehat{\boldsymbol{\beta}}^{tal}$估计模型参数。然后计算偏差和均方误差（MSE），如下所示

$$\text{Bias} = \frac{1}{\#\boldsymbol{\beta}^o}\left(\sum_{i=1,2;j=0}\left|\widehat{\beta}_{i,j} - \beta^o_{i,j}\right| + \sum_{i=1,2;j\geq 1} s_{i,j}\left|\widehat{\beta}_{i,j}\,\widehat{s}_{i,j} - \beta^o_{i,j}\right|\right) \tag{36}$$

$$\text{MSE} = \frac{1}{\#\boldsymbol{\beta}^o}\left(\sum_{i=1,2;j=0}\left(\widehat{\beta}_{i,j} - \beta^o_{i,j}\right)^2 + \sum_{i=1,2;j\geq 1} s^2_{i,j}\left(\widehat{\beta}_{i,j}\,\widehat{s}_{i,j} - \beta^o_{i,j}\right)^2\right) \tag{37}$$

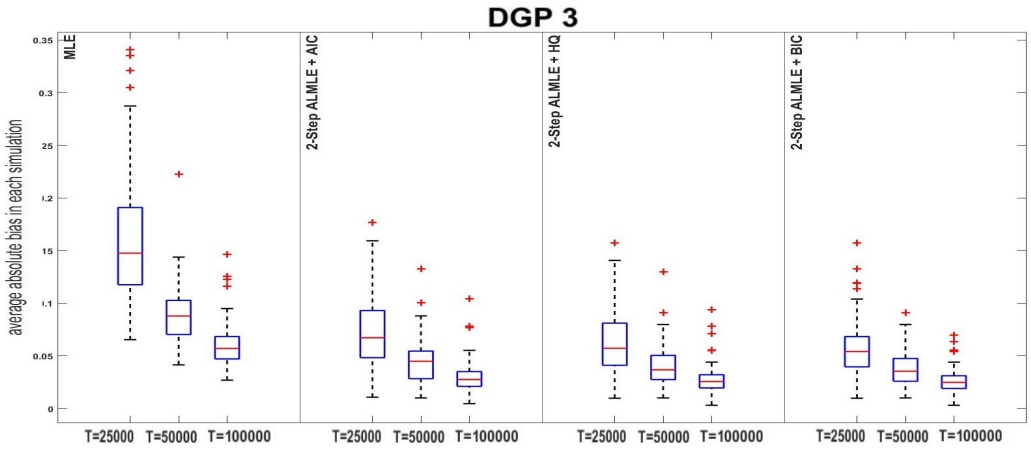
其中，$\#\boldsymbol{\beta}^o$ 表示 $\boldsymbol{\beta}^o$中的参数个数，$\widehat{s}_{i,j}$ 表示第（i,j）个预测变量的经验标准差；对于 I个（0）阶预测变量，$s_{i,j} = 1$ ；对于 I个（1）阶预测变量，$s_{i,j} = \sqrt{T}$ 。

表1展示了100次重复试验所得系数估计值的平均绝对偏差和平均MSE。结果表明，随着 T 增加， $\widehat{\boldsymbol{\beta}}^{mle}$ 和 $\widehat{\boldsymbol{\beta}}^{tal}$ 的偏差及MSE均降低，这与第3.2节和第4节所呈现的理论结果一致。此外，在各数据生成过程中，基于贝叶斯信息准则（BIC）的 $\widehat{\boldsymbol{\beta}}^{tal}$ 始终具有最低的偏差和MSE，这支持了在实证部分采用结合BIC的 $\widehat{\boldsymbol{\beta}}^{tal}$ 。图1中的偏差箱线图也支持这些结论。

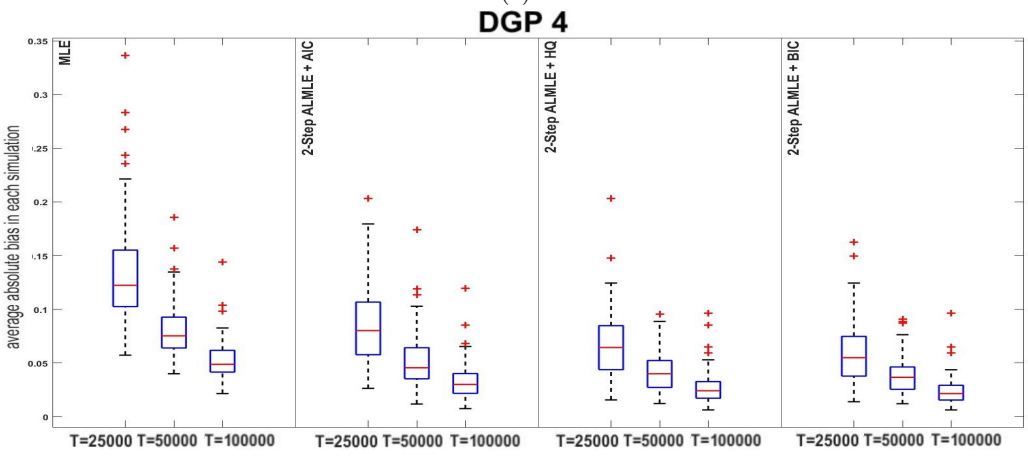
表 2 展示了 $\widehat{\boldsymbol{\beta}}^{tal}$ 和 $\widehat{\boldsymbol{\beta}}^{mle}$的变量选择结果。请注意，对于后者，我们根据与候选预测变量相关的 t统计量的显著性进行变量选择。为了衡量能否选出正确的预测变量，我们评估了形状参数和尺度参数中真实活跃预测变量与非活跃预测变量的平均选择率。此外，我们还计算了各估计量的正确分类率（CCR），即选中的真实活跃预测变量和未选中的真实非活跃预测变量占所有候选预测变量的比例。表 2 的结果表明，对于每种估计量，随着 T 增加，变量选择效果都会提升。对于 $\widehat{\boldsymbol{\beta}}^{mle}$ ，真实非活跃平稳预测变量的平均选择率趋近于显著性水平 $\alpha = 0.05$；而真实非活跃局部单位根预测变量的平均选择率则远高于 $\alpha = 0.05$，无论是对于 k 还是 σ，且不受数据生成过程影响。这些结果与第 3.2 节推导的渐近结果一致，也印证了第 4 节所讨论的使用 t检验选择非平稳预测变量时存在的尺寸扭曲问题。值得注意的是， $\widehat{\boldsymbol{\beta}}^{mle}$ 中真实活跃预测变量的平均选择率远低于 $\widehat{\boldsymbol{\beta}}^{tal}$。此外，由于估计 $\widehat{\boldsymbol{\beta}}^{mle}_{1\cdot}$存在不确定性，使用 $\widehat{\boldsymbol{\beta}}^{mle}_{1\cdot}$进行 t检验的检验功效低于使用 $\widehat{\boldsymbol{\beta}}^{mle}_{2\cdot}$进行检验的功效。最后，表 2 显示，采用贝叶斯信息准则的 $\widehat{\boldsymbol{\beta}}^{tal}$ 始终具有最高的 CCR，无论数据生成过程如何，其选择结果也最准确，这支持在实证应用中采用基于贝叶斯信息准则的 $\widehat{\boldsymbol{\beta}}^{tal}$ 。

6 实证研究

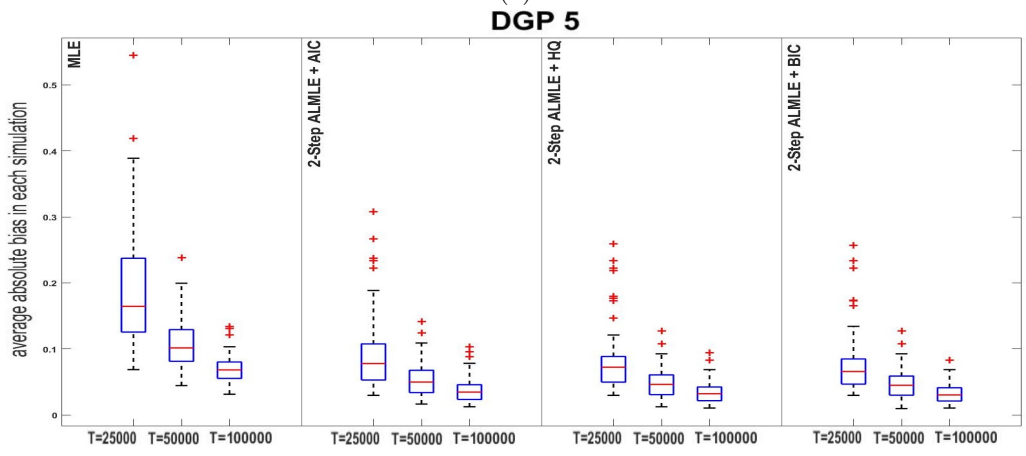
我们研究九只流动性较强的大型美国股票的高频超额损失分布：美国运通（AXP）、波音（BA）、通用电气（GE）、家得宝（HD）、IBM、强生（JNJ）、摩根大通（JPM）、可口可乐（KO）和埃克森美孚（XOM）。我们的数据涵盖了2006年1月至2014年12月期间观测到的所有交易。鉴于市场不确定性和流动性难以确切界定，我们使用多项高频波动性和流动性指标作为预测变量，研究它们对超额损失分布的影响，并选择



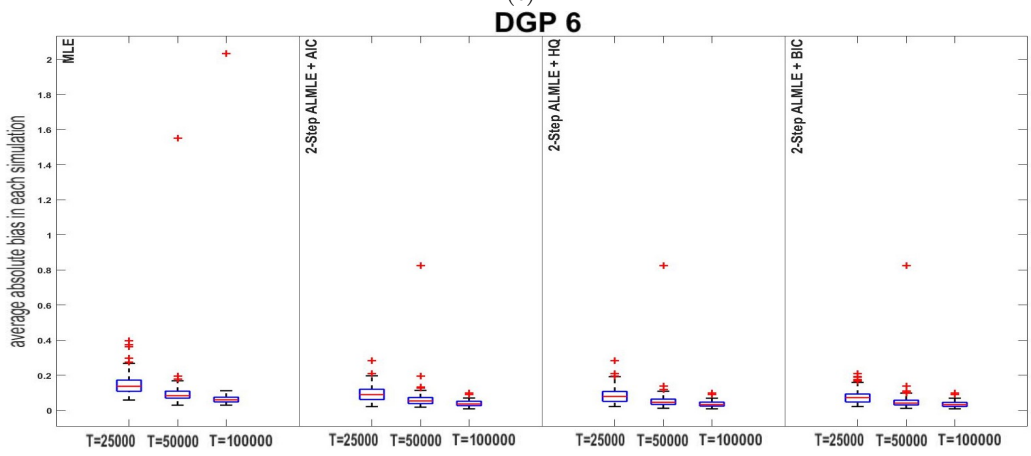
(a)



(b)



(c)



(d)

图1：使用 $\widehat{\beta}^{mle}$ 和 $\widehat{\beta}^{tal}$，并采用由赤池信息准则、Hannan-Quinn 信息准则和贝叶斯信息准则选出的最优调优参数进行 100 次重复试验所得的偏差 (36) 箱线图。12

表 1：平均值 在使用 $\widehat{\boldsymbol{\beta}}^{mle}$ 和 $\widehat{\boldsymbol{\beta}}^{tal}$ 并采用
最优调优参数时，基于 AIC、HQIC 和 BIC 准则选出的 100 次重复试验的绝对偏差（36）和均方误差（37）。

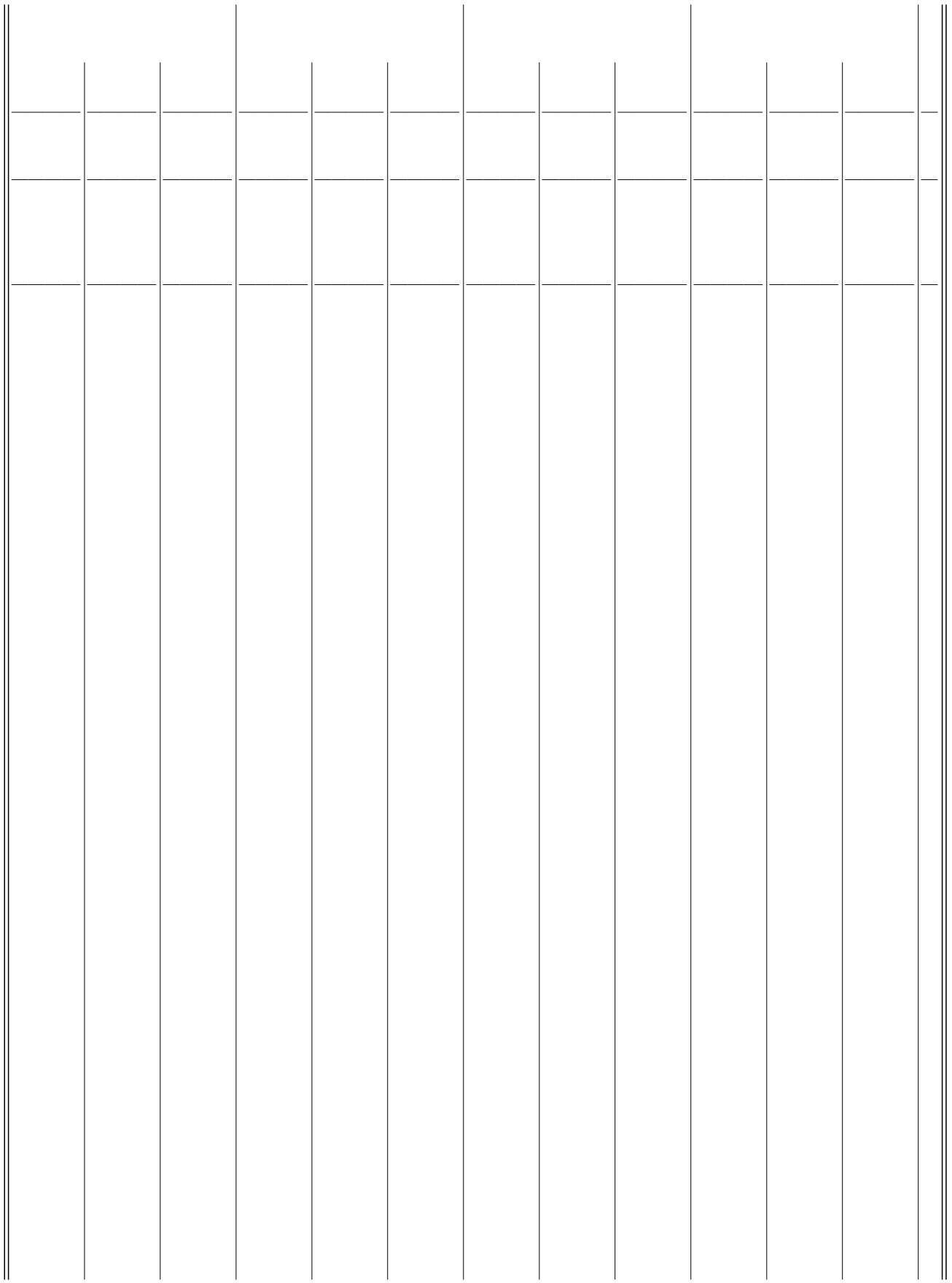
		Bias			MSE		
DGPs	T / 估计量	25,000	50,000	100,000	25,000	50,000	100,000
数据生成过程 1	$\widehat{\boldsymbol{\beta}}^{mle}$	0.031	0.015	0.008	0.105	0.031	0.016
	$\widehat{\boldsymbol{\beta}}^{tal}$ +赤池信息准则	0.014	0.007	0.004	0.036	0.014	0.007
	$\widehat{\boldsymbol{\beta}}^{tal}$ +Hannan-Quinn 信息准则	0.010	0.006	0.004	0.030	0.012	0.006
	$\widehat{\boldsymbol{\beta}}^{tal}$ +贝叶斯信息准则	0.010	0.005	0.004	0.026	0.011	0.006
数据生成过程 2	$\widehat{\boldsymbol{\beta}}^{mle}$	0.027	0.017	0.010	0.091	0.031	0.014
	$\widehat{\boldsymbol{\beta}}^{tal}$ +赤池信息准则	0.015	0.011	0.006	0.047	0.020	0.009
	$\widehat{\boldsymbol{\beta}}^{tal}$ +Hannan-Quinn 信息准则	0.012	0.009	0.004	0.034	0.012	0.007
	$\widehat{\boldsymbol{\beta}}^{tal}$ +贝叶斯信息准则	0.009	0.007	0.003	0.030	0.011	0.005
数据生成过程 3	$\widehat{\boldsymbol{\beta}}^{mle}$	0.047	0.023	0.009	0.197	0.051	0.023
	$\widehat{\boldsymbol{\beta}}^{tal}$ +赤池信息准则	0.021	0.015	0.005	0.082	0.024	0.013
	$\widehat{\boldsymbol{\beta}}^{tal}$ +Hannan-Quinn 信息准则	0.018	0.013	0.003	0.067	0.021	0.011
	$\widehat{\boldsymbol{\beta}}^{tal}$ +贝叶斯信息准则	0.017	0.012	0.003	0.060	0.020	0.010
DGP 4	$\widehat{\boldsymbol{\beta}}^{mle}$	0.029	0.025	0.019	0.128	0.165	0.263
	$\widehat{\boldsymbol{\beta}}^{tal}$ +赤池信息准则	0.018	0.017	0.006	0.071	0.099	0.013
	$\widehat{\boldsymbol{\beta}}^{tal}$ +Hannan-Quinn 信息准则	0.015	0.014	0.006	0.065	0.093	0.012
	$\widehat{\boldsymbol{\beta}}^{tal}$ +贝叶斯信息准则	0.015	0.014	0.005	0.057	0.091	0.011

采用第 4 节提出的两步 ALMLE，选出最合适的预测变量。我们进行样本内分析，对选定的预测变量对超额损失分布的影响作出经济解释；并开展样本外 VaR 预测分析，以评估预测的超额损失分布的拟合优度。

6.1 变量描述

所研究股票的原始日内数据包含以毫秒为单位的交易时间戳、每笔交易的每股交易价格，以及以股为单位的交易量。我们按照 Brownlees 和 Gallo（2006）以及 Barndorff-Nielsen 等（2009）提出的标准流程清理原始数据。由于交易数据的时间间隔不规则，我们需要在固定频率下定义一个等间隔网格，以便使用我们的模型分析损失。我们选择以五分钟频率分析损失。令 $P_{t,\,i}$ 表示第 t 个五分钟区间内第 i 笔交易的交易价格，令 $V_{t,\,i}$ 表示相应的成交股数，其中 $0 \le i \le n_t$，n_t 表示第 t 个五分钟区间内的交易笔数，且 $0 < t \le T$。我们将第 t 个五分钟区间内的交易价格中位数定义为五分钟价格 P_t，并将五分钟损失计算为第 t 个收益率的负值，$R_t := \log(P_t) - \log(P_{t-1})$。为获得超额损失时间序列，我们采用动态阈值，以反映高频损失随时间变化的特征。具体而言，时点 t 的阈值 u_t 定义为期间 $(t-1,\ t-h)$ 内观测到的损失的 90% 分位数，其中 $h > 1$ 为移动窗口大小。我们考虑 h 的 12 个可能取值，范围从一周到十二周。

流动性是指以较低的价格冲击、成本和延迟交易大量金融工具的能力。由于流动性可分解为不同维度（Harris et al., 1990），我们将若干流动性指标视为可能的预测变量。同样，为了刻画市场不确定性，我们考虑若干指标来衡量观测到的交易价格离散度。此外，为了区分不同频率下交易活动的影响，我们构建候选预测变量集时，既考虑 t个五分钟区间内的信息，也考虑 t个五分钟区间邻域中的信息。设$P_{t,\,BU} := [P_{t,1}, \ldots, P_{t,n_t}]'$和$R_{t,\,BU} := [R_{t,1}, \ldots, R_{t,n_t}]'$分别为在第 t个五分钟区间内观测到的成交价格和交易收益率向量，其中$R_{t,\,i} := \log(P_{t,\,i}) - \log(P_{t,\,i-1})$。设 T_w 为邻域大小，并定义$P_{t,\,T_w} := [P_t, \ldots, P_{t-T_w+1}]'$为邻域大小为 T_w的邻域内五分钟价格向量，以及$R_{t,\,T_w} :=$



$\log(P_{t,\ T_w})\ -\log(P_{t-1,\ T_w})$ 表示相应的收益率向量。令 $\text{dur}_{t,\ i}$ 表示i个交易在 t个五分钟区间内的执行时长，即订单执行时间与订单下达时间之间的时间差。表3列出了我们在分析中考虑的流动性预测变量。根据其频率（即在五分钟区间内或跨五分钟区间）以及性质（价格冲击或价差代理指标（Goyenko 等，2009）或流动性度量的波动率），对这些变量进行分类。表4列出了我们在分析中考虑的波动性预测变量，并根据其计算频率（即在五分钟区间内或跨五分钟区间）进行分类。

表3：使用五分钟区间内的信息（W）、大小为T_w的邻域内跨越5分钟观测值（A）或作为两种频率之间的比率（R）计算的流动性指标，分别作为价格冲击（PI）、价差（S）和流动性波动率（Vol）的代理指标。 △ 表示向量的一阶差分算子；cov(,) 和 cor(,) 分别表示两个输入变量之间的协方差和相关系数；var() 表示输入变量或向量的方差。

频率	**代理指标**	**流动性预测变量**	**公式**
W	PI	交易量	$\text{交易量}_t = \sum_{i=1}^{n_t} P_{t,i}\, V_{t,i}$
W	PI	交易数量	$\text{TQ}_t = \sum_{i=1}^{n_t} V_{t,i}$
W	Vol	微观交易量波动率	$\text{MTVV}_t = \sqrt{\frac{1}{n_t}\sum_{j=1}^{n_t}\left(P_{t,j}\,V_{t,j} - \frac{1}{n_t}\sum_{i=1}^{n_t} P_{t,i}\,V_{t,i}\right)^2}$
W	Vol	以股计交易数量的微观波动率	$\text{MTQV}_t = \sqrt{\frac{1}{n_t}\sum_{j=1}^{n_t}\left(V_{t,j} - \frac{1}{n_t}\sum_{i=1}^{n_t} V_{t,i}\right)^2}$
W	PI	Amihud非流动性度量	$\text{AM}_t = \frac{1}{n_t}\sum_{i=1}^{n_t} \frac{\lvert R_{t,i}\rvert}{P_{t,i}\,V_{t,i}}$
W	PI	扩展Amihud度量（Goyenko et al., 2009）	$\text{EAM}_t = \frac{\max(P_{t,BU}) - \min(P_{t,BU})}{\text{TV}_t}$
W	PI	交易持续时间	$\text{持续时间}_t = \frac{\sum_{i=1}^{n_t} \text{持续时间}_{t,\ i}}{n_t}$
A	S	Roll（Roll，1984年）	$\text{Roll}_t = \text{协方差}\ (\triangle P_{t,\ T_w},\ \triangle P_{t-1.T_w})$
A	S	修正Roll模型	$\text{RollMod}_t = \frac{\text{协方差}\ (\triangle P_{t,\ T_w},\ \triangle P_{t-1.T_w})}{P^m_t}$
A	S	负Roll模型	$\text{Roll}_t^- = \text{Roll}_t\, \mathbb{1}\{\text{Roll}_t < 0\}$
A	S	负向修正Roll指标	$\text{RollMod}_t^- = \text{RollMod}_t\, \mathbb{1}\{\text{RollMod}_t < 0\}$
A	S	收益自相关（Grossman和Miller，1988年）	$\text{RAC}_t = \text{cor}(R_{t,T_w}, R_{t-1.T_w})$
A	PI	Amihud非流动性度量	$\text{AMI}_t = \frac{1}{T_w}\frac{\sum_{j=0}^{T_w-1} \lvert R_{t-j}\rvert}{\text{TV}_{t-j}}$
A	Vol	交易量波动率	$\text{TVV}_t = \sqrt{\frac{1}{T_w}\sum_{j=0}^{T_w-1}\left(\text{TV}_{t-j} - \text{TV}_{t,\ T_w}\right)^2}$
A	Vol	相对交易量波动率	$\text{RTVV}_t = \frac{\text{TVV}_t}{\frac{1}{T_w}\sum_{j=0}^{T_w-1}\text{TV}_{t-j}}$
A	Vol	交易量波动率	$\text{TQV}_t = \sqrt{\frac{1}{T_w}\sum_{j=0}^{T_w-1}\left(\text{TQ}_{t-j} - \frac{1}{T_w}\sum_{j=0}^{T_w-1}\text{TQ}_{t-j}\right)^2}$
A	Vol	相对交易量波动率	$\text{RTQV}_t = \frac{\text{TQV}_t}{\text{TQ}_{t,\ T_w}}$
R	S	方差比（Hasbrouck 和 Schwartz，1988年）	$\text{VR}_t = \frac{T_w \cdot \text{var}(\log(P_{t,BU}) - \log(P_t))}{n_t\, \text{var}(R_{t,\ T_w})}$

6.2 样本内估计值

我们将每个时间序列划分为样本内期间和样本外期间：前者涵盖前90%的观测值，后者涵盖样本最后10%的观测值。我们使用式（3）—（4）中的时变GPD回归模型，对每只股票的超额损失 $\{y_t\}$ 进行建模，并采用表3和表4中定义的变量，其中 $T_w \in \{2, 6, 12\}$可作为尺度参数和形状参数的候选预测变量。使用两步 ALMLE 得到的系数估计值分别列于表5和表6，分别对应形状参数和尺度参数。

表5中形状参数的结果显示，股票的估计系数几乎总是符号相同。关于流动性预测变量，我们发现，几乎所有股票都选中了价格冲击代理变量，这表明它们能更好地捕捉流动性对极端损失的影响。具体而言，TV 和 TQ 的系数为正，而 AM 和 EAM 的系数为负，这意味着更大的极端

表4：波动性　　利用五分钟区间内（W）以及跨越（A）5-最少观测值数为 a T_w的邻域内，或表示为两种频率之间的比率（R）。

频率	波动率预测变量	公式
W	微观噪声收益波动率	$\text{MNRV}_t = \sqrt{\frac{1}{n_t}\sum_{i=1}^{n_t}(\log(P_{t,i}) - \log(P_t))^2}$ —
W	微观实现波动率	$\text{MRV}_t = \sqrt{\sum_{i=1}^{n_t}(\log(P_{t,i}) - \log(P_{t,i-1}))^2}$
A	已实现波动率	$\text{RV}_{t,\ T_w} = \sqrt{\sum_{j=0}^{T_w-1} R_{t-j}^2}$
R	MNRV2RV	$\text{MNRV2RV} = \frac{\text{MNRV}_t}{\text{RV}_t}$
R	MRV2RV	$\text{MRV2RV} = \frac{\text{MRV}_t}{\text{RV}_t}$

损失与最后五分钟内较高的流动性水平相关。乍看之下，这一结果可能有悖直觉，但结合其他选定变量来看，却十分有趣。关于流动性波动率，我们注意到，大多数股票都选中了 RTVV($T_w = 6$) 和 RTQV($T_w = 6$)，且其系数较大且为正，表明流动性波动率较高的时期往往伴随着更大的极端损失。几乎每只股票都选中了比率 MRV2RV($T_w = 12$)，该比率本质上捕捉了波动率的波动率或跳跃风险对极端损失的影响，并且其系数为正，说明不确定性较高的时期往往会出现更大的极端损失。总的来说，这些结果与 Brogaard et al. (2018) 的发现一致，即做市商会放大极端价格变动，而在较大的不确定性冲击导致其流动性供给无法满足流动性需求后，便会退出市场。

表6显示，尺度参数选取的变量更多，但这些变量在不同股票之间的模式较不稳定。总体而言，我们注意到自回归成分对动态变化有所贡献，而且在五分钟区间内计算的已实现波动率预测变量始终被选中，且系数为正。这与尺度参数能够捕捉数据中的时变异方差这一事实相符。

为便于比较，我们在表 7-8 中报告了通过最大似然估计（MLE）得到的形状参数和尺度参数的回归系数估计值。所有估计系数均不为零。由于得到的 MLE 费舍尔信息矩阵不是正定矩阵，因此无法计算相应的标准误。这使得变量解释极其困难，甚至不可能。

6.3 样本外风险价值预测

$\widehat{}$使用样本内期间得到的系数估计值 $\boldsymbol{\beta}$ ，计算样本外期间的一步风险价值预测。具体而言，在时间 t ，给定 $\boldsymbol{x}_{t-1}\widehat{}$和 $\boldsymbol{\beta}$ 时，风险水平为 α 的每只股票的风险价值为

$$\widehat{\text{VaR}_t}(\alpha) = \frac{\widehat{\sigma}_t(\boldsymbol{x}_{t-1}, \widehat{\boldsymbol{\beta}})}{\widehat{k}_t(\boldsymbol{x}_{t-1}, \widehat{\boldsymbol{\beta}})}\left(\left(1 - \frac{\alpha - F_t(\widehat{u}_t)}{1 - F_t(\widehat{u}_t)}\right)^{-\widehat{k}_t(\boldsymbol{x}_{t-1}, \widehat{\boldsymbol{\beta}})} - 1\right) + \widehat{u}_t. \tag{38}$$

其中， $F_t(\widehat{u}_t)$ 表示超过阈值 $\widehat{u}_t$ 的概率，并固定为 90%。样本外期间$\{\widehat{\text{风险价值}_t(\alpha)}\}_{t=T_{is}+1}^{T_{is}+T_{os}}$ 的覆盖率计算如下：

$$\text{Coverage Rate} = \frac{\sum_{t=T_{is}+1}^{T_{is}+T_{os}} \mathbb{1}\{l_t \le \widehat{\text{VaR}_t}(\alpha)\}}{T_{os}}. \tag{39}$$

表9展示了在风险水平 α 下，各种 $\alpha \in [90\%, 100\%)$ 对应的 $\{\widehat{\text{风险价值}_t(\alpha)}\}_{t=T_{is}+1}^{T_{is}+T_{os}}$ 覆盖率。我们采用 Kolmogorov–Smirnov （ K-S ）检验，检验样本外预测的广义帕累托分布的拟合优度

表5：通过两步 ALMLE 获得的形状参数回归系数的经验估计。

股票 / 预测变量	AXP	BA	GE	HD	IBM	JNJ	JPM	KO	XOM	按股票选择
TV	1.105	0.676		0.581	1.912	1.205	0.660	1.109	2.014	0.89
TQ	1.314	1.101	0.501	0.874	0.813	1.130	0.756		0.0004	0.89
AM	-0.345		0.719	-0.037			-0.522	-0.030		0.56
MTVV					0.067					0.11
EAM	-1.637	-1.760		-0.913	-1.789	-1.681	-1.396	-1.650	-0.201	0.89
MTQV	0.200		0.166		0.009	0.177				0.44
MRV			0.328		-0.901	-0.367	-0.274			0.44
MNRV		-0.794	-0.242	-0.491				0.020		0.44
dur										0.00
AMI（$T_w = 2$）										0.00
VR（$T_w = 2$）										0.00
RV ($T_w = 2$)										0.00
MNRV2RV ($T_w = 2$)								-0.084		0.11
MRV2RV ($T_w = 2$)								0.095		0.11
AMI ($T_w = 6$)								-0.003		0.11
Roll ($T_w = 6$)										0.00
Roll$^-$ ($T_w = 6$)										0.00
RollMod（$T_w = 6$）										0.00
RollMod$^-$（$T_w = 6$）										0.00
TVV（$T_w = 6$）										0.00
TQV（$T_w = 6$）										0.00
RTVV（$T_w = 6$）		0.068		0.298			0.159	0.218	0.264	0.56
RTQV ($T_w = 6$)		0.069		0.300			0.159	0.218	0.270	0.56
RAC ($T_w = 6$)										0.00
VR ($T_w = 6$)										0.00
RV ($T_w = 6$)			-0.203							0.11
MNRV2RV ($T_w = 6$)			-0.415					-0.185		0.22
MRV2RV ($T_w = 6$)							0.194	-0.010	0.0003	0.33
AMI ($T_w = 12$)			0.524					0.648		0.22
Roll ($T_w = 12$)										0.00
Roll$^-$ ($T_w = 12$)										0.00
RollMod ($T_w = 12$)										0.00
RollMod$^-$ ($T_w = 12$)										0.00
TVV ($T_w = 12$)			-0.851							0.11
TQV ($T_w = 12$)										0.00
RTVV ($T_w = 12$)								-0.090		0.11
RTQV ($T_w = 12$)								-0.090		0.11
RAC ($T_w = 12$)										0.00
VR ($T_w = 12$)	0.415						0.244		-0.00002	0.33
RV ($T_w = 12$)								-0.436		0.11
MNRV2RV ($T_w = 12$)								0.297		0.11
MRV2RV ($T_w = 12$)		0.994	0.608	0.287	0.554	0.548	0.487	0.544	0.766	0.89
所选变量总数	6	7	10	8	5	7	11	17	8	

表6：采用两步 ALMLE 得到的尺度参数回归系数经验估计。

股票 / 预测变量	AXP	BA	GE	HD	IBM	JNJ	JPM	KO	XOM	按股票选择
TV	-0.004		-0.005		0.042			-0.044	0.212 0.095	0.67
TQ	0.039	0.110	0.015	0.042		0.047		0.092	-0.102 -0.095	0.89
AM			0.024	0.050						0.22
MTVV		0.048			-0.029			-0.137	-0.260	0.44
EAM				-0.040	-0.012				-0.043	0.33
MTQV		-0.123	0.016		0.021			0.135	0.168 -0.003	0.67
MRV	0.049		0.032					0.057		0.33
MNRV	0.049	0.139	0.056	0.055	0.067	0.079		0.048	0.128 0.091	1.00
dur										0.00
AMI ($T_w = 2$)					-0.018					0.11
VR ($T_w = 2$)										0.00
RV ($T_w = 2$)					0.017			-0.021		0.22
MNRV2RV ($T_w = 2$)										0.00
MRV2RV ($T_w = 2$)										0.00
AMI ($T_w = 6$)					-0.014				-0.003	0.22
Roll ($T_w = 6$)						-0.054				0.11
Roll$^-$ ($T_w = 6$)				-0.035				-0.086		0.22
RollMod ($T_w = 6$)	-0.061				-0.048				-0.050	0.33
RollMod$^-$ ($T_w = 6$)	0.077		0.008	0.051	0.053	0.062		0.104	0.061	0.78
TVV ($T_w = 6$)				0.006	0.029					0.22
TQV ($T_w = 6$)				-0.015	-0.048	-0.006				0.33
RTVV ($T_w = 6$)					0.015					0.11
RTQV ($T_w = 6$)					0.0003					0.11
RAC ($T_w = 6$)										0.00
VR ($T_w = 6$)					-0.003					0.11
RV ($T_w = 6$)			0.017	0.038						0.22
MNRV2RV ($T_w = 6$)	-0.002	-0.056				-0.037			-0.082	0.44
MRV2RV ($T_w = 6$)	-0.028	0.041							0.044	0.33
AMI ($T_w = 12$)					0.039					0.11
Roll ($T_w = 12$)			-0.005					0.027		0.22
Roll$^-$ ($T_w = 12$)			0.006							0.11
RollMod ($T_w = 12$)	-0.005							-0.038	-0.014	0.33
RollMod$^-$ ($T_w = 12$)	0.004							0.006	0.022	0.33
TVV ($T_w = 12$)		-0.036		-0.009						0.22
TQV ($T_w = 12$)	-0.013				-0.001	-0.010		-0.011		0.44
RTVV ($T_w = 12$)									0.003	0.11
RTQV ($T_w = 12$)		0.046								0.11
RAC ($T_w = 12$)										0.00
VR ($T_w = 12$)					-0.004				-0.032	0.22
RV ($T_w = 12$)						0.027			0.034	0.22
MNRV2RV（$T_w = 12$）	0.014	0.007	0.032	0.037		0.061		0.023	0.103	0.78
MRV2RV（$T_w = 12$）	0.022				0.021	-0.005	-0.00005	0.014	-0.016	0.67
log（σ_{t-1}）	0.849	0.631	0.758	0.724	0.832	0.778		0.829	0.668 0.770	1.00
所选变量总数	14	10	12	12	20	11		16	7 18	

表7：通过最大似然估计得到的尺度参数回归系数经验估计。

预测变量 \ 股票	AXP	BA	GE	HD	IBM	JNJ	JPM	KO	XOM	按股票选择
TV	0.396	0.264	0.116	0.264	0.563	0.631	0.239	0.602	0.239	1.00
TQ	0.469	0.313	0.173	0.326	0.457	0.619	0.259	0.519	0.216	1.00
AM	-0.083	-0.167	0.303	-0.042	-0.287	-0.394	-0.067	-0.221	-0.080	1.00
TVV	0.249	0.017	0.064	0.094	0.135	0.115	0.055	0.090	0.059	1.00
EAM	-0.621	-0.406	-0.062	-0.451	-0.525	-1.095	-0.236	-0.966	-0.350	1.00
TQV	0.195	0.020	0.076	0.100	0.151	0.112	0.053	0.087	0.057	1.00
MRV	0.062	0.001	0.224	-0.003	0.136	0.188	0.051	0.086	0.048	1.00
MNRV	-0.055	-0.081	0.072	-0.140	-0.021	0.028	0.009	-0.121	-0.035	1.00
dur	-0.025	0.000	-0.026	0.016	0.000	0.033	0.020	0.021	0.000	1.00
AMI（$T_w = 2$）	-0.130	-0.048	0.036	0.129	-0.386	-0.194	-0.008	0.098	-0.239	1.00
VR（$T_w = 2$）	0.000	0.060	0.000	0.013	0.000	0.034	-0.001	0.016	0.000	0.56
RV ($T_w = 2$)	-0.047	-0.048	0.036	-0.178	0.047	0.006	0.027	-0.005	-0.070	1.00
MNRV2RV ($T_w = 2$)	0.000	0.048	0.000	0.012	0.000	0.084	-0.043	-0.066	0.000	0.78
MRV2RV ($T_w = 2$)	0.000	0.035	0.000	0.010	0.000	0.007	-0.025	0.036	0.000	0.78
AMI ($T_w = 6$)	0.005	-0.154	0.107	-0.059	-0.401	-0.188	0.000	0.111	0.019	1.00
Roll ($T_w = 6$)	0.043	0.035	-0.010	0.018	0.032	-0.045	-0.004	0.001	0.001	1.00
Roll$^-$ ($T_w = 6$)	0.052	0.019	-0.016	0.045	0.011	-0.066	-0.011	0.001	0.010	1.00
RollMod ($T_w = 6$)	0.026	0.024	-0.022	-0.014	0.029	-0.051	-0.007	0.001	0.005	1.00
RollMod$^-$ ($T_w = 6$)	0.024	0.013	-0.031	0.011	0.008	-0.076	-0.016	0.001	0.006	1.00
TVV ($T_w = 6$)	0.015	0.026	-0.048	0.063	0.004	0.010	0.059	-0.021	0.023	1.00
TQV ($T_w = 6$)	0.062	0.038	-0.004	0.095	-0.015	0.006	0.070	-0.049	0.020	1.00
RTVV ($T_w = 6$)	0.132	0.081	0.080	0.182	0.080	0.152	0.093	0.168	0.147	1.00
RTQV ($T_w = 6$)	0.131	0.081	0.080	0.182	0.080	0.152	0.092	0.168	0.147	1.00
RAC ($T_w = 6$)	-0.042	-0.063	-0.029	-0.071	-0.020	-0.013	0.053	-0.078	-0.101	1.00
VR ($T_w = 6$)	0.103	0.032	-0.010	0.115	-0.018	-0.020	0.033	-0.052	0.043	1.00
RV ($T_w = 6$)	-0.116	-0.104	0.116	-0.176	-0.036	-0.099	-0.015	-0.114	-0.128	1.00
MNRV2RV ($T_w = 6$)	0.011	0.090	-0.151	0.058	0.009	0.044	-0.004	-0.087	-0.018	1.00
MRV2RV ($T_w = 6$)	0.097	0.165	0.031	0.188	0.231	0.107	0.068	0.142	0.132	1.00
AMI ($T_w = 12$)	-0.032	-0.133	0.162	-0.036	-0.259	-0.030	-0.021	0.152	-0.015	1.00
Roll（$T_w = 12$）	-0.019	0.005	-0.023	0.000	0.047	0.009	0.011	0.002	0.009	1.00
Roll$^-$ （$T_w = 12$）	0.074	0.059	-0.013	0.081	0.041	0.031	0.014	0.003	0.040	1.00
RollMod ($T_w = 12$)	-0.021	0.012	-0.026	-0.040	0.045	0.006	0.009	0.002	0.013	1.00
RollMod$^-$ ($T_w = 12$)	0.026	0.051	-0.029	0.020	0.039	0.018	0.006	0.002	0.031	1.00
TVV ($T_w = 12$)	-0.125	-0.049	-0.143	-0.051	-0.132	-0.141	-0.019	-0.285	-0.069	1.00
TQV（$T_w = 12$）	-0.059	-0.031	-0.096	-0.021	-0.159	-0.150	-0.008	-0.308	-0.070	1.00
RTVV（$T_w = 12$）	-0.008	0.035	0.010	-0.001	0.006	0.033	0.030	-0.028	0.056	1.00
RTQV（$T_w = 12$）	-0.008	0.036	0.010	-0.002	0.005	0.032	0.030	-0.027	0.056	1.00
RAC ($T_w = 12$)	-0.097	-0.007	-0.063	-0.124	0.036	0.132	0.019	-0.054	0.031	1.00
VR ($T_w = 12$)	0.087	0.029	0.011	0.046	0.031	0.043	0.062	-0.003	0.085	1.00
RV ($T_w = 12$)	-0.171	-0.167	0.028	-0.241	-0.118	-0.267	-0.074	-0.227	-0.258	1.00
MNRV2RV ($T_w = 12$)	0.034	0.138	-0.050	0.043	0.161	0.217	0.050	0.136	0.091	1.00
MRV2RV ($T_w = 12$)	0.163	0.228	0.170	0.190	0.376	0.309	0.131	0.376	0.257	1.00
所选变量总数	39	42	39	42	41	42	42	42	41	

表 8：经验　　最大似然估计得到的尺度参数回归系数的估计值。

股票 / 预测变量	AXP	BA	GE	HD	IBM	JNJ	JPM	KO	XOM	按股票选择
TV	-0.033	-0.138	0.058	-0.064	0.114	-0.055	-0.075	0.314	0.253	1.00
TQ	0.094	0.325	-0.033	0.100	0.012	0.090	0.166	-0.190	-0.110	1.00
AM	-0.010	0.018	0.046	0.039	0.021	0.006	0.021	0.029	0.027	1.00
TVV	-0.045	0.157	-0.127	-0.018	0.051	-0.035	-0.224	-0.378	-0.043	1.00
EAM	-0.026	-0.039	-0.007	-0.055	-0.060	-0.018	-0.029	-0.008	-0.056	1.00
TQV	0.045	-0.312	0.142	0.011	-0.241	0.048	0.219	0.279	-0.148	1.00
MRV	0.028	0.024	0.021	0.018	-0.007	0.019	0.055	0.023	-0.015	1.00
MNRV	0.099	0.137	0.060	0.070	0.199	0.060	0.061	0.094	0.149	1.00
dur	-0.002	-0.031	0.000	-0.005	0.046	-0.007	-0.001	-0.005	0.003	1.00
AMI ($T_w = 2$)	0.030	0.009	-0.012	0.000	0.045	0.019	0.005	-0.045	0.043	1.00
VR（$T_w = 2$）	-0.005	-0.009	0.000	-0.007	0.014	0.001	0.020	-0.014	-0.007	1.00
RV（$T_w = 2$）	-0.037	-0.050	0.011	-0.006	-0.068	-0.007	-0.071	0.031	-0.015	1.00
MNRV2RV（$T_w = 2$）	0.010	0.001	0.001	-0.001	0.026	-0.010	-0.033	-0.015	-0.007	1.00
MRV2RV ($T_w = 2$)	-0.024	0.009	0.001	0.008	0.023	0.015	0.017	0.039	-0.007	1.00
AMI ($T_w = 6$)	0.009	-0.016	-0.009	-0.009	-0.047	-0.032	0.006	0.002	-0.106	1.00
Roll ($T_w = 6$)	0.069	0.014	-0.002	0.040	0.019	-0.087	0.020	0.050	-0.001	1.00
Roll$^-$ ($T_w = 6$)	-0.062	-0.013	-0.046	-0.090	-0.024	-0.022	-0.066	0.070	-0.009	1.00
RollMod ($T_w = 6$)	-0.114	0.022	-0.072	-0.027	-0.037	0.019	-0.058	0.040	-0.041	1.00
RollMod$^-$ ($T_w = 6$)	0.127	-0.016	0.139	0.096	0.049	0.113	0.125	0.058	0.068	1.00
TVV（$T_w = 6$）	0.016	0.045	-0.025	0.070	-0.033	0.055	-0.003	0.031	-0.011	1.00
TQV（$T_w = 6$）	-0.031	-0.041	0.024	-0.078	0.016	-0.074	-0.010	-0.027	-0.002	1.00
RTVV（$T_w = 6$）	-0.011	-0.010	-0.005	-0.004	0.014	0.006	-0.016	-0.003	-0.009	1.00
RTQV ($T_w = 6$)	-0.006	0.012	-0.005	0.004	0.013	0.008	0.005	-0.001	0.000	1.00
RAC ($T_w = 6$)	0.005	-0.010	0.008	-0.013	-0.012	-0.008	-0.004	-0.011	-0.002	1.00
VR ($T_w = 6$)	0.050	0.088	0.138	0.035	0.080	0.046	0.012	0.017	0.038	1.00
RV ($T_w = 6$)	0.020	-0.009	0.026	0.029	0.006	0.016	0.012	-0.064	0.031	1.00
MNRV2RV ($T_w = 6$)	-0.163	-0.190	-0.076	-0.080	-0.077	-0.114	-0.064	-0.112	-0.243	1.00
MRV2RV ($T_w = 6$)	0.081	0.054	0.014	0.010	-0.031	0.035	0.025	0.068	0.153	1.00
AMI ($T_w = 12$)	0.014	0.038	0.014	0.016	0.078	0.018	0.013	0.017	0.069	1.00
Roll ($T_w = 12$)	0.047	-0.052	0.097	0.041	-0.017	0.032	0.089	0.069	0.010	1.00
Roll$^-$ ($T_w = 12$)	-0.063	0.033	-0.071	-0.033	0.016	0.017	-0.114	0.063	-0.050	1.00
RollMod ($T_w = 12$)	-0.068	0.041	-0.077	-0.058	-0.023	-0.041	-0.094	0.063	-0.037	1.00
RollMod$^-$ ($T_w = 12$)	0.086	-0.024	0.047	0.051	0.043	-0.009	0.118	0.063	0.086	1.00
TVV ($T_w = 12$)	0.010	-0.092	0.006	-0.037	0.032	0.100	0.008	-0.029	0.001	1.00
TQV ($T_w = 12$)	-0.025	0.046	-0.015	0.031	-0.034	-0.095	-0.022	0.015	-0.011	1.00
RTVV ($T_w = 12$)	0.009	0.022	0.005	-0.001	0.007	-0.007	-0.002	0.003	0.013	1.00
RTQV ($T_w = 12$)	0.010	0.044	-0.001	0.008	-0.003	-0.003	0.016	0.011	0.012	1.00
RAC ($T_w = 12$)	0.000	0.016	-0.005	0.008	0.003	0.012	0.002	0.013	0.009	1.00
VR ($T_w = 12$)	-0.072	-0.205	-0.026	-0.029	-0.056	-0.063	-0.037	-0.022	-0.070	1.00
RV ($T_w = 12$)	0.038	0.041	0.001	0.024	0.066	0.054	0.022	0.097	0.042	1.00
MNRV2RV ($T_w = 12$)	0.183	0.149	0.083	0.073	-0.033	0.194	0.121	0.131	0.211	1.00
MRV2RV ($T_w = 12$)	-0.045	-0.019	-0.007	0.016	0.091	-0.060	-0.028	-0.071	-0.074	1.00
log(σ_{t-1})	0.752	0.480	0.736	0.674	0.445	0.739	0.761	0.645	0.545	1.00
所选变量总数	43	43	43	43	43	43	43	43	43	

$\widehat{\quad}$ $\widehat{\quad}$时期，即我们检验 $\{F(y_t|y_t > 0) = \text{GPD}(y_t;\ \boldsymbol{x}_{t-1}, \boldsymbol{\beta})\}$ 是否服从标准均匀分布。表9中的柯尔莫戈洛夫–斯米尔诺夫检验p值表明，在1%的显著性水平下，我们拒绝了九只股票中三只股票的回归模型。

表9：样本外VaR覆盖率及柯尔莫戈洛夫–斯米尔诺夫检验p值。

VaR风险水平 / 股票名称	0.9	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99	0.999	0.9999	柯尔莫戈洛夫–斯米尔诺夫检验p值
AXP	0.8999	0.9111	0.9224	0.9314	0.9414	0.9520	0.9631	0.9729	0.9827	0.9913	0.9986	0.9996	0.208
BA	0.9000	0.9098	0.9210	0.9309	0.9417	0.9522	0.9630	0.9738	0.9836	0.9930	0.9985	0.9996	0.0431
GE	0.8999	0.9171	0.9254	0.9376	0.9475	0.9622	0.9709	0.9811	0.9889	0.9943	0.9989	0.9999	2.68×10^{-12}
HD	0.9001	0.9097	0.9217	0.9339	0.9453	0.9551	0.9652	0.9756	0.9840	0.9934	0.9990	0.9998	0.0023
IBM	0.9003	0.9094	0.9204	0.9312	0.9414	0.9509	0.9614	0.9702	0.9809	0.9905	0.9980	0.9993	0.8954
JNJ	0.9000	0.9083	0.9195	0.9284	0.9382	0.9480	0.9571	0.9682	0.9786	0.9888	0.9977	0.9995	0.3322
JPM	0.9000	0.9094	0.9176	0.9294	0.9385	0.9500	0.9612	0.9711	0.9815	0.9904	0.9982	0.9997	0.3993
KO	0.8996	0.9109	0.9180	0.9273	0.9380	0.9489	0.9594	0.9709	0.9813	0.9909	0.9980	0.9994	0.3786
XOM	0.8988	0.9056	0.9136	0.9218	0.9303	0.9396	0.9502	0.9630	0.9742	0.9863	0.9972	0.9992	6.32×10^{-8}

7 结论

本文提出了一种新颖的极值回归框架，用于研究高频尾部风险的动态变化。所提出的模型纳入平稳和局部单位根预测变量，以捕捉高频极端损失的持久性。我们提出了一种两步正则化方法来自动进行变量选择，并证明相应的 ALMLE 在选择平稳和局部单位根预测变量时具有预言机性质。我们采用该方法，考察 42 项流动性和波动率指标对九只流动性较高的大型美国股票极端损失分布的预测能力。变量选择程序显示，在流动性波动率和波动率的波动率较高的时期，尾部风险的严重程度与较低的价格冲击密切相关。这些发现有助于及时提醒高频交易者风险水平正在上升，并推动其改进算法交易实践，以加强金融风险管理。此外，这些发现也能鼓励做市商在市场不稳定时期满足流动性需求。最后，本文提出了一组预测变量，供监管机构在调查交易活动时参考，帮助其制定适当的监管指引并维护金融稳定。

参考文献

Balkema, A. A. 和 De Haan, L.（1974）。高龄时的剩余寿命。《概率年刊》，2(5):792–804。Barndorff-Nielsen, O. E.、Hansen, P. R.、Lunde, A. 和 Shephard, N.（2009）。实践中的已实现核：交易与报价。《计量经济学杂志》，12:C1–C32。Bee, M.、Dupuis, D. J. 和 Trapin, L.（2019）。已实现超阈值峰值：一种采用基于高频度量的时变极值方法。《金融计量经济学杂志》，17(2):254–283。Billingsley, P.（2013）。概率测度的收敛。约翰·威立父子出版公司。Brogaard, J.、Carrion, A.、Moyaert, T.、Riordan, R.、Shkilko, A. 和 Sokolov, K.（2018）。高频交易与极端价格变动。《金融经济学杂志》，128(2):253–265。Brownlees, C. T. 和 Gallo, G. M.（2006）。超高频金融计量经济分析：数据处理方面的问题。《计算统计与数据分析》，51(4):2232–2245。Chavez-Demoulin, V. 和 Davison, A.（2012）。时间序列极值建模。*REVSTAT*，10(1):109–133。Chavez-Demoulin, V.、Embrechts, P. 和 Hofert, M.（2016）。一种用于根据协变量对操作风险损失建模的极值方法。《风险与保险杂志》，83(3):735–776。Chavez-Demoulin, V.、Embrechts, P. 和 Sardy, S.（2014）。金融时间序列的极端分位数跟踪。《计量经济学杂志》，181(1):44–52。Coles, S.（2001）。极值统计建模导论。Springer出版社。Cs¨org˝o, M.（1968）。关于鞅的强大数定律与中心极限定理。《美国数学会汇刊》，131(1):259–275。

Dionne, G.、Duchesne, P. 和 Pacurar, M.（2009）。使用逐笔数据计算日内风险价值（IVaR），并应用于多伦多证券交易所。《实证金融杂志》，16(5):777–792。Engle, R.（2001）。GARCH 101：ARCH/GARCH模型在应用计量经济学中的运用。《经济展望杂志》，15(4):157–168。Giot, P.（2005）。日内数据的市场风险模型。《欧洲金融杂志》，11(4):309–324。Goyenko, R. Y.、Holden, C. W. 和 Trzcinka, C. A.（2009）。流动性度量是否真正衡量了流动性? 《金融经济学杂志》，92(2):153–181。Grossman, S. J. 和 Miller, M. H.（1988）。流动性与市场结构。《金融杂志》，43(3):617–633。Hambuckers, J.、Groll, A. 和 Kneib, T.（2018）。理解操作损失严重程度的经济决定因素：一种正则化广义帕累托回归方法。《应用计量经济学杂志》，33(6):898–935。Hasbrouck, J. 和 Schwartz, R. A.（1988）。股票市场的流动性与执行成本。《投资组合管理杂志》，14(3):10。Hendershott, T.、Jones, C. M. 和 Menkveld, A.（2011）。算法交易能否改善流动性? 《金融杂志》，66(1):1–33。Hosking, J. R. 和 Wallis, J. R.（1987）。广义帕累托分布的参数与分位数估计。《技术计量学》，29(3):339–349。James, G.、Witten, D.、Hastie, T. 和 Tibshirani, R.（2013）。统计学习导论。Springer出版社。Kirilenko, A.、Kyle, A. S.、Samadi, M. 和 Tuzun, T.（2017）。闪电崩盘：电子市场中的高频交易。《金融杂志》，72(3):967–998。Kock, A. B.（2016）。在平稳与非平稳自回归模型中，使用自适应 LASSO 进行一致且保守的模型选择。《计量经济学理论》，32(1):243–259。Lee, J. H.（2016）。持久协变量下的预测分位数回归：IVX-QR方法。《计量经济学杂志》，192(1):105–118。Lee, J. H.、Shi, Z. 和 Gao, Z.（2022）。关于用于预测回归的 Lasso。《计量经济学杂志》，229(2):322–349。Massacci, D.（2017）。股票收益率的尾部风险动态：与宏观经济及全球市场连通性的联系。《管理科学》，63(9):3072–3089。Medeiros, M. C. 和 Mendes, E. F.（2016）。使用 $\mathcal{L}_1$ 正则化对具有非高斯和异方差误差的高维时间序列模型进行估计。《计量经济学杂志》，191(1):255–271。Nieto, M. R. 和 Ruiz, E.（2016）。VaR预测与回测前沿。《国际预测杂志》，32(2):475–501。Phillips, P. C.（1991）。协整系统中的最优推断。《计量经济学：计量经济学会杂志》，59(2):283–306。Phillips, P. C. 和 Durlauf, S. N.（1986）。含单整过程的多元时间序列回归。《经济研究评论》，53(4):473–495。Phillips, P. C. 和 Lee, J. H.（2013）。不同持久性程度下的预测回归与稳健长期回归。《计量经济学杂志》，177(2):250–264。Pickands, J.（1975）。利用极值次序统计量进行统计推断。《统计年刊》，3(1):119–131。Roll, R.（1984）。有效市场中有效买卖价差的一种简单隐含度量。《金融杂志》，39(4):1127–1139。Saikkonen, P.（1993）。单整过程的连续弱收敛与随机等度连续性结果及其在回归模型估计中的应用。《计量经济学理论》，9(2):155–188。Saikkonen, P.（1995）。单整与协整系统中最大似然估计渐近理论的问题。《计量经济学理论》，11(5):888–911。Schwaab, B.、Lucas, A. 和 Zhang, X.（2021）。极端事件建模：时变极端尾部形状。https://doi.org/10.2866/252648。欧洲央行工作论文，第2524号，ISBN 978-92-899-4524-0，欧洲中央银行（ECB），法兰克福。

Smith, R. L. (1985年). 非正则情形中的一类最大似然估计。*Biometrika*, 72(1):67–90。Tibshirani, R. (1996). 通过 Lasso 进行回归收缩与变量选择。英国皇家统计学会会刊：*B*辑, 58(1):267–288。Tibshirani, R., Bien, J., Friedman, J., Hastie, T., Simon, N., Taylor, J., and Tibshirani, R. J. (2012). 套索型问题中剔除预测变量的强规则。英国皇家统计学会会刊：*B*辑, 74(2):245–266。Weller, B. M. (2017年). 算法交易会减少信息获取吗? 金融研究评论, 31(6):2184–2226。Zou, H. (2006). 自适应 LASSO 及其预言机性质。美国统计学会会刊, 101(476):1418–1429。

附录

A 证明

A.1 最大似然估计-对数似然函数 $\mathcal{L}(\cdot)$ 的梯度和海森矩阵

$\mathcal{L}(\boldsymbol{\beta};\{y_t\},\ \{\boldsymbol{z}_{t-1}\})$ 的梯度函数为

$$\frac{\partial \mathcal{L}(\boldsymbol{\beta};\{y_t\},\{\boldsymbol{z}_{t-1}\})}{\partial \boldsymbol{\beta}} = \sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}), \tag{40}$$

$$\boldsymbol{\psi}_t(\boldsymbol{\beta}) := \mathbb{1}\{y_t > 0\} \left(\boldsymbol{X}_{t-1} \begin{bmatrix} g_{k,t}(\boldsymbol{\beta}) \\ \vdots \\ g_{k,t}(\boldsymbol{\beta}) \\ g_{\sigma,t}(\boldsymbol{\beta}) \\ g_{\sigma,t}(\boldsymbol{\beta}) \\ \vdots \\ g_{\sigma,t}(\boldsymbol{\beta}) \\ g_{\sigma,t}(\boldsymbol{\beta}) \end{bmatrix} + g_{\sigma,t}(\boldsymbol{\beta}) \sum_{i=0}^{t-1} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \beta_{2,p+1}^{i} \\ \beta_{2,p+1}^{i} z_{1,t-1-i} \\ \vdots \\ \beta_{2,p+1}^{i} z_{p,t-1-i} \\ \sum_{j=1}^{p} \beta_{2,j}\, i\, \beta_{2,p+1}^{i-1}\, z_{j,t-1-i} \end{bmatrix} \right), \tag{41}$$

$\boldsymbol{X}_t :=\text{diag}(\boldsymbol{x}_t)$,
$\boldsymbol{x}_t := [1, \boldsymbol{z}_t', 1, \boldsymbol{z}_t', \log(\sigma_t(\boldsymbol{\beta}))]'$,

$$g_{k,t}(\boldsymbol{\beta}) := \left(\frac{1}{k_t(\boldsymbol{\beta})} - 2\right) \log\left(1 + k_t(\boldsymbol{\beta}) \frac{y_t}{\sigma_t(\boldsymbol{\beta})}\right) + \left(\frac{1}{k_t(\boldsymbol{\beta})} - 1 - 2k_t(\boldsymbol{\beta})\right) \left(\frac{1}{1 + k_t(\boldsymbol{\beta}) \frac{y_t}{\sigma_t(\boldsymbol{\beta})}} - 1\right),$$

$$g_{\sigma,t}(\boldsymbol{\beta}) := \frac{1}{k_t(\boldsymbol{\beta})} \left(1 - \frac{1}{1 + k_t(\boldsymbol{\beta}) \frac{y_t}{\sigma_t(\boldsymbol{\beta})}}\right) - \frac{1}{1 + k_t(\boldsymbol{\beta}) \frac{y_t}{\sigma_t(\boldsymbol{\beta})}},$$

$$g_{A,i,t} := \left[0, \ldots, 0, \beta_{2,p+1}^{i}, \beta_{2,p+1}^{i} z_{1,t-1-i}, \ldots, \beta_{2,p+1}^{i} z_{p,t-1-i}, \sum_{j=1}^{p} \beta_{2,j}\, i\, \beta_{2,p+1}^{i-1}\, z_{j,t-1-i}\right]', \tag{42}$$

其中，$\text{diag}(\boldsymbol{x}_t)$ 表示一个方形对角矩阵，其主对角线上的元素为向量 $\boldsymbol{x}_t$ 的各元素。利用式（42）中的梯度信息，我们给出 $\mathcal{L}(\boldsymbol{\beta};\{y_t\},\{\boldsymbol{z}_{t-1}\})$的海森矩阵如下。

$$H_{\mathcal{L}}(\boldsymbol{\beta}) := \frac{\partial^2 \mathcal{L}(\boldsymbol{\beta};\{y_t\},\{\boldsymbol{z}_{t-1}\})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \sum_{t=1}^{T} \frac{\partial \psi_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}$$

$$\frac{\partial \psi_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} = \mathbb{1}\{y_t > 0\} \left(\boldsymbol{X}_{t-1} H_g + \sum_{i=0}^{t-1} \text{diag}\left(g_{A,i,t}\right) \left(\boldsymbol{I}_{(2p+3)\times 1} \bigotimes \frac{\partial g_{\sigma,t}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} \right) + g_{\sigma,t}(\boldsymbol{\beta}) H_A \right),$$

$$H_g := \begin{bmatrix} \frac{\partial g_{k,t}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} \\ \vdots \\ \frac{\partial g_{k,t}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} \\ \frac{\partial g_{\sigma,t}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} \\ \vdots \\ \frac{\partial g_{\sigma,t}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} \end{bmatrix}, \qquad H_A := \sum_{i=0}^{t-1} \begin{bmatrix} 0 \ldots 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 \ldots 0 & \vdots & & \ddots & \vdots & \vdots \\ 0 \ldots 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 \ldots 0 & 0 & 0 & \ldots & 0 & i\, \beta_{2,p+1}^{i-1} \\ 0 \ldots 0 & 0 & 0 & \ldots & 0 & i\, \beta_{2,p+1}^{i-1} z_{1,t-1-i} \\ 0 \ldots 0 & \vdots & & \ddots & \vdots & \vdots \\ 0 \ldots 0 & 0 & 0 & \ldots & 0 & i\, \beta_{2,p+1}^{i-1} z_{p,t-1-i} \\ 0 \ldots 0\ i\, \beta_{2,p+1}^{i-1} & i\, \beta_{2,p+1}^{i-1} z_{1,t} & \ldots & i\, \beta_{2,p+1}^{i-1} z_{p,t} & \sum_{j=1}^{p} \beta_{2,j}\, i\, (i-1)\, \beta_{2,p+1}^{i-2} z_{j,t-i-i} \end{bmatrix} \tag{43}$$

$$\begin{aligned} \frac{\partial g_{k,t}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} &= \left[\frac{\partial g_{k,t}(\boldsymbol{\beta})}{\partial \text{logit}(k_t)} \cdots \frac{\partial g_{k,t}(\boldsymbol{\beta})}{\partial \text{logit}(k_t)} \ \frac{\partial g_{k,t}(\boldsymbol{\beta})}{\partial \log(\sigma_t)} \cdots \frac{\partial g_{k,t}(\boldsymbol{\beta})}{\partial \log(\sigma_t)} \right] \text{diag}(\boldsymbol{X}_{t-1}) \\ \frac{\partial g_{\sigma,t}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} &= \left[\frac{\partial g_{\sigma,t}(\boldsymbol{\beta})}{\partial \text{logit}(k_t)} \cdots \frac{\partial g_{\sigma,t}(\boldsymbol{\beta})}{\partial \text{logit}(k_t)} \ \frac{\partial g_{\sigma,t}(\boldsymbol{\beta})}{\partial \log(\sigma_t)} \cdots \frac{\partial g_{\sigma,t}(\boldsymbol{\beta})}{\partial \log(\sigma_t)} \right] \text{diag}(\boldsymbol{X}_{t-1}) \end{aligned} \tag{44}$$

$$
\begin{aligned}
\frac{\partial g_{k,t}(\boldsymbol{\beta})}{\partial \text{logit}(k_t)} &= (2-\frac{1}{k_t})\log(1+k_t\frac{y_t}{\sigma_t}) + (1+k_t\frac{y_t}{\sigma_t})^{-1}((1-2\,k_t)^2 - (\frac{1}{k_t}-2\,k_t)(1-2\,k_t)) \\
&\quad + (1+k_t\frac{y_t}{\sigma_t})^{-2}\frac{y_t}{\sigma_t}(2k_t^2+k_t-1) \\
\frac{\partial g_{k,t}(\boldsymbol{\beta})}{\partial \log(\sigma_t)} &= k_t\frac{y_t}{\sigma_t}((1-\frac{1}{k_t})(1+k_t\frac{y_t}{\sigma_t})^{-1} + (\frac{1}{k_t}-1-2\,k_t)(1+k_t\frac{y_t}{\sigma_t})^{-2}) \\
\frac{\partial g_{\sigma,t}(\boldsymbol{\beta})}{\partial \text{logit}(k_t)} &= (1+k_t\frac{y_t}{\sigma_t})^{-2}\frac{y_t}{\sigma_t}(1+k_t)(1-2\,k_t) + ((1+k_t\frac{y_t}{\sigma_t})^{-1}-1)(1-2\,k_t)\frac{1}{k_t} \\
\frac{\partial g_{\sigma,t}(\boldsymbol{\beta})}{\partial \log(\sigma_t)} &= -\frac{y_t}{\sigma_t}(1+k_t\frac{y_t}{\sigma_t})^{-2}(1+k_t),
\end{aligned}
\tag{45}
$$

其中，$\bigotimes$ 表示克罗内克积算子。为简化记号，我们省略 $k_t := k_t(\boldsymbol{\beta})$、$\sigma_t := \sigma_t(\boldsymbol{\beta})$中的系数，以及 $\psi_t(\boldsymbol{\beta}) := \psi(\boldsymbol{\beta};\, L_{t-1}, \boldsymbol{Z}_{t-1}, L_{t-2}, \boldsymbol{Z}_{t-2}, ..)$中的条件变量。我们还特别记 $k_t^o := k_t(\boldsymbol{\beta^o})$和 $\sigma_t^o := \sigma_t(\boldsymbol{\beta^o})$，以简化记号。

A.2 最大似然估计—证明

命题1。 在假设 $M.1$, $M.2$, $M.3$, $M.5$*(1)* 和 $M.6$ 下，对于任意 $\varepsilon > 0$ 以及任意位于 Θ 内部的 $\boldsymbol{\beta}$，存在一个 $\delta > 0$，使得

$$
\sup_{\|\boldsymbol{\beta}-\widetilde{\boldsymbol{\beta}}\|<\delta} \left\| \boldsymbol{\psi}_t(\boldsymbol{\beta}) - \boldsymbol{\psi}_t(\widetilde{\boldsymbol{\beta}}) \right\| < \varepsilon, \tag{46}
$$

其中 $\|\cdot\|$ 是欧几里得范数。

证明。 该命题断言，$\boldsymbol{\psi}_t(\cdot)$ 关于 $\boldsymbol{\beta} \in \Theta$ 一致连续。为证明该命题，我们将在下文证明 $\boldsymbol{\psi}_t(\cdot)$ 在 $\boldsymbol{\beta} \in \Theta$ 中连续且紧。

首先，查看附录 A.1 中 $\boldsymbol{\psi}_t(\cdot)$ 的公式可知，$\boldsymbol{\psi}_t(\cdot)$ 是关于 $\boldsymbol{\beta}$ 的连续函数的复合，因此关于 $\boldsymbol{\beta}$ 连续。其次，根据假设 $M.1$, $M.3$, $M.5$(1) 和 $M.6$，可知 $\boldsymbol{\beta}$ 和 $\boldsymbol{z}_{t-1}$ 依概率有界，因此 $k_t(\cdot)$ 和 $\sigma_t(\cdot)$ 有界，且其下界依概率大于零。由此可得，$g_{\sigma,t}(\cdot)$ 和 $g_{A,i,t}(\cdot)$ 也依概率有界。此外，根据假设 $M.2$，可知 $|y_t|$ 依概率有界。由詹森不等式可得，$g_{k,t}(\cdot)$ 依概率有界。因此，我们得到 $\boldsymbol{\psi}_t(\cdot)$ 在 $\boldsymbol{\beta} \in \Theta$ 中依概率有界，从而完成证明。

□

命题2。 在假设 $M.1$, $M.2$, $M.3$, $M.5$*(1)* 和 $M.6$ 下，$\{\mathbb{E}[\boldsymbol{\psi}_t(\boldsymbol{\beta})], \boldsymbol{\beta} \in \Theta\}$ 在 $\boldsymbol{\beta} = \boldsymbol{\beta}^{o*}$ 处有唯一零点，且对任意 $\epsilon > 0$,

$$
\lim_{T\to\infty} P\left\{ \left\| \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*}) \right\| > \epsilon \right\} = 0, \tag{47}
$$

证明。 首先，等式（47）表明 $\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*})$ 依概率收敛于零。我们利用 Cs¨org¨o（1968）定理1中的条件证明等式（47）。

在假设 $M.1$ 和 $M.2$ 下，我们对 y_t 的真实条件概率函数取期望，得到

$$
\mathbb{E}\left[\boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*})|\mathcal{F}_{t-1}\right] = \mathbf{0}, \qquad \text{for } t = 1, \ldots, T. \tag{48}
$$

我们还有

$$
\begin{aligned}
\sum_{t=1}^{T} \frac{1}{t^2}\mathbb{E}\left[\boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*})\right] &\leq \sum_{t=1}^{T} \frac{1}{t^2} M\, \mathbf{1}_{2p+3} \\
&< \infty,
\end{aligned}
\tag{49}
$$

其中， $M \in \mathbb{R}$ 是一个很大的有限数， $\mathbf{1}_{2p+3}$ 表示由（$2p+3$）个 1 组成的向量；倒数第二个不等式由命题1中 $\boldsymbol{\psi}_t(\cdot)$ 的紧性得到；最后一个不等式由 $\sum_{t=1}^{T} \frac{1}{t^2} < \infty$ 得到。因此，应用 Cs¨org¨o（1968）的定理 1，我们得出 $\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*})$ 依概率收敛于零。其次，我们将通过反证法证明，在 Θ 中使得 $\mathbb{E}[\boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*})] = 0$ 的 $\boldsymbol{\beta}^{o*}$ 是唯一的。假设存在一个 $\boldsymbol{\beta} \in \Theta$ 和 $\boldsymbol{\beta} \neq \boldsymbol{\beta}^{o*}$，使得

$$\mathbb{E}[\boldsymbol{\psi}_t(\boldsymbol{\beta})] = 0. \tag{50}$$

另一方面，根据中值定理，有

$$\boldsymbol{\psi}_t(\boldsymbol{\beta}) - \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*}) = H_{\mathcal{L}}(\overline{\boldsymbol{\beta}})\,(\boldsymbol{\beta} - \boldsymbol{\beta}^{o*})\,, \tag{51}$$

其中 $\overline{\boldsymbol{\beta}}$ 介于 $\boldsymbol{\beta}$ 和 $\boldsymbol{\beta}^{o*}$之间。由于根据假设 $M.6$ 和$\boldsymbol{\beta} \neq \boldsymbol{\beta}^{o*}$， $H_{\mathcal{L}}(\cdot)$ 在 Θ 上几乎必然正定，因此我们得到 $\boldsymbol{\psi}_t(\boldsymbol{\beta}) \neq \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*})$ 几乎必然成立，这与（50）矛盾。因此，我们得出 $\boldsymbol{\beta}^{o*}$ 在 Θ中的唯一性。

□

定理1的证明

证明。 首先，我们证明 $\widehat{\boldsymbol{\beta}}^{\text{mle}}$ 属于 Θ。根据命题2，我们可以得到，对于任意 $\boldsymbol{\epsilon} \in \mathbb{R}^{2p+3}$ 和 $\boldsymbol{\epsilon} > \mathbf{0}$，存在 T_N，使得对于 $T > T_N$，有

$$\left| \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*}) \right| < \boldsymbol{\epsilon}. \tag{52}$$

在假设 $M.4$ 和 $M.6$下，我们知道 $\boldsymbol{\beta}^{o*}$ 是 Θ中的一个内点。由于 $\boldsymbol{\beta}^{o*}$ 是Θ 中的一个内点，并且根据命题1，$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\psi}_t(\cdot)$ 在 Θ 上一致连续，因此存在 $\delta > 0$ 和一个球 $B(\boldsymbol{\beta}^{o*},\ \delta) := \{\boldsymbol{\beta} \in \Theta \mid \|\boldsymbol{\beta}^{o*} - \boldsymbol{\beta}\| < \delta\}$，使得

$$\left| \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*}) - \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}) \right| < \frac{1}{2}\boldsymbol{\epsilon}. \tag{53}$$

此外，存在 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in B(\boldsymbol{\beta}^{o*}, \delta)$，使得如果 $\left|\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*})\right| \neq \mathbf{0}$，那么

$$\mathbf{0} < \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*}) - \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}_1) \leq \frac{1}{2} \left| \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*}) \right| \tag{54}$$

$$-\frac{1}{2} \left| \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*}) \right| \leq \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*}) - \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}_2) < \mathbf{0} \tag{55}$$

由此可得

$$\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*}) - \frac{1}{2} \left| \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*}) \right| \leq \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}_1) < \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*}) \tag{56}$$

$$\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*}) < \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}_2) \leq \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*}) + \frac{1}{2} \left| \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*}) \right|. \tag{57}$$

因此，我们得到 $-\left|\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}_1)\right| < -\left|\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*})\right| < \boldsymbol{\epsilon}$ 或 $\left|\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}_2)\right| < \left|\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*})\right| < \boldsymbol{\epsilon}$。继续这一过程，我们可以找到一列属于 Θ 的点，其 $\left|\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\psi}_t(\cdot)\right|$的值逐渐减小。由此得到的点列存在一个子列，且该子列的极限为 $\widehat{\boldsymbol{\beta}}^{\text{mle}}$，属于 Θ，并满足$\left|\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\psi}_t(\widehat{\boldsymbol{\beta}}^{\text{mle}})\right| = 0$。由于 $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*}) \equiv \mathbf{0}$，因此我们得到 $\lim_{T\to\infty} \widehat{\boldsymbol{\beta}}^{\text{mle}} = \boldsymbol{\beta}^{o*}$，从而完成证明。

□

定理 2 的证明

证明。 对 $\sum_{t=1}^{T} \boldsymbol{\psi}_t(\widehat{\boldsymbol{\beta}}^{\text{mle}})$ 在 $\boldsymbol{\beta}^{o*}$ 处作泰勒展开，并应用中值定理，可得

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{\psi}_t(\widehat{\boldsymbol{\beta}}^{\text{mle}}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*}) + \frac{1}{T} H_{\mathcal{L}}(\overline{\boldsymbol{\beta}}) \sqrt{T} \left(\widehat{\boldsymbol{\beta}}^{\text{mle}} - \boldsymbol{\beta}^{o*} \right), \tag{58}$$

其中 $\boldsymbol{\beta}$介于 $\widehat{\boldsymbol{\beta}}^{\text{mle}}$ 和 $\boldsymbol{\beta}^{o*}$之间。由于 $\sum_{t=1}^{T} \boldsymbol{\psi}_t(\widehat{\boldsymbol{\beta}}^{\text{mle}}) = 0$，上述展开得到 i n

$$\left(\frac{1}{T} H_{\mathcal{L}}(\overline{\boldsymbol{\beta}}) \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*}) = \sqrt{T} \left(\widehat{\boldsymbol{\beta}}^{\text{mle}} - \boldsymbol{\beta}^{o*} \right). \tag{59}$$

– $\frac{1}{T} H_{\mathcal{L}}(\cdot)$ 关于 Θ一致连续，这一点可通过类似命题1证明的方法证明，只需借助 $0 < k_t(\boldsymbol{\beta}^{o*}) < 0.5$，并知晓 $y_t^2 \in O_p(1)$。根据连续映射定理，并结合定理 1 中得到的 $\lim_{T\to\infty} \widehat{\boldsymbol{\beta}}^{\text{mle}} = \boldsymbol{\beta}^{o*}$，我们可得

$$\lim_{T\to\infty} \left(\frac{1}{T} H_{\mathcal{L}}(\overline{\boldsymbol{\beta}}) \right)^{-1} = \left(\frac{1}{T} H_{\mathcal{L}}(\boldsymbol{\beta}^{o*}) \right)^{-1}.$$

由假设 $M.7$ 和 $M.8$ 可知

$$\begin{cases} \dfrac{1}{\sqrt{T}} \displaystyle\sum_{t=1}^{T} \boldsymbol{\psi}_t(\boldsymbol{\beta}^{o*}) \overset{\mathcal{D}}{\sim} S_{\psi} \\ \dfrac{1}{T} H_{\mathcal{L}}(\boldsymbol{\beta}^{o*}) \overset{\mathcal{D}}{\sim} \Omega_H \end{cases} \tag{60}$$

因此，根据斯卢茨基定理，我们可得

$$\sqrt{T} \left(\boldsymbol{\beta}^{o*} - \widehat{\boldsymbol{\beta}}^{\text{mle}} \right) \overset{\mathcal{D}}{\sim} \Omega_H^{-1} \, S_{\psi}. \tag{61}$$

□

A.3 ALMLE- 证明

定理 3 的证明

证明。 我们用反证法证明该定理。在下面的证明中，我们将说明，如果不存在满足条件（22）的 $\lambda_{k,T}, \lambda_{\sigma,T} \in O(T^{\frac{1}{2}})$，那么真正的非活跃预测变量仍有非零概率被选中，因为 $\lambda_{k,T}, \lambda_{\sigma,T} \in O(T^{\frac{1}{2}})$，且不存在 $\widehat{\boldsymbol{\beta}}^{\text{al}}(\lambda_{k,T}, \lambda_{\sigma,T})$。

–

我们令 $w_{k,i}, w_{\sigma,j}$，例如使用第3节中的最大似然估计，使得 $\sqrt{T}(\frac{1}{w_{k,i}} - \beta_1^{*o}{}_{,i}) = O_p(1)$ 且 $\sqrt{T}(\frac{1}{w_{\sigma,j}} - \beta_2^{*o}{}_{,j}) = O_p(1)$，其中 $i = 1, \ldots, p$, $j = 1, \ldots, p+1$。

我们来看调优参数网格 $\{(\lambda_{k,T}, \lambda_{\sigma,T}) : \lambda_{k,T} \in S_{\lambda_{k,T}}, \lambda_{\sigma,T} \in S_{\lambda_{\sigma,T}}, \}$。$\lambda_{k,T,\max}$和 $\lambda_{\sigma,T,\max}$取值足够大，使得 $\widehat{\boldsymbol{\beta}}^{\text{al}}(\lambda_{k,T,\max}, \lambda_{\sigma,T,\max}) = \mathbf{0}$，即没有预测变量被选中。根据 Karush-Kuhn-Tucker（KKT）最优性条件，并结合 $(\lambda_{k,T,\max}, \lambda_{\sigma,T,\max})$，我们有

$$\begin{cases} \left| \dfrac{\partial \mathcal{L}}{\partial \widehat{\beta}_{1,i}^{\text{al}}} \right| \leq \lambda_{k,T,\max} \, w_{k,i}, \quad (1,i) \in \mathcal{A}_k \cup \mathcal{A}_k^c, \\ \left| \dfrac{\partial \mathcal{L}}{\partial \widehat{\beta}_{2,j}^{\text{al}}} \right| \leq \lambda_{\sigma,T,\max} \, w_{\sigma,j}, \quad (2,j) \in \mathcal{A}_{\sigma} \cup \mathcal{A}_{\sigma}^c, \end{cases} \tag{62}$$

其中，我们记 $\mathcal{A}_k := \{(1,i) : i \geq 1, \beta_1^{*o}{}_{,i} \neq 0\}$, $\mathcal{A}_k^c := \{(1,i) : i \geq 1, \beta_1^{*o}{}_{,i} = 0\}$, $\mathcal{A}_{\sigma} := \{(2,j) : j \geq 1, \beta_2^{*o}{}_{,j} \neq 0\}$, and $\mathcal{A}_{\sigma}^c := \{(2,j) : j \geq 1, \beta_2^{*o}{}_{,j} = 0\}$。根据中值定理，~~我们将 $\frac{\partial}{\partial \widehat{\boldsymbol{\beta}}^{\text{al}}} \lambda_{k,T,\max}(\,,\,,\lambda_{\sigma,T,\max})$~~ 改写为

$\mathcal{L}(;\{y_t\},\ \{\boldsymbol{z}_t^*\})$ 二阶连续可微，并将这一改写代入上述 KKT 条件，得到

$$\begin{cases} \left|\dfrac{\partial\mathcal{L}}{\partial\beta_{1,i}^{*o}}-\sum\limits_{i_g=0}^{p}\dfrac{\partial^2\mathcal{L}}{\partial\widehat{\beta}_{1,i}^{\mathrm{al}}\partial\widehat{\boldsymbol{\beta}}_{1,i_g}^{\mathrm{al}\prime}}\overline{\boldsymbol{\beta}}_{1,i_g}^{(2\,n_{\lambda_k}\,n_{\lambda_\sigma})}-\sum\limits_{j_g=0}^{p+1}\dfrac{\partial^2\mathcal{L}}{\partial\widehat{\beta}_{1,i}^{\mathrm{al}}\partial\widehat{\boldsymbol{\beta}}_{2,j_g}^{\mathrm{al}\prime}}\overline{\boldsymbol{\beta}}_{2,j_g}^{(2\,n_{\lambda_k}\,n_{\lambda_\sigma})}\right|\le\lambda_{k,T,\max}\,w_{k,i}, \\ \qquad\qquad (1,i)\in\mathcal{A}_k\cup\mathcal{A}_k^c, \\ \left|\dfrac{\partial\mathcal{L}}{\partial\beta_{2,j}^{*o}}-\sum\limits_{i_g=0}^{p}\dfrac{\partial^2\mathcal{L}}{\partial\widehat{\beta}_{2,j}^{\mathrm{al}}\partial\widehat{\boldsymbol{\beta}}_{1,i_g}^{\mathrm{al}\prime}}\overline{\boldsymbol{\beta}}_{1,i_g}^{(2\,n_{\lambda_k}\,n_{\lambda_\sigma}-1)}-\sum\limits_{j_g=0}^{p+1}\dfrac{\partial^2\mathcal{L}}{\partial\widehat{\beta}_{2,j}^{\mathrm{al}}\partial\widehat{\boldsymbol{\beta}}_{2,j_g}^{\mathrm{al}\prime}}\overline{\boldsymbol{\beta}}_{2,j_g}^{(2\,n_{\lambda_k}\,n_{\lambda_\sigma}-1)}\right|\le\lambda_{\sigma,T,\max}\,w_{\sigma,j}, \\ \qquad\qquad (2,j)\in\mathcal{A}_\sigma\cup\mathcal{A}_\sigma^c, \end{cases} \tag{63}$$

其中 $\underline{\boldsymbol{\beta}_1^{(2\,n_{\lambda_k}\,n_{\lambda_\sigma})}}$，$\overline{\boldsymbol{\beta}_1^{(2\,n_{\lambda_k}\,n_{\lambda_\sigma}-1)}}$，$i_g\in\left[\boldsymbol{\beta}_{1,i_g}^{*o},\widehat{\boldsymbol{\beta}}_{1,i_g}^{\mathrm{al}}(\lambda_{k,T,\max},\lambda_{\sigma,T,\max})\right]$，且 $\underline{\boldsymbol{\beta}_1^{(2\,n_{\lambda_k}\,n_{\lambda_\sigma})}}$，$\overline{\boldsymbol{\beta}_1^{(2\,n_{\lambda_k}\,n_{\lambda_\sigma}-1)}}$，$j_g\in[\boldsymbol{\beta}_{2,j_g}^{*o},\widehat{\boldsymbol{\beta}}_{2,j_g}^{\mathrm{al}}(\lambda_{k,T,\max},\lambda_{\sigma,T,\max})]$。事实上，我们可以在 $O(T)$ 中找到一组 $(\lambda_{k,T,\max},\lambda_{\sigma,T,\max})$。

降低 $(\lambda_{k,T,\max},\lambda_{\sigma,T,\max})$ 中的调优参数后，$\widehat{\boldsymbol{\beta}}^{\mathrm{al}}$ 能够筛选变量。真正不活跃的预测变量始终满足 KKT 条件的不等式，其中 $\frac{\lambda_k}{\sqrt{T}},\frac{\lambda_\sigma}{\sqrt{T}}\to\infty$ 以及 $\lambda_{k,T},\lambda_{\sigma,T}\in O(T^{\frac{1}{2}})$，可得，对于 $(1,i)\in\mathcal{A}_k$，$\lambda_{k,T}\,w_{k,i}=O_p(T^{\frac{1}{2}})$；对于 $(1,i)\in\mathcal{A}_k^c$，$\lambda_{k,T}\,w_{k,i}=O_p(T)$；对于 $(2,j)\in\mathcal{A}_\sigma$，$\lambda_{\sigma,T}\,w_{\sigma,j}=O_p(T^{\frac{1}{2}})$；对于 $(2,i)\in\mathcal{A}_\sigma^c$，$\lambda_{\sigma,T}\,w_{\sigma,i}=O_p(T)$。如果对于任意 $\lambda_{k,T}\in O(T^{\frac{1}{2}})$ 和 $\lambda_{\sigma,T}\in O(T^{\frac{1}{2}})$，都不存在满足条件 (22) 的 $\widehat{\boldsymbol{\beta}}^{\mathrm{al}}(\lambda_{k,T},\lambda_{\sigma,T})$，那么存在 $(1,i_a)\in\mathcal{A}_k$或 $(2,j_a)\in\mathcal{A}_\sigma$，使得

$$\left|\frac{\partial\mathcal{L}}{\partial\widehat{\beta}_{1,i_a}^{\mathrm{al}}}\right|\le\lambda_{k,T}\,w_{k,i_a} \tag{64}$$

or

$$\left|\frac{\partial\mathcal{L}}{\partial\widehat{\beta}_{2,j_a}^{\mathrm{al}}}\right|\le\lambda_{\sigma,T}\,w_{\sigma,j_a} \tag{65}$$

但对于任意 $\widehat{\beta}_{1,\,i_a}^{\mathrm{al}}=0$ 或 $\widehat{\beta}_{2,\,j_a}^{\mathrm{al}}=0$，相应地，对于任意 $\lambda_{k,\,T},\ \lambda_{\sigma,\,T}\in O(T^{\frac{1}{2}})$。

在条件 (22) 不成立时，若 $\left|\frac{\partial\mathcal{L}}{\partial\widehat{\beta}_{1,\,i_a}^{\mathrm{al}}}\right|\le\lambda_{k,\,T}\,w_{k,\,i_a}$，且 $\widehat{\beta}_{1,\,i_a}^{\mathrm{al}}=0$ 以及任意 $\lambda_{k,\,T},\ \lambda_{\sigma,\,T}\in O(T^{\frac{1}{2}})$，则对于 $(1,\ i)\in\mathcal{A}_k$，有 $\lambda_{k,\,T}\,w_{k,\,i}=O_p(T^{\frac{1}{2}})$；对于 $(1,\ i)\in\mathcal{A}_k^c$，有 $\lambda_{k,\,T}\,w_{k,\,i}=O_p(T)$。于是，形状模型中每个真正非活跃、但与第 i_a 个预测变量相关的预测变量，都有非零概率因相关性而被选中

$$P\left\{\left|\frac{\partial\mathcal{L}}{\partial\widehat{\beta}_{1,i}^{\mathrm{al}}}\right|\ge\lambda_{k,T}\,w_{k,i},i\in\mathcal{A}_k^c\middle|\widehat{\beta}_{1,i_a}^{\mathrm{al}}=0,\ \lambda_{k,T}\in O(T^{\frac{1}{2}}),\ \lambda_{\sigma,T}\in O(T^{\frac{1}{2}})\right\}\neq 0. \tag{66}$$

通过像（63）中那样应用中值定理重写 $\left|\frac{\partial\mathcal{L}}{\partial\widehat{\beta}_1^{\mathrm{al}},i}\right|$，即可轻松得到上述概率，并得出 $\left|\frac{\partial\mathcal{L}}{\partial\widehat{\beta}_1^{\mathrm{al}},i}\right|\in O_p(T)$。对于每个满足 $\lambda_{k,T}\in O(T^{\frac{1}{2}})$ 的 $\widehat{\beta}_{1,i}^{\mathrm{al}}\neq 0,i\in\mathcal{A}_k^c$，$\widehat{\beta}_{1,i}^{\mathrm{al}}$的偏差为 $O_p(1)$。降低$\lambda_{k,T}$后，可知 $\widehat{\beta}_{1,i}^{\mathrm{al}}=O_p(T^{-\gamma})$ 满足与 $\lambda_{k,\,T}\in O(T^{\frac{1}{2}}-\gamma)$ 相同的 KKT 条件等式，并且在 $\gamma\in[0,\frac{1}{2}]$ 中仍被选中。

同样的推理也适用于 $\left|\frac{\partial\mathcal{L}}{\partial\widehat{\beta}_2^{\mathrm{al}},j_a}\right|\le\lambda_{\sigma,T}\,w_{\sigma,j_a}$，其中 $\widehat{\beta}_{1,j_a}^{\mathrm{al}}=0$，且对于任意 $\lambda_{k,T},\lambda_{\sigma,T}\in O(T^{\frac{1}{2}})$，我们可以得出，任何实际上不活跃但与尺度模型第 j_a 个预测变量相关的变量，都有非零概率被选中，即:

$$P\left\{\left|\frac{\partial\mathcal{L}}{\partial\widehat{\beta}_{2,j}^{\mathrm{al}}}\right|\ge\lambda_{\sigma,T}\,w_{\sigma,j},j\in\mathcal{A}_\sigma^c\middle|\widehat{\beta}_{2,j_a}^{\mathrm{al}}=0\ \lambda_{k,T}\in O(T^{\frac{1}{2}}),\ \lambda_{\sigma,T}\in O(T^{\frac{1}{2}})\right\}\neq 0. \tag{67}$$

因此，如果条件（22）不满足，就无法实现模型选择一致性，证明完毕。

□

定理 4 的证明

证明。在此证明中，我们首先证明形状模型的真实活跃预测变量上的 $\widehat{\boldsymbol{\beta}}^{k,\,al}$ 的选择一致性，即

$$\lim_{T\to\infty} P\left\{\widehat{\beta}_{1,i}^{k,al}\neq 0, \forall(1,i)\in\mathcal{A}_k\right\}=1 \tag{68}$$

这也等价于 $\lim_{T\to\infty} P\left\{\mathcal{A}^{k,\,al}_{k,\,T}\supseteq\mathcal{A}_k\right\}=1$，其中 $\mathcal{A}^{k,\,al}_{k,\,T}:=\{(1,\ i):\ \widehat{\boldsymbol{\beta}}^{k,\,al}_{1,\,i}\neq 0\}$。由此可知，由于 $\widehat{\boldsymbol{\beta}}^{k,\,al}$ 导致的形状模型过度选择可能性已被证明可由 $\widehat{\boldsymbol{\beta}}^{tal}$抑制至零，最终获得了 $\widehat{\boldsymbol{\beta}}^{tal}$ 的预言机性质。

首先，我们重写目标函数，得到如下的 $\widehat{\boldsymbol{\beta}}^{k,\,al}$：根据假设 $L2$，存在$\boldsymbol{\beta}^{k,\,o}$，且

$$V^{(k)}(\boldsymbol{\nu}^{(k)})=-\mathcal{L}\left(\boldsymbol{\beta}^{k,o}+\frac{\boldsymbol{\nu}^{(k)}}{\sqrt{T}};\{y_t\},\{\boldsymbol{z}_t^*\}\right)+\lambda_{k,T}\sum_{i=1}^{p}\widetilde{w}_{k,i}\left|\beta_{1,i}^{k,o}+\frac{\nu_{1,i}^{(k)}}{\sqrt{T}}\right|, \tag{69}$$

其中 $\boldsymbol{\nu}^{(k)}:=[\nu_{1,0}^{(k)},\ldots,\nu_{1,p}^{(k)},\nu_{2,0}^{(k)},0,\ldots,0]'\in\mathbb{R}^{p+2}\times\mathbf{0}_{(p+1)}$。我们得到 $\widehat{\boldsymbol{\nu}}^{(k)}=\arg\min_{\boldsymbol{\nu}^{(k)}\in\mathbb{R}^{p+2}\times\mathbf{0}_{(p+1)}} V_{(k)}(\boldsymbol{\nu}^{(k)})-V^{(k)}(\mathbf{0})$，使得二者等价

$$\widehat{\boldsymbol{\beta}}^{k,al}=\boldsymbol{\beta}^{k,o}+\frac{\widehat{\boldsymbol{\nu}}^{(k)}}{\sqrt{T}}. \tag{70}$$

具体而言，

$$\begin{aligned}
&V^{(k)}(\boldsymbol{\nu}^{(k)})-V^{(k)}(\mathbf{0})\\
&=-\left(\mathcal{L}(\boldsymbol{\beta}^{k,o}+\frac{\boldsymbol{\nu}^{(k)}}{\sqrt{T}};\{y_t\},\{\boldsymbol{z}_t^*\})-\mathcal{L}\left(\boldsymbol{\beta}^{k,o};\{y_t\},\{\boldsymbol{z}_t^*\}\right)\right)\\
&+\frac{\lambda_{k,T}}{\sqrt{T}}\sum_{i=1}^{p}\widetilde{w}_{k,i}\sqrt{T}\left(\left|\beta_{1,i}^{k,o}+\frac{\nu_{1,i}^{(k)}}{\sqrt{T}}\right|-\left|\beta_{1,i}^{k,o}\right|\right).
\end{aligned} \tag{71}$$

根据假设 $L2$和定理 2，我们有 $\sqrt{T}(\widehat{\beta}^{k,\,mle}_{1,\,i}-\beta^{k,\,o}_{1,\,i})=O_p(1)$ 以及 $\sqrt{T}(\widehat{\beta}^{mle}_{1,\,i}-\beta^{*o}_{1,i})=O_p(1)$。考虑到

$$\lim_{T\to\infty}\sqrt{T}\left(\left|\beta_{1,i}^{k,o}+\frac{\nu_{1,i}^{(k)}}{\sqrt{T}}\right|-\left|\beta_{1,i}^{k,o}\right|\right)=\begin{cases}\nu_{1,i}^{(k)}\ \mathrm{sgn}(\beta_{1,i}^{k,o}), & \text{if}\,\beta_{1,i}^{k,o}\neq 0,\\ \left|\nu_{1,i}^{(k)}\right|, & \text{if}\,\beta_{1,i}^{k,o}=0,\end{cases} \tag{72}$$

并且在假设 $L1$ 下，我们得到

$$\frac{\lambda_{k,T}}{\sqrt{T}}\widetilde{w}_{k,i}\sqrt{T}\left(\left|\beta_{1,i}^{k,o}+\frac{\nu_{1,i}^{(k)}}{\sqrt{T}}\right|-\left|\beta_{1,i}^{k,o}\right|\right)=\begin{cases}O_p(T^{-\gamma_1})\,\nu_{1,i}^{(k)}\ \mathrm{sgn}(\beta_{1,i}^{k,o}), & \text{if}\,\beta_{1,i}^{*o}\neq 0,\\ O_p(T^{\frac{1}{2}-\gamma_1})\,\nu_{1,i}^{(k)}\ \mathrm{sgn}(\beta_{1,i}^{k,o}), & \text{if}\,\beta_{1,i}^{*o}=0\ \text{but}\ \beta_{1,i}^{k,o}\neq 0,\\ O_p(T^{1-\gamma_1})\ \left|\nu_{1,i}^{(k)}\right|, & \text{if}\,\beta_{1,i}^{*o}=0\ \text{and}\ \beta_{1,i}^{k,o}=0.\end{cases} \tag{73}$$

此外，我们知道

$$-\left(\mathcal{L}(\boldsymbol{\beta}^{k,o}+\frac{\boldsymbol{\nu}^{(k)}}{\sqrt{T}};\{y_t\},\{\boldsymbol{z}_t^*\})-\mathcal{L}\left(\boldsymbol{\beta}^{k,o};\{y_t\},\{\boldsymbol{z}_t^*\}\right)\right)=O_p(1)\ \max\{(\nu_{1,i}^{(k)})^2\} \tag{74}$$

根据泰勒展开以及定理 2 的证明中所示的 Θ 中 $\frac{1}{T}H_{\mathcal{L}}(\cdot)$ 的紧性。将式 (73) 代回式 (71)，并根据斯卢茨基定理，我们得到

$$V^{(k)}(\nu_{1,i}^{(k)})-V^{(k)}(\mathbf{0})=\begin{cases}O_p(T), & \text{if}\quad \frac{\nu_{1,i}^{(k)}}{\sqrt{T}}=O_p(1),\forall(1,i)\in\mathcal{A}_k^{k,al}\,\text{and}\,\nu_{1,i}^{(k)}=O_p(1),\forall(1,i)\notin\mathcal{A}_k^{k,al};\\ O_p(T^{1-2\gamma}), & \text{if}\quad \nu_{1,i}^{(k)}=O_p(T^{\frac{1}{2}-\gamma}),\forall(1,i)\in\mathcal{A}_k^{k,al}\,\text{with}\,0<2\gamma<\gamma_1\,\text{and}\,\nu_{1,i}^{(k)}=O_p(1),\forall(1,i)\notin\mathcal{A}_k^{k,al};\\ O_p(T^{1-\gamma_1}), & \text{if}\quad \nu_{1,i}^{(k)}=O_p(T^{\frac{1}{2}-\gamma}),\forall(1,i)\in\mathcal{A}_k^{k,al}\,\text{with}\,\frac{\gamma_1}{2}\leq\gamma\leq\frac{1}{2}\,\text{and}\,\nu_{1,i}^{(k)}=O_p(1),\forall(1,i)\notin\mathcal{A}_k^{k,al};\\ O_p(T^{1-2\gamma}), & \text{if}\quad \nu_{1,i}^{(k)}=O_p(T^{\frac{1}{2}-\gamma}),\forall(1,i)\in\mathcal{A}_k^{k,al}\,\text{with}\,0<\gamma<\gamma_1\,\text{and}\,\nu_{1,i}^{(k)}=0,\forall(1,i)\notin\mathcal{A}_k^{k,al};\\ O_p(T^{1-\gamma-\gamma_1}), & \text{if}\quad \nu_{1,i}^{(k)}=O_p(T^{\frac{1}{2}-\gamma}),\forall(1,i)\in\mathcal{A}_k^{k,al}\,\text{with}\,\gamma_1\leq\gamma\leq\frac{1}{2}\,\text{and}\,\nu_{1,i}^{(k)}=0,\forall(1,i)\notin\mathcal{A}_k^{k,al},\end{cases} \tag{75}$$

其中 $\mathcal{A}^{k,al}_k := \{(1,i): \beta^{k,o}_{1,i} \neq 0\}$。因此，我们有

$$\widehat{\nu}^{(k)}_{1,i} = O_p(1), \forall (1,i) \in \mathcal{A}^{k,al}_k \text{ and } \widehat{\nu}^{(k)}_{1,i} = 0, \forall (1,i) \notin \mathcal{A}^{k,al}_k, \tag{76}$$

最小化 $V_{(k)}(\boldsymbol{\nu}^{(k)}) - V_{(k)}(\mathbf{0})$，因此

$$\lim_{T\to\infty} P\left\{\widehat{\beta}^{k,al}_{1,i} = \beta^{k,o}_{1,i} + \frac{\widehat{\nu}^{(k)}_{1,i}}{\sqrt{T}} \neq 0, \forall (1,i) \in \mathcal{A}^{k,al}_k\right\} = P\left\{\beta^{k,o}_{1,i} \neq 0, \forall (1,i) \in \mathcal{A}^{k,al}_k\right\} = 1. \tag{77}$$

因此我们得到 $\lim_{T\to\infty} P\left\{\widehat{\beta}^{k,al}_{1,i} \neq 0, \forall (1,i) \in \mathcal{A}_k\right\} = 1$。

其次，我们展示 $\widehat{\boldsymbol{\beta}}^{tal}$ 的渐近行为。记 $\boldsymbol{\beta} = \boldsymbol{\beta}^{o*} + \frac{\boldsymbol{\nu}}{\sqrt{T}}$ 为满足 $\boldsymbol{\nu} := [\nu_{1,0}, \ldots, \nu_{1,p}, \nu_{2,0}, \ldots, \nu_{2,p+1}]' \in \mathbb{R}^{2p+3}$ 的 $\boldsymbol{\beta} \in \Theta$，并重写目标函数，得到如下的 $\widehat{\boldsymbol{\beta}}^{tal}$：

$$V(\boldsymbol{\nu}) = -\mathcal{L}\left(\boldsymbol{\beta}^{o*} + \frac{\boldsymbol{\nu}}{\sqrt{T}}; \{y_t\}, \{\boldsymbol{z}^*_t\}\right) + \widehat{\lambda}_{k,T} \sum_{i=1}^{p} \widetilde{w}_{k,i} \left|\beta^{o*}_{1,i} + \frac{\nu_{1,i}}{\sqrt{T}}\right| + \lambda_{\sigma,T} \sum_{j=1}^{p+1} \widetilde{w}_{\sigma,j} \left|\beta^{o*}_{2,j} + \frac{\nu_{2,j}}{\sqrt{T}}\right|. \tag{78}$$

我们得到 $\widehat{\boldsymbol{\nu}} := \arg\min_{\boldsymbol{\nu}\in\mathbb{R}^{2p+3}} V(\boldsymbol{\nu}) - V(\mathbf{0})$，使得

$$\widehat{\boldsymbol{\beta}}^{tal} = \boldsymbol{\beta}^{*o} + \frac{\widehat{\boldsymbol{\nu}}}{\sqrt{T}}. \tag{79}$$

具体而言，

$$\begin{aligned}
&V(\boldsymbol{\nu}) - V(\mathbf{0}) \\
&= -\left(\mathcal{L}(\boldsymbol{\beta}^{o*} + \frac{\boldsymbol{\nu}}{\sqrt{T}}; \{y_t\}, \{\boldsymbol{z}^*_t\}) - \mathcal{L}\left(\boldsymbol{\beta}^{o*}; \{y_t\}, \{\boldsymbol{z}^*_t\}\right)\right) \\
&+ \frac{\widehat{\lambda}_{k,T}}{\sqrt{T}} \sum_{i=1}^{p} \widetilde{w}_{k,i} \sqrt{T}\left(\left|\beta^{o*}_{1,i} + \frac{\nu_{1,i}}{\sqrt{T}}\right| - \left|\beta^{o*}_{1,i}\right|\right) + \frac{\lambda_{\sigma,T}}{\sqrt{T}} \sum_{j=1}^{p+1} \widetilde{w}_{\sigma,j} \sqrt{T}\left(\left|\beta^{o*}_{2,j} + \frac{\nu_{2,j}}{\sqrt{T}}\right| - \left|\beta^{o*}_{2,j}\right|\right).
\end{aligned} \tag{80}$$

类似于 (75)，我们还可以得到

$$\lim_{T\to\infty} V(\boldsymbol{\nu}) - V(\mathbf{0}) = \begin{cases} O_p(1), & \text{if } \nu_{1,i} = O_p(1), \nu_{2,j} = O_p(1), \text{ for } \forall (1,i) \in \mathcal{A}_k, \forall (2,j) \in \mathcal{A}_\sigma, \\ & \text{and } \nu_{1,i} = 0, \nu_{2,j} = 0, \text{ for } \forall (1,i) \notin \mathcal{A}_k, \forall (2,j) \notin \mathcal{A}_\sigma, \\ \infty, & \text{otherwise of } \quad \boldsymbol{\nu} = O_p(\mathbf{1}). \end{cases} \tag{81}$$

因此，最小化 $V(\boldsymbol{\nu}) - V(\mathbf{0})$ 等价于最小化 $-\left(\mathcal{L}(\boldsymbol{\beta}^{o*} + \frac{\boldsymbol{\nu}}{\sqrt{T}}; \{y_t\}, \{\boldsymbol{z}^*_t\}) - \mathcal{L}(\boldsymbol{\beta}^{o*}; \{y_t\}, \{\boldsymbol{z}^*_t\})\right)$，其中 $\nu_{1,i} = 0$, $\nu_{2,j} = 0$ 对任意 $(1,i) \notin \mathcal{A}_k$，$(2,j) \notin \mathcal{A}_\sigma$ 均成立，这便得到

$$\sqrt{T}\left(\widehat{\boldsymbol{\beta}}^{tal} - \boldsymbol{\beta}^{o*}\right) = \widehat{\boldsymbol{\nu}} = \underset{\boldsymbol{\nu}\in\mathbb{R}^{2p+3} \cap \{\nu_{1,i}=0. \nu_{2,j}=0, (1,i)\notin\mathcal{A}_k, (2,j)\notin\mathcal{A}_\sigma\}}{\arg\min} V(\boldsymbol{\nu}) - V(\mathbf{0}) \tag{82}$$

由于 $T \to \infty$，因此根据定理 2，在 $\lim_{T\to\infty} \widehat{\nu}_{1,i} = 0$，$\lim_{T\to\infty} \widehat{\nu}_{2,j} = 0$ 对任意 $(1,i) \notin \mathcal{A}_k$，$(2,j) \notin \mathcal{A}_\sigma$ 的限制下，我们得到了 $\widehat{\boldsymbol{\nu}}$ 的渐近行为。此外，

$$\begin{aligned}
&\lim_{T\to\infty} P\{\mathcal{A}^{tal}_T = \mathcal{A}\} = \lim_{T\to\infty} P\left\{\{\widehat{\beta}^{tal}_{1,i} = \beta^{*o}_{1,i} + \frac{\widehat{\nu}_{1,i}}{\sqrt{T}} \neq 0, \widehat{\beta}^{tal}_{2,j} = \beta^{*o}_{2,j} + \frac{\widehat{\nu}_{2,j}}{\sqrt{T}} \neq 0, \forall (1,i) \in \mathcal{A}_k, \forall (2,j) \right. \\
&\left. \in \mathcal{A}_\sigma\} \cap \{\widehat{\beta}^{tal}_{1,i} = 0, \widehat{\beta}^{tal}_{2,j} = 0, \forall (1,i) \notin \mathcal{A}_k, \forall (2,j) \notin \mathcal{A}_\sigma\}\right\} = P\{\{\beta^{*o}_{1,i} \neq 0, \beta^{*o}_{2,j} \neq 0, \forall (1,i) \in \mathcal{A}_k, \forall (2,j) \\
&\in \mathcal{A}_\sigma\} \cap \{\beta^{*o}_{1,i} = 0, \beta^{*o}_{2,j} = 0, \forall (1,i) \notin \mathcal{A}_k, \forall (2,j) \notin \mathcal{A}_\sigma\}\} = 1
\end{aligned} \tag{83}$$

由此可得，$\widehat{\boldsymbol{\beta}}^{tal}$ 模型选择一致。结合 $\sqrt{T}\left(\widehat{\boldsymbol{\beta}}^{tal} - \boldsymbol{\beta}^{o*}\right)$ 的极限分布，我们得出 $\widehat{\boldsymbol{\beta}}^{tal}$ 具有预言机性质。

□